



**T.C.
BURSA ULUDAĞ ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
İŞLETME ANABİLİM DALI
SAYISAL YÖNTEMLER BİLİM DALI**

SAĞLIK SEKTÖRÜNDE MÜŞTERİ KAYIP ANALİZİ VE VERİ MADENCİLİĞİ İLE BİR UYGULAMA

(DOKTORA TEZİ)

MELİH KARAHASANOĞLU

BURSA – 2023

ÖZET

Yazar Adı Soyadı : Melih Karahasanoğlu
Üniversite : Bursa Uludağ Üniversitesi
Anabilim Dalı : İşletme
Bilim Dalı : İşletme
Tezin Niteliği : Doktora
Mezuniyet Tarihi :/...../2023
Tez Danışmanı : Doç. Dr. Gül Gökay Emel

Sağlık Sektöründe Müşteri Kayıp Analizi ve Veri Madenciliği ile Bir Uygulama

İşletmeler süreklilikleri için yeni müşteriler edinirken, müşteri portföylerini de tutundurmaya çalışmaktadırlar. Ağızdan ağıza pazarlamanın, sosyal medyanın mutsuz-kayıp müşterilere kazandırdığı etki alanı, yeni bir müşteri edinmenin mevcut müşterileri elde tutmaktan 5 ila 10 kat daha maliyetli oluşu, Müşteri Kayıp Analizinin (MKA) önemini arttırmıştır. MKA çalışmaları, müşteri aboneliği, sürekliliği bulunan sektörler için sıkça yer alırken sağlık sektöründe uygulamaları pek görülmemektedir. MKA'nın sağlık sektöründe yapılabilmesi için öncelikle, hizmet alıcılarının hangi durumlarda devamlı ve kayıp kabul edildiğini belirleyecek ölçütler gereklidir.

Çocuk hastalar üzerine yapılan bu çalışmada, yıllık yaklaşık 5.000.000 hasta başvurusu olan bir özel sağlık kurumunun 2011-2018 yılları arasındaki 496.847 çocuk hastaya ait 3.400.000 başvurusu incelenmiştir. Öncelikle, ilgili yazın ve sağlık kuruluşunun büyük veri tabanı incelenmiş, kuruluşun uzmanları ile yarı yapılandırılmış mülakat yöntemi ile görüşmeler yapılmış, seçilen çocuk hastalar grubu için devamlı ve kayıp hasta kabul edilme ölçütleri yapılan nitel çalışma sonucunda belirlenmiştir. Veri madenciliği uygulama kısmında; çocuk hastaların yaşları, sosyal güvenceleri ve başvuru sıklıkları ile kayıp hasta kabul edilme durumu, ilgili hastaların sisteme kayıtlı şikâyetleri ile kayıp hasta kabul edilme durumları arasındaki örüntüler Karar Ağacı algoritmalarından CHAID ve CART ile analiz edilmiş ve tahminleme modelleri kurulmuştur. Anket sorularında Net Tavsiye Skorunu (NPS) etkileyen soruya verilen puanlandırma ile kayıp hasta kabul edilme durumu arasındaki örüntüler Ki-Kare testi ile incelenmiştir.

Yaş, sosyal grup ve başvuru sıklıklarına göre kayıp hasta tahminlemesinde kullanılan Veri Madenciliği teknikleri sonuçları karşılaştırılmış ve yaş gruplarına göre 0-1, 1-2 yaş grubunda CHAID algoritmasının, 2-6 ve 6-10 yaş gruplarında CART algoritmasının, şikâyetler ve kayıp hasta tahminlemesinde ise CHAID algoritmasının daha başarılı sonuçlar verdiği gözlenmiştir. Ayrıca, en yüksek Kayıp hasta (churn) oranına sahip 2-6 yaş çocuk hasta grubunu en iyi açıklayan/bölünmeyi veren değerin çocuk hastaların 5-6 yaş aralığındaki başvuruları arasındaki maksimum gün farkı olduğu ve ilgili farkın 177 günden fazla olması durumunda kayıp hasta oranının çok daha belirgin olduğu saptanmıştır.

Anahtar kelimeler: Müşteri Kayıp Analizi, Müşteri İlişkileri Yönetimi, Veri Madenciliği, Karar Ağaçları, Sağlık Hizmetleri Sunumu

ABSTRACT

Name & Surname : Melih Karahasanoğlu
University : Bursa Uludağ University
Institute : Institute of Social Sciences
Field : Business Administration
Subfield : Quantitative Techniques
Degree Awarded : *PhD*.
Date of Degree Awarded :/...../ 2023
Supervisor : Assoc. Prof. Dr. Gül Emel

Churn Prediction in Healthcare and an Application with Data Mining

While businesses aim to acquire new customers, they also focus maintaining existing customers. With increased impact of word-of-mouth and social media on unhappy-churned customers and as acquiring new customers costs 5-10 times more than retaining existing ones, increased the importance of churn analysis. While churn prediction is frequently used in sectors with customer subscription basis, examples in healthcare are very limited. To apply churn analysis in healthcare, firstly criteria for profiling regular and churned patients need to be defined.

3,400,000 visits of 496,847 pediatric patients between 2011-2018 in a private hospital chain, which has approximately 5,000,000 patient-visits annually, were examined. Firstly, literature and database of the hospital were reviewed, subsequently, via qualitative techniques / semi-structured interviews were conducted with the experts to define criteria for regular and churned patients. In the data-mining; patterns between age, social security and admission frequency of the pediatric patients and churned-patient status, the complaints of the same patients registered in the system and the churned- patient status were analyzed with decision tree algorithms CHAID, CART and predictive models were established. The patterns between the question affecting Net Promoter Score in surveys and churned-patient status were examined via Chi-Square Test.

The results of the techniques were compared and it was observed that CHAID in the 0-1, 1-2 age groups, CART in 2-6 and 6-10 age groups for social security, age groups, frequency of visits vs. churn and CHAID for complaints vs. churn showed better results in predicting churned patients. Also, 2-6 years old group which has highest churn rate is explained best by maximum day difference between admission on the 5-6 years old, it was seen that when the maximum difference between 2 visits are more than 177 days, churn rate has a significant increase.

Keywords: Customer Churn, CRM, Data Mining, Decision Trees, Provision of Healthcare Services

ÖNSÖZ

Araştırmanın fikir aşamasından tamamlanmasına kadar geçen tüm süreç boyunca desteğini her an yanımda hissettiğim, sabrı, ısrarcılığı ve yönlendirmeleri ile beni güçlendiren değerli danışman hocam Doç. Dr. Gül GÖKAY EMEL'e, araştırma fikrinin olgunlaşması, pratikleştirilmesi, uygulanmasında kendi yoğunluğu içerisinde ailesinden ve işinden ayırdığı uzun zamanlar için Prof. Dr. Sezgin IRMAK'a, zorlandığım anlarda yanımda olan ve motive eden yönlendirmeleri için Prof. Dr. Çağatan TAŞKIN'a gönülden teşekkür ederim.

Çalışmayı yapabilmem için gerekli veri ve zamanı bana sağlayan, her yeni fikrimde yanımda olan kurumuma, bilgisini ve desteğini hiç esirgemeyen iş arkadaşlarım Sn. Sami SEVGİLİ, Sn. Mehmet ŞEKER ve Sn. Mehmet Orkun SARIBIYIK'a teşekkürü bir borç bilirim.

Akademik hayatım boyunca her zaman manevi desteğini derinden hissettiğim kadim dostum Doç. Dr. Yusuf ACAR'a, yoğun iş hayatı ve uzun süren bu süreçte inancını hiç yitirmeyen, desteği ile hep yanımda olan aileme, sevgili can yoldaşım Ecem'e ve oğlum Ömer Ziya'ya şükranlarımı sunarım.

MELİH KARAHASANOĞLU

03.03.2023

İÇİNDEKİLER

ÖZET	i
ABSTRACT	ii
İÇİNDEKİLER	iv
TABLolar LİSTESİ.....	viii
ŞEKİLLER LİSTESİ.....	ix
KISALTMALAR	xi
GİRİŞ	1

1. BÖLÜM

MÜŞTERİ KAYIP ANALİZİ	5
1.1. Müşteri ve Müşteri İlişkileri Yönetiminin Kavramsal Çerçevesi	5
1.1.1. Müşteri Kavramı	5
1.1.2. Müşteri Tatmini, Sadakati ve Davranışları	5
1.1.3. Müşteri Yaşam Boyu Değeri.....	7
1.1.4. Müşteri İlişkileri Yönetimi.....	7
1.1.5. Müşteri Memnuniyeti Ölçümü ve Teknikleri.....	10
1.1.6. Pazarlama Karması.....	13
1.2. Müşteri Kaybı (Churn) Analizi.....	14
1.2.1. Müşteri Kayıp Çeşitleri	16
1.2.2. Müşteri Kayıp Tahmini	18

2. BÖLÜM

VERİ MADENCİLİĞİ KAVRAM VE YÖNTEMLERİ	20
2.1. Veri, Enformasyon, Bilgi ve Bilgelik Kavramları	20
2.2. Veri Tabanları	22
2.3. Veri Ambarları	23
2.4. Veri Madenciliği	23
2.4.1. VM Tarihsel Gelişimi	24
2.4.2. Veri Madenciliği Uygulama Alanları	25
2.4.3. Veri Tabalarında Bilgi Keşfi ve Veri Madenciliği Süreci.....	26
2.4.3.1. CRISP-DM	27
2.4.3.2. SEMMA	29
2.4.4. Veri Madenciliğinde Karşılaşılabilen Problemler	31

2.5. Veri Madenciliği Yöntemleri	31
2.5.1. Sınıflandırma ve Regresyon Modelleri	33
2.5.1.1. Karar Ağaçları	34
2.5.1.1.1. Karar Ağaçları Avantaj ve Dezavantajları	36
2.5.1.1.2. Bölme Kriterleri	37
2.5.1.1.2.1. DKM Kriteri	38
2.5.1.1.2.2. Bilgi Kazancı	38
2.5.1.1.2.3. Kazanç Oranı	40
2.5.1.1.2.4. Gini İndeksi	41
2.5.1.1.2.5. Twoning.....	42
2.5.1.1.2.6. Ki-Kare Testi	43
2.5.1.1.3. Karar Ağacı Algoritmaları.....	44
2.5.1.1.3.1. CHAID Algoritması	45
2.5.1.1.3.2. CART Algoritması	47
2.5.1.1.3.1. C5.0 Algoritması.....	48
2.5.1.2. Yapay Sinir Ağları.....	49
2.5.1.3. Lojistik Regresyon	51
2.5.1.4. Rassal Orman	53
2.5.1.5. Destek Vektör Makinaları	54
2.5.1.6. Naive-Bayes Yöntemi	55
2.5.1.7. K- En Yakın Komşu	55
2.5.1.8. Genetik Algoritmalar	56
2.5.2. Kümeleme Modelleri	57
2.5.3. Birliktelik Kuralları ve Modelleri	59
2.5.3.1. AIS Algoritması.....	60
2.5.3.2. Apriori Algoritması	60
2.5.3.3. CARMA Algoritması	61

3. BÖLÜM

MÜŞTERİ KAYIP ANALİZİ VE VERİ MADENCİLİĞİ	62
3.1. Veri madenciliği ve Müşteri Kayıp Analizi	62
3.2. Sağlık Sektöründe Veri Madenciliği	66
3.3. Sağlık Sektöründe Müşteri Kayıp Analizi.....	70
3.3.1. Türkiye'deki Sağlık Sektörünün ve Hasta Tercihlerinin Gelişimi	70
3.3.2. Sağlık Kurumlarında Müşteri Kayıp Analizi'nin Önemi	71

3.3.3. Sağlık Kurumlarında Müşteri ve Hasta Kavramsal Çerçeve	72
3.3.4. Sağlık Kurumlarında Müşteri Kayıp Analizinin Kullanım Alanları ve Örnekleri	74

4. BÖLÜM

SAĞLIK SEKTÖRÜNDE MKA İÇİN BİR ARAŞTIRMA VE METODOLOJİSİ..77

4.1. Araştırmanın Önemi	77
4.2. Araştırmanın Amacı ve Soruları	78
4.3. Araştırmanın Yöntemi ve Aşamaları	79
4.3.1. Geliş Sıklıkları ve Kayıp Hasta Tahminlemesi	80
4.3.2. Şikâyetler ve Kayıp Hasta Tahminlemesi	85
4.3.3. Anketler ve Kayıp Hasta Tahminlemesi	86
4.4. Araştırmanın Model Değerlendirme Kriterleri	87
4.5. Araştırmanın Kısıtları	89

5. BÖLÜM

ARAŞTIRMANIN BULGULARI VE DEĞERLENDİRME METRİKLERİ91

5.1. Devamlı Hasta/ Kayıp Hasta Belirleme Ölçütleri	91
5.1.1. Devamlı Hasta Kabul Edilme Ölçütleri ve Nedenleri	92
5.1.2. Çocuk Hasta Gruplamaları ve Geliş Sıklıkları	93
5.1.3. Kayıp Hasta Belirleme Ölçütleri ve Nedenleri	94
5.2. Geliş Sıklıkları ve MKA	94
5.2.1. 0-1 Yaş Grubu Geliş Sıklıkları ve MKA	96
5.2.1.1. CHAID Algoritması ile Bulgular ve Değerlendirme Metrikleri	96
5.2.1.2. CART Algoritması ile Bulgular ve Değerlendirme Metrikleri	103
5.2.1. 1-2 Yaş Grubu Geliş Sıklıkları ve MKA	110
5.2.1.1. CHAID Algoritması ile Bulgular ve Değerlendirme Metrikleri	110
5.2.1.2. CART Algoritması ile Bulgular ve Değerlendirme Metrikleri	118
5.2.3. 2-6 Yaş Grubu Geliş Sıklıkları ve MKA	123
5.2.3.1. CHAID Algoritması ile Bulgular ve Değerlendirme Metrikleri	124
5.2.3.2. CART Algoritması ile Bulgular ve Değerlendirme Metrikleri	142
5.2.4. 6-10 Yaş Grubu Geliş Sıklıkları ve MKA	149
5.2.4.1. CHAID Algoritması ile Bulgular ve Değerlendirme Metrikleri	149
5.2.4.2. CART Algoritması ile Bulgular ve Değerlendirme Metrikleri	161
5.3. Şikâyetler ve MKA	167

5.3.1. CHAID Algoritması ile Bulgular ve Değerlendirme Metrikleri	168
5.3.2. CART Algoritması ile Bulgular ve Değerlendirme Metrikleri.....	173
5.4. Hasta Memnuniyet Anketleri ve MKA	177
SONUÇ VE ÖNERİLER.....	181
KAYNAKÇA	191
EKLER.....	208
ÖZGEÇMİŞ.....	Error! Bookmark not defined.



TABLÖLAR LİSTESİ

Tablo 2.1.: Veri Madenciliği Modelleri ve Kullanım Alanları	33
Tablo 2.2.: Karar Ağacı Algoritmaları ve Özellikleri	45
Tablo 2.3.: Kümeleme Metotları ve Genel Özellikleri	58
Tablo 4.1.: Geliş Sıklıkları ve Churn Modeli Temin Edilen Değişkenler ve Tanımlamaları	81
Tablo 4.2.: Geliş Sıklıkları ve Churn Eklenen Değişkenler ve Tanımlamaları	82
Tablo 4.3.: Yaş Grupları ve Geliş Sıklıkları Grupları	83
Tablo 4.4.: Devamlı Hasta ve Kayıp Hasta Kabul Kriterleri.....	83
Tablo 4.5.: Karşıtlık Matrisi Tanıtımı	87
Tablo 5.1.: Model Yaş Grup Değişkeni Churn Etiketini	95
Tablo 5.2.: 1-12 Ay CHAID Poliklinik Adetleri ve Churn Durumu	97
Tablo 5.3.: 0-1 Yaş CHAID Değerlendirme Metrikleri	101
Tablo 5.4.: 0-1 Yaş CHAID Karşıtlık Matrisi.....	102
Tablo 5.5.: 0-1 Yaş CART Değerlendirme Metrikleri.....	108
Tablo 5.6.: 0-1 Yaş CHAID Karşıtlık Matrisi.....	109
Tablo 5.7.: 1-2 Yaş Poliklinik Başvurularına Göre Churn Durumu	111
Tablo 5.8.: 1-2 Yaş CHAID Değerlendirme Metrikleri	116
Tablo 5.9.: 1-2 Yaş CHAID Karşıtlık Matrisi.....	117
Tablo 5.10.: 1-2 Yaş CART Değerlendirme Metrikleri.....	122
Tablo 5.11.: 1-2 Yaş CART Karşıtlık Matrisi	122
Tablo 5.12.: 2-6 Yaş CHAID MAXG_5 Sınıfı Başvurular Arası Gün Farkı ve Kayıp hasta Durumu.....	124
Tablo 5.13.: 2-6 Yaş CHAID MAXG_4 Sınıfı Başvurular Arası Gün Farkı ve Kayıp hasta Durumu.....	131
Tablo 5.14.: 2-6 Yaş CHAID Son ve Sondan2. Başvurular Arası Gün Sayısı ve Kayıp hasta Durumu	134
Tablo 5.15.: 2-6 Yaş CHAID Değerlendirme Metrikleri	140
Tablo 5.16.: 2-6 Yaş CHAID Karşıtlık Matrisi.....	140
Tablo 5.17.: 2-6 Yaş CART Değerlendirme Metrikleri.....	147
Tablo 5.18.: 2-6 Yaş CART Karşıtlık Matrisi	148
Tablo 5.19.: 6-10 Yaş CHAID MAXG_6 Sınıfı Başvurular Arası Gün Farkı ve Kayıp hasta Durumu	150
Tablo 5.20.: 6-10 Yaş CHAID Değerlendirme Metrikleri	159
Tablo 5.21.: 6-10 Yaş CHAID Karşıtlık Matrisi.....	160
Tablo 5.22.: 6-10 Yaş CART Değerlendirme Metrikleri.....	165
Tablo 5.23.: 6-10 Yaş CART Karşıtlık Matrisi	165
Tablo 5.24.: Şikâyetler CHAID Değerlendirme Metrikleri.....	171
Tablo 5.25.: Şikâyetler CHAID Karşıtlık Matrisi	171
Tablo 5.26.: Şikâyetler CART Değerlendirme Metrikleri	176
Tablo 5.27.: Anketler ve Kayıp hasta Ki-Kare Testi Sonuçları	178

ŞEKİLLER LİSTESİ

Şekil 2.1.: Veri, Enformasyon, Bilgi ve Bilgeliğin Bağlantılılığı	21
Şekil 2.2: Veri Madenciliği Bilgi Keşfi Süreci.....	26
Şekil 2.3. : CRISP-DM Modeli	27
Şekil 2.4.: SEMMA Modeli.....	30
Şekil 2.5.: Örnek Karar Ağacı Modeli	34
Şekil 2.6.: Yapay Sinir Ağı Katmanları.....	50
Şekil 2.7.: Lojistik Regresyon Eğrisi	52
Şekil 2.8.: Destek Vektör Makinaları Grafikselsel İşleyişi	55
Şekil 2.9.: Genetik Algoritmaların İşleyişi	56
Şekil 2.10.: Koordinat Düzleminde Kümeleme Örneği	57
Şekil 2.11.: Birliktelik Kuralları İşlem Adımları	59
Şekil 4.1.: Veri Madenciliği ile Bilgi Keşfi Adımları	80
Şekil 5.1.: Geliş Sıklıkları ve Yaş Gruplarına Göre Churn Genel Bakış	95
Şekil 5.2.: 0-1 Yaş CHAID Genel Durumu	96
Şekil 5.3.: 0-1 Yaş CHAID 2. Döğüm Dalları	98
Şekil 5.4.: 0-1 Yaş CHAID 3. Döğüm Dalları	98
Şekil 5.5.: 0-1 Yaş CHAID 6. Döğüm Dalları	99
Şekil 5.6.: 0-1 Yaş CHAID 7. Döğüm Dalları	100
Şekil 5.7.: 0-1 Yaş CART Genel Durumu	103
Şekil 5.8.: 0-1 Yaş CART İlk Bölünme.....	104
Şekil 5.9.: 0-1 Yaş CART 1. Döğüm ve Alt Dalları	104
Şekil 5.10.: 0-1 Yaş CART 4. Döğüm ve Dalları	105
Şekil 5.11.: 0-1 Yaş CART 2. Döğüm ve Dalları	106
Şekil 5.12.: 0-1 Yaş CART 5. Döğüm ve Dalları	106
Şekil 5.13.: 0-1 Yaş CART 12. Döğüm ve Dalları	107
Şekil 5.14.: 0-1 Yaş CART 6. Döğüm ve Dalları	108
Şekil 5.15.: 1-2 Yaş CHAID Genel Durum	110
Şekil 5.16.: 1-2 Yaş CHAID 1. Döğüm ve Dalları	112
Şekil 5.17.: 1-2 Yaş CHAID 12. Döğüm ve Dalları	112
Şekil 5.18.: 1-2 Yaş CHAID 13. Döğüm ve Dalları	113
Şekil 5.19.: 1-2 Yaş CHAID 14. Döğüm ve Dalları	113
Şekil 5.20.: 12 Yaş CHAID 8. Döğüm ve Dalları	114
Şekil 5.21.: 1-2 Yaş CHAID 9. Döğüm ve Dalları	115
Şekil 5.22.: 1-2 Yaş CHAID 10. Döğüm ve Dalları	116
Şekil 5.23.: 1-2 Yaş CART Genel Durumu	118
Şekil 5.24.: 1-2 Yaş CART 1. Döğüm ve Dalları	119
Şekil 5.25.: 1-2 Yaş CART 3. Döğüm ve Dalları	119
Şekil 5.26.: 1-2 Yaş CART 15. Döğüm ve Dalları	120
Şekil 5.27.: 1-2 Yaş CART 17. Döğüm ve Dalları	121
Şekil 5.28.: 2-6 Yaş CHAID Genel Durumu	124
Şekil 5.29.: 2-6 Yaş CHAID 1. Döğüm ve Dalları	125
Şekil 5.30.: 2-6 Yaş CHAID 2. Döğüm ve Dalları	126
Şekil 5.31.: 2-6 Yaş CHAID 3. Döğüm ve Dalları	127
Şekil 5.32.: 2-6 Yaş CHAID 4. Döğüm ve Dalları	128
Şekil 5.33.: 2-6 Yaş CHAID 5. Döğüm ve Dalları	129
Şekil 5.34.: 2-6 Yaş CHAID 6. Döğüm ve Dalları	129
Şekil 5.35.: 2-6 Yaş CHAID 7. Döğüm ve Dalları	130
Şekil 5.36.: 2-6 Yaş CHAID 8. Döğüm ve Dalları	130
Şekil 5.37.: 2-6 Yaş CHAID 27. Döğüm ve Dalları	131

Şekil 5.38.: 2-6 Yaş CHAID 28. Döğüm ve Dalları	132
Şekil 5.39.: 2-6 Yaş CHAID 27. Döğüm ve Dalları	132
Şekil 5.40.: 2-6 Yaş CHAID 30. Döğüm.....	133
Şekil 5.41.: 2-6 Yaş CHAID 32. ve 33. Döğüm	134
Şekil 5.42.: 2-6 Yaş CHAID 40. Döğüm ve Dalları	135
Şekil 5.43.: 2-6 Yaş CHAID 41. Döğüm ve Dalları	136
Şekil 5.44.: 2-6 Yaş CHAID 42. Döğüm ve Dalları	137
Şekil 5.45.: 2-6 Yaş CHAID 43. Döğüm ve Dalları	138
Şekil 5.46.: 2-6 Yaş CHAID 44-45-46. Döğüm ve Dalları	139
Şekil 5.47.: 2-6 Yaş CART Genel Durumu	142
Şekil 5.48.: 2-6 Yaş CART 3 - 4. Döğüm ve Dalları	143
Şekil 5.49.: 2-6 Yaş CART 2. Döğüm ve Dalları	144
Şekil 5.50.: 2-6 Yaş CART 5. Döğüm ve Dalları	145
Şekil 5.51.: 2-6 Yaş CART 6. Döğüm ve Dalları	146
Şekil 5.52.: 6-10 Yaş CHAID Genel Durum	149
Şekil 5.53.: 6-10 Yaş CHAID 1. Döğüm ve Dalları	151
Şekil 5.54.: 6-10 Yaş CHAID 2. Döğüm ve Dalları	152
Şekil 5.55.: 6-10 Yaş CHAID 3. Döğüm ve Dalları	153
Şekil 5.56.: 6-10 Yaş CHAID 17. Döğüm ve Dalları	154
Şekil 5.57.: 6-10 Yaş CHAID 4. Döğüm ve Dalları	155
Şekil 5.58.: 6-10 Yaş CHAID 20. Döğüm ve Dalları	156
Şekil 5.59.: 6-10 Yaş CHAID 5. Döğüm ve Dalları	157
Şekil 5.60.: 6-10 Yaş CHAID 23. Döğüm ve Dalları	158
Şekil 5.61.: 6-10 Yaş CHAID 6. Döğüm.....	159
Şekil 5.62.: 6-10 Yaş CART Genel Durumu	161
Şekil 5.63.: 6-10 Yaş CART 1. Döğüm ve Dalları	162
Şekil 5.64.: 6-10 Yaş CART 2. Döğüm ve Dalları	164
Şekil 5.65.: Şikâyetler ve Kayıp hasta CHAID Uygulama Genel Durumu	168
Şekil 5.66.: Şikâyetler CHAID 1. Döğüm ve Dalları	169
Şekil 5.67.: Şikâyetler CHAID 2. Döğüm ve Dalları	170
Şekil 5.68.: Şikâyetler ve Kayıp hasta CART Uygulaması 1. Döğüm ve 2. Döğüm ve Dalları	173
Şekil 5.69.: Şikâyetler CART 3. Döğüm ve Dalları	174
Şekil 5.70.: Şikâyetler CART 4. Döğüm ve Dalları	175
Şekil 5.71.: Şikâyetler CART 5. Döğüm ve Dalları	176
Şekil 5.72.: Kültürel Boyut Ölçütünde Ülke Karşılaştırması	180

KISALTMALAR

AHP	: Analitik Hiyerarşi Prosesi
BV	: Büyük Veri
CART	: Classification & Regression Trees (Sınıflandırma ve Regresyon Ağaçları)
CES	: Customer Effort Score (Müşteri Efor Puanı)
CHAID	: Chi-Squared Automatic Interaction Detector (Ki-Kare Otomatik Etkileşim Dedektörü)
CRISP-DM	: Cross Industry Standard Process Model for Data Mining (Endüstriler Arası Standart İşleme – Veri Madenciliği)
CRM	: Customer Relationship Management
CSAT	: Customer Satisfaction Score (Müşteri Memnuniyet Puanı)
DVM	: Destek Vektör Makinaları
İVTYS	: İlişkisel Veri Tabanı Yönetim Sistemi
CLV	: Customer Lifetime Value
LR	: Lojistik Regresyon
MİY	: Müşteri İlişkileri Yönetimi
MKA	: Müşteri Kayıp Analizi
MYBD	: Müşteri Yaşam Boyu Değeri
NPS	: Net Promoter Score (Net Tavsiye Puanı)
OLAP	: On-line Analytical Processing (Çevrim İçi Analitik İşleme)
OLTP	: On-Line Transaction Processing (Çevrim İçi Hareket İşleme)
RO	: Rassal Orman
SEMMA	: Sample, Explore, Modify, Model and Assess (Örnekleme, Keşfetme, Değiştirme, Modelleme ve Değerlendirme)
SQL	: Structured Query Language (Yapısal Sorgu Dili)
VA	: Veri Ambarı
VM	: Veri Madenciliği
VTBK	: Veri Tabanlarında Bilgi Keşfi
VTYS	: Veri Tabanı Yönetim Sistemi
YSA	: Yapay Sinir Ağları

GİRİŞ

Günümüzdeki teknolojik gelişmeler ve kültürel değişimler, işletmelerin çalışma ortam ve biçimini etkilemiş, işletmelerin başarısı için ekosistemlerinin verimliliği çok önemli olmuş ve sorunlar ortak değer tabanlı geleceği inşa etmeye yönelik iş modelleri ile çözülür hale gelmiştir. Teknoloji ve müşteri yönetimi ile ilgili yeni birçok kavram iş yazınında yerini almış, işletmeler ise bu kavramlara dayalı iş modelleri için aksiyon aşamasına geçmişlerdir. İşletmeler için büyümenin ya da hayatta kalmanın en temel faktörlerinden birisi, hizmet sunumuna yeni müşterileri eklemek ya da mevcut müşterilerini koruyabilmektir. Yapılan araştırmalarda; yeni müşteri kazanmanın, mevcut müşteriye mutlu edip elde tutmaya göre beş ile on kat daha maliyetli olduğu saptanmıştır (Kotler ve Armstrong, 2017). Beal (2014), yeni müşteri kazanmanın mevcut müşteriye elde tutmaktan yedi kat daha maliyetli olduğunu ve memnuniyetsiz ayrılan müşterilerin %96'sının şikâyet başvurusunun olmadığını fakat deneyimlerini en az 15 kişi ile paylaştıklarını ortaya koymuştur. Genellikle, memnun ayrılan bir müşteri ise görüşlerini, önerilerini çevresindeki sadece iki-üç kişi ile paylaşmaktadır. Nilsen (2012), raporunda müşterilerinin %92'sinin ağızdan ağıza pazarlamayı diğer pazarlama-iletişim kanallarını kullanmaya tercih ettiklerini, tanıdıklarından duydukları ya da müşteri bloglarından okumuş oldukları müşteri deneyimlerini daha güvenilir bulduklarını paylaşmıştır. Özellikle, hizmet sektöründeki müşterilerin memnuniyetsizliklerini dile getirmedeki yatkınlığı da göz önünde bulundurulduğunda, ağızdan ağıza pazarlama kanalı olan mevcut müşteriler, bireysel bağlılıklarının yanı sıra işletmenin büyümesinde etkin bir rol almaktadırlar. Sernovitz (2012) ve Ruby'nin (2015) yaptığı çalışmalar, memnuniyetsiz-sorunu çözülmemiş bir müşterinin potansiyel etkisini kaldırmak için yaklaşık 12 tane yeni müşteri kazanılması ya da 12 tane olumlu yorum alınması gerektiğini göstermektedir.

Ağızdan ağıza pazarlama kanalını oluşturan mevcut müşterilerin memnuniyet ve şikayetlerini dile getirme şekli, sosyal medyanın yaygınlığı, hızı ve rakipler arasında müşteri kayması üzerindeki gücü, mevcut müşterileri elde tutmayı çok daha önemli hale getirmiştir. Bu gerçeğin farkında olan birçok hizmet sektöründe çalışan işletme ise satış kanallarını arttırmayı hedeflemekle birlikte, mevcut

müşteriyi mutlu etmenin ya da elde tutmanın yollarına daha fazla odaklanmaya başlamıştır. Dolayısıyla, işletmeler için yeni müşteri kazanmanın maliyeti, mevcut müşteriyi memnun etmenin maliyeti ile kıyaslanan ve takip edilen kritik bir performans göstergesi konumuna gelmeye başlamıştır. İşletmeler, müşteri kayıplarını minimize etmek için kimlerin kayıp müşteri olabileceğini önceden tahmin edebilmelidirler. İlgili tahminlerin yapılabilmesi için ise öncelikle hem hizmet alımına devam eden müşterilerin hem de başka bir hizmet sunucusuna geçen kaybedilen müşterilerin profil yapıları, segmentleri, talepleri, davranış ve eğilimlerinin iyi anlaşılmış olması gerekmektedir. Pazarlama ve müşteri ilişkileri yönetimi birimleri koordineli çalışarak kayıpları önlenmeye çalışmalıdırlar.

Yazında ve pratik uygulamalarda ise bu ihtiyacın Müşteri Kayıp Analizi (MKA) ile sağlanabileceği gözlenmiştir. Yazında, müşteri yıpranması (customer attrition), müşteri cayması (customer turnover), müşteri kaçması (customer defection) gibi terimlerle de ilişkili olsa da en kapsayıcı tanım kayıp müşteri (customer churn) olarak görülmüştür.

MKA; genellikle telekomünikasyon, bankacılık ve sigortacılık gibi müşteri sürekliliği bulunan sektörlerde uygulanmaktadır. Bu tarz arzın yoğun olduğu sektörlerde müşteriler, ellerindeki alternatifler doğrultusunda kolaylıkla hizmet sunucusu değişikliğine gidebilmektedir. Yeni müşteri bulmanın maliyeti de göz önünde bulundurulduğunda, mevcut müşterilerin elde tutulması, yani hizmet alımının devam etmesi, şirketlerin sürdürülebilirliği açısından kritik bir rol almaktadır. Dolayısıyla, mevcut müşterilerin tercih ve eğilimleri yakinen izlenmelidir. Temelde işletmeler, veri madenciliği teknikleri ile ellerindeki dağınık veriyi anlamlı hale getirmeye ve geçmişteki müşteri eğilimleri ile karşılaştırarak, müşteriler için kritik hizmet sunucusu değiştirme noktalarını saptamaya çalışmaktadırlar. Bu çalışma sonucunda da müşterilerinin hizmet alım yaşam döngülerini belirlemeye çalışmaktadırlar. İşletmeler ile ilgili sınıflandırma ve tahmin modelleri yapıldığında, müşterilerinin olası kaybedilme durumu öncesi çeşitli tutundurma faaliyetleri üzerinden olası tahminlenen kaybı önlemeye çalışabilmektedirler.

Yapılan bu çalışmada ise amaç; bankacılık, telekomünikasyon ve sigortacılık gibi sektörlere göre nispeten kurumsal olarak daha az gelişmiş sağlık ve özellikle hastane sektöründe bir uygulama örneği üzerinden kayıp müşterileri tanımlayabilmek, tutum-eğilimlerini saptayacak, gelecekteki eğilimlerini tahmin edebilecek bir model kurgulamak ve müşteri kayıplarını önlemeye çalışacak aksiyon planlarına ışık tutmaktır. Dünya genelinde sağlık sektöründe MKA uygulanmasının çok nadir olması ve Türkiye’de ise bu uygulamanın yazın taramasında saptanmaması çalışmayı farklı kılan unsurlardandır. Ayrıca bu çalışma ile (kayıp müşterilerinin nedenleri ile tanımlanıp, yeni kayıpların önlenmesinin) birçok açıdan sağlık hizmeti sunucusuna ve hastalara (hizmet kalitesi ve toplum sağlığı gibi) olumlu katkılarının olacağı yeni çalışmalara ışık tutması amaçlanmaktadır.

Ayrıca, çalışma kapsamında aşağıdaki bazı sorulara da yanıtlar aranacaktır:

- Sağlık sektöründe hasta mı müşteri mi daha kabul edilir bir tanımlamadır?
- Sağlık sektöründe sürekli hizmet alıcısı ve kayıp hizmet alıcısı tanımlanabilir mi?
- MKA her hastaya uygulanabilir mi?
- Sağlık sektöründe hastaların hizmet sunucularını değiştirmekteki eğilimleri, abonelik, müşteri sürekliliği bulunan diğer sektörlerde olduğu gibi MKA ile değerlendirilip önleyici faaliyetlerde bulunabilinir mi?
- Hastaların geliş sıklıkları ve kayıp hasta kabul edilme durumları arasında bir ilişki var mıdır?
- Hastaların ayrılma nedenleri hasta gruplarına göre değişiklik gösterir mi?
- Hastaların şikayetleri ile kayıp hasta kabul edilme durumları arasında bir ilişki var mıdır?
- Hastaların Net Tavsiye Puanını (NPS) etkileyen anket sorusuna verdiği cevaplar ile kayıp hasta kabul edilme durumu arasında ilişki kurulabilir mi?
- Hastaların hizmet sunucusundan ayrılmaları tahminlenebilir mi?

Bu doğrultuda; çalışmanın ilk iki bölümünde yazın taraması ile Müşteri Kayıp Analizi'nin ve Veri Madenciliği'nin kavramsal çerçevesi uygulama örnekleri ile incelenmiştir. Üçüncü bölümde, sağlık sektörü ve Veri Madenciliği ele alınmıştır. Müşteri-hasta, devamlı-kayıp hizmet alıcısı gibi kavramlar ve alanda yapılan çalışmalar, yazın ve sektörel bilgi ışığında irdelenmiştir. Dördüncü bölümde, çalışmanın amaç, yöntem ve aşamaları paylaşılmıştır. Beşinci bölümde, yapılan çalışmanın analiz bulguları paylaşılmıştır. Özel bir sağlık grubunun 500 bine yakın hastasının 3 milyona yakın hizmet alım başvurusu üzerinden Veri Madenciliği tekniklerinden CHAID (Chi-Squared Automatic Interaction Detector) ve CART (Classification And Regression Trees) algoritmaları aracılığı ile MKA yapılmış, modeller karşılaştırılmış ve bulgular değerlendirilmiştir. Ayrıca, çocuk hastalara ait şikayetler ve kayıp hizmet alıcısı kabul edilme durumu yine CHAID ve CART ile karşılaştırılmalı olarak değerlendirilmiştir. Çalışmanın uygulama bölümünde son olarak, anketler ve kayıp hizmet alıcısı kabul edilme örüntüleri Ki-Kare testi aracılığı ile incelenmiştir. Son bölümde ise; çalışmanın çıktıları sağlık sektörünün yapısı, pazarlama bilimi ve kültürel pencereden değerlendirilmiş ve gelecekteki olası çalışmalara önermeler ile tamamlanmıştır.

1. BÖLÜM

MÜŞTERİ KAYIP ANALİZİ

1.1. Müşteri ve Müşteri İlişkileri Yönetiminin Kavramsal Çerçevesi

Bu bölümde müşteri ve müşteri ilişkileri yönetimine ait kavramsal çerçeve paylaşılmıştır.

1.1.1. Müşteri Kavramı

Şirketlerin devamlılıkları için ürettikleri, sundukları mal ya da hizmetlere sürekli olarak alıcı bulmaları gerekmektedir. Buradaki alıcılar; müşteri olarak anılmakta, en basit tanımlama ile belirli bir mal veya hizmet sunucusunda düzenli alışverişte bulunan ya da yapma niyetinde olan kişiler olarak tanımlanmaktadır (Barış ve Odabaşı, 2002). Simmons (2014)'a göre müşteri, mal ve/veya hizmet sunucusuna satın aldıkları mal ve/veya hizmet ile gelir sağlayan kurum ya da insanlardır. Müşteri kavramı, alınan hizmet ya da malın aynı zamanda son kullanıcısı olduğu durumlar dışında tüketici kavramı ile ayrılmaktadır. Kotler ve Keller (2009), tüketiciyi ihtiyaçlarını karşılamak amacı ile bir ürün veya hizmet için ücret ödeyen belirli kişi veya taraf olarak tanımlamıştır. Dolayısı ile her ne kadar müşteri ve tüketici, ihtiyaç ve isteklerini karşılamak için alışveriş yapan kişiler olarak tanımlansa da, bir süpermarkette alışveriş yapan restoran müşteri iken, restoranda alınan malzemelerle yapılan yemeği yiyenler yani nihai kullanıcılar ise tüketici olarak tanımlanmaktadır (Majava, vd., 2013).

1.1.2. Müşteri Tatmini, Sadakati ve Davranışları

Kotler (2000) tarafından müşteri tatmini-memnuniyeti; bir kişinin bir ürünün algılanan performansı ile kendi beklentilerini karşılaştırmasından kaynaklanan zevk veya hayal kırıklığı duyguları olarak tanımlanmıştır. Hansemark ve Albinsson (2004)'a göre ise "memnuniyet; bir hizmet sağlayıcıya yönelik genel bir müşteri tutumu veya bazı ihtiyaçların, hedeflerin veya arzuların yerine

getirilmesi ile ilgili olarak müşterilerin, bekledikleri ile deneyimledikleri arasındaki farkla ilgili duygusal bir tepkidir". Bir başka tanımlama ile müşteri memnuniyeti; arzu edilen, beklenen mal ya da hizmetin alıcıya sunulması ve alıcı tarafından satın alınması sonrası beklentinin karşılanmasıdır (Pekmezci, vd., 2008). Müşterilerin tatmin olması, işletmeden mal ve hizmet alımına devam edilmesi anlamına gelmektedir. Yapılan çalışmalarda, mevcut müşteriyi elde tutma oranındaki %5'lik bir artışın birçok sektörde müşterilerin net mevcut değerlerinde %25 ile %85 arasında artışa olanak tanıdığı paylaşılmaktadır (Reichled ve Sasser, 1990; Reicheld, 1996). Bu doğrultuda, memnuniyetsiz müşterilerin birim olarak kaybının yanı sıra mahrum kalınan muhtemel ömür boyu katkıları ve deneyimlerini paylaşarak etkileyecekleri potansiyel müşteri kitlesi düşünüldüğünde, müşteri tatmini-memnuniyeti ölçülen ve yakın takip edilen bir konu haline gelmiştir. Sağlanan müşteri tatminini firmalara, İngilizce yazında "customer loyalty" olarak geçen müşteri sadakati konusunu gündeme getirmiştir.

Dick ve Basu (1994), sadakati müşterinin markaya bağlılığı veya markaya yaklaşımı (hizmet, ürün kategorisi vb.) olarak tanımlamıştır. Sadakat, belirli bir marka ile ilişkiyi sürdürme beklentisi olarak da yorumlanmaktadır. Aaker'e (1991) göre marka sadakati, "müşterinin bir markaya duyduğu bağlılık" olarak tanımlanmaktadır.

Müşteri sadakati oluşturabilmek için müşteri tatminini sağlamak, müşteri tatmini için de müşteri davranışlarının anlaşılması gerekmektedir. Müşteri davranışları, genel olarak bireylerin, grupların ve kuruluşların ihtiyaç ve isteklerini karşılamak için malları, hizmetleri, fikirleri veya deneyimleri nasıl seçtiği, satın aldığı, kullandığı ve elden çıkardığı olarak tanımlanabilmektedir (Kotler ve Keller, 2009). Dolayısıyla, müşterilerin davranışları iyi anlaşılmalı, olası müşteri kaybı önceden tespit edilmeli, engellenmeli, memnuniyetleri sağlanmalı ve sadakat düzeyleri üzerine çalışılmalıdır.

1.1.3. Müşteri Yaşam Boyu Değeri

İngilizce yazında Customer Lifetime Value (CLV) olarak yerini bulan Müşteri Yaşam Boyu Değeri (MYBD), bir işletmenin müşterisi ile bağından ileride kazanacağı değerin, bugünkü zamanda maddi karşılığı olarak tanımlanabilmektedir (Yörük ve Eşmekaya, 2018). Bir başka tanım da; bir şirketin kazançlarının ve dolayısı ile değerinin, müşterilerin şirkete sağladığı gelecekteki kârların iskonto edilmiş değeri olarak tanımlanmış toplam MYBD olarak yer bulmuştur (Gupta ve Lehmannand, 2004).

MYBD'nin amacı; şirket için müşterilerinin her birinin finansal değerini belirlemektir. MYBD geçmişteki verilerden yararlanarak çıkarım yapmakta ve gelecekte müşterinin firmaya sağlayacağı değeri tahmin etmektedir (Jain, Dipak ve Siddhartha, 2002). Bu açıdan, geçmişteki verilerin ölçülmesi ve değerlendirilmesi ile saptanan müşteri karlılığı birbirinden ayrılmaktadır. Yapılan çalışmalarla ulaşılmak istenen, müşterinin gelecekteki oluşturması beklenen toplam gelirleri ve ilgili müşterinin tutundurulması için toplam maliyet arasındaki farkın bugünkü değeridir. $MYBD = \text{Toplam Müşteri Gelirleri} - \text{Toplam Müşteri Maliyetleri}$ şeklinde formüle etmek mümkündür.

1.1.4. Müşteri İlişkileri Yönetimi

Müşteri İlişkileri Yönetimi (MİY), yabancı yazında Customer Relationship Management (CRM) olarak geçen, etkili müşteri ilişkileri ve etkileşimleri geliştirerek ve sürdürerek müşteri memnuniyetini, bağlılığını ve karlılığını artırmayı amaçlayan yenilikçi bir teknoloji olarak tanımlanmaktadır (Hung S.-Y, vd., 2010). MİY sistemi; bir işletmenin müşterileri hakkında ayrıntılı ve anlamlı bilgi edinmelerini sağlayacak yazılım ve donanıma sahip olması ve böylece müşterilerine daha kapsamlı, hedef odaklı ve kişiselleştirilmiş pazarlama sunabilmeleri anlamına gelmektedir (Tokay, 2019). MİY, müşteriler hakkında en ayrıntılı bilgileri edinmek, müşterileri çok küçük kesimler ile bölümlendirmek (mikro segmentasyon), ilgili segmentleri karlılıklarına göre ayırtmak, her segmente ait müşteriye ayrı ve özel pazarlama stratejisi uygulamak olarak da tanımlanabilmektedir (Kırım, 2001). Bir diğer tanıma göre de, müşteri değerinin

artışını ve değerli müşterilerin yaşam boyu sadık kalmasını sağlayarak tekrar satın almaya yönelmesini sağlayan, doğru araçları kullanmayı mümkün kılan bir alt yapıdır (Dyche, 2002). MİY yaklaşımı ile genel olarak müşteri tatmininin, sadakatinin artırılması, müşterinin elde tutulmasının geliştirilmesi, müşteri ilişkilerinin karlı hale getirilmesi, farklılaşmayı ve kişiselleştirmeyi sağlama, müşteri memnuniyeti oranının artışı amaçlanmaktadır (Seymen, 2022).

MİY, 1970'lerde (Buttle, 2004) şirketlerdeki satış gücü otomasyonunu yönetmek ve optimize etmek için yeni bir araç olarak ortaya çıkmıştır. Sadece satış ve pazarlama amaçları için değil, aynı zamanda daha etkili müşteri etkileşimi (King ve Burgess, 2007) ve müşteri bilgi yönetimi için kullanılan en popüler araçlardan biri haline gelmiştir. Chen ve Popovich (2003), MİY'ini, bir şirketin müşterilerini mümkün olan en iyi şekilde anlamaya çalışan, süreçlerin, insan sermayesinin ve teknolojinin entegrasyonu olarak tanımlamaktadır. Ayrıca, özellikle müşteriye elde tutma ve ilişkiyi yönetmeye odaklanılırsa, MİY ilişki yönetimi için mevcut en yeni bütünsel yaklaşımdır. Gelişimi incelendiğinde; önce 1990'lı yılların başında birinci nesil fonksiyonel olarak MİY satış gücü otomasyonu, satış öncesi işlevleri içeren ve satış sonrası işlevleri ele alan müşteri hizmetleri ve desteği olarak bulunduğu görülmüştür. 1996'dan sonra ikinci nesil MİY; müşteriye dönük yaklaşımı benimsemiştir. Tamamen müşteri deneyimine yönelmiş, satış gücü otomasyonu ve satış sonrası hizmetleri birleştirmiş ve müşteri deneyimini tüm müşterilere ayrımsız uygulamayı benimsemiştir. 2002'li yıllara geldiğimizde MİY stratejik bir form almıştır. Müşteri portföyüne stratejik yaklaşım, gelirlerin artırılması, maliyetlerin azaltılması irdelenmeye başlanmış ve internet teknolojileri ile MİY güçlendirilmiştir. 2008 sonrasında günümüze kadar ise stratejik MİY, çevik ve esneklik özelliği kazanmış, uzun dönemli, sürdürülebilir müşteri ilişkileri kurgulanmış, sosyal medya, E-MİY gündeme gelmeye başlamıştır (Kumar ve Reinartz, 2018). MİY boyutları açısından; analitik, operasyonel ve stratejik olarak üç başlıkta değerlendirilebilmektedir.

Operasyonel MİY; müşteri ile yüz yüze, telefon, faks, e-posta gibi birebir iletişimin getirdiği süreçleri kapsamaktadır. MİY sistemi ile birçok birim kendi alanına ait müşteri bilgilerini izleyerek aksiyon almakta ve pazarlama planını

oluşturmaktadır (Demir ve Kırdar, 2007). Operasyonel MİY kendi bünyesinde; müşterilerin mal/hizmet alım döngüsü içinde satış öncesi ve pazarlama aşaması MİY, satış alanında MİY ve satış sonrası MİY olarak üç farklı alt başlıkta incelenebilmektedir. Satış öncesi ve pazarlama aşaması; kampanyalar, mevcut ilişkilerin güçlendirilmesi, satış alanı aşaması; müşteri ile mal/hizmet satışı için ilk ilişkiden satışın olumlu yada olumsuz oluşuna değin süreçteki yaşananları, satış sonrası ise; memnuniyet ölçümleri ve şikâyet yönetimleri gibi başlıkları içermektedir (Amuso, 2000).

Analitik MİY; müşterilere ait ayrıntılı verinin toplanması, depolanması (veri ambarları), işlenerek analize hazır hale getirilmesi, veri madenciliği araçları ile analiz edilmesi, tahminlenmesi ve raporlaması işlemlerini kapsamaktadır (Payne,1993). Başka bir tanımlama ile verinin temini, kayıt altına alınması, derlenmesi-işlenmesi, bölümlendirilmesi, raporlanması, analiz ve tahminlemeye olanak tanıyacak süreçlerdir (Buttle ve Maklan 2015). MİY'in entegrasyon ve operasyon özellikleri üzerine analiz ve raporlama olanakları olunca; müşteri, müşteri iletişimleri, müşterinin ticari faaliyetleri ile ilgili tüm verilerin yorumlanarak müşteri ilişkilerinin düzenlenmesi için yapılan analiz çalışmaları analitik MİY başlığı altında toplanabilmektedir.

Stratejik MİY; temelde, katma değeri olan müşterileri tutundurmaya ve benzer durumdaki yeni müşterileri sisteme dahil etmeye çalışıldığı müşteri odaklı bir yaklaşımdır. Stratejik MİY'in amacı yukarıdaki tanıma ilave olarak işletmelerin tutundurmak istedikleri müşterilerinin yaşam boyu değerinin maksimize eden pazarlama prensiplerini uygulamaktır (Iriana ve Buttle, 2007). Dolayısıyla, stratejik MİY müşterilerin farklı segmentlere ayrıldığı, müşteri değeri kavramının analizinin yapıldığı, segment-müşteri bazlı yaşam boyu değer hesaplamalarının yer bulduğu, pazarlama – satış yatırımların maliyet etkinliğinin çalışılabilirdiği, kayıp müşterilerin gelişiminin ve zaman içindeki değişiminin tespiti ve olası kayıpların ve/veya üretilecek değerlerin tahminlenebilmesinde kullanılabilen bir yöntemdir (Saputra, vd., 2019).

1.1.5. Müşteri Memnuniyeti Ölçümü ve Teknikleri

İşletmeler, özellikle karlı müşterileri ellerinde tutmak, üretecekleri yaşam boyu değerden faydalanabilmek için MİY süreçlerini etkin bir şekilde gerçekleştirmeleri gerekmektedir. MYBD için müşteri sadakatinin, müşteri sadakati için de müşteri memnuniyetinin sağlanması gerektiği yapılan birçok çalışmada görülmüştür (Genoveva, 2015; Kotler, vd., 2003; Denk, vd., 2010).

Giriş bölümünde değinildiği üzere; Beal (2014), yeni müşteri kazanmanın mevcut müşteriyi elde tutmaktan yedi kat daha maliyetli olduğunu ve memnuniyetsiz ayrılan müşterilerin %96'sının şikâyet başvurusunun olmadığını fakat deneyimlerini en az 15 kişi ile paylaştıklarını ortaya koymuştur. Genellikle, memnun ayrılan bir müşteri ise görüşlerini, önerilerini çevresindeki sadece iki-üç kişi ile paylaşmaktadır. Nilsen (2012) raporunda; müşterilerinin %92'sinin ağızdan ağıza pazarlamayı diğer pazarlama-iletişim kanallarını kullanmaya tercih ettiklerini, tanıdıklarından duydukları ya da müşteri bloglarından okumuş oldukları müşteri deneyimlerini daha güvenilir bulduklarını paylaşmıştır. Özellikle, hizmet sektöründeki müşterilerin memnuniyetsizliklerini dile getirmedeki yatkınlığı da göz önünde bulundurulduğunda, ağızdan ağıza pazarlama kanalı olan mevcut müşteriler, bireysel bağlılıklarının yanı sıra işletmenin büyümesinde etkin bir rol almaktadırlar. Sernovitz (2012) ve Ruby'nin (2015) yaptığı çalışmalar, memnuniyetsiz-sorunu çözülmemiş bir müşterinin potansiyel etkisini kaldırmak için yaklaşık 12 tane yeni müşteri kazanılması ya da 12 tane olumlu yorum alınması gerektiğini göstermektedir. Ayrıca, yapılan çalışmalar müşterilerden gelen şikâyetlerin doğru yönetilmesinin genel müşteri memnuniyetine olumlu etkisinin olduğunu göstermektedir (Uzun ve Özgöz, 2022). Bu doğrultuda işletmeler hizmet/mal sundukları müşterilerinin memnuniyetlerini ölçmek için öncelikle yüz-yüze, basılı anket, telefona kısa anket mesajı, telefona karekod, e-posta, canlı sohbet, sosyal medya yorumları gibi çeşitli veri toplama metotları geliştirmiştir. Sonrasında ilgili veriler; Müşteri Memnuniyet Puanı (CSAT), Müşteri Efor Puanı (CES) ve Net Tavsiye Skoru (NPS) gibi belirlenen metrikler doğrultusunda ölçülmeye başlanmıştır. İlgili metotlara aşağıda kısaca değinilmiştir.

Müşteri Memnuniyet Puanı (CSAT); yaklaşımında işletmeler, müşterilerinin alınan mal/hizmet sonrası farklı adımlardaki deneyimlerini ölçmek için yönelttikleri tek bir soru üzerinden “son derece memnun değil” ile “son derece memnun” arasında bir skalaya göre yanıt toplamaktadırlar. CES ve NPS'den farklı olarak CSAT puanı, bir müşterinin bir şirket ile belirli bir etkileşimden sonra yaşadığı memnuniyet derecesini ölçtüğü için mal/hizmet sunum sürecinin farklı adımlarındaki deneyimi analiz etmek için CSAT anketlerini kullanır. İşletmeler, “Bizden tekrar mal/hizmet almayı ne kadar düşünür müsünüz?” 1 pek olası değil, 5 büyük ihtimalle veya “Aldığınız hizmetten ne kadar tatmin oldunuz?” 1 son derece memnuniyetsiz, 5 son derece memnun gibi anketler ile müşteri memnuniyet oranını tanımlamaktadırlar. İlgili yöntemde memnuniyet oranı temelde, memnunum ve son derece memnunum değerlendirmesinde bulunanların toplam yanıt veren müşterilere oranı olarak hesaplanmaktadır.

Müşteri Efor Puanı (CES); müşteri efor ya da çaba puanı, alınan hizmetten duyulan memnuniyet ya da hizmet sunucusunu başkalarına önerip önermeme durumundan ziyade müşterinin problemini çözmek için ne kadar efor sarfettiğini ölçmeye yönelik bir anket sorusudur. Genellikle, “Sorununuzu çözmek için ne kadar uğraştınız? Sorununuz kurumumuzda çözmek ne kadar kolaydı? Kurum sorununuzun çözümünü benim için kolaylaştırdı” tarzında sorulur ve 1-5 skalasına tekabül eden 1 çok kolaydı, 5 çok zordu ya da Likert skalası 1-7 arasında değerlendirme yapılması istenir. Birden fazla hesaplama yöntemi olsa da temelde kolay ve çok kolay işaretleyenlerin toplam anket katılımcısı oranından, zor ve çok zor işaretleyenlerin oranının çıkarılmasıyla tespit edilmektedir. İşletme, yüksek bir CES puanına sahipse müşteriler için deneyimi kolaylaştırdığını, eğer CES puanı düşükse müşteri deneyimini iyileştirmek için açık bir alan olduğunu yorumlayabilmektedir. Çünkü, yapılan bir çalışmada düşük bir CES puanı olması durumunda, her ne kadar CES ile CSAT arası korelasyon yüksek olmasa da ilgili deneyimin genel müşteri memnuniyetine negatif etki yaptığını, NPS'yi etkilediğini göstermekte ve nihai olarak müşterinin mal/hizmet sunucusunu değiştirmesine sebebiyet verebildiği görülmektedir (Bleuel, 2019).

Net Tavsiye Skoru (NPS); Harvard İş Okulu profesörlerinden Fred Reichheld'ın 2003 yılında yazına kazandırdığı, müşterilerin hangilerinin işletmenin aleyhine, hangilerinin pasif ve hangilerinin lehine aktif rol oynayacaklarını ölçmeye yönelik olarak geliştirmiştir. İşletmemizi bir yakınınıza tavsiye eder misiniz sorusunu 0-10'luk skala üzerinden 0 en düşük, 10 en yüksek ihtimalle tavsiye ederim ölçütleri ile değerlendirmektedir. Müşterilerin cevapları eğer 0-6 ise kötüleyenler, 7-8 pasifler, 9-10 destekleyenler olarak gruplanmaktadır. NPS, destekleyenlerin yüzdesinden, kötüleyenlerin yüzdesini çıkarmakla hesaplanmaktadır. Eğer NPS skoru sıfırın altında ise işletmenin müşteri memnuniyeti üzerine ciddi çalışmalar yapması gerektiğini göstermektedir. 0-50 arasındaki sonuçlar her ne kadar sektör pazar bazlı diğer şirketlere göre kıyaslamada farklılık gösterse de iyi bir sonuç olarak, 70 üzeri ise müşterilerin aslında birer pazarlama elamanı görülebileceği olarak yorumlanabilmektedir. NPS, ölçümlemesi günümüzde şirketler için kritik bir gösterge konumundadır. İşletmeler artan sosyal medya gücünü, ağızdan ağıza iletişime olan yüksek güvenin etkisi ile NPS puanını dikkatle takip etmektedirler. Diğer yandan denetlenmeyen ve işletmelerin kendilerince bildirdikleri ve popülaritesi artan NPS'nin, bazı pazarlarda yeni müşteriye etkilemek için de paylaşıldığı, müşterilerce ulaşılabilir olmasının NPS'yi zedeleyecek şekilde yanlış yönlendirmelere ve suistimallere yol açtığı görülmüştür. Bu doğrultuda Fred Reichheld ve Bain Company'den Darnell ve Burns (2021) yazına, **kazanılan büyüme oranı** (earned growth rate) yani mevcut müşterilerin tekrar hizmet alımlarının ve refere ettiklerinin oluşturduğu gelir olarak NPS'ye ek bir parametre tanımlamışlardır (Reichheld vd, 2021). Güncel bir gelişme olan bu yaklaşımın pratik uygulamada yerini yeterince bulabilmesi için işletmelerin topladıkları veri tipini ve toplama yöntemlerini de gözden geçirmeleri gerekmektedir.

Günümüzde sosyal medyanın iş ve sosyal hayatımızdaki artan yeri ile sosyal medya metrikleri de müşteri memnuniyetini ölçme ve değerlendirmede yerini almaya başlamaktadır. Müşterilerin yorumları google alerts, mention, socialmention gibi yazılımlar aracılığı ile sosyal medya taramaları yapılabilmekte ve müşterilerimizin ve/veya potansiyel kitlenin işletmeler hakkındaki yorumları toplanabilmektedir. Benzer şekilde, işletmeler SMS hizmeti almak istemeyen

müşteriler, reklamları web üzerinden engelleyen potansiyel müşteriler hakkında veri toplamak, analiz etmek ve bilgiye dönüştürme sonrası aksiyon planlarını şekillendirme imkânı bulabilmektedirler.

Müşteri memnuniyeti üzerine yapılan çalışmalar, işletmelere, müşterilerin aldıkları mal/hizmet karşılığındaki hissiyatlarını ölçme, mal/hizmet sunumunda nerelerde iyileştirme yapılması gerektiğini anlamada ve müşterilerin yaşam boyu değerini korumak için hangi örüntüler üzerinde çalışılmasını önceliklendirme de yardımcı olmaktadır. Bir hizmet işletmesi göz önünde bulundurulduğunda, çalışanların tutum ve davranışlarının müşterilerin aldıkları hizmet algısında önemli rol aldığı, çalışanların memnuniyetlerinin ise tutum ve davranışlarını etkileyen en önemli unsurlardan olduğu bilinmektedir (Türk, 2005). Böylelikle, MİY çalışmaları ile iş süreçlerinin daha iyi anlaşılması ve ilgili sunumun müşterilerdeki yansımalarının değerlendirilmesi sağlanabilmektedir.

1.1.6. Pazarlama Karması

İşletmelerin pazarlama iş akışlarını etkin yönetilebilmesi adına geliştirilen pazarlama karması, adını dört temel bileşenden alır ve 4P olarak adlandırılır. 1964 yılında McCarthy tarafından yazına kazandırılan, İngilizce Product (Ürün), Price (Fiyat), Promotion (Tanıtım), Place (Dağıtım) baş harflerinden oluşmaktadır. Firmanın hedef pazardaki pazarlama gayelerini gerçekleştirmek için kullandığı pazarlama araçları seti ve pazardaki ilgili müşterilere nasıl davranılması gerektiğine yardımcı olan bir yaklaşım olarak tanımlanmaktadır (Peppers ve Rogers, 2016). Buradaki ürün bileşeninde; üretilmekte olan ürünler ile müşterilerin istek ve ihtiyaçları karşılanabilmekte midir sorusuna cevap aranmakta, fonksiyonellik, kalite, görünüm, marka, satış sonrası hizmet gibi unsurlar ön plana çıkmaktadır. Fiyat bileşeninde; müşterilerin satın almak istedikleri ürün/hizmet için ne kadarlık bir ödeme yapmayı düşündükleri sorusuna cevap aranmakta ve fiyatlar, indirimler, ödeme yöntem/kolaylıkları ön plana çıkmaktadır. Tanıtım bileşeninde; hedef müşteri kitlesinin işletme ve ürünleri hakkında ne kadar bilgi sahibi oldukları değerlendirilmekte, reklam, satış ve pazarlama unsurları ön plana çıkmaktadır. Dağıtım bileşeninde ise; ürünlerin

doğru zamanda, doğru yerde ve doğru miktarda müşterilere sunulup sunulmadığı sorgulanmakta, yer, lojistik ve depo yönetimi önem kazanmaktadır (Kotler ve Keller, 2005).

Gelişen ve değişen ihtiyaçlar doğrultusunda pazarlama karmasına People (İnsan), Process (Süreç) ve Physical Evidence (Fiziksel Kanıt) bileşenleri eklenmiştir (Booms ve Bitner, 1982). İnsan bileşeninde; bir şirketin çalışanları, müşterilerle etkileşim kurarken, sorularını, siparişlerini ve şikâyetlerini, şahsen, çevrimiçi sohbet yoluyla, sosyal medyada veya çağrı merkezi aracılığıyla alıp işleme koyarken ön plandadır. Yolculukları boyunca müşterilerle etkileşime girmekte ve müşteri için kuruluşun "yüzü" haline gelmektedirler. Süreç bileşeni; üretim ya da hizmetin gerçekleştiği, ödeme sistemleri, dağıtım prosedürleri, müşteri ilişkilerini yönetmek gibi tüm süreci tanımlamaktadır. Gerçek prosedürleri, mekanizmaları ve akışları barındırmaktadır. Fiziksel kanıt bileşeni; marka bilinirliğini etkileyecek, şirketin sunduğu deneyimin kalitesine dair somut ipuçları sağlayan etmenlerdir. Bir kurumdaki kıyafet kodu, dokümanları, kartvizitleri, mağazanın kullandığı dekorlar, bir kafenin koltukları, oluşturduğu atmosfer ve benzeri fiziksel durumlardır.

1.2. Müşteri Kaybı (Churn) Analizi

Müşteri kaybı, müşterinin hizmet aldığı sunucusundan vazgeçmesi ve başka bir hizmet sunucusuna geçmesi ile kaybedilme durumunu ifade etmektedir (Gürsoy, 2009). İşletmelerin temelde müşterilerini kaybetme nedenleri, alınan ürünün ya da hizmetin deneyimi, kalitesi, fiyatı, müşteri ihtiyaçlarının değişmesi ve mevcut kurum tarafınca sunulamaması, rekabet ortamındaki promosyonlar vb. olduğu saptanmıştır (Blattberg, vd., 2008). Müşteri kaybı ise işletmeler için, gelir kaybı, müşterinin yaşam boyu değerinden yararlanamama, ağızdan ağıza pazarlama fırsatlarını kaybetme ve negatif pazarlama sonucu yeni müşteri arayışına girme ve daha fazla maliyete katlanma gibi ciddi sonuçlar getirmektedir (Kotler, 2000).

Deneyimlerinden algıladıkları beklentilerini karşılayamayan müşterilerde memnuniyetsizlikler ve buna bağlı olarak şikâyetler başlamaktadır. Müşteri şikâyetleri giderilmediğinde ise memnuniyetsizlik tatminsizliğe ve kayba

dönüşmektedir. Şikâyetleri giderilen müşterinin ise bağlılığı yenilemekte ve tekrar satın alma gerçekleşmektedir. Genellikle arzın yoğun olduğu hizmet sektörlerindeki müşteriler, ellerindeki seçenekler doğrultusunda kolaylıkla hizmet sunucusu değişikliğine gidebilmektedir. Hizmet sunucusundan ayrılan müşteri, kaybedilen müşteri (churn) grubuna dahil edilmekte ve bu durum alan yazınında “Churn oldu” olarak ifade edilmektedir. MKA ise müşterilerin hizmet sunucusundan ayrılış nedenlerini inceleyen analitik ve sistematik bir yaklaşımdır. Bu analizle; birçok veri madenciliği ve yapay zekâ aracından yararlanılarak potansiyel müşteri kayıplarının tahmini yapılmaya çalışılmaktadır. Bu analizi yapan bir işletmenin, kendisini terk etmiş müşterilerini ve terk etme nedenlerini saptama, bu müşteri tiplerinin özniteliklerini belirleme, terk etmesi olası müşterilerin kimler olduğunu tahmin etme, onları takip etme ve kendilerinin yapacakları aksiyonları belirleme gayreti, kısaca mevcut müşterilerini elinde tutabilme gayreti içinde olduğu söylenebilmektedir (Çiçek ve Arslan, 2020).

Mevcut müşterilerin hizmet alımlarının devam etmesi, yeni müşteri bulma maliyetleri açısından kritiktir. Bu müşterilerin bugün oluşturduğu kar ile birlikte ilerideki karları ve büyümeye etkileri de beraber göz önüne alınmalı ve MYBD'nin değerlendirilmesi gerekmektedir. MKA'nın finansal etkisinin değerlendirilmesinde hem kaybedilen (churn olan) müşterinin ürettiği karı hem de kaybedilen müşterinin etkisini giderebilmek için aynı segmentteki yeni müşteri bulmanın ve sisteme dahil etmenin maliyetinin göz önünde bulundurulması gerekmektedir. Bu nedenle, mevcut müşterilerin tercih ve eğilimleri yakinen izlenmelidir.

MKA ile işletmeler ellerindeki dağınık veriyi anlamlı hale getirmeye ve geçmişteki müşteri eğilimleri ile karşılaştırarak, müşteriler için kritik hizmet sunucusu değiştirme noktalarını saptamaya çalışmaktadırlar. İşletmeler, bu çalışmaların sonucunda da müşterilerinin hizmet alım yaşam döngüleri belirlemeye çalışmaktadırlar. Müşteri geri bildirimleri, şikâyetler, teknik destek geri dönüşleri, müşteri memnuniyet anketleri, fatura itirazları ve anlaşmazlıkları gibi durumlar tahmin modelleri için önemli veriler içermektedir (Şeker, 2016). Sağlık sektöründe ise hastalardan gelen şikâyetlerin ve uygulanan anketlerin analiz sonuçları çalışma için önemli destek sağlamaktadır.

1.2.1. Müşteri Kayıp Çeşitleri

Müşteri kayıp çeşitleri gönüllü kayıp (voluntary churn) ve gönülsüz kayıp (involuntary churn) olarak iki farklı başlık altında değerlendirilebilmektedir.

Gönüllü kayıp; müşterilerin kendi istekleri ile mevcut hizmet sunucusundan vazgeçerek farklı işletmelerin hizmetlerinden yararlanmayı tercih etmesi durumudur. Başka bir ifade ile rekabet nedeni ile müşterinin firmayı tercih etmeyi bırakması olarak da tanımlanabilmektedir (Nettleton, 2014). Gönüllü kayıp, MKA'ya konu olan gruptur. Örneğin; abonelik tarzındaki hizmet alım sürekliliğine dayalı bir telekomünikasyon müşterisinin hattını başka bir sunucuya taşıması, bir banka müşterisinin başka bir bankadan hizmet almaya başlaması ya da bir internet kullanıcısının başka bir sunucuya geçmesi durumunda müşteri kaybı gönüllü olarak yaşanmıştır denilebilmektedir. Bu durum genelde, müşterinin kendisi için daha avantajlı olduğuna inandığı ya da mevcut sunucunun hizmetinden memnuniyetsizlik duyduğu şartlarda gerçekleşmektedir.

Gönülsüz kayıp; müşterinin kendi tercihlerinin dışında gelişen ve genelde çevresel şartlardan etkilenen kayıp durumudur. Gönülsüz kayıp müşterinin hizmet sunucusundan yararlanma şansını ortadan kalktığı durumları tanımlamaktadır (Şeker, 2016). Müşterinin farklı bir ülkeye taşınması, sağlık problemleri, vefatı vb. durumlar örnek gösterilebilmektedir. Bu grup, kontrol edilebilir ya da önlenemez bir kayıp olmadığı için MKA'nın odağında olan bir çalışma alanı değildir. Bu tarz kayıpların göz ardı edilmesinin sebebi genelde müşteri kaybının önlenemez olmasıdır. Modellemede yanıltıcı çıktı üreteceği için model dışında bırakılmaktadır. Çoğu MKA çalışmalarında, gönülsüz kayıplar kolaylıkla saptanamadığı için genellikle zaman serisi şeklinde gönülsüz kayıplar tahmin edilmeye ve istatistiksel olarak modele eklenilmeye çalışılmaktadır (Şeker, 2016).

Kayıp müşteri oranı- kayıp analizi hesaplaması

Kayıp müşteri oranı (churn oranı), belirli bir zaman dilimi içerisinde hizmet alımını durduran müşteri oranını tanımlamaktadır. Kayıp müşteri oranı- Kayıp analizi hesaplaması aşağıda verilmiştir;

$$\text{Dönemlik Kayıp Oranı (\%)} = \frac{(\text{Dönem Başındaki Müşteri Sayısı} - \text{Dönem Sonundaki Müşteri Sayısı})}{\text{Dönem Başındaki Müşteri Sayısı}} \quad (1.1)$$

Bu oran bir örnek ile açıklanırsa; bir telekomünikasyon sunucusunun dönem (aylık) başındaki aktif müşteri sayısı 10 milyon ve dönem sonundaki müşteri sayısı 9.5 milyon olsun. Bu firmanın aylık müşteri kaybı oranı;

$$\text{Dönemlik Kayıp Oranı (\%)} = \frac{(10.000.000 - 9.500.000)}{10.000.000} = 5\% \quad (1.2)$$

olarak hesaplanmaktadır. Bu firmanın aylık 5%'lik müşteri kaybı yaşadığını gösterir. Kaybedilen müşterilerden doğan gelir kaybı hesaplanmak istendiğinde; örneğin, sunucunun Ocak ayı başındaki aylık tekrarlı gelir (her seferinde satış yaparak kazanılanlar değil aylık abonmanlık vs gibi her ay tekrarı olması beklenen gelirler) 10.000.000 pb, Mart ayı sonunda ise 8.500.000 pb olsun. Bu durumda; sunucunun dönemsel (çeyrek bazlı) gelir kaybı;

$$\text{Dönemlik Kayıp Oranı (\%)} = \frac{(10.000.000 - 8.500.000)}{10.000.000} = 15\% \quad (1.3)$$

olarak hesaplanır. Diğer yandan burada göz önünde bulundurulması gereken bir nokta hizmet alımına devam eden müşterilerin ek gelir üretip üretmediğidir. Eğer mevcut müşteriler ek paket alımları ile 1.250.000 pb lik ek bir gelir üretirse bu durumda firmanın dönemsel kaybı;

$$\text{Dönemlik Kayıp Oranı (\%)} = \frac{(10.000.000 - 8.500.000 + 1.250.000)}{10.000.000} = 2,5\% \quad (1.4)$$

olarak hesaplanacaktır. Bu durumlar bir telekomünikasyon firmasında ek paket alımları ya da satış fiyatlarına yapılan zam gibi değerlendirilebilir. Bir başka açı ise; eğer bu ek paket kullanımlarından bir satış maliyeti oluştu ise ilgili gider yukarıdaki gelir kaybına da eklenmelidir.

1.2.2. Müşteri Kayıp Tahmini

Yeni müşteri bulma ve mevcut müşteriye elde tutma arasındaki finansal farklılık, mutsuz ayrılan müşterinin deneyimlerini birçok mecrada mutlu olan müşteriye nazaran çok daha fazla kişi ile paylaşması, MKA'yı işletmenin sürekliliği açısından kritik bir çalışma haline getirmiştir. Yeni müşteri kazanmanın yüksek maliyeti, işletmeler için olası kayıpların önceden doğru olarak tahmin edilmesi gereğini ön plana çıkarmıştır. Bu çalışmanın yapılabilmesi için ise hangi müşterinin devamlı müşteri olduğu ve hangi müşterinin hangi durumda kayıp müşteri olduğu etmenlerinin çok iyi değerlendirilip belirlenmesi gerekmektedir.

Büyük müşteri portföyü olan ve dijitalleşme aksiyonlarını tamamlayan işletmeler, kayıpları önceden tahmin edebilmek amacı ile veri madenciliğine dayanan uygulamalar geliştirmektedir. Büyük veri içinden devamlı müşteri olarak tanımlanan ve belirli bir dönem içerisinde kayıp müşteri olarak kabul edilen müşterilerin segmentler halinde ve zaman serileri ile kayıp müşteri olma örüntüleri incelendiğinde, gelecekte hangi müşterinin hangi durumlarda kayıp müşteri olma riski barındırdığı, veri madenciliği modelleri aracılığı ile tahmin edilebilmektedir. Veri madenciliği kapsamında geliştirilen ilgili modeller; sınıflandırma ve regresyon, kümeleme ve birliktelik kurallarınca üç farklı kategoride değerlendirilmektedir (Özkan, 2013).

Sınıflandırma ve regresyon, önceden tanımlanmış etiketleri, öğelerin örneklerine atamak için kullanılan denetimli bir öğrenme tekniğidir. Örneğin, bir bankada müşteri veri tabanında yüksek gelirli müşteriler, yüksek karlı müşteriler, düşük karlı müşteriler, banka hizmetlerini aktif kullananlar vb. tanımlamalarda kullanılmaktadır. Sınıflandırmada en yaygın olan kullanılan teknikler; Karar Ağaçları (Decision Trees), Yapay Sinir Ağları (YSA, Artificial Neural Networks),

Rassal Orman (RO, Random Forest), Destek Vektör Makinaleri (DVM, Support Vector Machines), Lojistik Regresyon (LR, Logistic Regression), K-En Yakın Komşular (KNN, K-Nearest Neighbors), Bayes Ağları ve Genetik Algoritmalarıdır (Han, Kamber ve Pei, 2012).

Kümeleme, benzer örnekleri özelliklerine göre gruplandırmak için kullanılan, gruplar arasında örüntü bulunmasını sağlayan denetlenmeyen bir analiz tekniğidir. Bu modellerde genellikle, Kohonen Ağları, K-ortalama (K-means) yöntemi ve İki Adımlı (Two Step) Kümeleme teknikleri kullanılmaktadır.

Birliktelik Kuralları, büyük veri kümeleri arasındaki aynı zamanda gerçekleştirilen birliktelikleri saptamak için kullanılmaktadır. Bu çalışmalarda, örneğin bir giyim alışveriş merkezindeki ya da süpermarketteki müşterinin aynı zaman diliminde satın aldıkları arasındaki örüntüyü saptama, Pazar/Market Sepeti Analizi ve müşterinin satın alma davranış paternini anlamlandırmak amaçlanmaktadır. Bu modellerde; AIS, GRI, CARMA ve Apriori şeklinde kural belirleme teknikleri yer bulmaktadır (Ersöz, 2015). MKA'da yazınında ve pratik uygulamada sıklıkla kullanılan bu teknikler takip eden bölümde paylaşılmıştır.

Yapılan tahminler sonrasında, kayıp müşteri olma riski barındıran müşteriler için işletmeler indirim, hizmet sunumu ya da kapsam değişimi gibi tutundurma faaliyetleri önem kazanmakta ve kayıp müşteri olma durumunu önleyici faaliyetlerine ağırlık vermektedirler.

2. BÖLÜM

VERİ MADENCİLİĞİ KAVRAM VE YÖNTEMLERİ

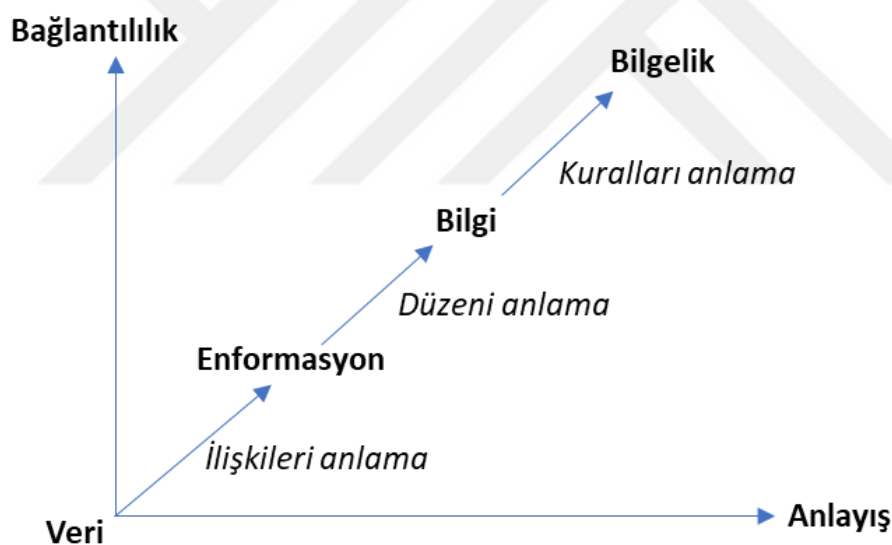
Bu bölümde, bu çalışmanın konusu olan MKA'da kullanılan Veri Tabanlarında Bilgi Keşfi (VTBK) ve Veri Madenciliği (VM) süreci teorik çerçevede incelenecektir. VTBK ve VM genel olarak ele alınırken, ilerki bölümlerde gerçekleştirilecek olan sağlık sektöründe MKA uygulamasında kullanılacak VM yöntem ve algoritmaları ise ayrıntılı olarak incelenecektir.

2.1. Veri, Enformasyon, Bilgi ve Bilgelik Kavramları

Yazında sıkça karşımıza çıkan ve bazen birbirinin yerine kullanılan veri, enformasyon, bilgi ve bilgelik kavramlarına açıklık getirmek, aralarındaki hiyerarşiyi incelemek önemlidir. Milan Zeleny (1987), makalesinde bu bilgi yönetimi hiyerarşisini irdlemiştir. Çalışmasında; veri, enformasyon, bilgi ve bilgelik kavramlarını know-nothing, know-what, know-how, know why, Türkçe sırası ile hiçbir şey bilmemek, ne olduğunu bilmek, nasıl olduğunu bilmek ve neden olduğunu bilmek olarak sınıflandırmıştır (Zeleny, 1987). Öncü örgütsel dönüşüm kuramcılarında Russell ve Ackoff (1989) da, ilgili bilgi yönetimi hiyerarşisine anlamak kısmını eklemiş ve bu hiyerarşiyi veri, enformasyon, bilgi, anlamak ve bilgelik şeklinde sınıflandırmıştır. Buradaki ilk dört aşama geçmiş ile, beşinci aşama ise görüş içerdiği için gelecek ile ilişkilidir.

Veri; sözlük anlamı ile olgu, kavram veya komutların, iletişim, yorum ve işlem için elverişli biçimde gösterimi olarak TDK' da yer bulmuştur. Veri, kendi başına anlam ifade etmeyen, ham, işlenmemiş gözlemlere ait nitelikleri, olayları ve ilişkili çevreleri tanımlayan sembollere, sorunu çözmemize ya da karar vermemize yardımcı olan sayı, renk, durum vb. her türlü olgu için yapılan tanımlamadır (Koçdar, 2018). Enformasyon (Information); kökeninde Latince ve İngilizce; taslak, görüş, düşünce ve haber verme anlamlarına gelmekte, Türkçe sözlük anlamı ile ise danışma, tanıtma, malumat, haber alma ve haber verme gibi anlamlar taşımaktadır. Bilişim alanında ise kurallardan yola çıkılarak veriye yöneltilen anlam, verinin organize edilmesi ve sınıflandırılması doğrultusunda

Bilgi hiyerarşisi, İngilizce’de data, information, knowledge, wisdom olarak geçen, Türkçe de sırası ile veri, enformasyon, bilgi, bilgelik olarak yer bulan verinin bilgeliğe olan dönüşümü aşağıdaki Şekil 2.1.’de gösterilmiştir (Bruyckere, 2018).



Kaynak: Bruyckere , <https://theeconomyofmeaning.com/2018/06/29/the-difference-between-data-information-knowledge-and-wisdom/>

21

2.2. Veri Tabanları

Veri tabanları; birbirleri ile bağlantısı olan verilerin dijital ortamda depolandığı, saklandığı alanlardır. Veri tabanları barındırdığı büyük çaptaki veriyi yapısına uygun biçimlerde kaydetmeye, güncellemeye ve bu veri üzerinde gelişmiş sorgulamalar yapmaya olanak sağlayan yapılardır. Bir veri tabanı, genellikle bir bilgisayar sisteminde elektronik olarak depolanan yapılandırılmış bilgi veya veriden oluşan düzenli bir koleksiyondur (Zygiaris, 2018).

Veri tabanları, genellikle bir Veri Tabanı Yönetim Sistemi (Database Management System) ile kontrol edilmektedir. Veri, Veri Tabanı Yönetim Sistemi (VTYS) ve aynı zamanda bunlarla ilişkili uygulama yazılımları bir araya getirildiğinde sıklıkla, yalnızca veri tabanı olarak kısaltılan Veri Tabanı Sistemi (VTS) olarak ifade edilmektedir (Zygiaris, 2018). Toplanan ilişkili çok sayıdaki veri, veri tabanlarında farklı tablo grupları halinde depolanmaktadır. İlgili tabloların mantıksal bir ilişki içerisinde tutulduğu veri tabanlarına ilişkisel veri tabanı denilmektedir. İlişkisel veri tabanı oluşturulması ve yönetilmesi için İlişkisel Veri Tabanı Yönetim Sistemlerine (Relational Database Management System) ihtiyaç duyulmaktadır. Bu veri tabanlarında İngilizcesi ile Structured Query Language (SQL), Türkçesi Yapısal Sorgu Dili kullanılmaktadır. SQL aracılığı ile veri tabanı üzerinde araştırma amaçlı sorgulamalar yapılabilmekte, kayıtlardan listeler oluşturulabilmekte, veri eklenip değiştirilebilmektedir.

Veri tabanındaki bu çok sayıda veri, İngilizce yazında Big Data olarak geçen Büyük Veri'den ayrılmaktadır. Veri tabanındaki veri, sisteme dahil etmek isteyerek işlediğimiz veriyi temsil ederken, Büyük Veri (BV) ise ilişkisel veri tabanlarında tutulan yapısal verinin dışında kalan, son dönemlere dek çok da kullanılmayan, yapısal olmayan veri yığınlarını temsil etmektedir. Yapılacak farklı analizler ile buradaki değişik veri yığınının önemli enformasyon ve bilgi üretilebilmektedir.

2.3. Veri Ambarları

İngilizce yazında Data Warehouse olarak geçen ve 1991 yılında William H. Inmon tarafından yazına kazandırılan Veri Ambarı (VA) kavramı; analiz amaçlı sorgulamalar yapmak için birçok farklı veri tabanından alınarak birleştirilen verinin toplandığı özellikli depolardır. Günümüzde yaygın olarak kullanılmaya başlanan VA'ları günlük kullanılan veri tabanlarının birleştirilmiş ve işlemeye daha uygun bir özetini saklamayı amaçlamaktadır (Türker, vd., 2020). İlişkisel veri tabanları, olaylar ve işlemler ile ilgili verileri saklar ve sürekli bir veri giriş çıkışı gerektirirler. Dolayısıyla, güncel veriyi içerirler. Veri Ambarları ise ilgili veri tabanlarındaki veri ile diğer dış kaynaklardan toplanan veriyi belirli aralıklarla derleyip arşivlemektedir. Bu doğrultuda işletmeler için önem arz eden dönemsel analizlerin yapılmasına olanak sağlamaktadır. Veri Ambarları değerlendirilirken, analiz-sorgu kısmında OLAP (On-line Analytical Processing - Çevrimiçi Analitik İşleme) ve OLTP (On-Line Transaction Processing - Çevrimiçi İşlem Gerçekleştirme) kavramları uygulamada yer almaktadır.

OLTP, çevrim içi çalışan, veri değişimi işlemlerinin çoğunlukta olduğu, sürekli ve anlık değişen operasyonel veri tabanlarındaki işlemleri kapsamaktadır. OLAP, 1993 yılında Dr. E.F. Codd tarafından geliştirilmiş ve veriler arasında fark edilemeyen ilişkileri bulup çıkartma, aynı veriyi çok boyutlu inceleme, veri ilişkilerini daha iyi görüntülemek için grafiksel sunum ve trend analizlerini gösterebilmektedir. Ayrıca, bazı fonksiyonları ile anlık karar destek mekanizmalarını güçlendirmektedir. Veri tabanlarında yapılan sorgular, günlük veya haftalık satışlar, cinsiyete göre en az / en çok satılan ürünler şeklindedir. OLAP ise; ilişkisel veri tabanlarını, rapor yazmayı ve VM'ni de kapsayan, daha ayrıntılı ve tahminlemede kullanabilen bir iş zekâsı türüdür.

2.4. Veri Madenciliği

İşletmelerin işletim sistemlerinde toplanan veri çoğaldıkça, işletmeler için kaydedilen verinin işlenmesi, ellerindeki verinin anlamlandırılması önem kazanmış ve karar almada yararlanılması için İngilizce'de "Data Mining" olarak geçen Veri Madenciliği kavramı ortaya çıkmıştır. Veri Madenciliği büyük hacimli,

belirgin olmayan verilerden, önceden bilinmeyen fakat potansiyel olarak kullanışlı bilgi ve örüntülerin çıkarılması olarak tanımlanmıştır (Türker vd., 2020). Bir diğer tanımlamada VM; önceden bilinmeyen, veri içinde gizli, anlamlı ve yararlı örüntülerin büyük ölçekli veri tabanlarından otomatik biçimde elde edilmesini sağlayan veri tabanlarında bilgi keşfi süreci içindeki bir adım olarak tanımlanmıştır (Provost ve Fawcett, 2013). Han, Kamber ve Pei (2012) göre VM; büyük miktarlardaki veri arasından, geleceği tahminlemeye katkı sağlayacak, anlamlı ve faydalı örüntü ve kuralların bilgisayar yazılım programları yardımı ile analizi olarak tanımlanmıştır (Han vd, 2012). Capri (2015) ise, VM farklı dağınık veri türlerinden çeşitli algoritmik uygulamalar ile kullanışlı bilgi üreten yaklaşım olarak tanımlanmıştır.

2.4.1. VM Tarihsel Gelişimi

VM'nin kökeni 1946 yılında tanıtılan ilk sayısal bilgisayara (ENIAC - Electrical Numerical Integrator And Calculator) kadar dayanmaktadır. 1950'li yıllarda basit düzeydeki sayma, aritmetik işlemler gibi konularda çalışabilen ilk bilgisayarlar geliştirilmiştir. Bilgisayarların yaygın kullanımda yer bulması ve veri toplanmaya başlanması ile 1960'lı yıllarda veri tabanı ve verilerin depolanması gibi kavramlar ortaya çıkmaya başlamıştır. 1960'lı yılların sonlarına doğru veri tabanları, veri depolamanın yanı sıra artan işlem yapma kabiliyetleri doğrultusunda da kullanılmaya başlanmış, bilgisayar tabanlı basit öğrenme algoritmaları kullanılmaya başlanmıştır. Minsky ve Papert (1969), yaptıkları çalışmalarda, günümüzde yapay sinir ağları olarak bilinen perseptronların basit düzeydeki kuralları öğrenebileceklerini göstermişlerdir.

1970'li yıllara gelindiğinde bilişim dünyası Codd (1970)'un, ilişkisel veri tabanları üzerine yayınlanan makalesi ile ilişkisel veri tabanları ve İlişkisel Veri Tabanı Yönetim Sistemleri (İVTYS) ile tanışmıştır. İlgili dönemde; yapısal bir sorgu dili olan SQL, Sistem R ve Sybase gibi sorgulama dilleri geliştirilmiş, 1980'li yıllarda SQL standart dil olarak kabul görmüştür. Bu yıllarda VTYS'lerin işletmelerin gelişen ihtiyaçlarında kullanımı yaygınlaşmıştır. 1990'lı yıllara gelindiğinde Visual Basic, Excel, Access gibi programlar iş dünyasının hizmetine sunulmuştur.

Verinin katlanarak artması sonucu faydalı bilginin mevcut sorgu ve raporlama sistemleri ile elde edilemeyeceği görülmüş ve 1989 yılında Gregory Piatetsky-Shapiro'nun önderliğinde ilk Uluslararası Veri Tabanlarında Bilgi Keşfi (Knowledge Discovery in Databases) çalışmayı gerçekleştirilmiştir. 1991 yılında VTBK çalıştayının sonuç bildirgesi sayılabilecek "Knowledge Discovery in Real Databases: A Report on the IJCAI-89 Workshop" makalesi VTBK ile ilgili temel tanım ve kavramları ortaya koymuştur. Takibinde, VTBK için ilk VM yazılımı 1992 yılında gerçekleştirilmiştir. 1990'lı yıllarda web ortamının hayata girmesi ile birlikte 2000 yılı ve sonrasında VM ciddi bir şekilde tüm alanlarda kullanımı yaygınlaşmıştır. Fortune 500 şirketlerinin ve devlet mercilerinin alana yaptıkları büyük yatırımlar, ilgili alana olan ilgiyi daha da arttırmıştır (Dolgun, 2014).

2.4.2. Veri Madenciliği Uygulama Alanları

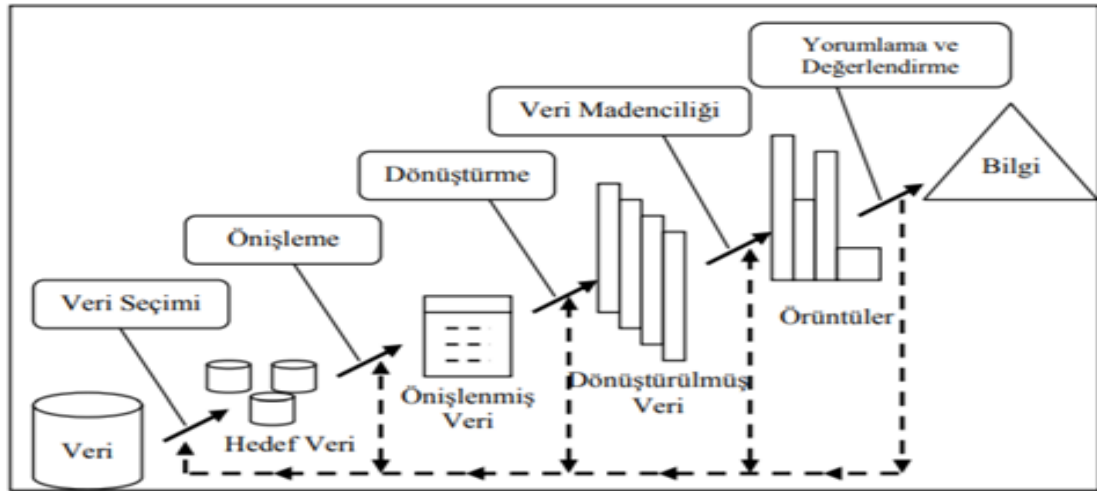
VM çalışmaları, büyük veri setlerinde veri kayıtları arası ilişkisel–nedensel bağların belirlenmesine, anlamlandırılmasına ve karar verme sürecine ihtiyaç duyulduğu tüm alanlarda kullanılabilmektedir. Veriden bilgiye olan dönüşümde yer alan VM, pazarlama, finans, sağlık, bankacılık, MİY, eğitim gibi iş sahalarında kullanılabilmektedir. Pazarlama alanında; müşterilerin satın alma alışkanlıklarının belirlenmesi, Pazar Sepet Analizi (hangi ürünleri birlikte aldığı), pazar araştırması ve satışlara–müşterilere yönelik tahminlemeler yapılabilmektedir. Örneğin; bankacılıkta kredi kartı kullanım alışkanlıklarına göre müşterilerin sınıflandırılmasında, sigortacılıkta sigorta risk gruplarının saptanmasında ve tazminat prim oranlarının tahminlemesinde kullanılabilmektedir.

MİY kapsamında; VM müşteri sadakatini artırılması, etkin pazarlama ve müşteri tutundurulması faaliyetlerinde sıkça yer almaktadır. Sağlık alanında da laboratuvar testleri ve hastalık riskleri, kalp verileri üzerinden kalp krizi riskleri, genetik yatkınlıktan kaynaklı kanser riskleri gibi VM uygulamalarından sıklıkla yararlanılmaktadır. Makine öğrenmesi, yapay zekâ gelişimleri ile iyice güçlenen tele-tıp, tele-monitörizasyon VM çalışmaları ile risk altındaki hastalara anlık gönderi ile destek sağlayabilmektedir. Ayrıca, veri- karar verme sürecinin olduğu

her sektörde veri tabanı analizi ve karar verme desteği, yönetim karar destek sistemleri VM aracılığı ile yapılabilmektedir (Lezki, 2019).

2.4.3. Veri Tabalarında Bilgi Keşfi ve Veri Madenciliği Süreci

VTBK; veriden faydalı bilginin, anlamlı örüntülerin keşfedilmesi sürecinin tamamını ifade ederken, VM bu sürecin bir adımına karşılık gelmektedir. Standard bir VM süreci; verinin gürültülü ve tutarsız veriden temizlenmesi, verinin birçok veri kaynağını birleştirerek bütünleştirilmesi, analize konu olacak verinin seçilmesi ve bir VM tekniğinin kullanabilecek hale dönüştürülmesi, kayıtlı ve seçilen veriden örüntülerin çıkarılması için VM yöntemlerinin uygulanması, örüntülerin değerlendirilmesi ve kullanıcıya bilginin sunulması aşamalarını kapsamaktadır (Clifton, 2022). Şekil 2.2.'de VTBK süreci; verinin seçimi, ön işleme, dönüştürülmesi, örüntülerin VM aracılığı ile çıkarılması ve sonuçları yorumlama- değerlendirme ile bilgiye dönüştürülmesi süreci olarak verilmiştir (Fayyad vd., 1946, 41).



Şekil 2.2: Veri Madenciliği Bilgi Keşfi Süreci

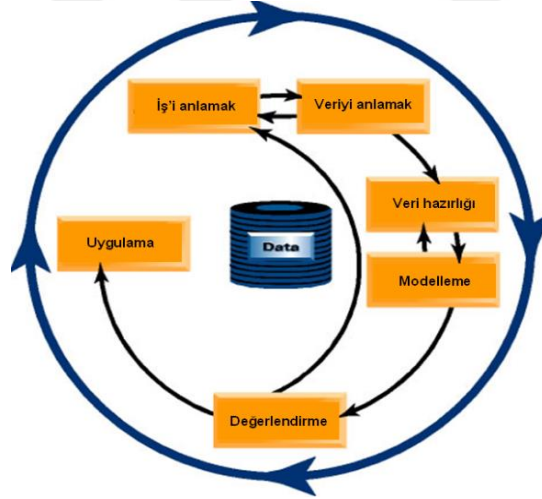
Kaynak: Fayyad, vd., 1996.

VTBK sürecinin, VM'nde yaygın kullanılan CRISP-DM (Cross Industry Standard Process Model for Data Mining) ve SEMMA (Sample, Explore, Modify, Model and Assess) olarak iki süreç modeli bulunmaktadır.

2.4.3.1. CRISP-DM

CRISP – DM, VM sürecinin standartlaştırılabilmesi için 1996 yılında tasarlanmış, 1997 yılında ESPRIT (European Strategic Programme on Research in Information Technology) tarafından Avrupa Birliği projesine haline gelmiş, Daimler, Ingetral Solutions gibi öncü firmaların da desteği ile 1999 yılında ilk metodolojisi hazırlanmıştır. 2008 yılında ise CRISP-DM, SIG (Special Interest Group) 2.0 olarak güncellenmiştir (Plumed vd., 2019).

Endüstriler Arası Standart İşleme –VM olarak, Türkçe yazında yer bulan CRISP-DM; iş problemlerinin veri tabanlı çözümler ile nasıl çözümlendiğini anlatmak ve iş uygulamalarının verimliliğini arttırmak amacı ile kullanılan standartlaştırılmış bir metodoloji ve süreç modeli olarak tanımlanmaktadır. CRISP-DM süreç modeli temelde; işin anlaşılması (business understanding), verinin anlaşılması (data understanding), verinin hazırlanması (data preperation), modelleme (modelling), değerlendirme (evaluation) ve uygulama-yayma (deployment) aşamalarından oluşmaktadır (Şekil 2.3).



Şekil 2.3. : CRISP-DM Modeli

Kaynak: Chapman vd., 2000.

İlk aşama olan **işin anlaşılmasında**; yapılacak veri madenciliği çalışmasının neye katkı sağlamanın amaçlandığı, proje hedefinin ne olduğu, problemin

tanımı ve ön planı belirlenmektedir. Bir banka örneği verilirse; problem tanımı ile eldeki geçmiş veri üzerinden potansiyel Müşteri Kayıp Analizi yapmak mı, pazarlama faaliyetlerine yön verebilmek için benzer davranış gösteren ya da benzer öz nitelikteki müşterilerini sınıflandırmak mı gibi sorulara netlik kazandırılmaktadır. İşin amacı belirlendikten sonra mevcut durum değerlendirmesi, donanım, kayıtlı veri ve yazılım gibi mevcut kaynaklar ve riskler değerlendirilmektedir. Sonrasında işin anlaşılması aşamasında VM hedefleri belirlenmekte ve proje planı geliştirilmektedir (Olson ve Delen, 2008).

İkinci aşama olan **verinin anlaşılmasında**; proje kapsamında kullanılacak verinin kalite, sayı, ulaşılabilirlik, güvenilirlik, geçerlilik gibi açılardan hangi durumda olduğu değerlendirilmektedir. Temelde bu süreç, veriyi tanımak, veri kalitesini ve sorunları belirleme adımlarından oluşmaktadır. Verinin tanınması aşamasında; eldeki verilerin dağılımı, ortalama değerleri veya varsa veriler arasındaki korelasyon hakkında saptamalar yapılmakta ve veri kalitesinin doğrulanması aşamasında eldeki verinin yeterli olup olmadığı, olası farklılaşan durumları kapsama durumu, eksik ve yanlış verilerin miktarı saptanmaktadır. Bu aşamada farklı veri tabanlarındaki veriler arasında örüntü aranacaksa, olası entegrasyon sorunlarının giderilmesi gerekebilir.

VM sürecinde en fazla zaman alan aşamalardan biri olan **verinin hazırlanmasında**; önceki aşamada verilerde saptanan eksik veriler, hatalar düzeltilir ve veri model oluşturma aşaması için hazırlanır. Bu aşama; verinin işlenmesi, sorun teşkil edebilecek eksik verilerin tahminlenerek giderilmesi, verinin ortak biçimde yapılandırılarak dönüştürülmesi, entegrasyonu, modellemelere uygun şekle getirilmesi adımlarından oluşmaktadır.

Modelleme aşaması; VM'nin doğrudan dahil olduğu kısımdır. Bu aşamada tanımlanan problem ve veri kaynakları üzerinde yapılmak istenen işe en uygun makine öğrenmesi modeli veya istatistiksel model geliştirilir. Modelleme aşamasında eldeki veri, mevcut kaynaklar ve proje hedefi doğrultusunda sınıflandırma, kümeleme, birliktelik kuralları ile mi çalışılacağı, eğer sınıflandırma çalışılacaksa karar ağaçları mı, yapısal sinir ağları mı kullanılacağı, karar ağaçları ise CHAID, CART, C5.0 gibi algoritmalarından hangisinin kullanılacağı

belirlenmektedir. Sonrasında, modelin kalitesinin, başarısının hangi testler ile doğrulanacağı, karşılaştırma yapılacaksa hangi algoritmaların VM açısından daha başarılı sonuçlar vereceği vb. incelenmekte ve hedefe en uygun olanı seçilmektedir (Şeker, 2018).

Modelleme evresindeki sonuçların VM açısından değerlendirilmesi sonrası CRISP-DM **değerlendirme aşamasında**, çıkan sonuçlar işin anlaşılması aşamasında belirlenen iş amaçları doğrultusunda incelenmektedir. Buradaki çıktılar, iş sahipleri yani yapılan çalışmayı kullanacaklar tarafından iş hedeflerini karşılama durumuna göre değerlendirilmekte, başlangıçta belirlenen başarı kriterlerince test edilmekte ve hedeflerin karşılanması durumunda, yapılacak son kontroller sonrası model gerçek koşullarda kullanılmak üzere uygulama aşamasına geçilmektedir.

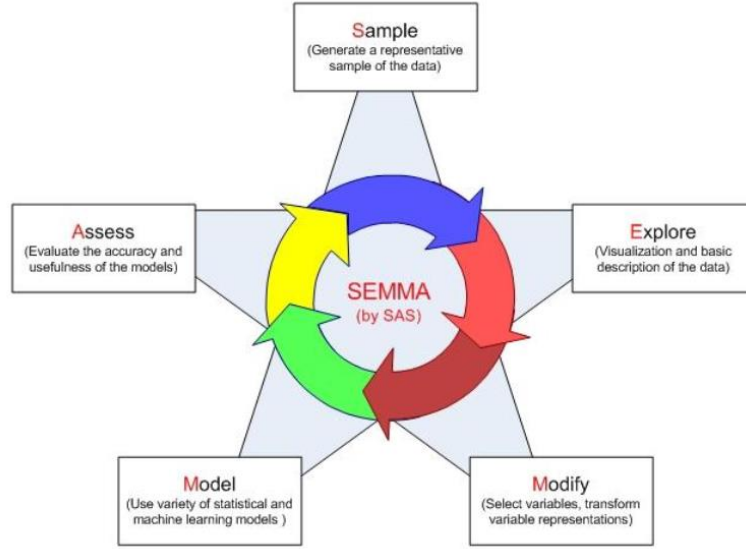
Modelin son aşaması olan **uygulama-yayma**; problemin çözümüne yönelik oluşturulan modelin performansının önceden belirlenen başarı kriterlerini sağlaması ve gerçek hayata uygunluğunun testi-değerlendirilmesi sonrası modelin canlıya alınması, günlük hayatta kullanılmaya başlaması aşamasıdır. Örneğin, bir sigorta şirketinde hangi müşterilerin hizmet sunucusu değiştirme ihtimalinin yüksek olduğunun belirlenmesi neticesinde, saptanan potansiyel kayıp müşterilere promosyon, satış sonrası servis, ek ürün vb. çeşitli tutundurma faaliyetleri uygulanabilmektedir.

2.4.3.2. SEMMA

Sample(Örnekleme), Explore(Keşfetme), Modify(Değiştirme), Model(Modelleme) ve Assess(Değer Saptama- Değerlendirme) kelimelerinin baş harflerinden oluşan, keşifsel istatistiksel ve görselleştirme tekniklerini uygulayan, önemli tahmin edilen değişkenleri, seçmeyi, dönüştürmeyi, sonuçları ortaya çıkarmak için değişkenleri kullanarak bir model oluşturan ve doğruluğunu kontrol etmeyi kolaylaştıran oldukça yinelemeli bir süreç yönetim modelidir.

Kronolojik olarak CRISP- DM'den önce ilk çıkan yöntem olan ve SAS Institute tarafından geliştirilen süreç modeli SEMMA, süreç adımları ile CRISP-DM ile

oldukça benzerlik göstermektedir. Aslında SEMMA yönteminin, bir anlamda CRISP-DM için geçiş süreci olduğu ve CRISP-DM'e dönüştüğü ifade edilebilmektedir. SEMMA; veri işlemenin kısıtlı imkanlar ile yapıldığı, verinin farklı amaçlara yönelik olarak bir seferde işlenmesi amacı ile geliştirilen bir geçiş modeli olarak görülmektedir (Şeker, 2018). CRISP-DM'den farkı, CRISP-DM tüm projenin metodolojisi iken SEMMA VM yapılan kısmın metodolojisidir.



Şekil 2.4.: SEMMA Modeli

Kaynak: Aliyeva, 2023.

Örnekleme aşaması veri örnekleme, modelleme için veri seçimini kapsamaktadır. **Keşfetme aşamasında**; beklenen ve beklenmeyen değişkenler arasında ilişkileri ve hataları keşfederek verinin anlaşılması hedeflenir. **Değiştirme aşamasında**; modelleme sürecine hazırlık amacı ile veri gürültü/hata, artık, eksik veri gibi sorunlardan temizlenerek, modelde kullanılmak üzere dönüştürülür. **Model aşamasında**; eğilim, örüntüleri, tahminleri en iyi sağlayacak model kurulup veriye uygulanmaktadır. SEMMA'nın son aşaması olan **Değer saptama– değerlendirme aşamasında**; uygulanan modelin planlanan hedeflere uygunluğu değerlendirilmektedir.

2.4.4. Veri Madenciliğinde Karşılaşılabilen Problemler

VM uzmanları büyük ölçekteki veri içeren veri tabanlarında çalışırken eksik, atık, aykırı, belirsiz veri gibi bazı sorunlarla karşılaşabilmektedirler. Başarılı bir veri madenciliği uygulaması için muhtemel sorunların giderilmesi gerekmektedir. Sistemlere veri girişi esnasında yapılan yanlış girişlerden kaynaklanan ya da veri toplanması sırasında oluşan sistem dışı hatalara, veri özellikleri ya da sınıflarındaki hatalara gürültü adı verilir. Veri tabanlarındaki eksik bilgi ve yanlışlardan dolayı VM amacına tam olarak ulaşamayabilir, ilgili gürültülerin düzeltilmesi ya da temizlenmesi gerekebilmektedir. Veri tabanlarındaki geçersiz veriler de sorun teşkil edebilmekte; değeri birincil tanımlamada yer almayan herhangi bir niteliğin değeri geçersiz ise o nitelik bilinmeyen ve uygulanamaz bir değere sahip olmaktadır. Bir başka potansiyel sorun, artık verilerdir. Verilen veri kümesi, eldeki probleme uygun olmayan, örneklem kümesi için gereksiz nitelikler içerebilir. Kurumsal çevrim içi veri tabanları dinamik olduğu için veri içeriği sürekli değişebilmektedir, bu durum bilgi keşfi için riskler barındırmaktadır. Kayıp değerler, veri toplanması aşamasında yanlış işlenen veri vb. durumlar çalışmaları etkilemektedir. Sadece sembolik ve kategorik değil tamsayı, kesirli sayılar, coğrafi bilgi vb. farklı tipteki verilerin ele alınması aşamasında da sorunlar ortaya çıkabilmektedir. Veri tabanı boyutu da çalışmaları etkileyebilen unsurlardandır (Albayrak, 2009).

2.5. Veri Madenciliği Yöntemleri

VTBK ve VM işleyişindeki süreç yönetim metotları değerlendirildiğinde, VM'nin asıl dahil olduğu aşama modelleme aşamasıdır ve burada farklı stratejiler, yöntemler ve algoritmalar bulunmaktadır. Veri Madenciliğinde, Denetimli (Supervised) ve Denetimsiz (Unsupervised) yöntemler kullanılmaktadır. İlgili bu yaklaşımlar; veriyi analiz etme ve modelleme metodolojisini belirlemede kullanılmaktadır. Temel olarak, VM'nde önceden net bir şekilde belirlenmiş ulaşılmak istenen bir hedef olduğunda Denetimli (supervised), elde edilmesi istenen sonuç için özel bir tanımlama yapılmamışsa veya belirsizlik söz konusu ise Denetimsiz (unsupervised) yöntemlerden yararlanılmaktadır (Hastie, vd., 2001).

Denetimli yöntemler, kayıtlı veriden önceden belirlenen hedef doğrultusunda bilgi ve sonuç çıkarmaya yönelik kullanılmaktadır. Bu yöntemlerde, veri üzerinden her bir sınıfa ilişkin özellikler ve ilgili özelliklerin kurallar halinde tanımlaması yapılmaktadır. Örneğin, bir telekomünikasyon şirketinde, geçmişteki veri nitelikleri ve onların nihai olarak devamlı müşteri ya da kayıp müşteri olma durumlarının analiz edilmesi ile kaybedilme olasılığı yüksek müşteriler listelenebilmektedir. Bu yöntemlere, Karar Ağaçları, RO, DVM ve YSA örnek verilebilir.

Denetimsiz yöntemler, özel bir tanımlama yapılmamışsa, elde edilmesi istenilen sonuç net değil ya da belirsiz ise kullanılmaktadır. Bu yöntemlerde çok sayıdaki verinin gözlenmesi ve özellikleri arasındaki benzerliklerden hareket ederek sınıfların tanımlanması amaçlanmaktadır. Bu yöntemlerde girdi ile çıktı arasında mantıksal bağ kurulabilecek bir amaçlanan değer, çıktı olmadığı için veri içindeki benzer patern gösterenler araştırılmaktadır. Denetimsiz yöntemlere ise Hiyerarşik Kümeleme, K-Ortalamlar gibi kümeleme ve Apriori, CARMA gibi birliktelik kuralları örnek verilebilir.

VM modelleri, amaçları ve gördükleri işlevlere göre temelde tahminleme ve tanımlama olarak iki ana amaç altında incelenmektedir. Tahminleyici veri madenciliği modelleri de Sınıflama (Classification) ve Regresyon (Regression) olarak ikiye ayrılmaktadır. Tanımlayıcı VM modelleri de Kümeleme (Clustering) ve Birliktelik Kuralları (Association Rules) olmak üzere iki alt başlık altında toplanabilir (Özekes, 2003). Her ne kadar modellerin kullanım alanları çok geniş olsa da genel olarak kullanımları Tablo 2.1.'de paylaşılmıştır (Karabulut, 2021).

Tablo 2.1.: Veri Madenciliği Modelleri ve Kullanım Alanları

Modeller	Kullanım Alanları
Sınıflandırma & Regresyon	Kayıp müşteri tahmini Satış tahminleri Üretim için sipariş tahminleri Dolandırıcılık tespiti
Kümeleme	Segmentasyon Sağlık alanında coğrafi risk faktörleri Ürün satış profili
Birliktelik	Pazar sepet analizi Promosyon analizi Katalog ve yerleşim düzeni

VM'nin veri tabanlarından sağlanan büyük veri ile çalışabilmesi, yukarıda değinilen tahminleyici, tanımlayıcı yöntemlerin uygulanabilmesi için çeşitli veri madenciliği yazılımlarına ve programlarına ihtiyaç duyulmaktadır. Endüstride farklı programlar bulunsa da yazında yapılan birçok çalışmada IBM SPSS Modeller, WEKA, Clementine, RapidMiner, Knime, Keel, ve Orange gibi paketlerin kullanıldığı görülmüştür (Yıldız ve Şeker, 2016).

2.5.1. Sınıflandırma ve Regresyon Modelleri

Sınıflandırma ve regresyon modelleri, sonuçları bilinen verilerden hareket edilerek bir model geliştirilmesini ve kurulan modellerden yararlanılarak, sonuçları bilinmeyen veri kümeleri için sonuç değerlerinin tahmin edilmesini amaçlanmaktadır. Örneğin bir banka, önceki dönemlerde vermiş olduğu kredilere ilişkin verilere ait kredi alan müşterilerinin özellikleri ve geri ödeme durumu arasındaki örüntüyü inceleyip, yeni kredi başvurusunu değerlendirirken, modelin tahminlemesi ile ilgili müşterisinin geri ödeme durumunu tahminleyebilmektedir. Tahminleyici VM yöntemleri arasındaki Sınıflandırma ve Regresyon modellerindeki temel fark, bağımlı değişkenin sürekli veya kategorik özellik barındırmasıdır. Sınıflandırmada tekniklerinden kategorik değerleri, regresyondan süreklilik özelliği taşıyan değerleri tahminlerken yararlanılmaktadır (Öztürk, 2014). Aşağıda sınıflandırma tekniklerinden Karar Ağaçları (çalışmada

Karar ağaçları çözümlemelerinin genellikle aşağıdaki amaçlar için kullanıldığı görülmektedir (<https://www.ibm.com/docs/tr/cloud-paks/cp-data/4.5.x?topic=palette-decision-tree-nodes>, 2023):

- Belirli bir sınıfın muhtemel üyesi olacak öğelerin saptanmasında (segmentation),
- Çeşitli vakaların risk grupları gibi farklı kategorilere atanmasında (stratification),
- Öngörü amacı ile gelecekteki olayların tahmin edilebilmesi için kurallar oluşturulmasında,
- Tahmine dayalı öznitelikleri sürekli bir değişkenin değerlerinde ilişkilendirmede,
- Veri azaltmada yani parametrik modellerin kurulmasında kullanılmak üzere çok miktardaki değişken ve veri kümesinden yararlı olması beklenenlerin seçilmesinde,
- Sadece belirli alt gruplara özgü olan ilişkilerin tanımlanmasında,
- Kategorilerin birleştirilmesinde ve sürekli değişkenlerin kesikli değişkene en az bilgi kaybına neden olacak şekilde dönüştürülmesinde kullanıldığı görülmektedir.

Karar Ağacı yöntemi ile verinin sınıflandırılması öğrenme ve sınıflama olmak üzere iki aşamada yapılmaktadır (Han ve Kamber, 2000). Öğrenme adımı önceden belirli bir eğitim verisi kullanılmaktadır. Sınıflama aşamasındaki test verisi, sınıflama kurallarının veya karar ağacının doğruluğunu ortaya çıkarmak amacı ile kullanılmaktadır (Boyacı, vd., 2018). Temelde, Karar Ağacı bir öğrenme kümesi oluşturur, belirlenen hedefi en iyi bölen kestirici yani öğrenme kümesini en iyi ayıran nitelik belirlenir ve ağacın ilk düğümü oluşturulur. Bu düğümden aynı mantıkla, yeni oluşan düğüm en iyi bölen kestirici ile dallanır ve dallar da aynı mantıkla yapraklara ayrılır. Bu aşamalar, kalan örneklerin hepsinin aynı sınıfa ait olması, örnekleri bölecek nitelik kalmaması ve kalan niteliklerin değerini taşıyacak örnek olmaması durumuna kadar devam etmektedir.

2.5.1.1.1. Karar Ağaçları Avantaj ve Dezavantajları

Denetimli sınıflandırma tekniklerinden Karar Ağaçları çeşitli avantajlar ve dezavantajlar içermektedir. Bazı avantajlar aşağıdaki şekilde listelenebilmektedir (Schmid, 2013; Çelik, vd., 2017):

- Hem sınıflandırma hem de regresyon problemlerinde kullanılabilir,
- Doğrusal olmayan ilişkileri yakalayabilir, doğrusal olarak ayıramayan verileri sınıflandırmak için kullanılabilir,
- Karar Ağacında normalleştirme gerekmez ve hem sayısal hem de kategorik verilerle çalışabilir,
- Anlaşılması ve yorumlaması kolaydır, model sonucu çıkan ağaçlar görselleştirilerek desteklenebilir,
- Kullanılan ağacın maliyeti, ağacı eğitmek için kullanılan veri değişkenlerinin sayısı ile logaritmiktir,
- Verilerdeki eksik değerlere sahip olabilecek veri kümelerini işleme yeteneğine sahiptir,
- Çok çıktılı problemleri ele alabilir,
- İstatistiksel testler kullanılarak bir modelin doğrulanması mümkündür.

Bazı dezavantajları aşağıdaki şekilde listelenebilmektedir (Schmid, 2013; Çelik vd., 2017):

- Veriyi iyi bir şekilde açıklamayan aşırı karmaşık ağaçlar üretilir. Bu durumda karar ağacı çok büyük olacağı için dallanması takip edilemeyebilir,
- Over-fitting (ezbere öğrenme) olarak geçen, sistemin aşırı eğitilmesi sonucunda esnekliğini kaybetmesi ve gereksiz ayrıntı bulundurma durumun ortaya çıkabilir. Modelin parametrelerinde kısıtlamalar ya da budama yöntemi ile iyileştirilebilir,
- Veri setinde birkaç benzer özellik tanımlı ise başarılı performans sergileme eğilimindedir. Diğer yandan, karmaşık birkaç etkileşim varsa performansında düşüş gözlenmektedir. Bu durumun sebeplerinden birisi,

diğer sınıflandırıcıların, temsili zor olacak bir sınıflandırıcıyı kompakt olarak tanımlayabilmesidir,

- Sürekli değişkenleri tahmin etmede daha az kullanışlıdır. Çok sayıda sürekli değişken ile çalışılması gereken durumlarda diğer yöntemlere göre hesaplama maliyeti daha fazladır,
- Değişken girdi miktarının az olması durumunda ağaç yeterli bilgiyi sağlayamayabilir.

Karar Ağaçları algoritmalarında karşılaşılan sorunlardan biri olan aşırı öğrenme (over-fitting); modelin ezbere öğrenmesi, esnekliğini kaybedip gereksiz ayrıntıları bulundurması ve algoritmanın eğitim verisi üzerinden en alt kırılıma kadar çalışıp, sonuçları ezberlemesi ve sadece o veriler üzerinde başarı elde edebilmesidir. Bu test verisi üzerinde benzer başarının sağlanamaması durumu olarak tanımlanmaktadır. Bu tarz sorunlar için budama (pruning) işlemi devreye girmektedir. Budama temel olarak sınıflandırma içinde gereksiz ya da tekrarlı yerlerin çıkarılması işlemidir. Dallar budanarak en yalın yapıya ulaşmak amaçlanmaktadır. Bu yaklaşımda, Karar Ağacı'nın daha sade ve anlaşılır olması amaçlanmaktadır. Bir karar ağacı oluşturulduğunda, dalların barındıracağı gürültü veya aykırı değerler eğitim verilerindeki anormallikleri yansıtacağı için en az güvenilir dalları kaldırmak amacı ile istatistiksel yöntemler kullanılmaktadır. Budama; ön budama ve sonradan budama olarak ikiye ayrılmaktadır.

Ön budama (pre-pruning) işlemi, ağacın oluşum aşamasındadır. Örneğin, belirli bir düğümdeki eğitim kümelerinin alt kümesini, daha fazla bölmemeye karar vererek “budar” ve ağaç yapısının dışında bırakır, ilgili düğüm yaprak olur. Diğer yaklaşım olan sonradan budama (post-pruning) ise tamamen büyümüş ağaç üzerinden, belirli bir düğümdeki bir alt ağaç, dalları çıkarılarak ve bir yaprakla değiştirilerek budanır. Yaprak, değiştirilen alt ağaç arasında en sık kullanılan sınıfla etiketlenir ve böylelikle budama gerçekleştirilmiş olur (Duran, 2022).

2.5.1.1.2. Bölme Kriterleri

Karar ağacı yönteminde, dallandırmaya ilk hangi değişkeni baz alarak başlanacağı önemli bir husustur. Karar Ağacı alt kümelerine dallanırken, en iyi

kestirici, en iyi bölen nitelik tarafınca işlem görmektedir. Bu doğrultuda algoritmalar başlangıçta bazı değerler hesaplamakta ve sonuçlara göre ağaç oluşturma sürecine girmektedirler (Kuzey, 2012). Buradaki en iyi bölen niteliğin seçiminde Gini İndeksi, Twoing, Bilgi Kazancı (Information Gain), Kazanç Oranı (Gain Ratio), Ki-Kare (Chi-Square) testi, F-Testi, Olabilirlik Oran Sınanması Testi (Likelihood Ratio Test) ve DKM kriterleri rol oynamaktadır. İlgili kriterlerden bazıları aşağıda paylaşılmıştır.

2.5.1.1.2.1. DKM Kriteri

Sınıf özniteliğinin ikili olması durumunda kullanılan DKM bölünme kriteri, adını 1996 yılında keşfedicileri olan Dietterich, Kearns, ve Mansour'dan almaktadır. DKM kriteri safsızlık tabanlı (impurity based) bir bölünme kriteridir ve Eşitlik (2.1)'deki gibi tanımlanmaktadır.

$$DKM(S_{d_i}) = 2\sqrt{p_1 \cdot p_2} \quad (2.1)$$

Burada; y seçilen nitelik-değişkenini, $S_y=d_{i,j}$ seçilen niteliğin $y = d_i$ durumunda iken S veri kümesinin alt kümelerini, i seçilen niteliğin durum sayısını, j ise sınıflara ayırmak istenilen niteliğin sınıf sayısını belirtmektedir. p_1 , seçilen nitelikte d_i durumu için ilk sınıfın gerçekleşme olasılığını, p_2 ise ikinci sınıfın gerçekleşme olasılığını göstermektedir.

Rokach ve Maimon (2014); DKM ölçütünün, Bilgi Kazancı ve Gini İndeksine göre belirli bir doğruluk seviyesi elde etmek için daha küçük bir ağaç yapısına ihtiyaç duyduğunu paylaşmışlardır.

2.5.1.1.2.2. Bilgi Kazancı

Bilgi Kazancı (Information Gain) kriteri bir veri kümesindeki özniteliğin sınıflandırmaya ne düzeyde katkıda bulunduğunu anlatan bir ölçüttür ve 0-1 arasında değer almaktadır. Bilgi kazancı değerinin hesaplanması entropi

kavramına dayanmakta ve entropi bir sistemdeki düzensizliğin ya da belirsizliğin ölçüsünü, rastgeleliği, belirsizliği, beklenmeyen durumun ortaya çıkma olasılığını göstermektedir. Bir olayın oluşma olasılıklarına göre, veri setinin homojenliğini hesaplayan Entropi, 0-1 arasında bir değer almaktadır. Veri kümesindeki entropi değeri yüksek olan öznitelikler sınıflandırma problemlerindeki sınıf etiketlerini belirlemede diğer özniteliklere göre daha çok belirsizlik ve karasızlık içermektedir, dolayısıyla ağacın kökünde Entropi ölçüsü en az olanlar yani Kazanım Değeri en yüksek olanlar kullanılmaktadır (Bilgin, 2018). Kazanım oranı, Eşitlik 2.2 'de ki gibi hesaplanabilmektedir;

$$E(C|A_k) = \sum_{j=1}^{M_k} p(a_k, j) \left[- \sum_{i=1}^N p(c_i | a_k, j) \log_2 p(c_i | a_k, j) \right] \quad (2.2)$$

Burada; $E(C/A_k)$ = A_k alanın sınıflama özelliğinin Entropi değerini, $p(a_k, j)$ = a_k alanının j değerinde olma olasılığını, $p(c_i | a_k, j)$ = a_k alanı j değerinde iken sınıf değerinin c_i olma olasılığını, M_k = a_k alanının içerdiği değerlerin sayısı; $j = 1, 2, \dots, M_k$, N = farklı sınıfların sayısı; $i = 1, 2, \dots, N$, K =alanların sayısı; $k = 1, 2, \dots, K$ ifade etmektedir.

Eğer bir S kümesindeki elemanlar, kategorik olarak $C_1, C_2, C_3, \dots, C_i$ sınıflarına ayrıştırılırlarsa, S kümesindeki bir değişkenin sınıfını saptamada Eşitlik 2.3 kullanılmaktadır;

$$I(S) = -(p_1 \log_2(p_1) + p_2 \log_2(p_2) + \dots + p_i \log_2(p_i)) \quad (2.3)$$

Bu formülde p_i , C_i sınıfına ayrılma olasılığıdır. Entropi hesaplaması ise Eşitlik (2.4) deki gibi gösterilebilir;

$$E(A) = \sum_{i=1}^n \frac{|S_i|}{|S|} \times I(S_i) \quad (2.4)$$

Böylelikle, A alanı kullanılarak yapılacak dallandırma adımında, bilgi kazancı Eşitlik 2.5 deki şekilde belirlenmektedir:

$$\text{Kazanç (A)} = I(S) - E(A) \quad (2.5)$$

Kazanç (A), A alanının değerini bilmekten kaynaklanan entropideki azalmadır. Akış olarak Bilgi Kazancı kriterinde; tüm bağımsız değişkenler için kazanım değeri hesaplanıp birbirleri ile kıyaslanmakta ve en yüksek kazanım değerine sahip bağımsız değişken kök düğüm kabul edilmektedir ve sınıflara ayrılma en iyi bölen kabul edilmektedir (Çalış, vd., 2014).

2.5.1.1.2.3. Kazanç Oranı

Bilgi Kazanç Oranı (Gain Ratio), bilgi kazanımının normalize edilmiş halidir. C4.5, C5.0 da kullanılmaktadır. En yüksek bilgi kazanım oranına sahip olan bağımsız değişken, kök düğüm olarak seçilerek ağaçtaki dallanma başlatılmakta ve Eşitlik 2.6 da gibi formüle edilmektedir (Aggarwal, 2015).

$$\text{GainRatio}(a_i, S) = \frac{\text{InformationGain}(a_i, S)}{\text{Entropy}(a_i, S)} \quad (2.6)$$

Burada payda yani entropi değeri 0 olduğunda, sınıflandırma olanağı yoktur ve Kazanç Oranı tanımsız olup kullanılamamaktadır. Diğer yandan payda değeri düşük olduğunda değişkenler-öznitelikler lehine eğilim gösterebilmektedir. Bu doğrultuda Kazanç Oranında öncelikle tüm değişkenler için bilgi kazancı hesaplanmakta, sonrasında en iyi kazanç oranını sağlayan öznitelik seçilmektedir. Quinlan, yaptığı çalışmalarda Kazanç Oranının, Bilgi Kazancı kriterine göre doğruluk ve sınıflandırma kolaylığı açısından daha iyi bir performans sergilediğini paylaşmıştır (Rokach ve Maimon, 2014; Demir, 2021).

2.5.1.1.2.4. Gini İndeksi

1912’de Corrado Gini tarafından gelir dağılımlarının istatistiki ölçümü üzerine geliştirilen Gini İndeksi yaygınlıkla kullanılan bir bölünme kriteridir. Temelde aynı popülasyondan rastgele seçilen iki değişkenin aynı sınıfta olma olasılığını hesaplamaktadır (Berry ve Linoff, 2004). Gini indeksi, hedef özniteliklerin değerlerinin olasılık dağılımları arasındaki farklılıkları ölçen safsızlık tabanlı bir kriterdir (Ercan, 2016: 29). Gini indeksi hesaplaması eşitlik 2.7 de verilmiştir.

$$Gini(y, S) = 1 - \sum_{c_j \in dom(y)} \left(\frac{| \sigma_{y=c_j} S |}{|S|} \right)^2 \quad (2.7)$$

a_i özniteliğini seçmek için değerlendirme kriteri Eşitlik 2.8 deki gibi tanımlanmaktadır;

$$GiniGain(a_i, S) = Gini(y, S) - \sum_{v_{i,j} \in dom(a_i)} \frac{| \sigma_{a_i=v_{i,j}} S |}{|S|} \cdot Gini(y, \sigma_{a_i=v_{i,j}} S) \quad (2.8)$$

Gürültülü (aykırı) veriler için başarılı sonuçlar üreten Gini İndeksi, CART algoritmalarında da bölünme kriteri olarak yer almaktadır (Aggarwal, 2015). Gini indeksi, düğümler saf hal geldiğinde, Karar Ağacı maksimum derinliğe ulaştığında ya da minimum düğüm boyutuna ulaşıldığında tüm veri setlerinde başarılı sonuç vermeyebilmektedir. Bunun sorunların üstesinden gelmek için ise Eşitlik (2.9) daki işlemler uygulanabilmektedir. Öncelikle ilgili değişkenin-niteliğın sol ve sağ Gini değeri hesaplaması yapılmaktadır.

$$Gini_{sol} = 1 - \sum_{i=1}^k \left[\frac{L_i}{|T_{sol}|} \right]^2 \quad (2.9)$$

$$Gini_{sağ} = 1 - \sum_{i=1}^k \left[\frac{R_i}{|T_{sağ}|} \right]^2$$

Burada, k: sınıfların sayısını, T: Düğümdeki örnek sayısını, T_{sol}: sol taraftaki örneklerin sayısını, T_{sağ}: Sağ taraftaki örneklerin sayısını, L_i: sol taraftaki i kategorisindeki örneklerin sayısı ve R_i: sağ taraftaki i kategorisindeki örneklerin sayısını simgelemektedir. Sonrasında bir değişkenin-niteliğın Gini değeri Eşitlik 2.10 daki şekilde saptanmaktadır.

$$Gini_j = \frac{1}{n} (|T_{sol}| Gini_{sol} + |T_{sağ}| Gini_{sağ}) \quad (2.10)$$

Gini değeri her bir değişken için hesaplandıktan sonra, en küçük olanı seçilmekte ve bölünmeye ilgili değişkenden başlanmaktadır. Geriye kalan veri seti üzerinde ilgili adımlar yinelenmekte ve diğer bölünmeler saptanmaktadır (Adak ve Yurtay, 2014).

2.5.1.1.2.5. Twoning

Twoning, Sınıflandırma ve Regresyon Ağaçları (CART) uygulamalarında, çok sınıflı problemlerin çözümünde kullanılmaktadır. Hedef özniteliğın alanı nispeten geniş olduğunda Gini indeksi bazı problemlerle karşılaşabilmektedir. Bu doğrultuda Twoning'in Gini İndeksine alternatif olarak geliştirildiği söylenebilmektedir. Twoning algoritmasında eğitim kümesi her adımda iki parçaya bölünmektedir (t_{sol} ve t_{sağ}). Sonrasında, her bir bölüm için ayrı ayrı olasılıkları hesaplanmaktadır. Üçüncü adımda aday bölünmesinin uygunluk değeri aşağıdaki şekilde hesaplanmakta ve en yüksek olanı seçilmektedir.

$$\Phi(s | t) = 2P_{sol}P_{sag} \sum_{j=1}^n |P(j | t_{sol}) - P(j | t_{sag})| \quad (2.11)$$

Sonrasında, ilk adıma dönülerek ağacın alt kümesine aynı işlemler uygulanmaktadır.

2.5.1.1.2.6. Ki-Kare Testi

Karl Pearson tarafından 1912'de keşfedilen Ki-Kare (Chi-Squared) testi, iki veya daha fazla değişkenin-niteliğin birbiri ile bağımsızlık durumu araştırmakta kullanılmaktadır. Yapılan bu tez çalışmasında da kullanılan Ki-kare testi uygulama kolaylığı nedeni ile çalışmalarda tercih edilmektedir (Suner ve Demirarslan, 2021). CHAID algoritmasında da Ki-Kare, en iyi bölen kestiricinin belirlenmesinde kullanılmıştır. En önemli koşul verilerin kategorik olması ve grupların birbirinden bağımsız olmasıdır.

Ki-Kare testi temelde gözlenen ve beklenen frekanslar arasındaki farkın anlamlı olup olmadığı temeline dayanmaktadır. Genellikle iki ya da daha çok grup arasında fark olup olmadığının testinde, iki değişken arasında bağ olup olmadığının testinde, gruplar arası homojenlik testinde ve örneklemden elde edilen dağılımın istenen herhangi bir teorik dağılıma uyup uymadığının testinde (uyum iyiliği testi) kullanılmaktadır. Niteliksel olarak belirtilen verilerin analizinde kullanılmaktadır. Ki-Kare testi gözlenen frekans değeri ile beklenen frekans değerlerinin karşılaştırılmasına dayanır.

Ki-kare testinde; (H0) Sıfır Hipotezi: iki kriterin bağımsız olduğunu, (HA) Araştırma Hipotezi: iki kriterin bağımlı olduğunu ifade eder. Formül Eşitlik 2.12 verilmiştir.

$$X^2 = \sum \frac{(G-B)^2}{B} \quad (2.12)$$

Burada G gözlenen değeri, B ise beklenen değeri sembolize etmektedir. Ki-Kare değerinin anlamlı bir fark olup olmadığını anlamak için serbestlik değerine ihtiyaç duyulur. Serbestlik değeri kategorik değişkenlerin kategori sayılarından 1

çıkartılarak bu sayıların birbiriyle çarpılarak hesaplanmaktadır. Serbestlik derecesi a satır sayısı ve b sütun sayısı olmak üzere $(a-1) \times (b-1)$ olarak hesaplanabilmektedir (Kasım, 2022).

2.5.1.1.3. Karar Ağacı Algoritmaları

Algoritma, belirlenen bir problemi çözmek veya belirli hedefe ulaşmak amacı ile tasarlanan yol, başlangıç durumundan başladığında, açıkça belirlenmiş bir son durumunda sonlanan, sonlu işlemler kümesi olarak tanımlanmaktadır. Bir başka ifade ile mantıksal adımlar içeren çözüm yoludur. VM’de algoritmaların temelde büyük veri yığını arasından anlamlı örüntülere ulaşmayı sağlayacak kuralları oluşturmayı hedeflemektedir. Model başarısını önemli ölçüde etkileyen algoritma seçimi, modelin ve veri setinin özelliklerine bağlıdır. Algoritmasına göre değişmekle birlikte, algoritmalar nicel, nitel ve 0-1 değerlerinden oluşan veri setlerinde çalışmaktadırlar (Gedleç ve Yılmaz, 2020).

1970’li yılların başlarında Morgan ve Sonquist tarafından geliştirilen Otomatik Etkileşim Belirleme (Automatic Interaction Detector- AID), Karar Ağacı bazlı algoritma olarak yazındaki yerini almıştır. AID’i, Breiman vd. (1984) tarafından yazına kazandırılan Sınıflandırma ve Regresyon Ağaçları (CART) takip etmiştir. Quinlan (1986), karar ağaçlarına yeni bir algoritma eklemiş ve entropi ve bilgi kazancına dayalı Yinemeli İkili Ağaç (Iterative Dichotomiser 3- ID3) algoritmasını geliştirmiştir. Bu algoritma C4.5 ve C5.0’ın temellerini oluşturmaktadır. 1980 yılında Kass tarafından geliştirilen bir diğer algoritma ise Ki-Kare Otomatik Etkileşim Detektörü (CHAID) olmuştur. Onları SLIQ (Supervised Learning in Quest), SPRINT (Scalable Parallelizable Induction of Decision Trees), QUEST (Quick, Unbiased, Efficient Statistical Tree), MARS (Multivariate Adaptive Regression Splines) ve Exhaustive CHAID gibi algoritmalar takip etmiştir (Emel ve Taşkın, 2005). Karar ağaçları algoritmaları kök, düğüm ve dallanma ölçütlerinin belirlenmesinde benimsenen yaklaşım açısından birbirinden farklılaşmaktadır. Tablo 2.2.’de genel yaklaşımlar paylaşılmıştır (Palmer, vd., 2011). Diğer yandan, çalışmada kullanılan CHAID, CART ve ayrıca C5.0 algoritmaları daha ayrıntılı incelenmiştir. .

Tablo 2.2.: Karar Ağacı Algoritmaları ve Özellikleri

Algoritma	Girdi Değişkeni	Çıktı Değişkeni	Tahmin Türü	Dal Sayısı	Bölünme Ölçütü
CHAID	Kategorik, Sayısal	Kategorik, Sayısal	Sınıflandırma, Regresyon	≥ 2	Ki Kare, F-Testi, Likelihood Ratio
CART	Kategorik, Sayısal	Kategorik, Sayısal	Sınıflandırma, Regresyon	≥ 2	Gini İndeksi, Twoing, En Küçük Kare Deviasyonu
C4.5 & C5.0	Kategorik, Sayısal	Kategorik	Regresyon	≥ 2	Kazanç Oranı
QUEST	Kategorik, Sayısal	Kategorik	Regresyon	≥ 2	Ki Kare, F

Kaynak: Palmer vd., 2011 : 379

2.5.1.1.3.1. CHAID Algoritması

1980 yılında Kass tarafından geliştirilen Ki-Kare Otomatik İterasyon Detektörü (CHAID) hem regresyon hem de sınıflama amacı ile kullanılmaktadır. CHAID, kategorik bağımlı değişkenler için geliştirilmiş AID analizinin uzantısı olarak kabul edilmektedir. CHAID algoritmasındaki temel mantık, her bir bağımsız değişkenin bağımlı değişkenle ilişkisinin Ki-kare testi ile ölçüldüğü ve tüm gözlemlerin homojen kalana dek sürecin devam ettiği dallanma kuralları oluşturmaktır (Depren, vd., 2017). CHAID algoritması büyük veri kümelerinin kolay ve anlaşılır şekilde yorumlanabilmesi için bağımsız değişkenlerle bağımlı değişkeni alt gruplara ayırmaya çalışmaktadır. Dolayısıyla, buradaki amaç veriyi daha homojen olan bir alt gruba ayırmaktır. Algoritmanın tercih edilme nedenlerinin başında kategorik ve sürekli tüm değişken tipleriyle çalışabilmesi ve ağaçtaki her düğümü ikiden çok alt parçaya bölebilmesi gibi nedenler yer almaktadır (Pehlivan, 2006). Değişkenler arası örüntü doğrusallıktan daha karmaşık ise veride gizli olan ilişkileri saptayabilmek için verinin belli kısımlarını eleme tekniği olarak CHAID kullanılmaktadır. “Ki-kare” ismini almasının nedeni algoritmasında birçok çapraz tablonun kullanılması ve istatistiksel önem oranları ile çalışmasıdır.

Bir niteliği–değişkeni birden fazla bölmeleri CHAID’i, CART’den farklılaştırmaktadır. Sınıflandırma çalışmalarında, her adımda, düğümde, en iyi bölünmeyi belirlemek için Ki-Kare’den yararlanılmaktadır ve dalların sayısı iki ile tahmin edicinin kategori sayısı arasında değişmektedir. Regresyon çalışmalarında, yani hedef değişkenin sürekli olduğu durumlarda, en iyi bölünme aşaması için F-Testi kullanılmaktadır (Nisbet, vd., 2009). Nominal değişkenler için Pearson Ki- Kare kriterleri, Sıralı-Rank değişkenleri için Olabilirlik-Oran testi kullanılmaktadır (Ercan, 2016).

CHAID analizi, çok kategorili değişkenlerin yer aldığı büyük bir veri kümesini, benzer kategorileri birleştirip, önemli sayılan değişkenlere göre bölerek sadeleştirmektedir. Her bir bağımlı değişken için kategorilerin anlamlı bir şekilde birleştirilmesinden sonra, bağımlı değişkene göre kontenjans tabloları oluşturularak, Bonferroni p değerleri ile χ^2 istatistikleri hesaplanmaktadır. Açıklayıcı değişkenler birbirleri ile karşılaştırılıp, en küçük Bonferroni p değerine sahip olan açıklayıcı değişkenin kategorilerine göre, veriler alt gruplara ayrılmaktadır. CHAID analizinde her bir açıklayıcı değişken için en iyi bölünme tespit edilmektedir. Daha sonra açıklayıcı değişkenler en iyi seçilene kadar karşılaştırarak ve seçilen en iyi açıklayıcı değişkene göre yeniden bölünmeler yapılmaktadır. Tüm alt bölümler bağımsız olarak yeniden analiz edilip, her bir açıklayıcı değişken kategorilerini izin verdiği mümkün bölünmeler gerçekleştirilerek χ^2 testindeki önem derecesine göre kontenjans tabloları oluşturulmaktadır.

İkili Karar Ağacı sentezinin durdurulması maksimum ağaç derinliğine ulaşıldığında, bütün terminal düğümlerin belirlenen eşik değerinin altında sayıları içerdiğinde ve kayıp değişken değerleri ayrı grupta birleştirebildiğinde gerçekleşmektedir (Donskoy, 2013).

CHAID; her bir girdi özneliği, bağımlı değişkeni için hedef özneliğine göre en az anlamlı derecede farklı olan değerler çifti saptamaktadır. Anlamlı farklılık bir istatistik testinden elde edilen “p” değeri yolu ile ölçülmekte ve belirlenen her bir çift için, elde edilen “p” değerinin belli bir birleştirme eşik değerinden büyük olup olmadığını kontrol edilmektedir. Eğer büyük ise, değerleri birleştirir ve ilave potansiyel çiftleri birleştirmek için arama yapmaktadır. Bu süreç anlamlı çiftler

bulunamayana kadar tekrarlanmaktadır (Ercan, 2016). Bu sayede mevcut düğümü en iyi bölecek olan bağımlı değişken tespit edilmektedir. Eğer en iyi bağımlı değişkeninin düzeltilmiş “p” değeri belli bir bölme eşik değerinden küçük değil ise bölme gerçekleşmemektedir. CHAID, eksik değerlerin hepsine tek bir geçerli kategori gibi muamele ederek işlem yapar ve budama işlemi gerçekleştirilmemektedir (Kuzey, 2012: 77-78).

Bu tezde yapılacak sağlık sektöründe MKA çalışmasında da kullanılan CHAID algoritmasının; hızlı olması, geniş karar ağaçları oluşturabilmesi, tek dala bağlı çok sayıda terminal düğüm noktaları üretebilir olması ve kolay anlaşılır olması gibi avantajları bulunmaktadır. Diğer yandan her bir teknikte olduğu gibi, bağımlı değişkenleri birden fazla bölme ile alt kategorilere ayırmasından dolayı, güvenilir sonuçlar elde edebilmek için büyük miktarda veriye ihtiyaç duymaktadır. Ayrıca çoklu bölünmenin gerçek iş ortamı ilişkilendirilmesinin zor olması nedeni ile yorumlaması zor olan çok kısa tablolar da üretebilmektedir (Nisbet vd., 2009:147).

2.5.1.1.3.2. CART Algoritması

Classification and Regression Trees, CART ya da C&RT olarak yazında geçen Sınıflandırma ve Regresyon Ağaçları algoritması, Breiman, vd. (1984) yılında Friedman, Olshen ve Stone tarafından geliştirilmiş ve yazına kazandırılmıştır.

CART algoritmasında, karar ağacını oluştururken her bir düğümde ve aşamada ilgili grup, kendinden daha homojen iki alt gruba ayrılmaktadır. En iyi dallara ayırma, en iyi bölen kestiriciyi belirlemede ID3, C4.5 ve C5.0'da olduğu gibi entropiden yararlanmaktadır. Diğer yandan her aşamada, kendisinden daha homojen iki alt grup oluşturarak, bölünerek çalışması CART'ı farklılaştırmaktadır. Bu işlemlerde en iyi bölen kestirici seçilirken, bağımlı değişkenleri kategorik ise Gini İndeksi ve Twoing, sürekli değişkenler ise En Küçük Kare Sapması (Least-Squared Deviation) kullanılmaktadır (Atılğan, 2011). Ayrıca eğer bağımlı değişkenler kategorik ise CART algoritması sınıflandırma ağacı, eğer bağımlı değişkenler sürekli ise regresyon ağacı oluşturmaktadır (Yohannes ve Hoddinott, 2018).

CART algoritmasında örneğin, bağımlı değişkenin kategorik olduğu, Gini indeksi ile en iyi kestiricinin saptandığı durumda, işleyiş olarak ilk önce dallanması muhtemel her bağımlı değişken ve değişkene ait kategorileri için, ağacın aday ikili alt grupları belirlenmektedir. İkinci adımda; en düşük Gini değerine haiz aday bölünme tespit edilmekte ve ilgili bölünmeye ait bağımsız değişken dallanarak iki alt düğüm oluşturmaktadır. Bu aşama, oluşan alt düğümlerin homojenliğe ulaşarak sınıflandırılmayacak duruma gelinceye kadar devam etmektedir. Son adımda, homojen ve sınıflandırılmayan düğüme ulaşıldığında, düğüm terminal düğüm olarak etiketlenmekte ve ağaç sonlanmaktadır. Bu süreç ağacın tüm alt düğümleri terminal olarak etiketleninceye kadar devam etmektedir. Çalışma sırasında budama işlemi yapılabilmektedir. Bu aşamada en yüksek tahmin hata oranlı dallar belirlenip, karmaşıklığı azaltmak ve tahmin başarısını arttırmak için temizlenmektedir (Sezer, vd., 2010).

CART, günümüzde firmalar tarafından analiz hızı yüksek olduğu için sıklıkla kullanılmaktadır (Hadden, 2008). CART algoritmasının avantajları değerlendirildiğinde, aykırı (gürültü) değerlerden etkilenmeyen, kategorik, sürekli bağımlı değişkenlerle ve bileşimleri ile çalışabilen, eksik verileri işleyebilen ve çok karmaşık örüntüleri analiz edebilen bir algoritma olarak kabul görmektedir (Nisbet vd., 2009).

2.5.1.1.3.1. C5.0 Algoritması

ID3 algoritması barındıran, C.4.5'un bir gelişmiş versiyonu olan C5.0 Quinlan tarafından 1993 yılında yazına kazandırılmıştır. İlk versiyonları kategorik veri türleri ile sınırlı iken, C5.0 sürekli veri türleri üzerinde de uygulama olanağı sağlamaktadır. Büyük veri tabanlarında kullanımı amaçlanmıştır. Tahminlemedeki başarısından dolayı sıklıkla kullanılan karar ağacı algoritmalarındandır.

C5.0'ın içerdiği ID3 algoritması bir veri kümesi içindeki belirsizliği ve rastgeleliği ölçebilen entropi aracılığıyla değişkenler arasından sınıflamada en ayırıcı özelliğe sahip değişkeni bulmayı hedeflemektedir (Dunham, 2003). Analize konu müşteri grubunun örneğin aynı sosyodemografik gruba mensup olduğu ve aynı

davranışları sergilediğini düşünüldüğünde entropi değeri 0, bir rastgelelik, düzensizlik yani düzensiz ve sistematik olmadığı durumda, veriler sayısallaştırıldığında 0-1 arasında bir değer almaktadır.

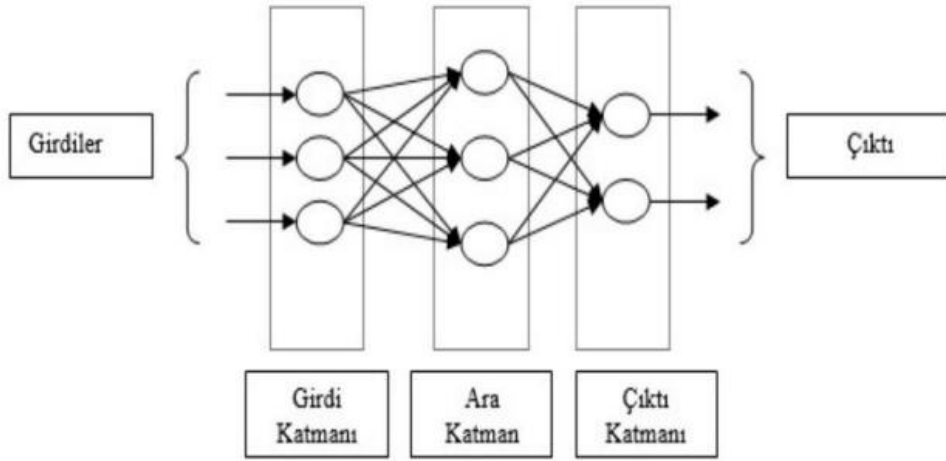
C5.0 her işlem adımında bütün özellikler kontrol edilmektedir. İlk olarak veri setindeki tüm girdi gözlemlerine ait entropi değeri hesaplanmaktadır. İkinci adımda, dallanması olası bağımsız değişkenler için Bilgi Kazanç değeri ve Kazanç Oranları hesaplanmaktadır. Üçüncü adımda, en yüksek kazanım oranlı bağımsız değişken, kendi kategorisi kadar dala ayrılmakta ve alt düğümlerini oluşturmaktadır. Dördüncü adımda, eğer oluşan alt düğümler tekrar sınıflandırılabilir olacak homojen olmayan yapıda ise, ilgili düğümlerdeki gözlemler için tekrar entropi değeri hesaplanmakta ve işlem ikinci adımdan itibaren yinelenmektedir. Geriye kalan düğümler homojen yapıda oldukları için terminal düğüm olarak belirlenmektedir. Algoritma ağacın tüm düğümleri için terminal düğüm saptanıncaya kadar devam etmektedir (Bardi ve Can, 2021).

C5.0 çok sayıdaki veri üzerinde hızlı işlem yapıp analizleri sunabilmektedir. Yaptığı analizler ve kullandığı ağaç yapısı, YSA gibi diğer yöntemlere nispeten çok daha kolay anlaşılabilen ve yorumlanabilmektedir. Sınıflandırma aşamasında boosting, isabet oranını artırıcı güçlendirme fonksiyonelliği bulunmaktadır. Başarı oranı yüksek ve hızlı çalışan bir algoritma yapısı bulunduğu için dolayı en çok kullanılan Karar Ağacı algoritmalarındandır (Ruggieri, 2002). Regresyon ve YSA gibi modeller sadece kategorik ya da sadece sayısal verilerle çalışırken, C5.0 kategorik ve sayısal verileri birlikte çalışabilmektedir. Herhangi bir eksik veri olsa da tahmin yapıp karar ağacı oluşturabilmektedir (Gorunescu, 2011).

2.5.1.2. Yapay Sinir Ağları

Yapay Sinir Ağları (Artificial Neural Network, YSA) insan beyninin öğrenme prensiplerini baz alarak geliştirilmiş, yaygın kullanılan kestirimsel Veri Madenciliği yöntemlerinden birisidir. Tahmin, sınıflandırma, veri ilişkilendirme gibi amaçlarla kullanılmaktadır. İnsanlarda tıpkı biyolojik sinir ağlarında gözlemlendiği gibi

YSA'da da temel öğrenme ve süreç elemanı olarak yapay sinir hücresi olan nöronu kullanmaktadır. Her bir nöron çalışma sonucunda bir ağırlıklandırma aracılığıyla birleştirilen girdi setini barındırmaktadır (Tolon, 2007). Bu algoritma hem kategorik hem de sürekli bağımlı değişkenler için kullanılabilir. YSA genel olarak lineer yapıda olmayan ve girdi-çıkı eşleştirmesi baz alınarak çalışan kullanışlı bir sınıflandırma tekniğidir. YSA, kalite kontrol, finansal-ekonomik tahminleme, kredi derecelendirme gibi alanlarda sıklıkla kullanılmaktadır. Sinir ağının girdi katmanı, gizli katman ve çıktı katmanı olmak üzere üç katmanı bulunmaktadır (Şekil 2. 6).



Şekil 2.6.: Yapay Sinir Ağı Katmanları

Kaynak: Kutlu ve Badur, 2009.

Girdi katmanında bağımsız değişkenler yer almaktadır ve mevcut veri seti sinir ağları ile eşlenerek sisteme dahil edilmektedir. Burada girdi değerlerinde herhangi bir işlem yapılmamaktadır. Ara katmanda yer alan her bir düğüm kendinden önceki katmandan gelen değişkenler veya düğümlerin değişik ağırlıklarla birleşmiş kombinasyonlarını içermektedir. Girdi katmanından gelen bilgileri işleyerek Çıktı Katmana iletmekle yükümlüdür. Çıktı katmanı ise ara katmandan gelen bilgileri işler, girdi katmanındaki veriler için uygun çıktı bilgisi üretip, raporlama sorumluluklarını taşımaktadır (Karabulut, 2021).

Yapay sinir hücreleri katmanlar halinde birleşerek yapay sinir ağlarını oluşturmaktadırlar. Oluşturulan yapay sinir ağı, veri setindeki yapıyı öğrenip beklenen sorumluluğun yerine getirmek için genelleştirmeler yapmaktadır. İlgili olayın örnekleri ile eğitilerek ağa genelleme yapabilme yetisi kazandırmakta ve benzer olayların çıktı setlerini belirlemektedir. Öğrenme işlemi sağlandıktan sonra yeni gelen bir verinin, hesaplanmış ağırlıklar yardımıyla bağlı olduğu sınıf hesaplanıp tanımlanabilmektedir (Atalay ve Çelik, 2017). Yapay sinir ağları modelleri Tek Katmanlı, Çok Katmanlı algılayıcılar ve İleri, Geri Yayımlı yapay sinir ağları olmak üzere dört farklı grupta incelenebilmektedir.

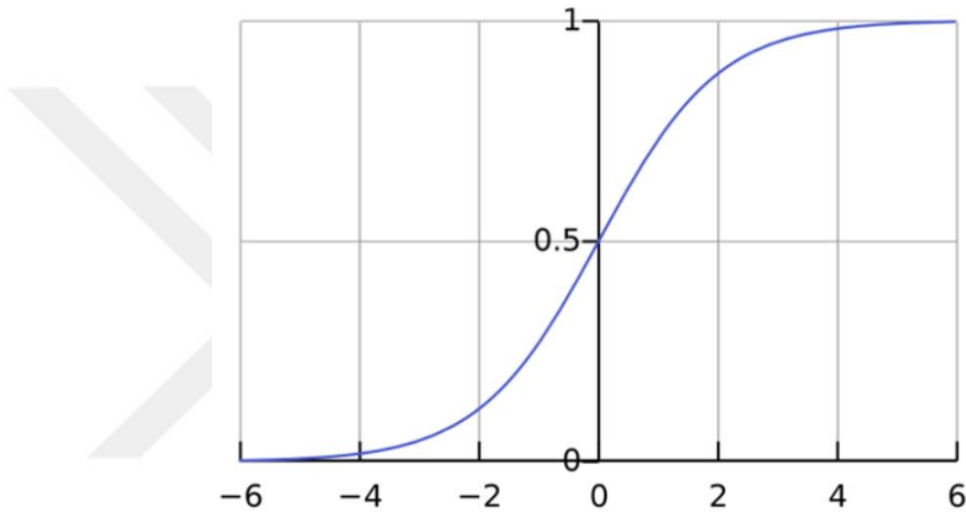
2.5.1.3. Lojistik Regresyon

Hosmer ve Lemeshow tarafından 1989 yılında geliştirilen Lojistik Regresyon (Logistic Regression- LR), müşteri kayıp yönetimi ve müşteri değeri gibi çalışmalarda sınıflama işlemlerinde, iki veri faktörü arasındaki örüntüyü bulmada ve ayırt edici olasılıklı sınıflandırma için yaygın olarak kullanılan bir istatistiksel modelleme tekniğidir. LR, gerçekleşen önceden belirlenmiş bir olayın olasılığını tahminlemektedir. Bir web alışveriş sitesi örneği düşünüldüğünde, müşterilerin geçmişteki ziyaretlerinin süresi, sepete attıkları ürün sayısı ile satın alma işleminin gerçekleşmesi durumunu analiz edip, olasılıklar doğrultusunda gelecekteki müşterilerin davranış kalıpları tahminlenebilmektedir. LR, bağımlı değişkenlerinin tahmini değerleri için olasılık hesaplayıp uygun sınıflama yapabilme olanağı tanımaktadır (Gorunescu, 2011). Doğrusal regresyonda kesikli bağımlı değişken ile tahmin değişkeni arasındaki ilişkiye bakılmaktadır ve eğer bağımlı değişkenin kategorik, ikili veya çoklu olduğu problemlerde doğrusal regresyon yetersiz kalacağı için açıklayıcı değişkenlerle neden-sonuç ilişkisini saptamada LR'dan faydalanılmaktadır (Larose, 2006). Doğrusal regresyondan ayrıldığı bir diğer nokta ise normal dağılım ve süreklilik varsayımı ön şartını gerektirmemesidir (Çokluk,2010). Genel anlamda LR, bağımlı (yanıt) ve bağımsız (açıklayıcı) her iki değişkenin geçmiş veri değerlerine bakarak bağımsız değişkenlerin bir bağımlı değişkeni nasıl etkilediğini araştırmaktadır.

LR fonksiyonu, matematikte iki faktör arasındaki ilişkinin bulunmasında matematikten, logit fonksiyonundan yararlanan istatistiksel bir modeldir.

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2.13)$$

olarak ifade edilmektedir. Denklem bir grafik üzerinde görselleştirildiğinde Şekil 2.7’de olduğu gibi, doğrusal regresyondaki doğrunun aksine S görünümüne bir eğri elde edilmektedir.



Şekil 2.7.: Lojistik Regresyon Eğrisi

Kaynak: <https://aws.amazon.com/tr/what-is/logistic-regression>

Bu fonksiyon bağımsız değişkenin değerlerinden bağımsız olarak, bağımlı değişken için yalnızca 0 ile 1 arasındaki değerleri oluşturmakta ve LR bağımlı değişkenin değerini bu doğrultuda tahminlemektedir. Lojistik regresyon yöntemleri aynı zamanda çoklu bağımsız değişken ile bir bağımlı değişken arasındaki denklemleri de modellemede kullanılmaktadır. Bu durumda birden fazla bağımsız değişken bağımlı değişkenin değerini etkilemektedir. LR’da bağımlı değişkenin sonuçlarına göre üç farklı analiz türünden bahsedilebilmektedir. Bunlar, ikili lojistik regresyon, çok terimli lojistik regresyon ve sıralı lojistik regresyondur (Taya, 2022).

İkili LR; sadece bağımlı değişkenin, 0,1 ya da evet, hayır gibi iki değeri aldığı, yani yalnızca iki muhtemel sonucun olduğu, ikili sınıflandırma tekniğidir. Değer hesaplamaları 0-1 arasında olsa da hesaplanan değer hangisine daha yakınsa ona yuvarlanmakta ve o doğrultuda sınıflandırma yapılmaktadır.

Çok terimli LR, bağımlı değişkenin kategorik olduğu, en az üç gruba sahip olduğu ve gruplar arasında doğası gereği sıralama olmadığı durumlarda kullanılmaktadır.

Sıralı LR ise, McCullagh'un 1980 yılında yazına kazandırdığı, yanıt değişkenin kategorilerinin sıralı ölçekle ölçüldüğü durumlarda uygulanan bir tekniktir. Bağımlı değişken en az üç kategoriye sahiptir ve "Kesinlikle Katılmıyorum, Katılmıyorum, Kararsızım, Katılıyorum, Kesinlikle katılıyorum" gibi kategoriler doğal bir sıraya göre ve küçükten büyüğe doğru kodlanmakta ve sayıların gerçek değerlerdense sıralamaları temsil ettiği problemler için özel birçok terimli regresyon türüdür.

Yaygın bir kullanım alanına sahip olan LR, üretim, sağlık, finans ve pazarlama gibi alanlarda kullanılmaktadır. Sağlık alanında örneğin, hastaların medikal öyküsünün veya genlerin hastalıklar üzerindeki etkisini karşılaştırmak için lojistik regresyon modelleri kullanabilmektedirler. Pazarlama alanında; online reklamcılık araçları, web ziyaretçilerinin bir reklama tıklayıp tıklamayacağını tahmin etmek için lojistik regresyon modelini kullanabilmekte ve yapılan metinsel ya da görsel reklamlara web ziyaretçilerinin davranışlarını şekillerini inceleyip, etkileşim oranlarını arttıracak reklamlar yapmaya odaklanabilmektedirler.

2.5.1.4. *Rassal Orman*

2001 yılında Breiman tarafından yazına kazandırılan Rassal Orman (Random Forest, RO) denetimli sınıflandırma tekniklerinden olup, birden çok karar ağacı üzerinden her bir karar ağacını farklı bir gözlem örneği üzerinde eğiterek çeşitli modeller üretilip, sınıflandırma yapmaya olanak tanımaktadır (Breiman, 2001).

Toplu sınıflandırma yöntemlerinden olan RO, bir sınıflandırıcı yerine birden çok sınıflandırıcı üreten ve sınıflandırma işleminde tüm ağaçların ayrı ayrı tahminlemekte ve oy çokluğu yöntemi ile sınıflayıcı kararı belirlenmektedir (Müller

ve Guido, 2017). Sınıflandırma ve regresyon çalışmalarında yüksek doğruluk oranları nedeniyle sıklıkla tercih edilmektedir.

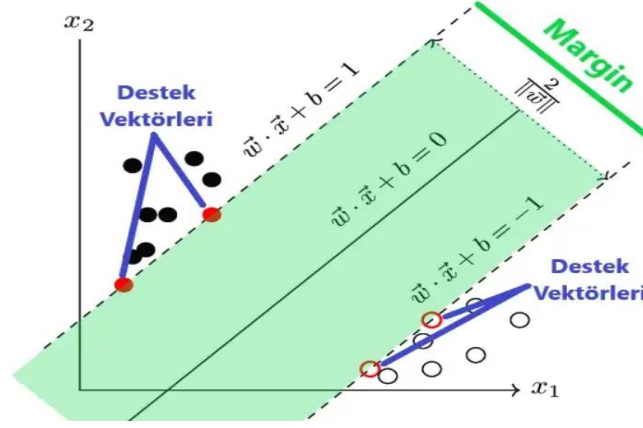
RO, diğer karar ağaçları yöntemlerinde olduğu gibi tüm değişkenler arasından en iyi dalı kullanarak her bir düğümü dallara ayırmak yerine, her bir düğümde rastgele olarak seçilen değişkenler arasından en iyisini kullanarak her bir düğümü dallara ayırır. Her bir veri seti ilk veri setinden yer değiştirilerek üretilir ve rastgele özellik seçimi kullanılarak ağaçlar geliştirilir (Akar ve Güngör, 2012).

İş akışı olarak; analiz edilecek veri seti hazırlandıktan sonra her bir örnek için karar ağacı oluşturulmakta ve her bir karar ağacı için tahmini değer çıkartılmaktadır. Tahmin sonucu oluşan her değer için yapılan oylama sonrası, en çok oyu alan üzerinden sonuç oluşturulmaktadır.

2.5.1.5. Destek Vektör Makinaları

1963 yılında Vladimir Vapnik ve Alexey Chervonenkis tarafından yazına kazandırılan Destek Vektör Makinesi (Support Vector Machines-SVM), etiketli bir eğitim verisi kümesinden girdi-çıkı haritalama fonksiyonları üreten, regresyon veya sınıflandırma amaçlı kullanılabilen denetimli bir öğrenme yöntemidir (Uncuoğlu, 2018).

Bir başka ifadeyle, eğitim verilerindeki herhangi bir noktadan en uzak olan iki sınıf arasında bir karar sınırı bulan vektör uzayı tabanlı makine öğrenme yöntemi, N boyutlu bir uzayda optimal bir hiper düzlemi tahmin etmek için eğitim veri setini kullanan ikili bir sınıflandırma tekniğidir (Schlkopf, 2018).



Şekil 2.8.: Destek Vektör Makinaları Grafiksel İşleyişi

Kaynak: <https://medium.com/deep-learning-turkiye>

Modelde, bir düzlem üzerine yerleştirilmiş noktaları ayırmak için bir doğru çizip, iki sınıfının noktaları için de maksimum uzaklıkta olmasını amaçlamaktadır. Tahmine dayalı kontrol, ses-yüz tanıma, yazı tanıma, metin-görsel sınıflandırma gibi birçok alanda kullanılmaktadır (Gour, 2019).

2.5.1.6. Naive-Bayes Yöntemi

Temin edilmiş verilerin, önceden belirlenen sınıflara ait olma olasılıklarını baz alan bir algoritmadır. Bu algoritma, bir rassal değişken için olasılık dağılımı içinde koşullu olasılıklar ile marjinal olasılıklar arasındaki ilişkiyi gösteren, istatistiki Bayes teorime dayanmaktadır. Bir başka ifadeyle, hangi verinin hangi sınıfa hangi olasılıkla ait olduklarının tahminlemesi üzerinde çalışan bir algoritmadır.

Bir sonucun gözlenmesinde, birden fazla bağımsız değişkenin olduğu durumlarda, hangi değişkenin daha etkin rol almasıyla ilgili sonucun doğması arasındaki ilişkiyi incelemektedir (Tontuş, 2020).

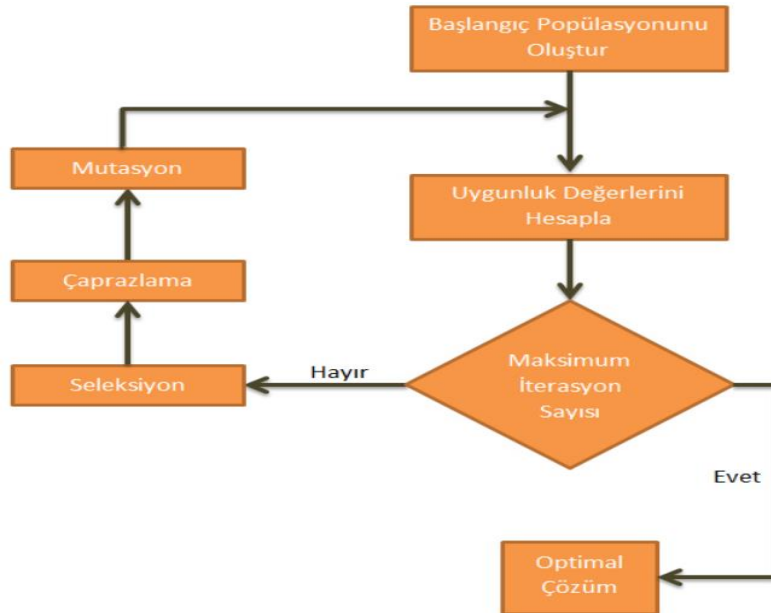
2.5.1.7. K- En Yakın Komşu

Sınıflandırma tekniklerinden, denetimli (supervised) öğrenme algoritmalarından K-En Yakın Komşu (K-Nearest Neighbor), 1967 yılında T. M. Cover ve P. E. Hart tarafında yazına kazandırılmış, örnekseme yolu ile öğrenmeye dayanan bir tekniktir. Bu yöntemde, sınıflara verilerin öğrenme kümesindeki normal davranış

verilerine benzerlikleri hesaplanarak, en yakın olduğu düşünölen k verinin ortalaması ile belirlenen eşik değere göre önceden net bir şekilde belirlenen sınıflara atamaları yapılır. Yöntemin performansını, k en yakın komşu sayısı, eşik değeri, benzerlik ölçümü ve öğrenme kümesindeki normal davranışların yeterli sayıda olması kriterleri etkilemektedir (Çalışkan ve Soğukpınar, 2008). Sınıflandırılmak istenen nesnenin ait olduğu kümeyi, en yakınında yer alan k birim nesneden en fazla birime ait olanların aynı kümede sınıflandırması mantığına dayanmaktadır (Shah ve Kursak, 2004).

2.5.1.8. Genetik Algoritmalar

1975 yılında John Holland tarafından yazına kazandırılan Genetik algoritmalar, çok boyutlu uzayda en iyileştirme amacıyla tekrarlar yapan ve her tekrarda en iyi sonucu üreten kromozomun doğal şartlara en fazla uyum sağlayanın hayatta kalması prensibine dayanan, en iyi çözümü arama yöntemi olarak tanımlanmıştır (Holland, 1992; Çiftci, 2006; Emel ve Taşkın, 2002). Genetik algoritmalara ait işleyiş akışı Şekil 2.9.'da paylaşılmıştır (Avuçlu, 2019).



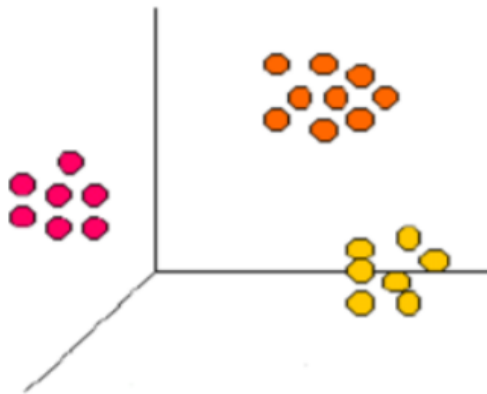
Şekil 2.9.: Genetik Algoritmaların İşleyişi

Kaynak: https://tr.wikipedia.org/wiki/Genetik_algoritma#cite_note-1

Bu teknikte öncelikle popülasyon adı verilen rastgele bir başlangıç çözüm kümesi atanarak sürece başlanmaktadır. Sonrasında alternatif çözümler ortaya konulup, ilgili çözümler birbirleriyle kıyaslanarak daha iyi performans sergileyen çözümler oluşturulmaktadır. Bu işlem tekrarlandıkça, iyi çözümlerin bulunduğu kümelerden daha iyi çözüm kümeleri hedeflenmektedir. Bu aşama, optimal çözümün, yani daha iyi bir çözüm kümesi bulunamayana kadar devam etmektedir. İlgili evrim süreci tamamlandığında örüntü kuralları veya sınıf modelleri saptanmış olması beklenmektedir.

2.5.2. Kümeleme Modelleri

Tanımlayıcı veri madenciliği yaklaşımları içinde yer alan, denetimsiz öğrenme yöntemlerinden Kümele (Clustering) temel olarak, dağınık yapıdaki bir nesne kitlesinin, nesneler arasında belirli ilişkiler kurarak daha homojen yapıdaki alt gruplara ayrılması olarak tanımlanabilmektedir. Kümeleme algoritmalarında amaç, belirli bir küme grubu içindeki üyelerin benzer olduğu fakat diğer grup kümelerinden ayrıştığı, ayrışan küme üyelerin ise kendi içlerinde benzeştiği şekilde bölünmesini sağlamaktır. Bu sayede her grup birbirinden farklı ancak grup içi nesneler birbirine benzer olacak şekilde bölme işlemi gerçekleştirilmektedir. Şekil 2.10.'da koordinat düzleminde kümeleme örneğinde görüldüğü gibi, her küme kendi içindeki üyelerce benzemekte fakat diğer kümelerden ayrışmaktadır.



Şekil 2.10.: Koordinat Düzleminde Kümeleme Örneği

Kaynak: Sarıman, 2014.

Kümeleme, verilerde bilinmeyen grupların bulunmasına olanak sağlayacağı için tanımlama ve öğrenmeye katkı sağlamaktadır. Küme analizi, iş zekâsı, görüntü örüntü tanıma, web arama, biyoloji ve güvenlik gibi birçok alanda sıklıkla tercih edilmektedir. Yazında kümeleme algoritmaları genellikle; hiyerarşik (hierarchical), yoğunluk-tabanlı (density-based) ve ızgara-tabanlı (grid-based) olarak gruplandırılabilir. İlgili yaklaşımların bazı özellikleri Tablo 2.3’de verilmiştir (Han vd., 2012).

Tablo 2.3.: Kümeleme Metotları ve Genel Özellikleri

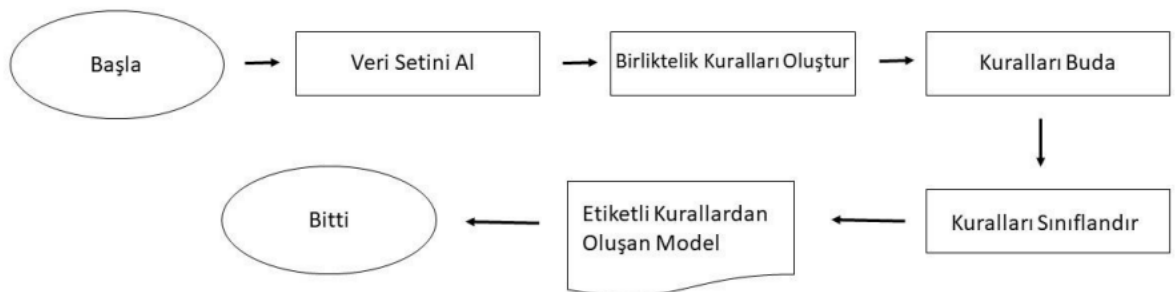
Metot	Genel Özellikleri
Hiyerarşik (Hierarchical)	- Kümeleme, hiyerarşik bir ayrışmadır (yani, çoklu düzeyler) - Hatalı birleştirmeler veya bölmeler düzeltilemez - Mikro-kümeleme gibi diğer teknikleri içerebilir veya nesne “bağlantılarını” (“linkages”) göz önünde bulundur
Yoğunluk-tabanlı (Density-based)	- Rastgele şekillendirilmiş kümeleri bulabilir - Kümeler, uzayda düşük yoğunluklu bölgelerle ayrılmış, nesnelerin yoğun bölgeleridir. - Küme yoğunluğu: Her noktanın “komşu” içinde minimum sayıda noktası olmalıdır. - Aykırı değerleri filtreleyebilir
Izgara-Tabanlı (Grid-based)	- Çok çözünürlüklü bir ızgara veri yapısı kullanır - Hızlı işlem süresi (tipik olarak dosya sayısından bağımsızdır). - Veri nesneleri, ancak ızgara boyutuna bağlıdır

Kümele yöntemleri, sınıflandırma yöntemlerinden prensip olarak farklılaşmaktadır. Sınıflama modelinde, veriler için önceden belirli sınıflar bulunmakta ve yeni gelen verinin hangi sınıfa dahil olabileceği öngörülmektedir. Kümeleme yönteminde ise, önceden tanımlı bir sınıf bulunmamakta, veriler bir model vasıtasıyla belirlenen benzerlik-yakınlık kriterlerine göre gruplar halinde kümelere ayrılmaktadır. K- Ortalamalar (K-Means), Bulanık Kümele (Fuzzy Clustering), Gauss karışım modeli gibi birçok algoritma kümeleme çözümlerinde yer almaktadır. Büyük veri tabanlarındaki verilerin kümeleneğinde etkin bir algoritma olan K-Ortalamalar sıklıkla kullanılmaktadır. Temelde işleyişi; tüm

nesneler arasından rastgele k kadar nesne seçilmesi ile başlayıp seçilen her bir nesnenin yeni oluşturulan kümelerin merkezi kabul edilmesi devam etmektedir. Sonrasında, geriye kalan tüm nesneler ilgili kümelerin merkezine yakınlıklarına göre dağıtılmaktadırlar. Üye sayısı değişen her küme için ortalama değer hesaplanması yinelenmekte ve hesaplanan yeni değer ilgili kümenin merkezi kabul edilmektedir. Bu işlem dışarıda herhangi bir değişken, nesne kalmayıncaya kadar devam etmektedir (Anand, 2003).

2.5.3. Birliktelik Kuralları ve Modelleri

Birliktelik kuralları analizi, farklı olayların birlikte gerçekleşme durumlarını inceleyen tanımlayıcı veri madenciliği metotlarındandır Birliktelik kuralları, geçmiş verilerin analiz edilerek, ilgili veriler içindeki birliktelik davranışlarının tespiti ve birlikte olma kurallarını belirli olasılıklarla paylaşarak geleceğe yönelik yapılacak tahminlemeleri desteklemeye çalışmaktadır (Agrawal vd., 1993). Birliktelik kuralı uygulaması en yaygın Pazar/Market Sepeti Analizinde yer almaktadır. Buradaki yaklaşım, Birliktelik Kuralındaki amaç; alışveriş esnasında müşterilerin satın aldıkları ürünler arasındaki bağı saptamak ve bu bağlamda müşterilerin satın alma alışkanlıklarını tespit etmektir. Mal ve hizmet sunucuları da keşfedilen değerli örüntüleri kullanarak daha etkili pazarlama, konumlandırma ve satış olanağı bulabilmektedirler (Agrawal ve Srikant, 1995). Birliktelik kuralları temelde aşağıda paylaşılan adımları kapsamaktadır:



Şekil 2.11.: Birliktelik Kuralları İşlem Adımları

Kaynak: Büyüktanır, 2022.

İlgili analizlerin yapılabilmesi için Apriori, GRI, Carma ve Sequence gibi algoritmalar bulunmaktadır. Aşağıda bazılarına kısaca değinilmiştir.

2.5.3.1. AIS Algoritması

AIS algoritması ismini, algoritmayı 1993'te yazına kazandıran Agrawal, İmielinski ve Swami'den almakta ve veri tabanları üzerinden karar vermede kullanılacak bilgiyi desteklemek amacıyla yapılan sorgulamalar için veri tabanlarındaki fonksiyonalliteyi arttırmayı hedeflemiştir.

AIS, iteratif şekilde çalışmakta, nesneleri alfabetik olarak sıralamakta, her taramada tüm işlemleri okumakta, frekansları yüksek olanları belirleyip yoğun nesne kümeleri olarak etiketlemektedir. Yoğun nesne kümelerinin etiketlenmesi sonrasında, önceki taramada frekansı yüksek nesne kümeleriyle, yapılan işlemin nesneleri arasındaki ortak nesne kümeleri saptanmaktadır. Belirlenen ortak nesne kümeleri saptanan diğer nesne kümeleriyle birleştirilerek yeni aday kümeleri oluşturmaktadır. Aday kümeleri belirlemek için de aday kümeler grubu içindeki gereksiz kümeler belirlenip silinmektedir. Geriye kalan tüm aday kümeler için destek değerleri hesaplanmakta ve önceden belirlenen minimum destek seviyesine eşit ve üstünde olanlar yaygın nesne kümesi olarak tanımlanmakta ve bu etiketi alanlar bir sonraki aday kümelerin saptanmasında yer almaktadır (Çınar ve Silahtaroglu, 2015).

2.5.3.2. Apriori Algoritması

Apriori algoritması, birliktelik kuralları arasında en çok kabul gören ve kullanılan algoritmadır. Apriori, yaygın nesnelerin bilgilerini bir önceki (prior) adımdan alması mantığıyla işlemektedir (Agrawal ve Srikant, 1994). Temel olarak iteratif bir niteliğe sahiptir ve hareket bilgileri içeren veri tabanlarında sık geçen öge kümelerinin keşfedilmesinde, verilerden bir dizi kural çıkarmayı ve en yüksek bilgi içeriğini sağlamaya çalışmaktadır. Modelde girdi ve çıktıların kategorik olması gerekmektedir. Aday nesnelerin oluşturulmasında sadece bir önceki taramada

geniş olarak etiketlenen nesne kümelerini kullanmaktadır. Buradaki mantık, geniş bir nesne kümesinin yine geniş bir nesne kümesinden sağlanabileceği kabulüne dayanmaktadır. Bu doğrultuda ilk adımda, veri setinden her bir nesnenin frekans ve destek değerini hesaplanmaktadır. İkinci adımda, daha önce belirlenen minimum destek değerine eşit ya da üstünde destek değerine sahip nesneler ile yeni bir tablo oluşturulmaktadır. Son aşamada ise oluşturulan son tablo üzerinden kurallar çıkarımı yapılmaktadır (Döşlü,2008).

2.5.3.3. CARMA Algoritması

Christian Hidber (1999) tarafından yazına kazandırılan CARMA (Continuous Association Rule Mining Algorithm) Sürekli Bağlantı Kuralı Veri Madenciliği Algoritması, çevrim içi küme hesaplamalarını yapan ve temelde kullanıcıya mevcut birliktelik kurallarını gösteren bir tekniktir. Dizi modeli – ilişkilendirme kurallarını sırasız veya zamana dayalı olarak keşfeder. Dizi, öngörülebilir bir sırada meydana gelme eğiliminde olan öge kümelerinin bir listesidir. Örneğin, bir tıraş bıçağı ve tıraş losyonu satın alan bir müşteri, bir sonraki alışverişinde tıraş kremi satın alabilir. İşleyiş olarak her taramadan sonra oluşan küme kurallarını kullanıcıya sunar ve veri tabanının ilk taramasındaki bir işlemde minimum destek ve güven seviyelerini güncelleme fırsatı tanır. Veri tabanı üzerinden 2 taramaya izin verir. Birinci taramada nesne kümelerini ortaya koyar ve ikinci taramada her nesne kümesi için destek değeri belirlenmekte ve yoğun kabul edilmeyenleri elemekte ve ortaya çıkan toplam nesne kümelerinin sayılmasını gerçekleştirmektedir (Karabatak, 2008: 37).

3. BÖLÜM

MÜŞTERİ KAYIP ANALİZİ VE VERİ MADENCİLİĞİ

Bu bölümde VM ve MKA, sağlık sektöründe VM ve sağlık sektöründe VM aracılığıyla MKA konuları yazın ışığında irdelenmiştir.

3.1. Veri madenciliği ve Müşteri Kaybı Analizi

MKA'nın özellikle müşteri hizmet alım devamlılığı bulunan bankacılık, sigortacılık ve telekomünikasyon gibi sektörlerdeki uygulamaları uzun yıllardır ve yaygın olarak yabancı ve yerli yazında yer almaktadır. Ayrıca son yıllarda e-ticaret, otelcilik ve perakendecilik gibi diğer sektörlerde de MKA çalışmaları yer almaya başlamıştır. Bu çalışmalar:

Yabancı yazında;

Telekomünikasyon alanında yapılan bir çalışmada, rastgele seçilen 5000 abonenin 11 farklı parametre ile (3G, Toplam Kullanım, Fatura bedelleri, SMS yerel-uluslararası vs.) 3 aylık zaman dilimlerinde churn olma durumu yapay sinir ağları aracılığı ile incelenmiştir. Müşteri kayıplarında, uluslararası arama hakları, ortalama konuşma süresi ve 3G hizmetinin belirleyici olduğu saptanmıştır (Adwan, vd., 2014). Yine telekomünikasyon alanında yapılan bir çalışmada; müşterilerin demografik özellikleri, telekom hizmetlerini kullanım durumları, paketler, ödemeler ve churn durumları arasındaki ilişkiler, kümeleme ve karar ağaçları algoritmalarından CHAID (Ki-Kare Otomatik Etkileşim Detektörü) aracılığı ile değerlendirilmiş ve tahmin modelleri çalışılmıştır (Pejić, Pivar ve Jaković, 2021).

E-ticaret alanında faaliyet gösteren bir işletme için yapılan bir çalışmada ise müşterilerin ilgili uygulamayı kullanma süreleri, yaptıkları son iki siparişleri arasındaki süre farkı, demografik özellikleri, sistemdeki şikayetleri ile kayıp müşteri kabul edilme durumları karar ağaçları, LR ve RO algoritması ile incelenmiştir. RO algoritmasının en iyi tahminlemede bulunduğunu saptanmıştır (Alshamsi, 2022). Müzik yayını sağlayan bir işletme için ise müşteri kayıplarının en iyi tahminini yapacak VM modelini belirleme çalışması yapılmıştır. Çalışmada; ödeme metodu, ödeme planları, otomatik yenileme seçeneğinin tercih edilip

edilmemesi, paket kayıt günü, paket iptal günü, üyelik geçerlilik bitiş tarihi gibi etmenler LR, naive bayes, K-en yakın komşu ve RO algoritmaları ile karşılaştırılmıştır. En iyi tahmini sağlayan algoritmanın RO algoritması olduğu görülmüştür (Gaddam ve Kadali, 2022).

Sigorta sektöründe, araç sigortaları üzerine yapılan 2014-2019 yılları için 97.696 müşteriyi içeren bir çalışmada; müşterilerin yaşı, medeni durumu, yaşadığı bölge gibi demografik özellikler ile aracın markası, araç tarzı, aracın sisteme dahil olduğu zamanki yaşı, aracın finanse edilme biçimi, poliçe ayrıntıları vb. değişkenler kullanılmıştır. Müşterilerin churn olma durumları için kümeleme algoritmalarından K-ortalama ile çalışmış ve churn olma durumu ile değişkenler arasındaki bağlar incelenmiştir. Hangi müşterilerin sistemde daha uzun süre kaldıkları, müşterilerin kaybedilmesindeki risk faktörlerinin neler olduğu saptanmıştır (Bravente ve Robielos, 2022).

Etopya Ticaret Bankasının müşterilerine ait cinsiyet, mesleki alan, mobil bankacılık durumu, kredi kartı işlemleri, bankanın kaç yıldır müşterisi olduğu gibi 11 farklı değişken ile toplam 204,161 veri üzerinden kayıp müşteri kabul edilme örüntüleri denetimli veri madenciliği metotlarından LR, RO ve DVM, K-en yakın komşu, derin öğrenme / YSA aracılığıyla incelenmiş ve en iyi tahminlerin derin öğrenmede sağlandığı görülmüştür (Seid ve Woldeyohannis, 2022).

Yerli yazında;

Yerli yazında da çeşitli çalışmalara rastlanılmaktadır. Kişioğlu ve Topçu tarafından 2011 yılında Türkiye'deki telekomünikasyon operatörleri üzerine bir çalışma yapılmış, Bayes ağları yöntem olarak seçilmiş ve oluşturulan nedensel dizilimde, abonelerin fatura bedelleri, ortalama konuşma süreleri, internet paket hak ve kullanımları, diğer operatörleri arama süreleri gibi parametrelerin mevcut tarife iptallerinin en belirgin nedenleri olduğunu saptanmıştır. Benzer bir çalışma V. Gülpınar (2013) tarafından yapılmıştır. Çalışmasında, özellikle numara taşıma serbestliği sonrası, Türkiye'de telekomünikasyon operatörleri üzerine, müşteri kaybını tahminleme ve sonrasında müşteri ağını analiz etmeyi amaçlamıştır. Tahminleme de YSA, müşteri iletişim ağına analizinde ise Sosyal Ağ Analizi (SAA) yöntemlerini kullanmıştır. Yaş, cinsiyet, medeni durum, eğitim durumu,

gelir düzeyi, hangi GSM operatörünü kullandığı, kaç yıldır mevcut operatörden hizmet aldığı, paket türü, ortalama görüşme süresi vb. girdiler ile müşterilerin gelecekte aynı GSM operatörü ile devam etmeyi düşünüp düşünmedikleri çıktısı ile arasındaki bağları çalışmıştır.

Dur R., Koçer S. ve Dünder Ö. (2021) ise çalışmalarında bankacılık sektöründe kısa mesaj hizmeti (SMS) ile müşterilerin SMS hizmetini bloklama durumunu, 29.635 müşteriye ait 22 öz nitelikler (yaş, eğitim, gelir, medeni durumu, meslek vb.) kullanarak sınıflandırma amacı ile incelemiştir. Müşteri veri setini regresyon, YSA ve DVM ile sınıflandırmış ve performanslarını doğruluk, kesinlik ve duyarlılık ölçütleri ile karşılaştırmışlardır. İlgili algoritmalar yakın sonuçlar verse de LR'ın daha iyi sonuç verdiğini gözlemlemiştir. Bankacılıkta yapılan diğer bir çalışmada ise bankaların personel sayısı, şube sayısı, bilançolarındaki toplam mevduatları ve kredileri girdi olarak, bankaların mudi sayıları ise çıktı olarak ele alınmıştır. Aralarındaki ilişki Bankacılık Düzenleme ve Denetleme Kurumu tarafından paylaşılan Haziran 2010 ve Ekim 2020 arasındaki aylık dönemleri kapsayan 2.500 adet veri üzerinden YSA ve DVM regresyonu ile incelenmiştir (Yetiz, vd., 2021).

Sigortacılık sektöründe yapılan bir çalışmada ise şirketin var olan müşterilerine ait yaş, cinsiyet, doğum yeri, kullanılan araç marka ve model bilgilerinin de bulunduğu öznitelikler ile kayıp müşteri kabul edilme durumu arasındaki ilişkiler Karar Ağaçları, RO ve k-en yakın komşu algoritmaları ile incelenmiştir. En başarılı sonucu veren algoritmanın RO algoritması olduğu gözlemlenmiştir (Akyiğit ve Taşcı, 2022). Bir diğer çalışma; Kilimci (2022) tarafından bankacılık, sigortacılık ve telekomünikasyon sektörlerindeki kayıp müşterileri tahminlemede hangi algoritmanın daha güvenilir tahminleme modeli, sonuçları verdiği üzerine yapılmıştır. Kurulan modelin en yüksek sonuçları bankacılıkta %89,93, sigortacılıkta %77,53 ve telekomünikasyonda %95,90 ile rassal orman algoritmasının sağladığı saptamıştır.

Çallı ve Kasım (2022), genellikle işletme ve müşteri arasındaki (B2C) bağları inceleyen churn çalışmasını, işletmeler arasındaki (B2B) kaybı tahminlemeye yönelik olarak, bir ERP (Kurumsal Kaynak Planlama) şirketi üzerinden, 10 farklı

algoritma üzerinden çalışmışlar ve RO algoritmasının en iyi sonuç verdiğini, ürün çeşitliliğinin ve müşteri karakteristiklerinin en belirleyici unsur olduğunu saptamışlardır.

Bir telekomünikasyon şirketinden SQL aracılığı ile elde edilen iki milyon müşteriye ait ilgili operatördeki ömrü, cinsiyet yaşı, medeni durum, telefon cihaz bilgisi gibi öznitelikler ile tarifenin ömrü, paket aşım miktarı, müşterinin veri indirme hızı, kapsama alanı dışında kaldığı süre vb. birçok değişken ile kayıp müşteri kabul edilme durumu LR, RO, YSA ve XGBoost modelleri ile Python ortamında çalıştırılmıştır. Yapılan karşılaştırma sonrası en başarılı algoritmanın XGBoost'a ait olduğunu, RO ve YSA'nın LR'a göre daha başarılı sonuçlar verdiğini saptamışlardır (Şenyürek, 2019).

Donat (2021), farklı bir telekomünikasyon şirketi üzerine yaptığı çalışmada; abone yaşı, abonelik yaşı paket aşımı, mobil tv kullanımı, son 6 aylık ortalama internet kullanımı ve son 1 aylık şikâyet sayısı gibi 26 değişken üzerinden kayıp müşteri olma durumunu LR, derin öğrenme, DVM, bayes ağları ve YSA algoritmaları karşılaştırılmış ve en başarılı tahmin sonucunu veren algoritmanın derin öğrenme olduğunu saptamıştır.

Kaba, A. (2021) yılında faizsiz finans sektöründeki öncü bir işletme üzerinde yaptığı çalışmasında, 18.507 müşteriye ait medeni durum, meslek grubu, sisteme dahil olduğu zaman, kampanya bedeli, kampanya tipi, vade gibi 14 farklı değişken ile kayıp müşteri kabul edilme durumunu LR, K-en yakın komşu ve DVM aracılığı ile incelemiş ve en başarılı tahminlemeyi LR algoritmasının verdiğini gözlemlemiştir.

Cengizci A.D., (2020), 3 farklı şubesi bulunan bir zincir otel işletmelerinde kayıp müşteri tahminlemesi üzerine çalışma yapmıştır. Müşterilere ait satın alımlar arasında geçen zaman, son üç yıldaki gecelik harcama tutarları, konaklama süresi ve sosyodemografik öz nitelikler ile LR ve RO algoritmaları aracılığı ile incelemiş ve RO'nın daha başarılı tahmin sonuçları verdiğini saptamıştır.

3.2. Sağlık Sektöründe Veri Madenciliği

Sağlık alanında, veri tabanlarında depolan verinin miktarı ve kalitesi gittikçe artmış ve teknolojik gelişmeler ışığında veri madenciliği kullanımı bir hayli yaygınlaşmıştır. Tanımlama ve tahminleme olanağı sunan modeller ve algoritmalar, veri tabanlarına işlenen veriden, anlamlı, faydalı örüntüler çıkarabilmekte ve geleceğe ışık tutabilecek kestirimlerde bulunabilmektedir. Büyük veri üzerinden yapılan veri madenciliği çalışmaları doğrultusunda aşağıdaki bazı konularda çalışmalar yapılabilmektedir (Çeltikçi, 2018):

- Aynı tip hastalığa sahip insanlar ve öznitelikleri arasında bağ kurulması ve sağlıklı bireyler için hastalık olasılıklarının tahminlenmesi,
- Erken evrede tanılama ve dolayısıyla tedavi fırsatı sunulması, örneğin, kronik hastalıklar için erken uyarı sinyallerinin veri madenciliği ile tespit edilmesi,
- Klinik deneylerin ve hasta kayıtlarının analizi ile ürünler pazara verilmeden önce endikasyonların daha iyi ve hızlı tanımlanabilir ve yan etkileri daha net anlaşılabilir olması,
- İleriye yönelik hastalık tahminleri yapılarak bulaşıcı ya da salgın hastalıklara ilişkin koruyucu tedbirlerin (aşı, ilaç) hızlandırılması,
- Klinik, cerrahi ve genomik veriler birleştirilerek hastalık riski ve en uygun tedavi yöntemlerin belirlenebilmesi,
- Reçete ve diğer hizmetlerin suistimalini önleyici tedbirler geliştirilmesi,
- Laboratuvar testleri, reçeteler ve diğer hizmetler için hata ve suistimal tespiti ve tedbirler geliştirilmesi.

Koyuncugil ve Özgülbaş (2010), yayınladıkları makalelerinde yukarıda geçen potansiyel veri madenciliği çalışma alanlarına hangi yöntem ve modellerini uygulanabileceği üzerinde durmuşlardır. Örneğin, kronik hastalıklar çalışmasında erken uyarı sistemi için Önemli Bileşenler Analizi, Faktör Analizi veya LR kullanım önerilerinde bulunmuş ve örüntülerin saptanması sonrasında risk sinyali olarak kullanılabileceğini paylaşmıştır. Benzer şekilde VM Klinik Karar Destek

Sistemlerinin geliştirilmesinde kullanılmaktadır. Burada, veri tabanındaki büyük ölçekteki veriden, uygulanan tedaviler, hastalara ait sosyodemografik öznitelikler ile tedavi sonuçları arasındaki örüntüler saptanabilir ve tanı, teşhis aşamasında ve tedavi seçenekleri aşamasında doktorlara destek sağlayabilmektedir. Karar destek sistemleri makina öğrenmesinin de gelişimiyle yönetsel süreçlerde de karşımıza çıkabilmektedir. Burada, sağlık işletmelerinde yönetsel amaçlı kullanılabilecek tüm değişkenler çok boyutlu olarak ele alınabilmekte, optimal değerler ve yol haritaları belirlenebilmektedir. Yöntem ve model olarak da tanımlama yaklaşımlarında; Önemli Bileşenler Analizi, Faktör Analizi, Regresyon, YSA kullanımı önerilmektedir. Verimlilik, kalite ve risk göstergelerinin belirlenmesinde, Karar ağaçları, Birliktelik Kuralları, K-ortalamlar Kümeleme Analizini takiben LR veya YSA kullanılabilmektedir.

Sağlık alanında VM yukarıdaki yaygın kullanımı yanı sıra, işlem başı maliyetler ve maliyete etki eden faktörler, geleceğe yönelik hasta yoğunluk tahminlemeleri, finansal performans üzerinden riskler için erken uyarı sistemi geliştirilmesi gibi alanlarda da kullanılmaya başlanmıştır. Sağlık alanındaki sigortacılıkta, sosyodemografik, prim tutarları, hastanın cepten yaptığı ödemeler, tazminat bedelleri gibi geçmiş veriler ve hizmet alımı sonlanan sigortalılar arasındaki örüntüler LR, karar ağaçları ile incelenebilmektedir. Yazında yer alan bazı çalışmalar aşağıda paylaşılmıştır.

VM'de kullanılan modellerin akciğer kanseri veri seti üzerinde başarıları incelemek için bir yapılan bir çalışmada, öncelikle onkoloji alanında çalışan hekimlerin görüşleri alınmış, akciğer kanseri alanında uyumlu veri seti için hazırlık yapılmış ve buna uygun olabilen veri madenciliği basamaklarını uygulamışlardır. Adım adım ilk önce ön işleme ve veri temizleme, veri indirgeme, veri dönüştürme ve veri madenciliği işlemlerini uygulanmıştır. Sonuç olarak, bu çalışma akciğer kanseri teşhisinde tahminde bulunabilecek ve bu sayede hastalık teşhis süresi kısalabileceğine dair verileri sunmuş, WEKA veri madenciliği yazılımı ile farklı algoritmalar (Naive Bayes, BayesNet, LR, Multilayer Perceptron, KStar, Bagging, OneR, ZeroR, J48 ve Random Tree) uygulanmıştır. Yapılan değerlendirmede en başarılı algoritmanın Naive Bayes olduğu saptanmıştır (Sebik ve Bülbül, 2018).

Bir diğerk çalışmada, sedef hastalığı, seboreik dermatit, liken planus, pityriasis rosea, kronik dermatit gibi dermatoloji hastalıklarını, beş tür veri madenciliği yöntemi (CART, RO, Karar Ağacı, DSM, Degrade Artırıcı Karar Ağacı) kullanılarak analizi yapılmıştır. Analizlerde, 35 cilt hastalığı değişkeni ve 360 örneklem kullanılan çalışmada girdi verileri değiştirilmiş, üç kez yöntemler tekrarlanmış ve en yüksek doğruluk oranı %98,64 olarak çoklu öğrenme algoritmaları ile saptanmıştır (Verma, vd., 2019).

Karabulut (2021) çalışmasında, 2010- 2020 yılları arasında özel bir hastanenin 23 farklı şubesinden 340.900 hastaya ait cinsiyet, yaş, sigorta bilgisi, medeni durum, ikamet ettiği şehir gibi 10 değişkenden oluşan ayrıntılı veri setini kullanarak, VM Karar Ağacı algoritmalarından CHAID, Quest ve C5.0 ile analiz çalışması gerçekleştirilmiştir. Hizmet alan bireylerin ve hizmet aldığı şubelerin özellikleri kategorize edilerek, hasta profillerini belirlemeye çalışmıştır.

Özyazar (2019) yaptığı retrospektif çalışmasında 1.160 tip 2 diyabetli hastanın demografik verileri, biokimyasal ve antropometrik verileri, insülin kullanımı, özgeçmiş, soygeçmiş, hipertansiyon hastalığı teşhisi, makro ve mikro hastalıkların teşhisi, hastalık süresi gibi nitelikleri değişken olarak belirlemiş ve C4.5 karar ağacı algoritması ile sınıflandırma yapılarak makrovasküler ve mikrovasküler komplikasyonlar için risk modelleri oluşturmayı amaçlanmıştır.

Bir hastanede yoğun bakım hastalarının mortalite ve yoğun bakımda yatış sürelerine bu süreçte hangi faktörlerin etkili olduğunu analiz etmeyi amaçlamış, 4.233 hastanın verisi üzerinden, veri madenciliği tekniklerinden, YSA, Naive Bayes, RO, C4.5, CART, LR, DVM ile incelenmiş ve en iyi sonuç veren modelin RO olduğu saptanmıştır (Sülekli, 2019).

Cissé ve Yılmaz (2022), yaşlıların günlük yaşamlarında destekleyici, hatırlatıcı mobil sağlık uygulaması geliştirilmesi çalışmasında Java yazılım ve SQLite veri tabanında faydalanmıştır. Kullanıcılardan topladığı veriler doğrultusunda, yaşlılara ve özellikle alzheimer gibi unutkanlık sorunu yaşayan yaşlılara ilaç kullanım zamanlarını, muayene kontrol zamanlarını hatırlatan bir program üzerinde çalışmışlardır.

Yapılan başka bir çalışmada sağlıklı bireylerin ilerleyen zamanlarda bazı kanserlere (meme, kolon, akciğer) yakalanma riski üzerinde yaş, cinsiyet, ergenlik yaşı, menapoz yaşı, ilk anne olma yaşı, alkol tüketimi gibi öznitelikler ve bulanık mantık yöntemi ile tahminlerde bulunmuş, meme kanserinde %80,83, akciğer kanserinde %80, kolon kanserinde %82,72 oranlarında başarı elde etmişlerdir. Çalışmanın sonucuna bağlı olarak, insanların bu riskleri göz önünde bulundurarak kansere yakalanmadan önce bazı tedbirler alabileceklerini ve bu sayede kanser olma olasılıklarını düşürebileceklerini ortaya koymuşlardır (Atınç ve Ayan, 2013).

Veri kaynağı olarak, SEER (Surveillance Epidemiology and End Results) kümesinden alınan 60.948 örnek üzerinde yapılan kanser türleri hakkında yapılan çalışmada verilerin %66'sı eğitim kümesi ve %34'ü test kümesi olmak üzere sınıflandırma modellemesinde kullanılmıştır. Scikit-learn ile modelleri oluşturmak için python programı kullanılmıştır. Scikit-learn ile elde edilen en iyi sonuca %91,84 doğrulukla Karar Ağaçları algoritması kullanılarak ulaşılmıştır. Tanagra ile en iyi sonuç 87.75 doğrulukla RO ile ulaşılmıştır. Weka, Tanagra ve Scikit-learn'de aynı bilgisayarda yedi sınıflandırma algoritması çalıştırıldığında en iyi sonucun 91,84 doğruluk oranıyla karar ağacı algoritması ile Scikit-learn aracı ile yapılan test olduğu görülmüştür. Ardından Tanagra ve WEKA takip etmiştir. Bu çalışmada kullanılan kaynaklar karşılaştırıldığında Scikit-learn en iyi sonuç veren araç olduğu görülmektedir (Çoğun, vd., 2022).

Alan (2019) çalışmasında, hastanede kayıt altında tutulan veriler arasında birliktelik kuralı olup olmadığını araştırmıştır. WEKA programı kullanarak hastanedeki servislere başvuran hasta sayısı ve hangi servislere başvurdıklarını Apriori algoritması ile analiz etmiştir ve yönetici kararlarına destek sağlamak üzere bulguları paylaşmışlardır.

Yapılan bir çalışmada hastanelerin gelecekteki hasta yoğunluklarının veri madenciliği yöntemleri ile tahminlenmesi amaçlanmış, hastane veri tabanı aracılığı ile aylık hasta kabul başvurusu için veri transferi, filtreleme ve veri ön-işleme faaliyetleri gerçekleştirilmiş sonrasında zaman serileri ve YSA teknikleri kullanılarak modellemeler yapılmıştır. YSA, ARIMA, üstel düzgünleştirme önce

kendi alt modelleri ile sonrasında her yöntemin tahminlemedeki en başarılı sonucu veren modelleri kendi aralarında karşılaştırılmıştır. En başarılı tahminleme sonucunu veren modelin üstel düzgünleştirme yöntemlerinden Winters Additive olduğu saptanmıştır (Irmak vd., 2012).

Gülsevin ve Türkan (2012), Sağlık Bakanlığı'na bağlı hastanelerin 2011 yılı verilerini kullanarak verimlilik performanslarını tanımlamak amacıyla çalışmayı yürütmüşlerdir. Bu çalışmada veri zarflama analizi kullanılmıştır. 15 hastanede analizler yapılmış ve performansı daha düşük olan hastaneler için önerilerde bulunulmuştur.

Bir devlet hastanesinde mevcut kadro yönetiminin iyileştirilmesi üzerine yapılan çalışmada 7 yıllık kayıt altına alınmış veriler üzerinden hastaların kaç kez hastanenin poliklinik ve klinik bölümlerine başvurduğu, hangi bölümlere yönlendirildiği Access veri tabanında analiz edilmiştir. Çalışmanın sonucunda, mevcut sağlık kadrosu ile maksimum faydanın nasıl sağlanacağına dair yönlendirmede bulunulmuştur (Özdemir, vd., 2010).

3.3. Sağlık Sektöründe Müşteri Kayıp Analizi

Sağlık kurumlarında MKA'nın yapılabilmesi için hizmet alanlara ait öznitelikler, hangi hizmet alıcılarının devamlı, hangilerinin kayıp kabul edilme durumunun belirlenmesi gerekmektedir. Ayrıca, MKA'nın kavramsal çerçevesi doğrultusunda, aşağıda öncelikle Türkiye'de hasta tercih eğilimlerinin gelişimi değerlendirilmiş ve sağlık hizmet alıcılarının müşteri mi yoksa hasta mı olarak anılmasının gerektiği kavramsal tartışması verilmiştir. Sonrasında sağlık kurumlarında MKA ve uygulama alanları paylaşılmıştır.

3.3.1. Türkiye'deki Sağlık Sektörünün ve Hasta Tercihlerinin Gelişimi

Sağlıkta dönüşüm programı (2003-2011) ve özel hastanelerin önlerinin açılmasıyla, (teşvik edilmesiyle birlikte) sektörün ve özel hastanelerin yapısında ve hastaların talep ve beklentilerinde değişimler gözlenmeye başlanmıştır. Devlet hastanelerinin, hasta taleplerini yeterince karşılayamaması, özel sektörün devlet politikalarıyla önünün açılmasıyla birlikte, ülkemizdeki özel sağlık hizmet

sunucularının sayısının hızla artması, tıbbi olarak başarılı hekimlerin devlet hastanelerinin yanı sıra, çalışma koşullarındaki avantajlar nedeniyle özel kurumları tercih etmeye başlaması vb. değişimler hastalar açısından gelirleri doğrultusunda farklı alternatiflere ulaşma imkânı sağlamıştır. Bu değişimlerle birlikte, önceleri hastaların özel hastaneleri tercih etmelerindeki en temel neden hekim tercihiydi. Özel sağlık hizmet sunucularının çoğalması, rekabetin artması ile birlikte hasta tercihlerinde hekimin yanı sıra kullanılan tıbbi teknoloji de hasta tercihlerindeki en önemli parametrelerden olmaya başlamıştı. Özel hastanelerin mali açıdan da güçlenmesi, hasta tercihindeki önemi nedeniyle, tıbbi teknoloji de birçok hastane için standart bir hizmet sunum aracı haline gelmişti. Hekim (çok spesifik hastalıklar ve geçmişten gelen takipte olunanlar hariç) ve tıbbi teknoloji beklentisi karşılanan hastaların, gelir düzeylerine göre özel hastaneleri seçmelerindeki etkin faktör ise hastanelerin fiziki ferahlığı, konforu ve tıbbi ve idari personel ile olan iletişim başlıkları olmuştur.

Günümüzde geldiğimizde; artık binaların fiziki koşulları, konforu gibi başlıklar büyük ölçüde rekabet gücünü gidermiş, tıbbi –idari personel ile olan iletişim (hasta karşılama, uğurlama vb.) yavaş yavaş sağlık hizmet rakipleri arasında standarda yaklaşmıştır. Dolayısıyla, artık hasta tercihlerini belirleyecek olan rekabet avantajının; hastaların tedavilerinin – aşamalarının bireysel takibi, medikal koçluk, kişisel sağlık sunumu vb. alanlarla sağlanacağı öngörülmektedir. Burada; hasta tercihinin belirleyecek etmenlerin azalması, alternatiflerin kat ve kat artarak devam etmesi (Özel kurumlar, şehir hastaneleri vb.) nedeniyle mevcut hastaların memnun edilmesi, elde tutulması işletmelerin devamlılığı için çok büyük bir önem arz etmektedir. Özel sağlık sektörü için, hastaların doğru tanımlanması, taleplerinin tahmin edilmesi ve kayıplarının önlenmesi çok kritiktir ve ancak sistematik bir yaklaşım ile sürekliliğinin sağlanabileceği düşünülmektedir.

3.3.2. Sağlık Kurumlarında Müşteri Kayıp Analizinin Önemi

Müşteri kaybı tahmin modeli, temelde özellikle hizmet odaklı endüstrilerde ve rekabetçi bir iş ortamında başarılı bir müşteri ilişkileri yönetimi gerçekleştirmenin etkili bir yolu olarak kabul edilmektedir. Günümüzde daha çok telekomünikasyon,

bankacılık ve sigortacılık gibi abonelik esasında çalışan endüstrilerde yaygın olarak kullanılmaktadır. Turizm, otelcilik ve sağlık gibi sektörlerde sistematik bir analiz olarak MKA henüz uygulamada yerini almamıştır. Özellikle hastanecilik alanında; sağlık hizmet sunucuları, hastalarının tercihlerini, ne zaman, hangi durumlarda hizmet sunucusu değiştirdiklerini analiz edememektedirler. Kimi kuruluşlar; hastalarına aldıkları hizmet sonrası memnuniyet anketleri göndererek, kimisi telefonla arayarak, kimisi de basılı formların doldurulmasını isteyerek hastalarını anlamaya çalışmaktadır. Her özel sağlık kurumu; mevcut hastalarını muhafaza etmek, memnuniyetleri ölçmek ve sağlamak için farklı politikalar izliyor olsalar da temelde reaktif uygulamaların yani hasta geri bildirimleri sonrası alınan aksiyonların ötesine çok fazla geçemediği görülmüştür. Dolayısıyla, mevcut hastalarını korumakta ve onların ömür boyu katacağı değerlerden mahrum kalma riski ile karşılaşmaktadırlar. Ayrıca, yeni müşteri kazanmanın, mevcut müşteriye mutlu edip elde tutmaya göre beş ile on kat daha maliyetli olduğu (Kotler ve Armstrong, 2017; Gallo, 2014), memnuniyetsiz ayrılan müşterilerin %96'sının şikayet başvurusunun olmadığını fakat deneyimlerini en az 15 kişi (sağlık alanında çok daha fazla olduğu öngörülmektedir) ile paylaştıklarını (Beal, 2014) ve müşterilerinin %92'sinin ağızdan ağıza pazarlamayı diğer pazarlama-iletişim kanallarını kullanmaya tercih ettiklerini, tanıdıklarından duydukları ya da müşteri bloglarından okumuş oldukları müşteri deneyimlerini daha güvenilir bulduklarını ortaya koyan çalışmalar (Nilsen, 2012), MKA'yı sağlık kurumları için daha da kritik hale getirmektedir.

3.3.3. Sağlık Kurumlarında Müşteri ve Hasta Kavramsal Çerçeve

Türkçemizde “client” ve “customer” kavramları her ne kadar “müşteri” olarak belirlenmiş ve kullanılıyor olsa da aslında “client” karşılığı hukuk, muhasebe-finans gibi resmi ve profesyonel hizmet alıcıları için, “customer” ise diğer iş yerlerinden veya mağazalardan mal ve hizmet alıcıları için kullanılmaktadır. “Consumer”- “tüketici” kavramı ise bir ürün ya da hizmeti nihai kullanım amacı ile satın alan kişiyi ifade etmektedir. Tüketici son kullanıcıdır, müşteri ve tüketicinin aynı kişi olduğu durumlar olsa da farklılık gösterebilmektedir. Müşteri, genel

anlamı ile belirli bir işletmeden düzenli aralıklarla mal ve/veya hizmet alan kişiyi ifade etmektedir.

Sağlık hizmet sunucularından hizmet alanlar hasta mı yoksa müşteri olarak mı anılmalı sorusu önemlidir. Sağlık sektöründe, medikal hizmet için başvuran hizmet alıcıları, genel olarak hasta olarak anılmaktadır. Diğer yandan, hastaların da bir hizmet alım devamlılığı arz etmesi, seçenekli sunucular arasından seçim yapmaları hastaların “müşteri” olarak da anılmasında/değerlendirilmesinde rol oynamaktadır. Sağlık sektöründeki hizmet alıcılarının kavramsal tanımlanması üzerine yapılan bir çalışmada (Costa, vd., 2019); İngilizce’de “customer, consumer, client, survivor, patient” kelimeleri üzerine 13 ülkede (Amerika, İngiltere, Kanada vs.) 1.522 makale üzerinden incelemeler yapılmış, tarama ve uygunluk aşamaları sonrası 105 çalışma uygun bulunmuş ve 47 çalışma niceliksel araştırmaya dahil edilmiştir. Yapılan çalışmada nihai olarak; “patient (hasta)” kavramının sağlık hizmet alıcıları için en çok kabul gören kavram olduğu saptanmıştır.

Türkiye’de yapılan birçok çalışmada (Soysal vd., 2017; Kısa ve Tokgöz, 2007); hastalar müşteri olarak değerlendirilse de hastanelerin karşılama-kabul birimleri günümüzde hasta hizmetleri, misafir hizmetleri, hasta kayıt birimi gibi isimler ile anılmaya devam etmektedir. Sağlık hizmet alıcıları açısından değerlendirildiğinde ise medikal açıdan müşteri olarak görülmek istemedikleri, beklentilerinin onları dinleyen, maddi kazanç sağlamaktan ziyade sağlıklarını iyileştirmeyi amaç edinen, doğru yönlendirmeyi yapan, özetle içinde bulundukları rahatsızlıktan veya hastalıktan kurtulmaları için gereken empati içerikli yaklaşım olduğu görülmektedir (Aydan, 2020).

Diğer yandan; hastalar hizmet alım deneyimleri açısından değerlendirildiğinde hizmet sunum ağının iyileştirilmesi, oda hizmetleri, güler yüzlü karşılama- hizmet sunum, konforlu alanlar vb. rekabet ortamında farklılık getirecek başlıklarda hizmet sunumunda hastaları müşteri perspektifinden değerlendirmek, deneyimleri geliştirecek, memnuniyetlerini arttıracak faaliyetlerde bulunmak gerekmektedir. Bu yaklaşımlar, nihai olarak medikal hizmet sunumuna katkı sağlayacak ve hastanın tedavi bütünlüğünde rol oynayacaktır. Ayrıca, medikal

hizmet alan hastalar dışındaki, refakatçi, kafeteryadan hizmet alan hasta yakını, anlaşmalı kurumlar gibi diğer paydaşlar hastanenin müşterileri olarak da değerlendirilebilmektedir (Kavuncubaşı, 2000). Çalışmamızda ise sağlık hizmet sunucularından medikal hizmet alanların “hasta” olarak, medikal hizmet dışındaki diğer hizmetlerdeki memnuniyetinin sağlanmasında müşteri yaklaşımının benimsenmesinin etkin olacağı ve hizmet alım devamlılığı olanların “devamlı hasta” olarak anılması daha doğru bulunmuş olup ileriki kısımlarda müşteri kavramı yerine hasta kavramı kullanılacaktır.

3.3.4. Sağlık Kurumlarında Müşteri Kayıp Analizinin Kullanım Alanları ve Örnekleri

Müşteri devamlılığı bulunan sektörlerde, müşteri belirleme ölçütleri ve kayıp müşteri kabul etme ölçütlerini net olarak tanımladıktan sonra belirli zaman serileri içerisinde, veri tabanları üzerinden, kayıp kabul edilen müşteriler ile nedenleri arasında bir örüntü ağı çıkarmaya çalışılmakta ve bir müşterinin ne zaman, hangi durumlarda hizmet alımını kestiğini haritalandırabilmektedirler. Sonrasında ise belirlenen ağlar ile bir müşterinin gelecekte kaybedilme riskini çeşitli veri madenciliği araçları ile tahmin etmektedirler.

Sağlık alanında, özellikle özel hastane işletmeciliğinde, hasta tutundurma çok önemli hale gelmiştir. Diğer yandan genellikle, hastayı elde tutma süreçleri gelen şikayetlere yönelik reaktif aksiyonlar ile yönetilmektedir. MKA'nın sağlık uygulamalarının yazında pek yer almadığı görülmektedir. Sağlık kurumlarında hastaların sürekli hizmet alması söz konusu olmadığı, kısa veya uzun dönemli aldıkları tedavi bitiminde hizmet alımları durduğu için kimlerin kayıp hasta olduğunu belirleme problemi vardır. MKA'nın sağlık alanı yazınında uygulama örnekleri pek olmamasına rağmen, hizmet devamlılığı gerektiren birçok alt alanında benzer çalışmaların yapılabileceği düşünülebilir. Bu durumda; kalp anomalileri, diyabet, zekâ geriliği vb. konjenital, kalıtsal, kronik hastalıklar, kanser gibi uzun süreli tedavi ve takip gerektiren hastalıklar ile hamilelik-doğum (dönemsel devamlılık gerektiren durumların) ve çocuk gelişimi (sağlıklı rutin takip/aşı ve hastalık takibi) gibi durumların MKA kapsamında

değerlendirilebileceği düşünülebilir. Bu alanların her birinde bir hizmet devamlılığı ve bir hizmet alıcısı bağlantısı bulunmaktadır. Mücbir sebepler, gönülsüz kayıplar elendikten sonra devamlı hasta kabul edilen bir hizmet alıcısının kayıp hasta olması durumunda, aldığı tıbbi ve idari hizmetten memnuniyeti, ödediği ücret gibi birçok nedene dayalı örüntüler çalışabilecektir. Örneğin hastaneler, gebelik tanısı almış hastanın poliklinik takibini ve doğumunu ilgili kurumda yapıp yapılmadığı, ilgili kurumda doğan bebeğin düzenli çocuk takiplerini aynı kurumda yaptırıp yaptırmadıklarını takip etmeye çalışmakta fakat takipten çıkan hastalar ve nedenleri üzerinden sistematik kapsamlı veri madenciliği çalışmaları yapmamaktadırlar. İlgili veri madenciliği çalışmaları ile MKA yapıldığında, hastaların sosyodemografik, geliş sıklıkları, sigorta güvence durumu, fatura ödem ayrıntıları, sistemde bulunan şikayetler, anket sorularına cevapları ile hizmet sunucusundan çıkma nedenleri arasındaki örüntüler saptanabilir ve geleceğe yönelik tahminlerle hastaların tutundurulması amaçlanabilmektedir.

Her ne kadar yazın taramasında sağlık alanında doğrudan örnekler ile nadiren karşılaşılsa da Kwoon, vd. (2021) tarafından, Kore’de yemek, egzersiz ve kilo verme gibi alanlarda interaktif mesaj gönderiminde bulunan bir dijital sağlık uygulaması churn kapsamında incelenmiştir. Çalışmada; 1 Ocak 2017 ile 1 Ocak 2019 tarihleri arasında programa kaydolup, ödediği programdan erken çıkış yapan ve yedi günlük ücretsiz deneme süresi sonrası devam etmeyen müşteriler, kayıp müşteriler olarak kabul edilmiştir. Bu doğrultuda, 1868 uygun müşterinin YSA aracılığı ile hangi durumlarda kayıp müşteri oldukları ve bunda en belirleyici etmenlerin nelerin olduğu (günlük atılan adım sayısı, alkol tüketimi, fazla yeme/geç yeme) analiz edilmiştir.

Yapılan bir diğer çalışmada; İran’da büyük bir kamu hastanesinin en fazla hasta sayısına sahip kardiyovasküler, nöroloji ve pediatrik hematoloji bölümlerinde, 3 yıllık süreyi kapsayan hastaların davranışlarının takip edilmesi (geliş sıklığı, kalış süresi, ödeme vb.) MYBD olarak RFML modeli ile özel grupları ve eğilimlerini tespit etmeye çalışılmıştır. Çalışmadaki ana gaye, ilişkilerinin güçlendirilmesi için potansiyel sadık hastaların tanımlanması, potansiyel kayıp hastaların

tanımlanması ile kayıp gerçekleşmeden önce tutundurma faaliyetlerine, müşteri kaybıyla ilişkili maliyetleri azaltılabilmesine olanak sağlamaktır. İlk olarak, hedef müşteri ve grupların belirlenmesi için K-means kümeleme algoritması uygulanmış ve daha sonra kayıp tahmini olarak karar ağacı sınıflandırıcısı kullanılmıştır. Sadık ve kayıp hasta sayısına göre üç kliniğin performansını karşılaştırılmış, en iyi sonucu veren pediatrik hematoloji kliniğindeki iyi uygulama örnekleri diğer birimlerde de değerlendirilmek üzere hastane yönetimiyle paylaşılmıştır (Mohammadzadeh, vd., 2017).

Hastaneler mevcut müşterilerini elde tutabilmek için çaba sarfetmektedirler. Diyabet hastaları üzerine 1999-2015 yılları arasındaki farklı hastanelere ait 91.000 gözlem üzerinden veri madenciliği araçlarıyla kayıp hasta oranını tahmin edecek bir çalışmalar yapılmıştır. Çalışmada; yaş, hastanede geçirilen zaman, kullanılan ilaç, tanı kodu, taburculuk kodu, ayaktan hasta başvuru sayısı, kritik kriz sayısı gibi hastanın kondisyonunu, özniteliklerini, kullandığı ilaç ve yapılan tetkikleri içeren 54 farklı diyabet ilişkili değişken kullanılmıştır. Verilerin %80'i eğitimde, %20'si test aşamasında kullanılmış ve LR, karar ağaçları ve RO ile tahminleme çalışmaları yapılmıştır. Çalışma bulgularıyla, hastaların profillere ayrılması, yeni müşterilerin kazandırılması ve mevcut hastaların modelin çıktılarını kullanarak tutundurulmaya katkı sağlayacağı öngörülmüştür (Chauhan, vd., 2020).

Yapılan bir diğer çalışmada hastaneler ve kronik hastalık hastaneleri, hangi hastaların tesislerinden bakım alacağını tahmin etmek için elektronik sağlık kayıtlarından yapılandırılmış verileri ve reçeteleri kullanılmıştır. Uygulama amaçları doğrultusunda, hasta kayıplarını tahmin etmek için toplu öğrenme teknikleri ve teşhis aciliyetine dayalı kümeleme teknikleri kullanılmıştır. Bulguların, klinik karar vericiler için ilgi çekici olduğu ve teşhiste yardımcı olacağını öngörmüşlerdir (Singh, vd., 2023).

4. BÖLÜM

SAĞLIK SEKTÖRÜNDE MKA İÇİN BİR ARAŞTIRMA VE METODOLOJİSİ

Bu bölümde, öncelikle sağlık sektöründe MKA için yapılan araştırmasının önemi, amacı, soruları paylaşılmıştır. Sonrasında, araştırmanın yöntemi, aşamaları, model değerlendirme kriterleri ve kısıtları paylaşılmıştır.

4.1. Araştırmanın Önemi

Günümüzde artan rekabet koşulları, gelişen teknoloji ve değişen müşteri beklentileri nedeni ile işletmelerin stratejileri ve iş yapış modelleri yeniden sorgulanmaktadır. Ayrıca, artan hammadde maliyeti ve ithalat yükleri, işletmelerde verimliliği ve finansal kaynakların etkin kullanımını daha da önemli hale getirmiştir. İşletmeler için büyümenin ya da hayatta kalmanın en temel etmenlerinden birisi, hizmet sunumuna yeni müşterileri eklemek ya da mevcut müşterilerini koruyabilmektir. Kotler ve Armstrong, (2017), çalışmalarında yeni müşteri kazanmanın, mevcut müşteriye mutlu edip elde tutmaya göre beş ile on kat daha maliyetli olduğunu belirtmişlerdir. Yine çalışmalarda; yaptıkları alımlardan memnuniyetsiz ayrılan müşterilerin %96'sının şikâyet başvurusunun olmadığı fakat deneyimlerini en az 15 kişi (sağlık alanında çok daha fazla olduğu öngörülmektedir) ile paylaştıkları (Beal, 2014) ve müşterilerinin %92'sinin ağızdan ağıza pazarlamayı diğer pazarlama-iletişim kanallarını kullanmaya tercih ettikleri, tanıdıklarından duydukları ya da müşteri bloglarından okumuş oldukları müşteri deneyimlerini daha güvenilir buldukları yer almıştır (Nilsen, 2012). Bu bilgiler doğrultusunda, memnuniyetsizlik sonucu kaybedilen hastaların YBMD'si de düşünüldüğünde, MKA sağlık kurumları için daha da kritik hal almaktadır.

Yapılan yerli ve yabancı yazın taramasında MKA analizlerinin abonelik ve/veya müşteri sürekliliği bulunan bankacılık, telekomünikasyon, sigortacılık vb. alanlarda yaygın olarak yapıldığı görülmektedir. Fakat sağlık alanında son birkaç yılda görülmeye başlandığı, özellikle hastanecilik dalında dünya genelinde çok nadir çalışmaların yapılmış olması ve Türkiye'de ise yazın taramasında

saptanmaması çalışmayı farklı kılan unsurlardandır. Konu sağlık olunca da insanların olumsuz deneyimlerini diğer sektörlere nispeten çok daha fazla paylaştığı düşünülmektedir. Dolayısıyla, çalışmanın özgünlüğünün yanı sıra MKA'nın hastanecilik için bir ihtiyaç olduğu, uygulanması durumunda müşteri tutundurma ile birlikte birçok farklı katkısının da olabileceği, farklı çalışmalara ışık tutabileceği araştırmayı önemli kılmaktadır.

4.2. Araştırmanın Amacı ve Soruları

Yapılan bu çalışmada amaç; sağlık hizmet alıcılarının müşteri mi hasta mı olarak anılması, devamlı hasta ve kayıp hasta belirleme ölçütlerinin neler olabileceği, MKA'nın sağlık alanı için uygulanabilir olup olmadığı, yüksek tahmin başarısı ve yorumlama kolaylığı bulunması nedeniyle seçilen VM tekniklerinden (CHAID ve CART) hangisinin daha başarılı sonuçlar verdiğini araştırmak ve gelecekte yapılabilecek katma değer üretecek çalışmalara ışık tutmaktır. Bankacılık ve telekomünikasyon gibi diğer sektörlere göre nispeten kurumsal olarak daha az gelişmiş sağlık ve özellikle hastanecilik hizmetleri sektöründe bir uygulama örneği üzerinden kayıp müşterileri tanımlayabilmek, tutum ve paternleri saptayacak, gelecekteki eğilimlerini tahmin edebilecek bir model kurgulamak ve müşteri kayıplarını önlemeye çalışacak aksiyon planlarına ışık tutmaktır. Çalışma kapsamında aşağıdaki sorulara yanıtlar aranmaktadır:

- Sağlık sektöründe hasta mı müşteri mi daha kabul edilir bir tanımlamadır?
- Sağlık sektöründe sürekli hizmet alıcısı ve kayıp hizmet alıcısı tanımlanabilir mi?
- MKA her hastaya uygulanabilir mi?
- Sağlık sektöründe hastaların hizmet sunucularını değiştirmekteki eğilimleri, abonelik, müşteri sürekliliği bulunan diğer sektörlerde olduğu gibi kayıp olma analizi ile değerlendirilip önleyici faaliyetler alınabilir mi?
- Hastaların geliş sıklıkları ve kayıp hasta olma durumları arasında bir ilişki var mıdır?

- Kayıp hastaların tahminlenmesinde tahmin başarısı ve yorumlama kolaylığı nedeniyle seçilen sınıflandırma tekniklerinden karar ağacı algoritmalarından olan CHAID ve CART arasından hangisi ile daha başarılı sonuçlar elde edilebilir?
- Hastaların şikâyetleri ile kayıp hasta olma durumları arasında bir ilişki var mıdır?
- Hastaların anket sorularına verdiği cevaplar ile kayıp hasta olma durumu arasında ilişki kurulabilir mi?

4.3. Araştırmanın Yöntemi ve Aşamaları

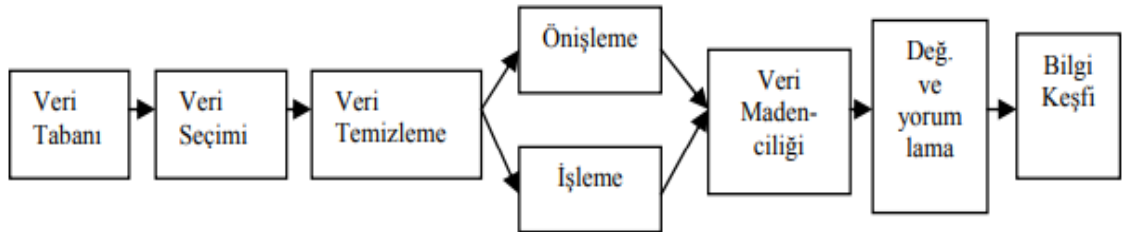
Araştırma, uluslararası bir şirketin parçası olan, Türkiye ve Avrupa’da, dört farklı ülkede toplam 24 hastanesi, 14 tıp merkezi, 25.000’e yakın çalışanı ve yıllık 5.000.000 hastası olan bir özel sağlık kuruluşunda yapılmıştır.

Hastane sektöründe MKA’nın uygulanabilirliği ele alınırken hizmet alım devamlılığı içeren hasta grupları düşünülmüş konjenital, kalıtsal, kronik rahatsızlıklar ya da hamilelik-doğum dönemi gibi düzenli takip gerektiren birçok tıbbi branş MKA ile değerlendirilebilir olduğu fikri oluşmuştur. Fakat söz konusu hasta grupları için birbirinden farklı ölçütler kullanılması gerekmektedir. Bu nedenle, araştırmada sadece çocuk hastalar seçilmiştir. Çocuk hastaların seçilmesinde rol oynayan ana etmenler aşağıda verilmiştir:

- Sağlıklı çocuk takibinde (ulusal kabul görmüş aşı takviminin de yer alması) aşıların özel hastanelerde uygulanabilir olması,
- Anne-babaların çocuklarının sağlıklarına (hastalık hali ve planlı takiplerde) kendilerine olduğundan daha fazla duyarlı olması,
- Çocukların hizmet alımının devamlılık arz etmesi, yaş grupları ve hizmet alım sıklıklarına göre gruplandırılabilmesi,
- Anne-babaların, konu çocukları olduğunda aldıkları hizmet sonrası gözlem, şikâyet ve önerilerini daha sık dile getirmeleri.

Bu doğrultuda, çocuk hastaların hastaneye başvuru sıklıkları ve kayıp hasta kabul edilme durumu, sistemde bulunan kayıtlı şikâyetleri ve kayıp hasta durumu ve anket sorunlarında NPS'yi etkileyen soruya verilen puanlandırma ile kayıp hasta olma durumu arasındaki örüntüler VM teknikleri incelenmiş ve tahmin modelleri geliştirilmiştir. Aşağıdaki kesimlerde söz konusu bu üç ayrı tahmin modeli için yapılan araştırmanın süreci sırası ile verilmiştir.

Araştırma için öncelikle ilgili sağlık kurumunun farklı veri tabanlarından, geliş sıklıkları çalışmasının değişkenleri SQL yapısal sorgu dili aracılığı ile MİY sistemine kayıtlı şikâyetler ve anketlere verilen cevaplar ise MİY yazılım programında tanımlı otomatik raporlar aracılığı ile temin edilmiştir. Yapılan her üç çalışmada, Şekil 4.1'de paylaşılan, veri seçimi, veri temizlemesi, veri işleme-dönüştürme, VM uygulaması, bulguların değerlendirilmesi ve yorumlanması neticesinde keşfedilen bilgi aşamaları takip edilmiştir.



Şekil 4.1.: Veri Madenciliği ile Bilgi Keşfi Adımları

4.3.1. Geliş Sıklıkları ve Kayıp Hasta Tahminlemesi

Çocuk hastalar; 28597 sayılı Sosyal Güvenlik Kurumu Sağlık Uygulama Tebliği'nde 0-18 yaş aralığındaki bireyler olarak tanımlanmıştır. Bu doğrultuda 01.01.2012 ve 31.12.2018 tarihleri arasındaki 3.400.000'i aşkın çocuk hasta başvuruları listelenmiştir. 496.847 çocuk hastanın başvuruları devamlı hasta kabul kriterlerince değerlendirilmiş ve 211.464 tane çocuk hasta devamlı hasta kabul edilip MKA'ya konu edilmiştir. 2020 başı itibari ile, 10 yaş üzerinde olan

11.164 adet çocuk hastalar, rutin takibin dışına çıktıkları için çalışma dışında bırakılmıştır. Çalışma geriye kalan 200.350 hasta üzerinden yapılmıştır.

Çalışma verisinde 2019 ve sonrasının dahil edilmemesinin nedeni, hastaların kayıp hasta kabul edilebilmesi için bazı yaş gruplarında minimum 1 yıl başvuru olmama durumunun gözetilmesi ve 2020'nin ilk ayları itibari ile ülkemizde de görülen Covid-19 salgının hasta davranışlarını değiştirmesidir. Covid-19 başlangıcı ile birlikte aileler zorunlu haller dışında çocuklarını hastanelere götürmemişler ve hekimleri ile fiziksel muayene yerine whatsapp, viber vb. online iletişim kanalları aracılığı ile iletişim kurma yolunu seçmişlerdir. Veri tabanından SQL aracılığı ile belirlenen çalışmada kullanılacak değişkenlerin başlıkları ve açıklamaları Tablo 4.1.'de verilmiştir.

Tablo 4.1.: Geliş Sıklıkları ve Kayıp Hasta Modeli İçin Kullanılan Değişkenler ve Tanımlamaları

Başlık	Anlamı
Hastano	Benzersiz Tanımlayıcı / Anahtar
Dogum_Tarihi	Hastanın Doğum Tarihi
Son_Gelis_Sube	Başvurunun Olduğu Son Şube
Son_Gelis_Sube_Istanbul	Başvurunun Olduğu Son Şubenin Kategorisi- İstanbul
Son_Gelis_Sube_Anadolu	Başvurunun Olduğu Son Şubenin Kategorisi- Anadolu
Son_Kabul_Kurumu_Grubu	Son Kullanılan Sosyal Güvence Grubu (Nakit, Özel Sigorta, Sgk Vs.)
Son_Gelis_Tipi	Son Başvurunun Tipi (Ayaktan, Yatan, Günübirlik)
Uyruk	Uyruk Belirteci
Son_Adres_Ili	Son Başvurudaki Hastanın Adresinin Kayıtlı Olduğu İl
Pol_Mua_Basvuru_Tarihi	Poliklinik Muayene Başvuru Tarihi
Kontrol_Mua_Basvuru_Tarihi	Kontrol Muayene Başvuru Tarihi
Diger_Basvuru_Tarihi	Poliklinik Muayene ve Kontrol Muayene Dışındaki Diğer Ayaktan Ya Da Yatan Başvurular

Çalışmanın veri temizleme aşamasında, hasta adı soyadı gibi gizlilik içeren veriler KVKK (Kişisel Verilerin Korunması Kanunu) doğrultusunda, hasta

ödemeleri, kurum ödemeleri, cinsiyet gibi başlıklar çalışma kapsam dışında bırakıldığı için çalışma veri setinden çıkarılmıştır. Benzersiz tanımlayıcı olan hastano'nun eksik olduğu durumlara ait gözlemler de gürültü, kayıp değer üretmemesi amacı ile çıkarılmıştır. Hedeflenen çalışma kapsamında; veri işlenmesi, dönüştürülmesi, düzenlemesi aşamasında eklenen ve modelde kullanılan yeni değişkenlerden bazıları Tablo 4.2'de verilmiştir. Modelde kullanılan tüm değişkenler Ek 2.'de verilmiştir.

Tablo 4.2.: Geliş Sıklıkları ve Churn Modeli Eklenen Değişkenler ve Tanımlamaları

Başlık	Anlamı
GA_1-2_Yas_Diger_Adet	1-2 Yaş Aralığındaki Çocuk Hastanın Poliklinik ve Kontrol Muayene Dışındaki Başvuru Adetleri
GA_1-2_Yas_Kontrol_Adet	1-2 Yaş Aralığındaki Çocuk Hastanın Kontrol Muayene Adedi
GA_1-2_Yas_Poliklinik_Adet	1- 2 Yaş Aralığındaki Çocuk Hastanın Poliklinik Muayene Adedi
GA_Toplam_Gelis_Adedi	Toplam Geliş Adedi (Poliklinik, Kontrol Ve Diğer)
Son-Sondan2_Arasi_Gun_Sayisi	Son Geliş- 1 Önceki Geliş Gün Farkı
Sondan2-Sondan3_Arasi_Gun_Sayisi	Bir Önceki- 2 Önceki Geliş Gün Farkı
Sondan3-Sondan4_Arasi_Gun_Sayisi	Sondan 2 Önce Geliş- 3 Önceki Geliş Gün Farkı
Sondan4-Sondan5_Arasi_Gun_Sayisi	Sondan 3 Önce Geliş- 4 Önceki Geliş Gün Farkı
Sondan5-Sondan6_Arasi_Gun_Sayisi	Sondan 4 Önce Geliş- 5 Önceki Geliş Gün Farkı
Sondan6-Sondan7_Arasi_Gun_Sayisi	Sondan 5 Önce Geliş- 6 Önceki Geliş Gün Farkı
Sondan7-Sondan8_Arasi_Gun_Sayisi	Sondan 6 Önce Geliş- 7 Önceki Geliş Gün Farkı
Sondan8-Sondan9_Arasi_Gun_Sayisi	Sondan 7 Önce Geliş- 8 Önceki Geliş Gün Farkı
Sondan9-Sondan10_Arasi_Gun_Sayisi	Sondan 8 Önce Geliş- 9 Önceki Geliş Gün Farkı
GS_1-2_Yas_Gelis	1-2 Yaş Toplam Geliş Sayısı
GS_1-2_Grubu	Geliş Sayısına Göre Segmenti (A, B, C, N/A)
MAXG_1	1-2 Yaş Arasında İki Başvuru Arasındaki Max Gün Sayısı
MAXG_2	2-3 Yaş Arasında İki Başvuru Arasındaki Max Gün Sayısı
Churn_1-2_Yas	1-2 Yaş Aralığındaki Churn Durumu (1 Evet / 0 Hayır)
2020_Churn_Durumu	01.01.2020 İtibariyle Churn Durumu (1 Evet / 0 Hayır)

Yapılan veri işleme, dönüştürme, birleştirme aşamasında, çocuk hastalar ilgili yaş gruplarındaki geliş sayılarına göre A, B ve C olarak gruplandırılmışlardır. Tablo 4.3'de ayrıntıları paylaşılmıştır.

Tablo 4.3.: Yaş Grupları ve Geliş Sıklıkları Grupları

Yaş Grubu	Geliş Sıklığı Grubu
0-1 Yaş	A: 6+ başvuru B: 3-5 başvuru C:3'den az başvurular N/A: İlgili yaş grubunda başvuru yok fakat diğer yaş gruplarında devamlı/bizim hastamız kabul edilmişse
1-2 Yaş	A: 4+ başvuru B: 2-3 başvuru C: 1 başvuru N/A: İlgili yaş grubunda başvuru yok fakat diğer yaş gruplarında devamlı/bizim hastamız kabul edilmişse
2-6 Yaş	A: 10+ başvuru, B: 6-9 başvuru, C:5 ve 5'den az başvurular (3-5 Başvuru), eğer daha önceki yaş grubunda bizim hastamız kabul edildiyse 1-5 başvuru N/A: İlgili yaş grubunda başvuru yok fakat diğer yaş gruplarında devamlı/bizim hastamız kabul edilmişse
6-10 Yaş	A: 8+ başvuru, B: 5-7 başvuru, C:4 ve 4'den az başvurular (3-4 Başvuru), eğer daha önceki yaş grubunda bizim hastamız kabul edildiyse 1-4 başvuru N/A: İlgili yaş grubunda başvuru yok fakat diğer yaş gruplarında devamlı/bizim hastamız kabul edilmişse

Devamlı hasta ve kayıp hasta kabul edilme kriterleri tablo 4.4.'de paylaşılmıştır.

Tablo 4.4.: Devamlı Hasta ve Kayıp Hasta Kabul Kriterleri

Devamlı Hasta Kabul Kriterleri	0-1 ve 1-2 yaş için 1 yıl içinde en az 3 poliklinik muayenesi, 2-6 ve 6-10 yaş aralığı için 2 yıl içinde en az 3 poliklinik muayenesi varsa devamlı hasta kabul edilmiştir.
Kayıp Hasta Kabul Kriterleri	0-1 yaş: 2 ay süre ile herhangi bir yeni başvurusu yok ise 1-2 yaş: 4 ay süre ile herhangi bir yeni başvurusu yok ise 2-6 yaş: 6 ay süre ile herhangi bir yeni başvurusu yok ise 6-10 yaş: 1 yıl ve üzeri herhangi bir yeni başvurusu yok ise 10 yaş ve üzeri: 2 yıl ve üzeri herhangi bir yeni başvurusu yok ise

Farklı endüstrilerde MKA çalışmaları doğrultusunda yazın örnekleri bulunmaktadır. Perakendecilikte yapılan kayıp müşteri kabul edilme süresinin ardışık 3 dönem olduğu (Braun ve Schweidel, 2011; Burez ve Poel, 2008; Oliveira,

2012), telekomünikasyon sektöründe (Chen, vd., 2012; Hadden, 2008) ve bankacılık sektöründe de 3 dönem olduğu (Prasad, 2012) görülmektedir. Yapılan çalışmalarda her bir dönemin 3 ayı temsil ettiği kabul edilmiştir. Diğer yandan sağlık alanında belirli bir kriter olmadığı için çalışmanın hedef değişkeni olan devamlı hasta – kayıp hasta etiketlemesi için ayrıca bir araştırma yapılmıştır. Araştırmada veri toplama yöntemi olarak, nitel çalışmalarda veri toplamada kullanılan yöntemlerden biri olan görüşme yöntemi seçilmiştir. Görüşmeler yarı yapılandırılmış mülakat tekniği ile yapılmıştır. Görüşme formu anket tarzı sorular içermektedir. Daha derinlikte veri elde etmek için ise sorulara şık eklenerek katılımcıların ek görüşünün alınması sağlanmıştır. Aşağıdaki sorulara yanıt aranmıştır:

- Hangi durumda ve neden bir çocuk hastayı kendi hastaları kabul edebilecekleri,
- Çocuk hastaları beklenen geliş sıklıklarına göre nasıl gruplandırabilecekleri,
- İlgili yaş gruplarında ne kadar sık bir başvuru öngördükleri,
- Belirlenen yaş grupları için ne kadar süre sonra (eğer çocuk hastanın yeni bir başvurusu yoksa) kayıp hasta olarak kabul edebilecekleri

İlgili sağlık kuruluşunun üç farklı şehirdeki beş farklı hastanesinde görev yapan 15 farklı çocuk hekimi ile medikal açıdan ve yönetim açısından da sağlık kuruluşunun 6 üst düzey yöneticisi ile görüşmeler gerçekleştirilmiştir. Yapılan mülakatlar sırasında her bir uzmana ilk önce konunun çıkış noktası, amaçlanan nihai hedef anlatılmış, Covid-19 kısıtlamaların çalışma kapsamındaki yeri açıklanmış ve sonrasında sorulara geçilmiştir. Çalışma doğrultusunda; devamlı hasta kabul etme ve kayıp hasta kabul etme kriterleri belirlenmiştir. Belirlenen kriterler MKA modellerinde kullanılan hedef değişken olan kayıp hasta durumunun etiketlemesini belirlemiştir.

Veri madenciliği uygulaması aşamasında; seçilen, temizlenen, işlenip birleştirilerek analize hazır hale getirilen veri SPSS Modeler 17.0 programı

aracılığıyla, Karar Ağaçları algoritmalarından CHAID ve CART algoritmaları ile geliş sıklıkları – kayıp hasta kabul edilme durumu analiz edilmiştir. Analiz aşamasındaki düğümler, dallar halinde gösterilmiş Pearson Ki-Kare testi ile belirlenen en iyi kestirimciler ve ilgili düğümdeki kayıp müşteri oranları paylaşılmıştır.

Karar ağaçları algoritmalarından olan CHAID, Ki-Kare Otomatik Etkileşim Dedektörü, değişkenlerin birbirleri ile olan ilişki, etkileşim, kombinasyonları ve belirlenen nihai hedef değişkeni arasındaki ilişkiler üzerine çalışmaktadır. CHAID algoritmasında; karar ağacı veri setindeki bağımsız değişken adedince dallanır, ilk dalda en uygun ön kestirici değişkeni saptayarak sonraki düğümlere odaklanır. CART ise ikili ağaç yapısından oluşur, her grubun her adımda kendinden daha homojen olan 2 alt gruba ayrılması mantığı ile çalışmaktadır.

Yapılan VM çalışması sonucundaki bulgular paylaşılmış ve değerlendirilerek yorumlanmıştır.

4.3.2. Şikâyetler ve Kayıp Hasta Tahminlemesi

İlgili sağlık kurumunun veri tabanından, MİY yazılımında tanımlı Ş.Ö.T (Şikâyet, Öneri, Teşekkür) standart rapor formatları ile 01.01.2013 ile 31.12.2018 tarihleri arasındaki sisteme kayıtlı 162.059 adet şikâyet edinilmiştir. Veri setinde, çocuk hastalara ait olan şikâyetler, sistemdeki hastaların ait olduğu branşların Excel üzerinde filtrelenmesi ile, 17.672 farklı çocuk hasta ve toplam 23.063 adet şikâyete indirilmiştir. MKA çalışmasına konu olmayan, devamlı hasta kriterini sağlamayan çocuk hastalar kapsam dışına çıkarılmış ve geriye kalan 10.046 hastaya ait 14.233 kayıtlı şikâyet çalışmaya dahil edilmiştir.

Gelen rapordan örüntüsü araştırılacak başlıklar seçilmiş, değişken olarak model eklenmiş ve aşağıda verilmiştir:

- Hasta numarası,
- Birden fazla şikâyeti olma durumu,
- İlgili hasta numarasına kayıtlı toplam şikâyet adedi,

- Devamlı hasta kriterini sağlama durumu,
- Şube,
- Şikâyet departmanı,
- Şikâyet ana grup (tıbbi, operasyonel, çalışanlar),
- Şikâyet alt grup (İlgi eksikliği, ücret, tıbbi vs.),
- Şikâyet yönetim süreci sonucu (şikâyet yöneticisini görüşme sonrası etiketlediği hastanın memnuniyet durumu, memnun, memnun değil vb.),
- Kayıp hasta durum etiketi.

Veri temizleme, işleme ve dönüştürme aşamasında; çalışmanın ilk yıllarındaki kayıtlarda görünen, hasta numarası tanımlanmamış çocuk hastalar temizlenmiş, katma değer üretmesi ön görülmeyen şikâyeti kaydeden, şikâyeti yönetip kapatan kişi, şikâyet bildiriminde bulan kişi (hasta mı hasta yakını mı, çocuk hastalarda hasta yakınının bildirimi beklendiği için) veri setinden kaldırılmıştır. İlgili hastalar için, sistemde birden fazla şikâyeti olup olmadığı, kaç şikâyeti olduğu, şikâyetlerinin arasındaki gün farkı, nihai hasta kayıp durum etiketi gibi değişkenler veri işlemesi aşamasında Excel aracılığı ile eklenmiş ve veri seti format olarak IBM SPSS Modeler 17.0'a yüklenebilecek hale getirilmiştir. Çalışılan ilgili çocuk hastalar için sisteme kayıt edilmiş şikâyetler, şikâyetlerin adedi, sıklığı ve türü ile kayıp hasta kabul edilme durumu arasındaki örüntüler karar ağacı algoritmalarında CHAID ve CART ile incelenmiş ve tahminlenmiştir.

4.3.3. Anketler ve Kayıp Hasta Tahminlemesi

Anket sorularından NPS skorunun belirleyicisi olan “Hastanemizi bir yakınınıza veya arkadaşınıza tavsiye eder misiniz / etme ihtimaliniz nedir? (0-10 arası değerlendirme) ile kayıp hasta sınıflandırmasında;

İlgili sağlık kurumunun veri tabanından MİY yazılımında tanımlı anketler raporu aracılığıyla 01.01.2016 ile 31.12.2018 tarihleri arasındaki sisteme kayıtlı 25.885 adet çocuk hastalar için doldurulmuş anket edinilmiştir. Devamlı hasta kriterlerini sağlayan ve MKA kapsamında değerlendirilen çocuk hastalar ile eşleştirilerek 13.545 adet ankete indirilmiştir.

Veri temizleme, işleme ve dönüştürme aşamasında, NPS ile ilişkili sorunun boş bırakıldığı durumlar veri setinden çıkarılmış ve 13.307 adet hasta NPS sorusu ve MKA değerlendirmesine dahil edilmiştir. NPS sorusuna 0-6 arası puan verenler “kötüleyenler”, 7-8 verenler “pasifler” ve 9-10 verenler “destekleyenler” olarak etiketlenmiştir. Birden fazla anket cevabı olan hastalarda en son dolduran anket baz alınmıştır. Soruya verilen cevaplar 0-10 ve grupları ile kayıp müşteri durumu IBM SPSS Modeler 17.0 programında kategorik değişkenler arasındaki ilişkiyi ölçmek için Ki-Kare testi aracılığı ile incelenmiştir.

4.4. Araştırmanın Model Değerlendirme Kriterleri

Veri madenciliğinde kullanılan sınıflandırma modellerinin performansını değerlendirmek için hedef niteliğe ait tahminlerin ve gerçek değerlerin karşılaştırıldığı hata matrisi / karışıklık matrisi (Confusion Matrix) ve karışıklık matrisi üzerinden elde edilen değerlendirme metriklerinden doğruluk, duyarlılık ve kesinlik hesaplamaları ile modelin doğruluğu ve geçerliliği değerlendirilmiştir. Karışıklık matrisi mantığı ve ilgili değerlendirme metrikleri ayrıntıları Tablo 4.5. de paylaşılmıştır.

Tablo 4.5.: Karışıklık Matrisi Tanıtımı

		Gerçekleşen	
		Doğru	Yanlış
Tahminlenen	Doğru	Gerçek Doğru	Yanlış Doğru
	Yanlış	Yanlış Yanlış	Doğru Yanlış

Doğruya doğru demek (True Positive – TP) **DOĞRU**

Yanlışta yanlış demek (True Negative – TN) **DOĞRU**

Doğruya yanlış demek (False Positive – FP) **YANLIŞ**

Yanlışta doğru demek (False Negative – FN) **YANLIŞ**

Yapılan çalışmada kayıp kabul edilecek hastaların ne kadar doğru tahmin edildiği gerçek veriye bakılarak değerlendirildiğinde, modelin geçerliği de sınanmış olacaktır (Seymen, 2022).

Teze konu olan hastaların kayıp hasta (churn) kabul edilme durumunu karşılık matrisince değerlendirildiğinde;

- Eğer hastamızı model çerçevesinde kayıp hasta olarak tahminlenmişsek ve hastamız gerçekten de kayıp hasta olmuş ise Gerçek/Doğru Doğru – True Positive,
- Eğer hastamız kayıp hasta olmamış, devamlı hasta olarak modelce tahminlenmiş ve gerçekte kayıp hasta olmadıysa Doğru Yalınış – True Negative
- Eğer kayıp hasta olmuş bir hastamızı kayıp hasta olmamış olarak tahmin ediyorsak, Yalınış Doğru, False Positive
- Eğer kayıp hasta olmamış hastamızı, kayıp hasta olmuş olarak tahmin ediyorsak, Yalınış Yalınış, False Negative olarak sınıflandırılmaktadır.

Değerlendirme metriklerinden çalışma kapsamında doğruluk, hassasiyet ve kesinlik metrikleri seçilmiştir.

Doğruluk (Accuracy): Doğru sınıflandırmanın, doğru tahminlenen doğruların ve doğru tahminlenen yanlışların toplamının, tüm gözlem toplamına bölümüdür.

$$\text{Doğruluk} = \text{TN (doğru yalınışlar)} + \text{TP (doğru doğrular)} / \text{Toplam Gözlem}$$

Duyarlılık/Hassasiyet (True Positive Rate–Recall): Tüm pozitif tahminlenenlerin ne kadarın gerçekte pozitif olduğunun incelemesidir.

$$\text{Hassasiyet} = \text{TP} / (\text{TP} + \text{FN})$$

Kesinlik/Precision: Gerçekten pozitif olanların kaç tanesinin pozitif olarak tahminlendiğini inceler.

$$\text{Precision} = \text{TP} / (\text{FP} + \text{TP})$$

4.5. Araştırmanın Kısıtları

496.847 çocuk hastaya ait 3.400.000'ü aşkın başvuru arasından gerekli devamlı hasta ölçütlerini sağlayan 200.350 hasta için yapılan çalışmada, hastaların vefat etmiş olması durumunun bilinmesi haricindeki mücbir sebeplerle sistemden çıkışlar, yani gönülsüz kayıplar nihai kayıp hasta olma durumundan ayıklanamamıştır. Örnek ile açıklanırsa, eğer bir çocuk hastanın hizmet alımı ailesinin ilgili sağlık grubunun şubesi bulunmayan başka bir şehre taşınması halinde kesildi ise ya da çok nadir olmakla birlikte daha önceden yapılmış kurumsal anlaşmalarda değişiklik olması, anlaşmalı kurumun ilgili özel sağlık hizmet sunucusunu ödeme listesinden kapsam dışına çıkarması gibi durumlar çalışmadan elemine edilememiştir. İleriki çalışmalarda benzer sorunların yaşanmaması için hastane işletim sistemine (HİS) ilgili kayıpların işlenebileceği bir alan oluşturulması önerilmektedir.

NPS skorunu belirleyen soru ve kayıp hasta durumu analizinde, hastalara uygulanan anketlerdeki sorunun doğru algılanıp algılanmadığı, manuel sorgu ile veri temininin sağlanması ve ilgili verilerin farklı veri tabanlarında olması nedeniyle, hastaların anket doldurma tarihi ile kayıp hasta kabul edilme tarihleri arasında karşılaştırma yapılamamıştır. Ayrıca, MKA'ya dahil edilen (devamlı hasta kriterlerini karşılayan) çocuklardan kaç tanesine anket gönderildi bilgisi olmadığı için, kaç tanesinin anketlere cevap verdiği bilinmesine rağmen, hasta yakınlarının geribildirim verme eğilimlerini, yanıt oranları saptanamamıştır. Dolayısıyla, sadece MKA'ya dahil edilen tüm çocuk hastaların, sistemde doldurulmuş anketi olanlarının karşılaştırılması yapılmıştır.

Şikâyet ve kayıp hasta durumu analizinde, şikâyet ana ve alt başlıklarının, görüşme sonrası hastanın memnuniyet durumu etiketlemesi geniş bir sistem (24

hastane, 14 tıp merkezi) üzerinde farklı kullanıcılar tarafından yapıldığı için, yazılı olmayan belirli işleyiş kuralları olsa da kullanıcı atamasıyla şekillendiğinden subjektif değerlendirmeler de içerebilmektedir.



5. BÖLÜM

ARAŞTIRMANIN BULGULARI VE DEĞERLENDİRME METRİKLERİ

Bu bölümde yıllık beş milyon hasta başvurusunun olduğu bir özel hastane grubunun çocuk hastaları için kayıp hasta analizleri (MKA) gerçekleştirilecektir. Bunun için öncelikle, devamlı/kayıp çocuk hasta belirleme kriterleri incelenecektir. İkinci aşamada; çocuk hastaların geliş sıklıkları ile churn (kayıp hasta olma) durumları yaş gruplarına (0-1, 1-2, 2-6, 6-10) göre ayrı ayrı CHAID ve CART aracılığı ile karşılaştırmalı olarak analiz edilecektir. Üçüncü aşamada, çocuk hastalar için kayıtlı şikayet adet ve tipleri ile kayıp hasta olma durumu CHAID ve CART algoritmaları kullanılarak karşılaştırmalı olarak incelenecektir. Son olarak, çocuk hastalar için toplanan anketlerdeki NPS puanını belirleyen soru ile kayıp hasta olma durumu arasındaki ilişki Ki-Kare testi aracılığı ile belirlenecektir.

5.1. Devamlı Hasta/ Kayıp Hasta Belirleme Ölçütleri

Devamlı hasta/ kayıp hasta belirleme ölçütlerini belirleyebilmek için araştırmada veri toplama yöntemi olarak nitel veri toplama yöntemlerden görüşme yöntemi kullanılmıştır. Görüşmeler yarı yapılandırılmış mülakat tekniği ile yapılmıştır. Konuya odaklanması açısından, önceden yazından ve uygulamanın yapıldığı sağlık grubunun hasta veri tabanından elde edilen çerçeve içinde görüşme formu hazırlanmıştır. Form anket tarzı sorular içermektedir. Daha derinlikte veri elde etmek için sorulara şık eklenerek katılımcıların ek görüşünün alınması sağlanmıştır. Mülakata temel teşkil eden anketin hazırlanmasındaki ana amaç aşağıda verilen sorulara cevap aramaktır:

- Hangi durumda ve neden bir çocuk hastayı kendi hastamız kabul edebileceğimiz,
- Çocuk hastaları beklenen geliş sıklıklarına göre nasıl gruplandırabileceğimiz,

- İlgili yaş gruplarında ne kadar sık bir başvuru öngördüğümüz,
- Belirlenen yaş grupları için ne kadar süre sonra eğer çocuk hastanın yeni bir başvurusu yoksa kayıp hasta olarak kabul edebileceğimiz

Mülakat için hazırlanan görüşme formu Ek 1.'de verilmiştir. Sağlık grubunun üç farklı şehirdeki beş farklı hastanesinde görev yapan 15 farklı çocuk hekimi ile medikal açıdan ve altı üst düzey yöneticisi ile yönetim açısından mülakatlar gerçekleştirilmiştir. Yapılan mülakatlar sırasında her bir uzmana ilk önce konunun çıkış noktası, amaçlanan nihai hedef anlatılmış, Covid-19 kısıtlamaların çalışma kapsamındaki yeri açıklanmış ve sonrasında sorulara geçilmiştir. Bu bireysel görüşmelerle; devamlı hasta tanımı, gruplandırılması, devamlı / kayıp hasta kabul edilme ölçütleri değerlendirilmiştir. Yöntem zaman alıcı ve maliyetli olmasına rağmen çocuk hastalarda devamlı / kayıp hasta ölçütlerinin tespitinde faydalı olmuştur.

5.1.1. Devamlı Hasta Kabul Edilme Ölçütleri ve Nedenleri

Kuruluşun 15 hekimi ile yapılan mülakatlar sonucunda, bir hastanın devamlı hasta olarak kabul edilmesi için ilgili yaş grubunda tanımlanan beklenen geliş sıklığı doğrultusunda üst üste üç kez poliklinik muayene başvurusu olması gerektiği ortak görüşüne varılmıştır. Beklenen geliş sıklıkları ve yaş grupları açısından ele alındığında ise 0-1,1-2 yaş gruplarında üst üste üç kez poliklinik muayenenin 1 yıllık dönem içinde, 2-6 ve 6-10 yaş gruplarında ise 2 yıllık dönemde gerçekleştiğinde devamlı hasta olacağı yönetici görüşleri sonucunda belirlenmiştir.

Bir kez yapılan poliklinik muayene başvurusunun devamlı hasta kabul ölçütü sayılmamasının nedeni, toplanan görüşlere göre çoğunlukla ebeveynlerin sabit bir doktor seçinceye kadar doktor arayışı içinde olmaları, farklı doktorları denemelerinden ileri gelmektedir. Özel sağlık kuruluşlarında çoğunlukla ödemeli ilk muayene sonrası ücretsiz bir kontrol hakkının bulunuyor olması da bedeli ödenen hizmet olarak algılandığı için ebeveynler ilgili doktora devam etmeyecek olsalar bile çoğunlukla bu haklarını kullandıkları bilgisi edinilmiştir.

Üst düzey yönetici mülakatlarından, iki poliklinik (ödemeli) muayene sonrası kendi hastaları olarak kabul edebileceği görüşü hâkim olmuştur. Diğer yandan bazen bir hastanın aynı hastane içinde başka doktorları sıra ile denemesi de göz önünde bulundurulduğunda, bir hastaya devamlı hastadır denilebilmesi için ilgili zaman diliminde en az üç poliklinik muayene başvurusunun olması gerektiği ortak fikrine varılmıştır.

5.1.2. Çocuk Hasta Gruplamaları ve Geliş Sıklıkları

Yapılan yarı yapılandırılmış mülakat ve anket sonuçları doğrultusunda; çocuk hastaların geliş sıklıkları göz önünde bulundurulduğunda, katılımcılarca, 0-1 yaş, 1-2 yaş, 2-6 yaş, 6-10 yaş ve 10 yaş üzeri olarak gruplandırılmasının daha uygun olduğu saptanmıştır. Sonrasında, her bir yaş grubu için tahmini öngörülen ortalama geliş adedi-sıklığı tanımlanmıştır. Bunlar ile ilgili açıklamalar aşağıda verilmiştir:

- **0-1 yaş:** 0-1 ay, 1-12 ay sıklık açısından ayrı değerlendirilmiştir. 0-1 ay yeni doğan olarak tanımlanmış ve ilgili dönemde üç ile dört kez geliş öngörülmektedir. Sonrasında ise ayda 1, yılda 8-10 poliklinik muayene gelişini uygun görülmüştür.
- **1-2 yaş:** Temel aşıların tamamlandığı, ilave birkaç aşının kaldığı ve daha çok rutin kontrol ve/veya ilk çocukluk hastalıklarının yer aldığı dönem olarak tanımlanmıştır. Yılda ortalama beş kez poliklinik muayene başvurusu öngörülmüştür.
- **2- 6 yaş:** Okul öncesi dönem olarak değerlendirilmiştir. Yılda dört kez poliklinik başvurusu öngörülmektedir.
- **6-10 yaş:** İlkokul dönemi olarak değerlendirilmiştir. Yılda üç kez poliklinik muayene başvurusu beklenmektedir.
- **10 yaş ve üzeri:** Çocuk hastaların genellikle hastalandıklarında geldiği 18 yaşına kadarki dönemi kapsamaktadır. Yılda bir kez poliklinik muayene öngörülmektedir. Yapılan çalışmada diğer yaş gruplarına göre başvuru sıklığının az olması neden ile kapsam dışında bırakılmıştır.

5.1.3. Kayıp Hasta Belirleme Ölçütleri ve Nedenleri

Hekimler ile yapılan görüşmeler sonrasında;

- **0-1 yaş:** 2 ay süre ile herhangi bir yeni başvurusu yok ise
- **1-2 yaş:** 4 ay süre ile herhangi bir yeni başvurusu yok ise
- **2-6 yaş:** 6 ay süre ile herhangi bir yeni başvurusu yok ise
- **6-10 yaş:** 1 yıl ve üzeri herhangi bir yeni başvurusu yok ise
- **10 yaş ve üzeri:** 2 yıl ve üzeri herhangi bir yeni başvurusu yok ise (çalışmada kapsam dışı)

Kayıp hasta olarak kabul edilmesi gerektiğine karar verilmiştir. Yöneticiler ile yapılan görüşmeler de aynı tarih aralıklarını desteklemektedir. Kayıp hasta kabul etme geliş sıklık aralığının tahmininin, beklenen geliş sıklığından daha geniş olmasının nedeni, ebeveynlerin yaptıkları iş, yıllık izin programları, yaz tatilleri, doktorların izin programları gibi devamsızlıkları telafi etmek içindir.

5.2. Geliş Sıklıkları ve MKA

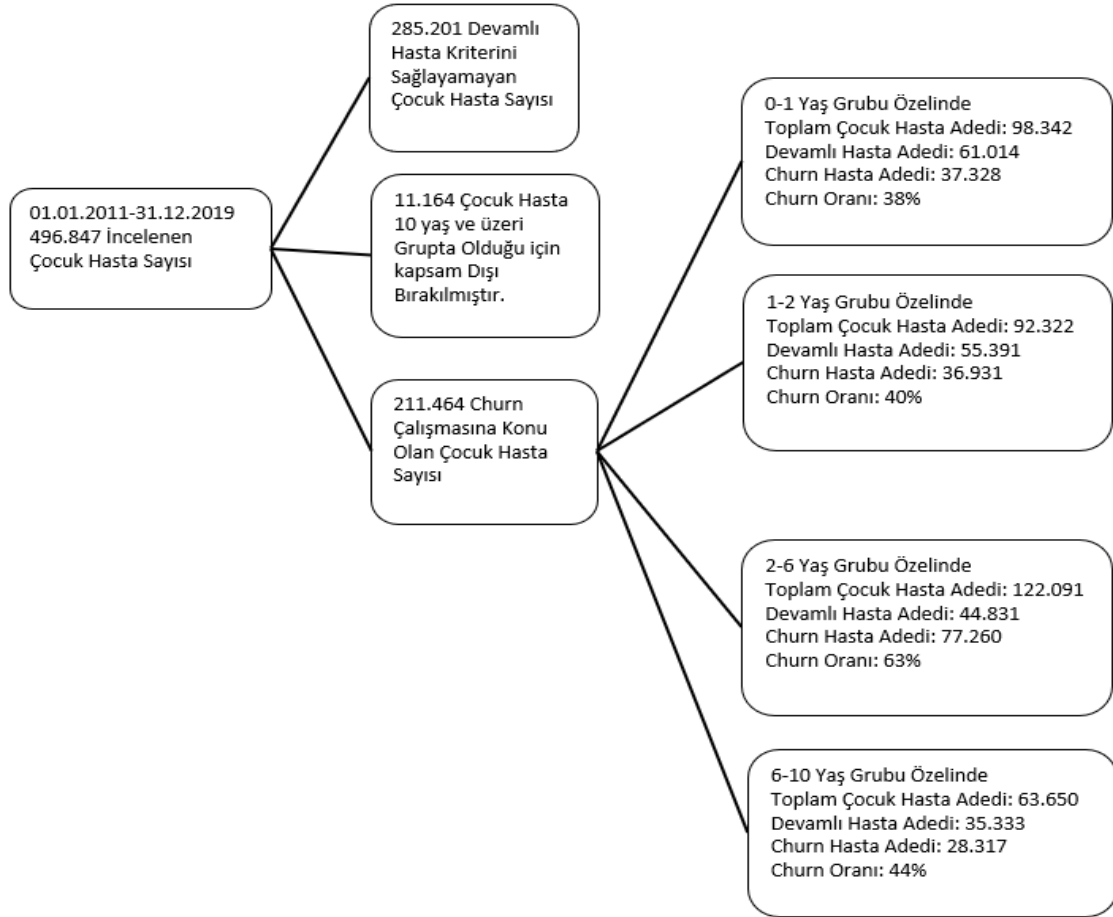
01.01.2012 ve 31.12.2018 tarihleri arasındaki 3.400.000'i aşkın çocuk hasta başvuruları listelenmiştir. 496.847 çocuk hastanın başvuruları devamlı hasta kabul kriterlerince değerlendirilmiş ve 211.464 tane çocuk hasta devamlı hasta kabul edilip MKA'ya konu edilmiştir. 2020 başı itibari ile 10 yaş üzerinde olan 11.164 adet çocuk hastalar rutin takibin dışına çıktıkları için çalışma dışında bırakılmıştır. Çalışma geriye kalan 200.350 hasta üzerinden yapılmıştır.

Çalışma verisinde 2019 ve sonrasının dahil edilmemesinin nedeni, kayıp hasta kabul edilebilmesi için bazı yaş gruplarında minimum 1 yıl başvuru olmama durumunun gözetilmesi ve 2020'nin ilk ayları itibariyle ülkemizde görülen Covid - 19 salgının hasta davranışlarını değiştirmesidir. Covid-19 başlangıcıyla birlikte aileler zorunlu haller dışında çocuklarını hastanelere götürmemişler ve hekimleri ile fiziksel muayene yerine whatsapp, viber vb. online iletişim kanalları aracılığı ile danışma yolunu seçmişlerdir.

Veri madenciliği uygulama kısmında; çocuk hastaların yaşları, sosyal güvenceleri ve başvuru sıklıkları ile kayıp hasta kabul edilme durumu karar ağacı algoritmalarından CHAID ve CART ile analiz edilmiştir. Hekimlerle yapılan mülakatlar sonrasında çocuk hastalar geliş sıklıkları açısından değerlendirmede aşağıdaki tablo 8’de yaş gruplarına ayrılmış ve churn değerlendirmesi ilgili gruplar ve kuralları dahilinde değerlendirilmiştir.

Tablo 5.1.: Model Yaş Grup Değişkeni Churn Etiketleri

CHURN_0-1_Yas	0-1 Yaş Aralığındaki Churn Durumu (1 Evet/ 0 Hayır)
CHURN_1-2_Yas	1-2 Yaş Aralığındaki Churn Durumu (1 Evet/ 0 Hayır)
CHURN_2-6_Yas	2-6 Yaş Aralığındaki Churn Durumu (1 Evet/ 0 Hayır)
CHURN_6-10_Yas	6-10 Yaş Aralığındaki Churn Durumu (1 Evet/ 0 Hayır)



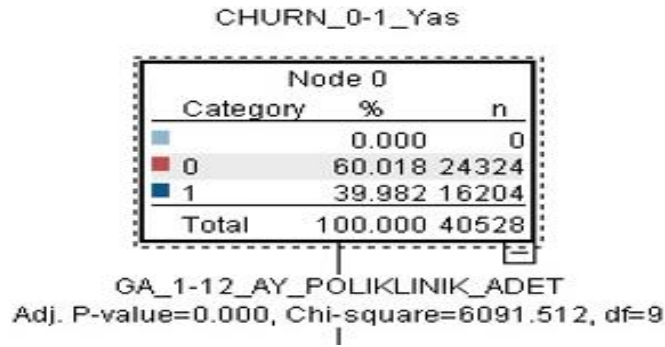
Şekil 5.1.: Geliş Sıklıkları ve Yaş Gruplarına Göre Churn Genel Bakış

Çocuk hastalarında 0-1 yaş grubunun genel churn oranının %38, 1-2 yaş grubunun %40, 2-6 yaş grubunun %63 ve 6-10 yaş grubunun %44 olduğu görülmüştür ve aşağıda her bir grup kendi içerisinde CHAID ve CART algoritmalarınca irdelenmiştir.

5.2.1. 0-1 Yaş Grubu Geliş Sıklıkları ve MKA

0-1 yaş aralığında 3 poliklinik muayene şartını sağlayan çocuk hastaların kayıp hasta kabul edilme durumu karar ağaçları algoritmalarından CHAID ve CART ile analiz edilmiştir.

5.2.1.1. CHAID Algoritması ile Bulgular ve Değerlendirme Metrikleri



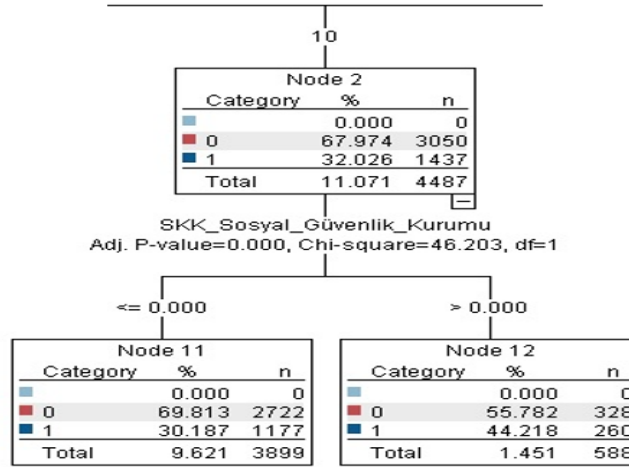
Şekil 5.2.: 0-1 Yaş CHAID Genel Durumu

0-1 yaş aralığında churn analizine konu olan 40.528 adet çocuk hastadan 16.204 hasta ile %39,98'inin churn olduğu, 24.324 hasta ile %60'ının ise devamlı hasta statüsünde hizmet alımına devam ettiği gözlenmiştir. CHAID algoritmasının ki-kare/df kombinasyonunda en iyi bölünmeyi veren değer GA_1-12_ay_poliklinik_adet yani çocukların 1-12 aylık dönemdeki poliklinik muayene başvuru adedi olduğu saptanmıştır.

Tablo 5.2.: 1-12 Ay CHAID Poliklinik Adetleri ve Churn Durumu

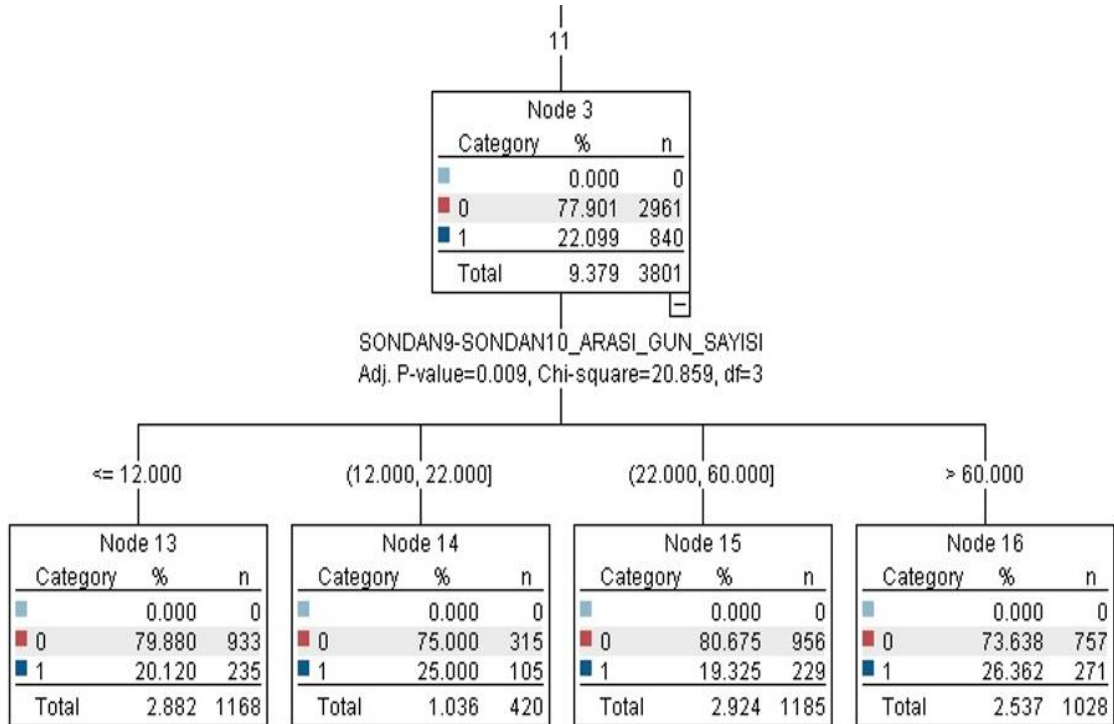
1-12 Ay Poliklinik Adet	Devamlı Hasta Adedi	Churn Adedi	Toplam Gözlem Adedi	Devamlı Hasta Oranı	Churn Oranı
2	743	160	903	82,3%	17,7%
3	2.009	657	2.666	75,4%	24,6%
4 ve 9	4.178	3.837	8.015	52,1%	47,9%
5 ve 8	3.167	4.387	7.554	41,9%	58,1%
6 ve 7	2.517	4.215	6.732	37,4%	62,6%
10	3.050	1.437	4.487	68,0%	32,0%
11	2.961	840	3.801	77,9%	22,1%
12	2.251	392	2.643	85,2%	14,8%
13	1.533	166	1.699	90,2%	9,8%
1, 14 ve üzeri	1.915	113	2.028	94,4%	5,6%
Toplam	24.324	16.204	40.528	60,0%	40,0%

0-1 yaş arasındaki churn oranı %60 olarak saptanmış iken, İlk düğümde; çocuk hastaların 1-12 ay poliklinik başvuru adedi; 2 ise chun oranının 903 gözlem arasından %17,7, 3 ise 2.666 gözlem, %24,6 churn oranı, 4 ve 9 ise, 8.015 gözlem, %47,9 churn oranı, 5 ve 8 ise 7.554 gözlem, %58,1 churn oranı, 6 ve 7 ise 6.732 gözlem %62,6 churn oranı, 10 ise 4.487 gözlem %32 churn oranı, 11 ise 3.801 gözlem %22,1 churn oranı, 12 ise 2.643 gözlem %14,8 churn oranı, 13 ise 1.699 gözlem, %9,8 churn oranı ve son olarak 1,14 ve üzeri ise 2.028 gözlem ve %5,6 churn oranı olarak bölündüğü görülmüştür. Burada özellikle 6 ve altı poliklinik muayene başvuru olan çocuk hastaların ne zaman, hangi aylarında devamlı hasta kabul kriterini sağladığı bilinmemektedir. Dolayısıyla, eğer bir çocuk hasta, devamlı hasta kriterini 9 aylık iken sağladıysa, 1 yaşını dolduruncaya kadar 2 ya da 3 poliklinik muayenesi olması beklenmektedir.



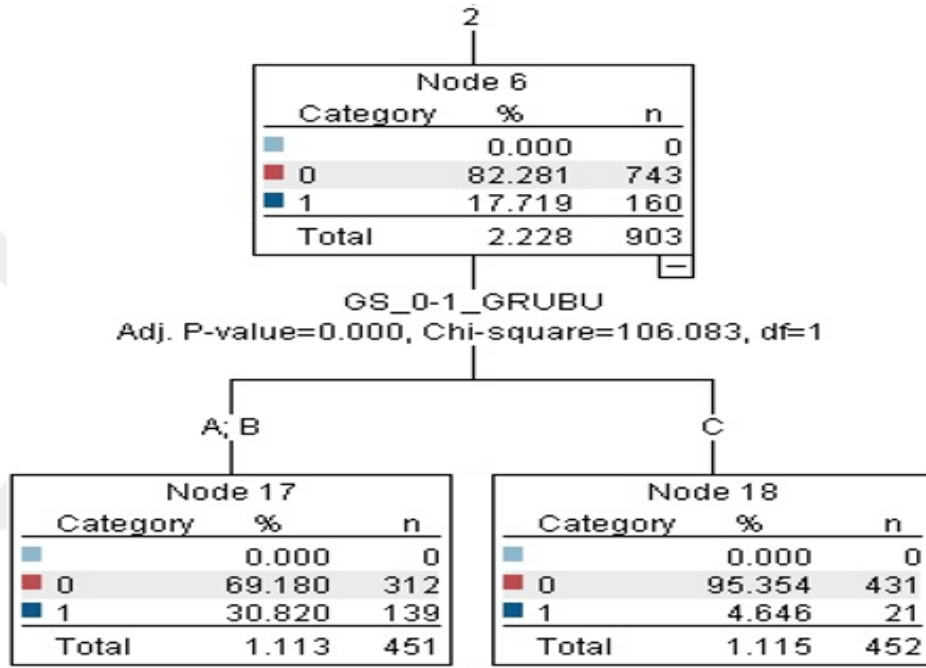
Şekil 5.3.: 0-1 Yaş CHAID 2. Düğüm Dalları

1-12 ay arasında poliklinik muayene adedinin 10, 4.487 gözlemin bulunduğu ve kayıp hasta oranının %32 olduğu 2. düğüm kendi içinde SGK (Sosyal Güvenlik Kurumu - son kabul kurumu) ile en iyi bölündüğü ve hastaların son kabul kurumu SGK olduğunda 588 gözlemin kayıp hasta oranının %44,2, SGK olmadığında ise 3.899 gözlemin kayıp hasta oranının %30,2 olduğu gözlenmiştir.



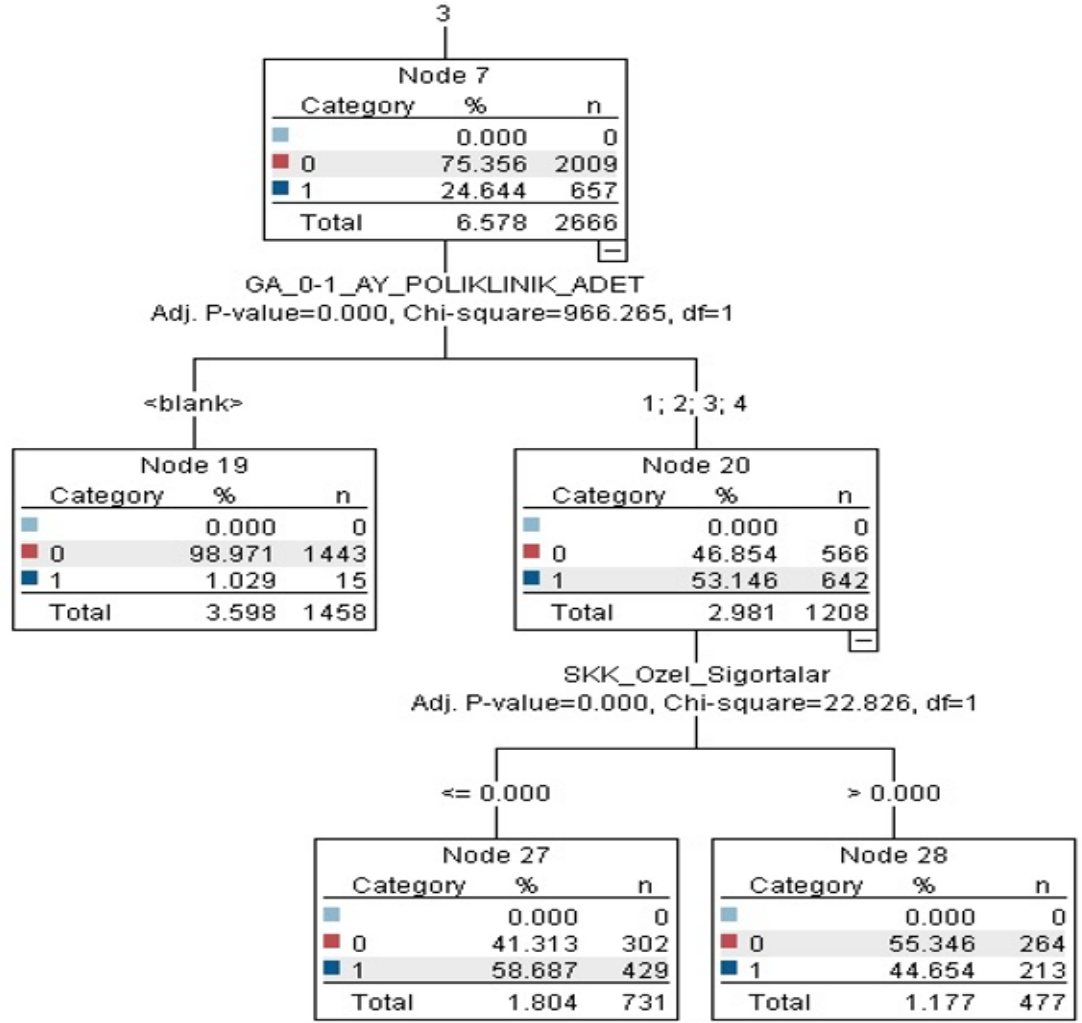
Şekil 5.4.: 0-1 Yaş CHAID 3. Düğüm Dalları

3.düğüm olan 1-12 ay arasında 11 poliklinik muayene dalını en iyi bölen kestiricinin Sondan9_Sondab10_arası_gün_sayısı olduğu ve eğer ilgili başvurular arasındaki gün sayısı 12 ve altında ise 1.168 gözlem arasından %20,1'nin, 12 ile 22 arasında ise 420 gözlemin %25 i, 22 ile 60 arasında ise 1.185 gözlem arasından %19,3'ünün ve son olarak 60 günden fazlaysa 1.028 gözlem arasından %26,4'ünün kayıp hasta olduğu saptanmıştır.



Şekil 5.5.: 0-1 Yaş CHAID 6. Düğüm Dalları

1-12 ay arasında 2 poliklinik muayene başvurusu bulunan ve 903 gözlem içeren 6. düğümü kendi içinde en iyi bölenin GS_0-1_Grubu (geliş grubu) olduğu, A (6+ poliklinik muayene) ve B (3-5 başvuru) gruplarında ise 451 gözlem arasından %30,8'inin, C (3'den az başvuru) grubundan ise 452 gözlem üzerindne %4,6'sının kayıp hasta olduğu gözlenmiştir.



Şekil 5.6.: 0-1 Yaş CHAID 7. Düğüm Dalları

1-12 ay arasında 3 poliklinik muayene başvurusu bulunan ve 2.666 gözlem içeren 7. düğümü kendi içinde en iyi bölenin GA_0-1_ay_poliklinik_adet olduğu ve ilgili dönemde eğer herhangi bir poliklinik muayene başvurusu yoksa 1.458 gözlem arasından %1'inin, eğer 1 ile 4 arası poliklinik muayene başvurusu varsa 1.208 gözlem arasından %53,1'inin kayıp hasta olduğu gözlenmiştir. Ayrıca, 1-4 poliklinik başvurusu bulunan 20.düğümü en iyi bölen kestiricinin de SKK_Özel_sigortalar olduğu ve eğer hastanın son kabul kurumu / güvencesi özel sigorta değilse 731 gözlem arasından %58,7'sinin, özel sigorta ise %44,7'sinin kayıp (churn) olduğu gözlenmiştir.

Bu grubun incelenmesinde; ebeveynler doğum sonrası genellikle kendilerini ait hissettikleri çocuk doktorunu buluncaya kadar farklı hekimleri denemektedirler. Diğer yandan artık neonatal dönem sonrası ebeveynler bir çocuk doktorunda geçici/göreceli kanaat getirmiş olmaktadır. Yenidoğan süreci sonrasında sisteme dahil olan çocukların tutunma oranının doğum ile birlikte sisteme dahil olanlara göre daha yüksek olduğu gözlenilmektedir. Diğer düşümlerden yeterince anlamlandırılacak sonuçlar çıkmamıştır.

Değerlendirme metriklerinde;

Çalışmada 0-1 yaş arasındaki 57.904 devamlı hasta kabul edilen çocuk hastanın churn durumları CHAID ile değerlendirilmiştir. Toplam veri setinin 40.528 hasta ile %70'i modelin eğitiminde, 17.376 hasta ile %30'u test için kullanılmıştır.

Tablo 5.3.: 0-1 Yaş CHAID Değerlendirme Metrikleri

CHAID 0-1 Yaş Churn Değerlendirme Metrikleri					
	Doğru	Yanlış	Doğruluk	Hassasiyet	Kesinlik
Eğitim	70.262	15.295	70,24%	64,12%	58,05%
Test	29.980	6.554	69,98%	63,16%	57,33%

Modelin eğitim veri setinde, modelin %70,24 oranında doğru tahminlemede, test aşamasında da %69,98'lik doğru tahminde bulunduğu saptanmıştır. 0-1 yaş çocuk hastalarının CHAID algoritması ile churn olma durumun değerlendirmesinde, churn olmayanlar ve churn olanlar için hatalı ve doğru tahmin sayıları hem eğitim veri seti için hem de test veri aşaması için aşağıdaki tablo 12'de karışıklık (confusion) matrisi olarak verilmiştir.

Tablo 5.4.: 0-1 Yaş CHAID Karşıtlık Matrisi

CHAID 0-1 Yaş Churn Karşıtlık Matrisi				
Eğitim		Değerler	Gerçekleşen	
			0	1
	Tahminlenen	0	19.061	5.263
		1	6.797	9.407
Test		Değerler	Gerçekleşen	
			0	1
	Tahminlenen	0	8.229	2.292
		1	2.925	3.930

Model eğitim veri setinde; 19.061 çocuk hastanın churn olmayacağı doğru şekilde tahminlenmiş, 5.263 çocuk hasta churn olmaması hatalı olarak tahminlenmiş, churn olduğu saptanmıştır. 9.407 çocuk hastanın churn olacağı model tarafından doğru tahminlenmiş, 6.797 çocuk hasta ise, churn olacağı tahminlenmişken, churn olmadığı gözlenmiştir.

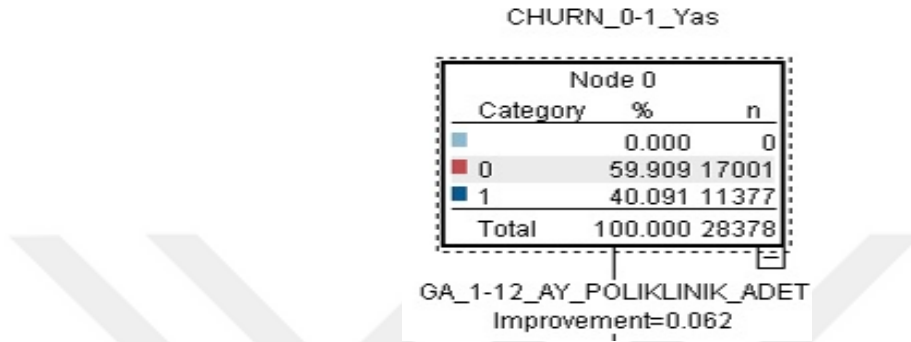
Benzer şekilde, **test veri aşamasında;** 8.229 çocuk hastanın churn olmayacağı modelce doğru şekilde tahminlenirken, 2.292 çocuk hasta churn hatalı olarak tahminlenirken, churn olduğu saptanmıştır. 3.930 çocuk hastanın churn olacağı model tarafından doğru tahminlenmiş, 2.925 çocuk hasta ise, churn olacağı tahminlenmişken, churn olmadığı gözlenmiştir.

Modelin değerlendirilmesinde; **hassasiyet (recall) metriği**, model dahilinde tüm pozitif tahminlenenlerin, yani churn olacağı öngörülenlerin, ne kadarının gerçekte pozitif / churn olduğunu incelemektedir. Hassasiyetin eğitim verisinde %64,12, test aşamasında %63,16 olduğu görülmüştür. Bir başka deyişle, 0-1 yaş çocuk grubunun CHAID algoritması ile tahminlemesinde, modelin eğitim verisi için tüm pozitif/churn tahminlenen hastalardan aslında %64,12'sinin churn olduğunu, test veri setinde ise ilgili oranın %63,16 olduğunu göstermektedir.

Kesinlik (Precision) metriği, gerçekte pozitif churn olanların aslında kaç tanesinin doğru tahminlendiğinin değerlendirilmesinde; eğitim veri setinde %58,05, test veri aşamasında %57,33 olarak gerçekleştiği saptanmıştır.

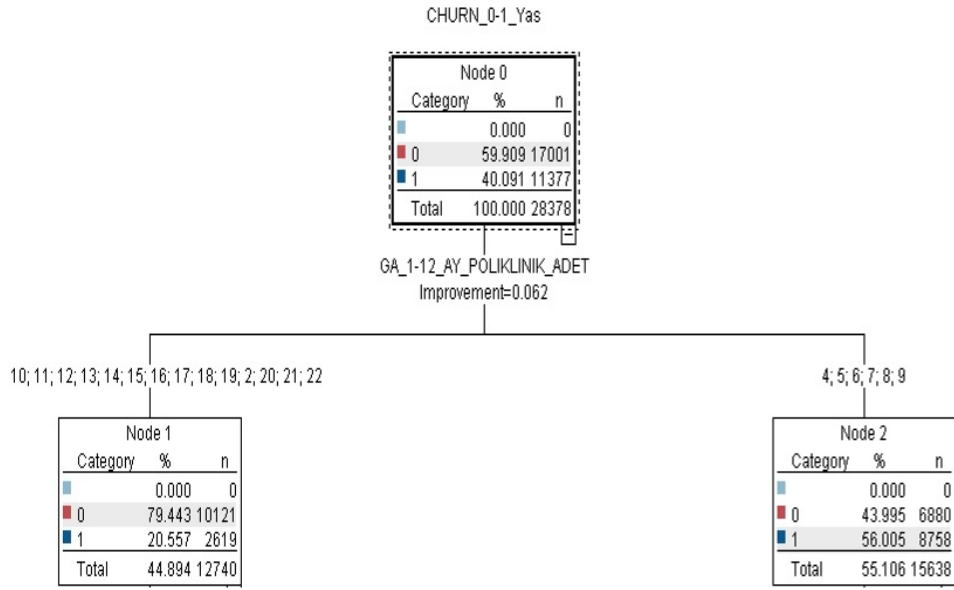
Bir başka ifade ile, eğitim veri setinde; gerçekte churn olan 100 hastanın 58,05 tanesinin pozitif/churn olan olarak tahminlendiği, test veri aşamasında ise gerçekte churn olan 100 hastanın 57,33'ünün churn/pozitif olarak tahminlendiği görülmüştür.

5.2.1.2. CART Algoritması ile Bulgular ve Değerlendirme Metrikleri



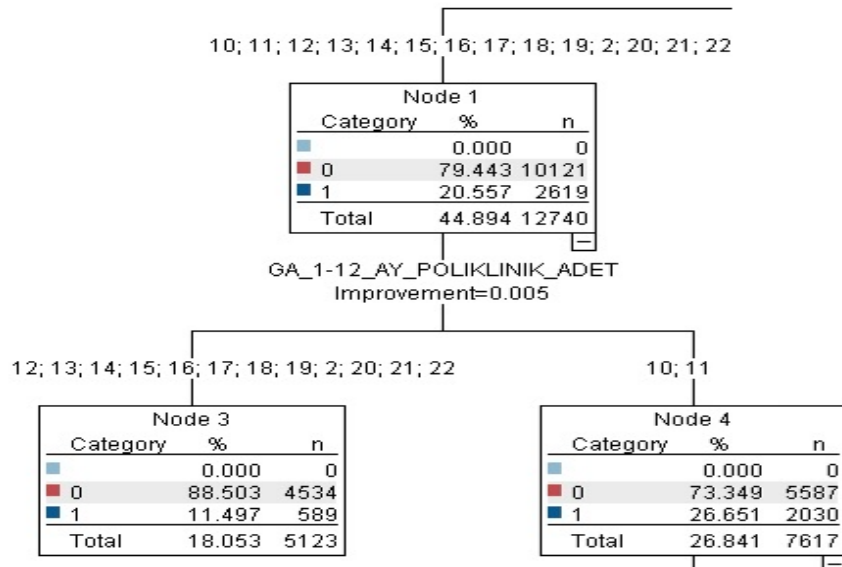
Şekil 5.7.: 0-1 Yaş CART Genel Durumu

0– 1 yaş aralığında churn analizine konu olan 28.378 adet çocuk hastadan 11.377 hasta ile %40'ının churn olduğu, 17.001 hasta ile %59,9'unun ise devamlı hasta statüsünde hizmet alımına devam ettiği gözlenmiştir. CART algoritmasında en iyi bölünmeyi veren değer GA_1-12_ay_poliklinik_adet, yani çocuk hastalarının 1-12 ay aralığındaki poliklinik muayene başvuru sayılarının olduğu saptanmıştır.



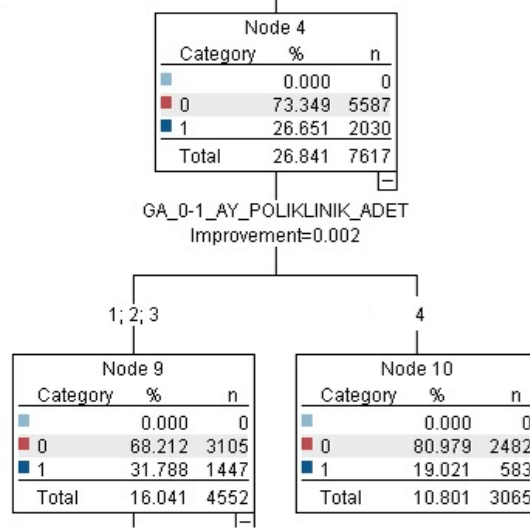
Şekil 5.8.: 0-1 Yaş CART İlk Bölünme

Devamlı hasta kabul edilme kriterince, kayıp hasta çalışmasına konu olabilmesi için en az 3 poliklinik muayene şartının sağlanması gerekmektedir. En iyi bölen olan 1-12 ay aralığındaki poliklinik muayene başvuru adetleri 4 ve 9 adet olması durumunda 15.638 gözlem arasından %58'inin kayıp hasta olduğu, 10 ve üzeri poliklinik muayene başvurusu olması durumunda 12.740 gözlem arasından %20,6'sının kayıp hasta olduğu gözlenmiştir.



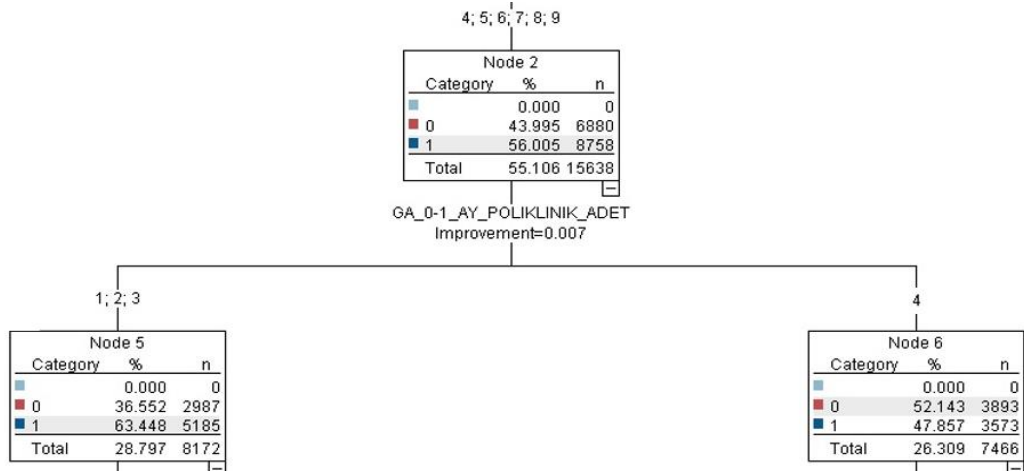
Şekil 5.9.: 0-1 Yaş CART 1. Düğüm ve Alt Dalları

12.740 gözlem barındıran 1-12 ay arasında 10 ve üzeri poliklinik muayene bulunduran grup kendi içinde 10-11 poliklinik muayene 7.617 gözlem, %26,7 churn oranı ve 12 ve üzeri poliklinik muayene 5.123 gözlem %11,5 kayıp hasta n oranı olarak ayrılmıştır.



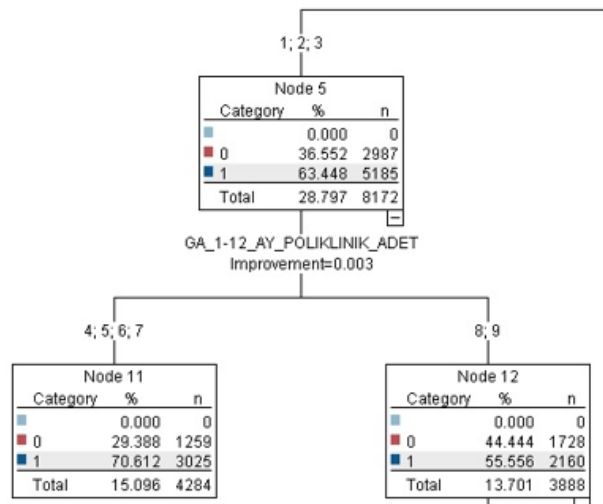
Şekil 5.10.: 0-1 Yaş CART 4. Düzüm ve Dalları

4. düğüm olan 1-12 ay arasında 10-11 poliklinik muayene grubunun en iyi bölen kestiricinin GA_0-1_ay_poliklinik_adet olduğu ve eğer poliklinik başvuru adedinin 1-3 arasındaysa 4.552 gözlem arasından %31,8'inin, eğer poliklinik başvurusu 4 ise, 3.065 gözlem arasından %19'unun churn olduğu saptanmıştır. 1 aylık süre zarfında poliklinik muayene sonrası 1 ya da 2 kontrol hakkı bulunmaktadır, dolayısıyla, 2 ve üzeri poliklinik muayene girişlerinin nedeninin sektör bilgisi ile çocukların ailelerinin doktor değiştirmeleri olarak yorumlanmıştır.



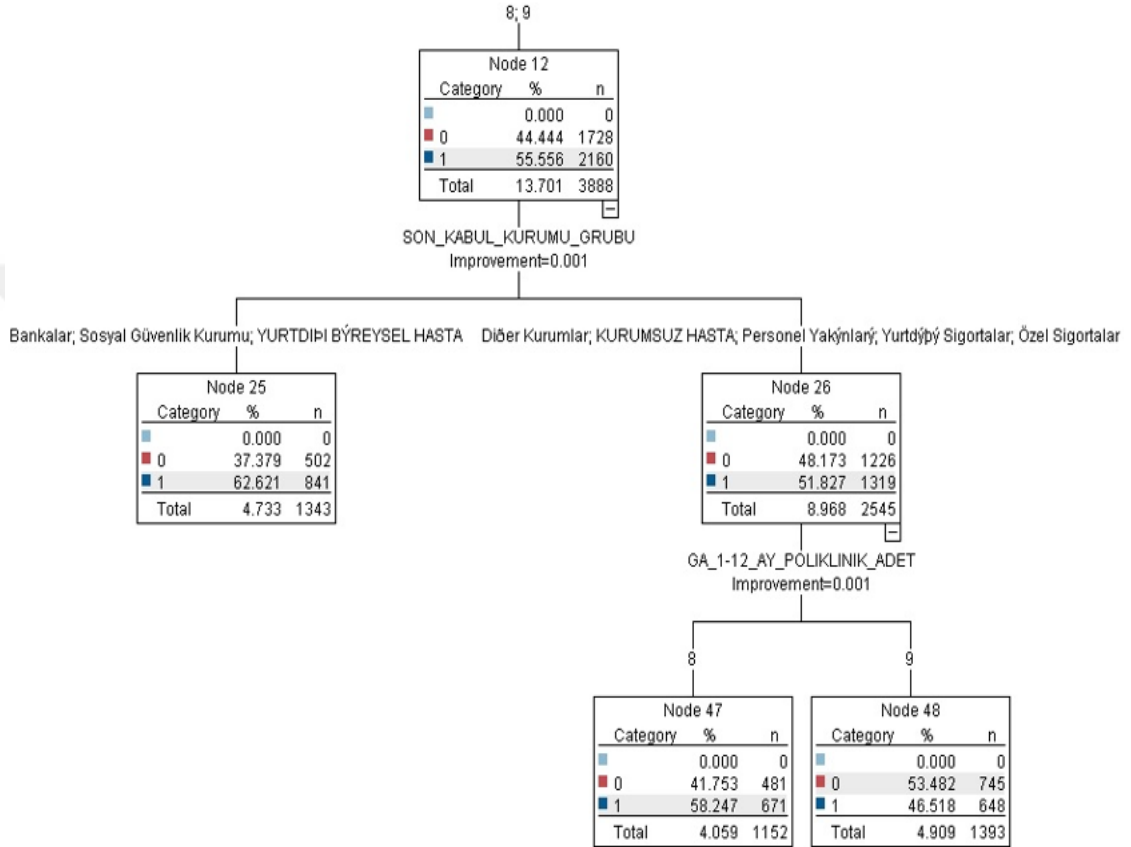
Şekil 5.11.: 0-1 Yaş CART 2. Düğüm ve Dalları

15.638 gözlem barındıran 1-12 ay arasında 4 ila 9 arasında poliklinik muayene bulunduran grubunu en iyi bölen kestiricinin GA_0-1_poliklinik_adet olduğu görülmüştür. Eğer poliklinik muayene başvuru sayısı 1,2,3 ise 8.172 gözlem arasından %63,5'inin, eğer poliklinik muayene başvurusu sayısı 4 ise 7.466 gözlem arasından %47,9'unun kayıp hasta olduğu saptanmıştır.



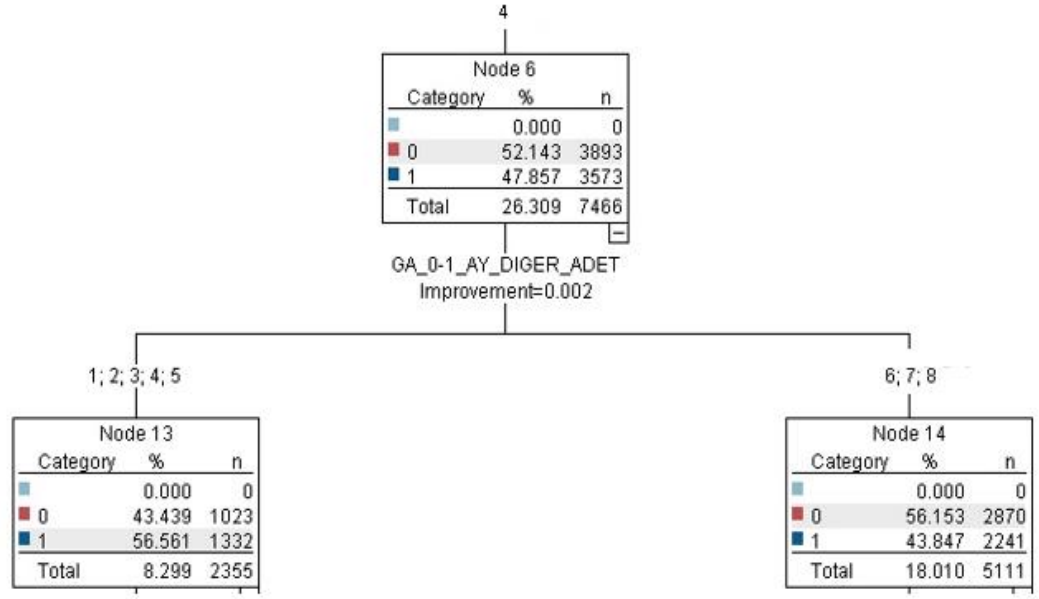
Şekil 5.12.: 0-1 Yaş CART 5. Düğüm ve Dalları

8.172 gözlem barındıran 0-1 ay poliklinik başvuru adedi 1, 2 ve 3 olan 5. düğümü en iyi bölen kestiricinin 1-12 ay poliklinik adet olduğu ve ilgili aralıkta poliklinik muayene başvuru adedinin 4-7 arasında olması durumunda 4.284 gözlem arasından %70,6'sının kayıp hasta olduğu, eğer poliklinik başvuru adedi 8-9 ise, 3.888 gözlem arasından %55,6'sının kayıp hasta olduğu saptanmıştır.



Öekil 5.13.: 0-1 Yaş CART 12. Döğüm ve Dalları

1-12 ay arasında 8-9 poliklinik muayenesi bulunan 3.888 çocuk hasta ise, son kabul kurumunca en iyi şekilde bölünmüştür. Bankalar, SGK, yurtdiÖı bireysel hasta olması durumunda 1.343 gözlem üzerinden %62,6'sının, kurumsuz hasta, özel sigortalar, yurtdiÖı sigortalar olması durumunda 2.545 gözlem arasından %51,8'inin kayıp hasta olduÖu görölmüştür. İlgili grup da eÖer 1-12 aylık dönemde 8 poliklinik muayenesi varsa 1.152 gözlem, %58,2 kayıp hasta oranı ve 9 poliklinik muayene varsa 1.393 gözlem ve %46,5 kayıp hasta oranı olarak bölünmüştür.



Şekil 5.14.: 0-1 Yaş CART 6. Düğüm ve Dalları

1-12 ay aralığında 4-9 poliklinik başvurusu bulunan, 0-1 ayda ise 4 poliklinik muayene başvurusu bulunan, 7.466 gözlem içeren grup 0-1 ay diğer adet başvurularca (poliklinik ve kontrol muayeneleri haricindeki tüm başvurular) en iyi bölünmüş ve eğer diğer başvuru adedi 1-5 arasında ise 2.355 gözlem arasından %56,6'sının, eğer 6-8 arasında ise 5.111 gözlem üzerinden %43,8'inin kayıp hasta olduğu saptanmıştır.

Değerlendirme metrikleri;

Çalışmada 0-1 yaş arasındaki 57.904 devamlı hasta kabul edilen çocuk hastanın churn durumları CART ile değerlendirilmiştir. Toplam veri setinin 40.528 hasta ile %70'i modelin eğitiminde, 17.376 hasta ile %30'u test için kullanılmıştır.

Tablo 5.5.: 0-1 Yaş CART Değerlendirme Metrikleri

CART 0-1 Yaş Churn Değerlendirme Metrikleri					
	Doğru	Yanlış	Doğruluk	Hassasiyet	Kesinlik
Eğitim	70.262	15.295	70,25%	65,09%	55,18%
Test	29.980	6.554	69,75%	63,87%	53,70%

Modelin eğitim veri setinde, modelin %70,25 oranında doğru tahminlemede, test aşamasında da %69,75'lik doğru tahminde bulunduğu saptanmıştır. 0-1 yaş çocuk hastalarının CART algoritması ile kayıp hasta olma durumu değerlendirmesinde, kayıp hasta olmayanlar ve kayıp hasta olanlar için hatalı ve doğru tahmin sayıları hem eğitim veri seti için hem de test veri aşaması için aşağıdaki Tablo 5.6.'da karışıklık (confusion) matrisi olarak verilmiştir.

Tablo 5.6.: 0-1 Yaş CART Karışıklık Matrisi

CART 0-1 Yaş Churn Karışıklık Matrisi				
Eğitim		Değerler	Gerçekleşen	
			0	1
	Tahminlenen	0	19.529	4.795
		1	7.263	8.941
Test		Değerler	Gerçekleşen	
			0	1
	Tahminlenen	0	8.439	2.082
		1	3.174	3.681

Model eğitim veri setinde; 19.529 çocuk hastanın churn olmayacağı doğru şekilde tahminlenmiş, 4.795 çocuk hasta churn olmaması hatalı olarak tahminlenmiş, churn olduğu saptanmıştır. 8.941 çocuk hastanın kayıp hasta olacağı model tarafından doğru tahminlenmiş, 7.263 çocuk hasta ise, kayıp hasta olacağı tahminlenmişken, kayıp hasta olmadığı gözlenmiştir.

Benzer şekilde, **test veri aşamasında;** 8.439 çocuk hastanın kayıp hasta olmayacağı modelce doğru şekilde tahminlenirken, 2.082 çocuk hasta churn hatalı olarak tahminlenirken, kayıp hasta olduğu saptanmıştır. 3.681 çocuk hastanın kayıp hasta olacağı model tarafından doğru tahminlenmiş, 3.174 çocuk hasta ise, kayıp hasta olacağı tahminlenmişken, kayıp hasta olmadığı gözlenmiştir.

Modelin değerlendirilmesinde; **hassasiyet (recall) metriği**, model dahilinde tüm pozitif tahminlenenlerin, yani churn olacağı öngörülenlerin, ne kadarının gerçekte pozitif/kayıp hasta olduğunu incelemektedir. Hassasiyetin eğitim verisinde %65,09,

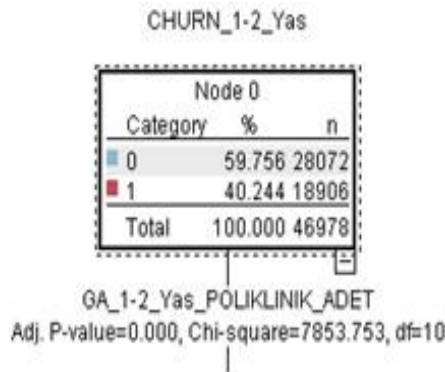
test aşamasında %63,87 olduğu görülmüştür. Bir başka deyişle, 0-1 yaş çocuk grubunun CART algoritması ile tahminlemesinde, modelin eğitim verisi için tüm pozitif/kayıp hasta tahminlenen hastalardan aslında %65,09'sinin kayıp hasta olduğunu, test veri setinde ise ilgili oranın %63,87 olduğunu göstermektedir.

Kesinlik (Precision) metriği, gerçekte pozitif churn olanların aslında kaç tanesinin doğru tahminlendiğinin değerlendirilmesinde; eğitim veri setinde %55,18, test veri aşamasında %53,70 olarak gerçekleştiği saptanmıştır. Bir başka ifade ile eğitim veri setinde, gerçekte churn olan 100 hastanın 55,18 tanesinin pozitif/kayıp hasta olan olarak tahminlendiği, test veri aşamasında ise gerçekte kayıp hasta olan 100 hastanın 53,7'sinin kayıp hasta /pozitif olarak tahminlendiği görülmüştür.

5.2.1. 1-2 Yaş Grubu Geliş Sıklıkları ve MKA

1-2 yaş aralığında ya da 0-1 yaş dönemlerinde 3 poliklinik muayene şartını sağlayan 1-2 yaş aralığındaki çocuk hastaların kayıp hasta kabul edilme durumu karar ağaçları algoritmalarından CHAID ve CART ile analiz edilmiştir.

5.2.1.1. CHAID Algoritması ile Bulgular ve Değerlendirme Metrikleri



Şekil 5.15.: 1-2 Yaş CHAID Genel Durum

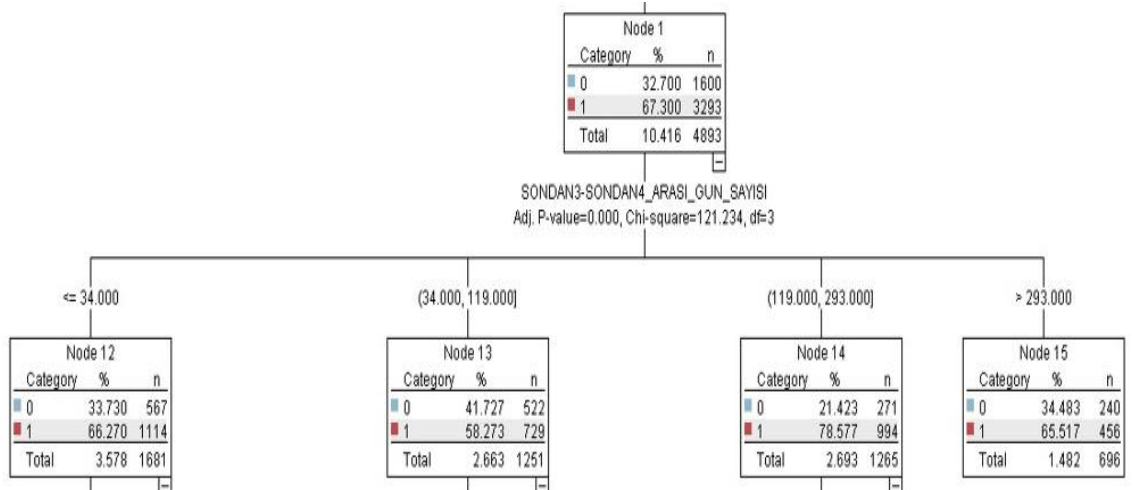
1– 2 yaş aralığında churn analizine konu olan 46.978 adet çocuk hastadan 18.906 hasta ile %40,24'ünün churn olduğu, 28.072 hasta ile %59,76'sının ise devamlı hasta statüsünde hizmet alımına devam ettiği gözlenmiştir. CHAID

algoritmasının ki-kare/df kombinasyonunda en iyi bölünmeyi veren değerin GA_1-2_Yaş_Poliklinik_adet yani, 1-2 yaş aralığındaki toplam poliklinik muayene başvuru sayısının olduğu saptanmıştır.

Tablo 5.7.: 1-2 Yaş Poliklinik Başvurularına Göre Churn Durumu

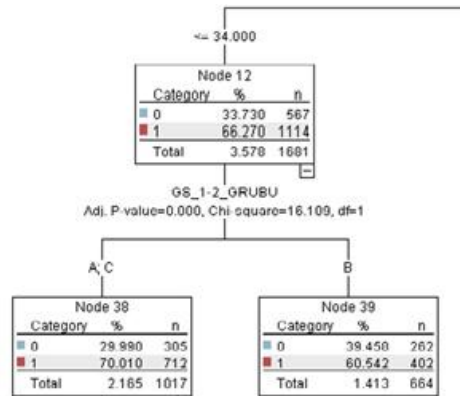
1-2 Yaş Poliklinik Adet	Devamlı Hasta Adedi	Churn Adedi	Toplam Gözlem Adedi	Devamlı Hasta Oranı	Churn Oranı
1	1.187	2.771	3.958	30,0%	70,0%
2	1.600	3.293	4.893	32,7%	67,3%
3	3.174	3.479	6.653	47,7%	52,3%
4	3.582	3.354	6.936	51,6%	48,4%
5	3.762	2.673	6.435	58,5%	41,5%
6	3.814	1.700	5.514	69,2%	30,8%
7	3.349	887	4.236	79,1%	20,9%
8	2.564	402	2.966	86,4%	13,6%
9	1.888	188	2.076	90,9%	9,1%
10	1.124	86	1.210	92,9%	7,1%
10 ve üzeri	1.928	73	2.001	96,4%	3,6%
Toplam	27.972	18.906	46.878	59,7%	40,3%

1-2 yaş arasındaki churn oranı %40,24 olarak saptanmış iken, yukarıdaki tabloda en iyi kestirici olan 1-2 poliklinik adedi başvurusuna göre gözlem adetleri ve churn oranları paylaşılmıştır. Çocuk hastaların 1-2 yaş arasındaki poliklinik muayene başvuru adedi 1 olduğunda, 3.958 gözlem arasından %70'inin churn olduğu, başvuru adedi 2 olduğunda 4.893 gözlem arasından %67,3'ünün churn olduğu, başvuru adedi 3 ise, 6.653 gözlem arasından %52,3'ünün churn olduğu, başvuru adedi 4 ise, 6.936 gözlem arasından %48,4'ünün churn olduğu, başvuru adedi, 5 ise %41,5'inin, poliklinik başvuru adedi 6 ise 5.514 gözlem arasından %30,8'inin, başvuru adedi 7 ise, 4.236 gözlem arasındna %20,9'unun, başvuru adedi 8 ise, 2.966 gözlem arasından %13,6'sının, başvuru adedi 9 ise, 2.076 gözlem arasından %9,1'inin, başvuru adedi 10 ise, 1.210 gözlem arasından %7,1'inin ve son olarak başvuru adedinin 10 ve üzeri olması durumunda 2.001 gözlem arasından %3,6'sının churn olduğu saptanmıştır.



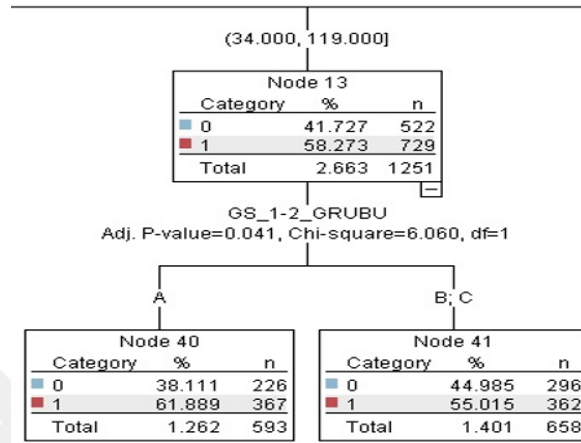
Şekil 5.16.: 1-2 Yaş CHAID 1. D ğ m ve Dalları

1-2 yař arasındaki 2 poliklinik muayene adedi ve 4.893 g zlem sayısının en iyi b len kestiricinin Sondan3-Sondan4_Arası_g n_sayısı olduėu g r lm ř ve eėer ilgili bařvurular arasındaki g n sayısı 34 ve altında ise 1.661 g zlem arasından %66,3' n n, 34 ile 119 g n arasında ise 1.251 g zlem arasından %58,3' n n, 119 ile 293 g n arasında ise 1.265 g zlem arasından 78,6'sının ve 293 g n  zerinde ise 696 g zlem arasında %65,5'inin churn olduėu g zlenmiřtir.



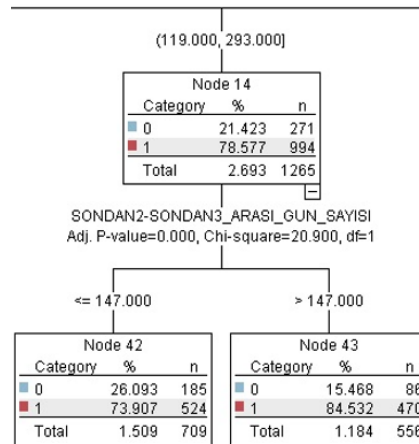
Şekil 5.17.: 1-2 Yaş CHAID 12. D ğ m ve Dalları

Sondan3_Sondan4_arası_gün_sayısının 34 ve altında olduğu 1.661 gözlem içeren 12. düğümü en iyi bölen kesitiricinin GS_1-2_Grubu yani, 1-2 yaş aralığındaki geliş grubu (A,B,C) olduğu, A (4 ve üzeri başvuru) ve C (1 başvuru) gruplarının 1.107 gözlem üzerinden %70'inin, B (2-3 başvuru) grubuna ait çocukların 664 gözlem üzerinden %60,5'inin churn olduğu görülmüştür.



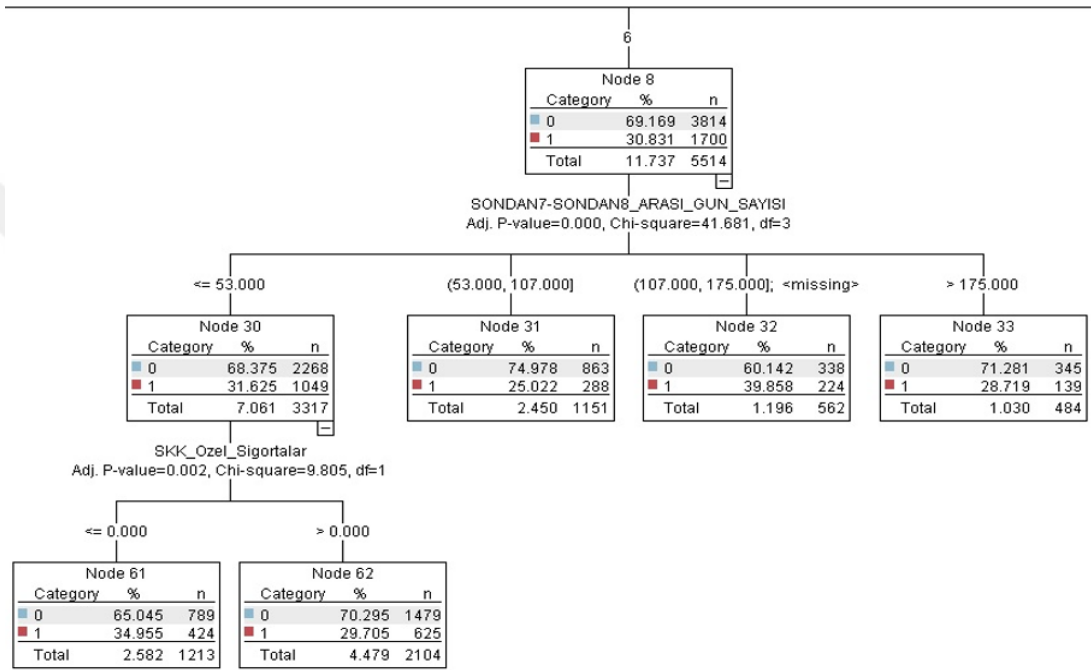
Şekil 5.18.: 1-2 Yaş CHAID 13. Düğüm ve Dalları

Sondan3_Sondan4_arası_gün_sayısının 34 -119 gün olduğu 1.251 gözlem içeren 13. düğümü en iyi bölen kesitiricinin benzer şekilde GS_1-2_Grubu yani, 1-2 yaş aralığındaki geliş grubu (A,B,C) olduğu, A (4 ve üzeri başvuru) grubunun 593 gözlem üzerinden %61,9'unun, B (2-3 başvuru) ve C (1 başvuru) grubuna ait çocukların 658 gözlem üzerinden %55'inin churn olduğu görülmüştür.



Şekil 5.19.: 1-2 Yaş CHAID 14. Düğüm ve Dalları

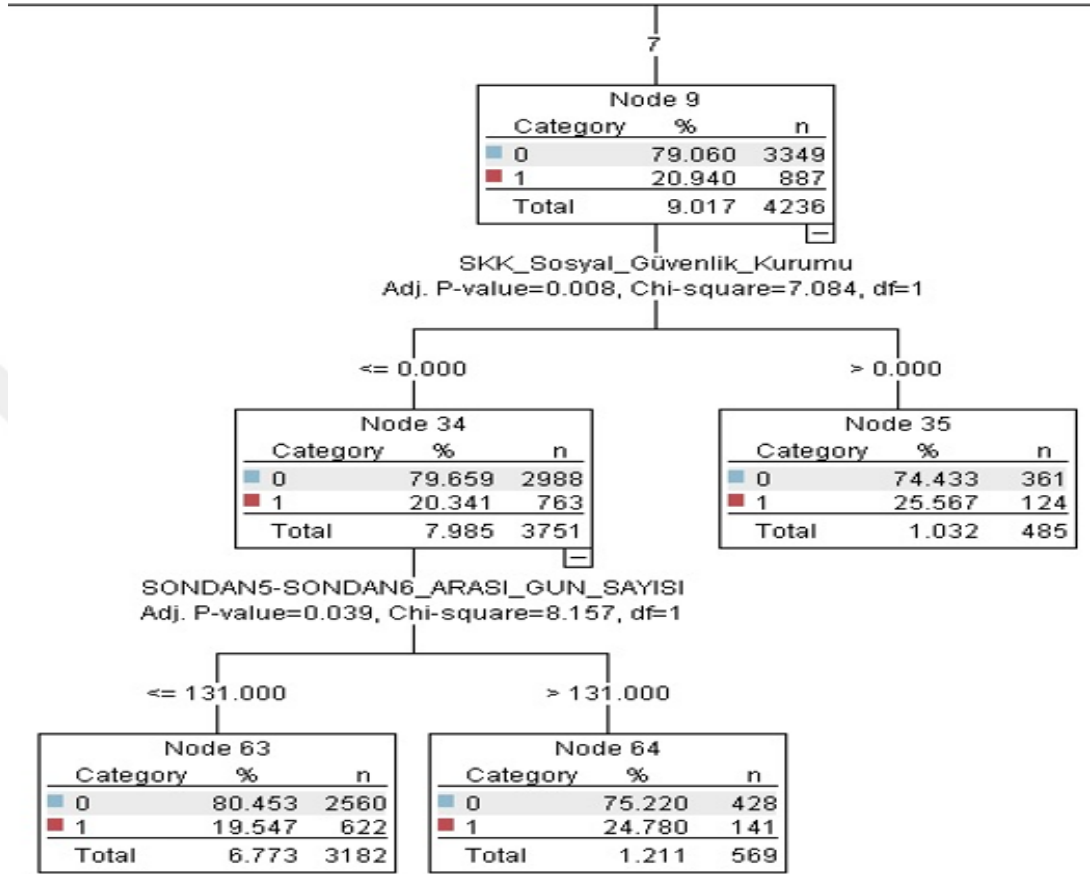
Sondan3_Sondan4_arası_gün_sayısının 119 -293 gün olduğu 1.265 gözlem içeren 14. düğümü en iyi bölen kesitiricinin Sondan2-Sondan3_arası_gün_sayısı olduğu, eğer başvurular arasındaki gün farkı 147 günden az ise 709 gözlem üzerinde %73,9'unun, eğer 147 günden fazla ise 556 gözlem arasından %84,5'inin churn olduğu saptanmıştır.



Şekil 5.20: 12 Yaş CHAID 8. Düğüm ve Dalları

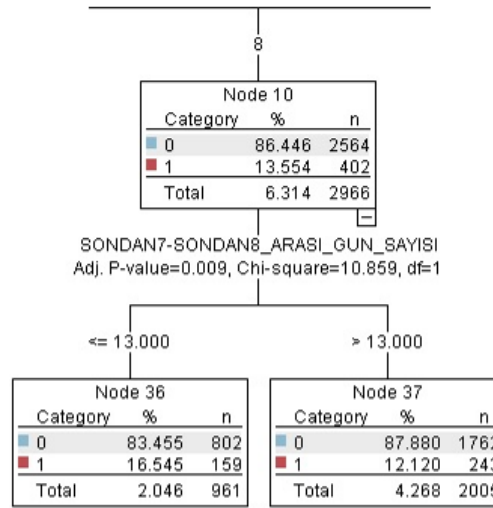
1-2 yaş aralığında 6 poliklinik muayene başvurusu olan 5.514 gözlem ve %30,8 churn oranına sahip olan 8.düğüm de kendi içinde Sondan7_Sondan8_arası_gün_sayısınca en iyi bölünmüş ve ilgili başvurular arasındaki gün sayısı 53 ve altında ise 3.317 gözlem arasından %31,6'sının, 53 ile 107 gün arasında ise 1.151 gözlem üzerinden %25'inin, 107 -175 gün ve kayıp değerler (ilgili yaş aralığında sondan7-sondan8 bulunmuyan, diğer yaş gruplarında bulunan) ise 562 gözlem arasındna %39,9 ve 175 gün üzeri durumunda 484 gözlem üzerinden %28,7'sinin churn olduğu gözlenmiştir. Ayrıca, CHAID modeli, ilgili aralıktaki başvuru gün sayısı farkının 53 gün ve altı durumunu SKK_Ozel_Sigortalar (son kabul kurumu özel sigortalar) olarak bölmüş ve eğer

çocuk hastanın son kabul kurumu özel sigortalar değilse 1.213 gözlem arasından %35'inin, özel sigortalar ise 2.104 gözlem üzerinden %29,7'sinin churn olduğunu saptanmıştır.



Şekil 5.21.: 1-2 Yaş CHAID 9. Düzüm ve Dalları

1-2 yaş aralığında 7 poliklinik muayene başvurusu olan 4.236 gözlem ve %20,9 churn oranına sahip olan 9.düzüm de kendi içinde SKK_Sosyal_Güvenlik_Kurumu, son kabul kurumu sosyal güvenlik kurumunca en iyi bölünmüş ve son kabul kurumu eğer SGK değilse 3.751 gözlem arasından %20,3'ünün, eğer SGK ise 485 gözlem arasından %25,6'sının churn olduğu görülmüştür. Son kabul kurumu SGK olmayan 3.751 gözlem de kendi içerisinde Sondan5_Sondan6_arası_gün_sayısınca bölünmüş ve eğer başvurular arasındaki gün sayısı 131 ve altında ise 3.182 gözlem arasından %19,5'inin, 131 gün üzerinde ise 569 gözlem arasından %24,8'inin churn olduğu saptanmıştır.



Şekil 5.22.: 1-2 Yaş CHAID 10. D ğ m ve Dalları

1-2 yař aralıęında 8 poliklinik muayene bařvurusu olan 2.966 g zlem ve %13,6 churn oranına sahip olan 10. d ğ m de kendi iinde Sondan7_Sondan8_arası_g n_sayısınca en iyi b l nm ř ve eęer ilgili bařvurular arasındaki g n sayısı 13 ve altında ise 961 g zlem arasından ocuk hastaların %16,5'inin, 13 g n  zerinde ise 2.005 g zlem arasından %12,1'inin churn olduęu g r lm řt r.

Deęerlendirme metriklerinde;

alıřmada 1-2 yař arasındaki 67.106 devamlı hasta kabul edilen ocuk hastanın churn durumları CHAID ile deęerlendirilmiřtir. Toplam veri setinin 46.978 hasta ile %70'i modelin eęitiminde, 20.128 hasta ile %30'u test iin kullanılmıřtır.

Tablo 5.8.: 1-2 Yaş CHAID Deęerlendirme Metrikleri

CHAID 1-2 Yaş Churn Deęerlendirme Metrikleri					
	Doęru	Yanlıř	Doęruluk	Hassasiyet	Kesinlik
Eęitim	70.262	15.295	70,86%	64,35%	61,87%
Test	29.980	6.554	70,42%	63,81%	61,66%

Modelin eęitim veri setinde, modelin %70,86 oranında doęru tahminlemede, test ařamasında da %70,42'lik doęru tahminde bulunduęu saptanmıřtır. 1-2 yař

çocuk hastalarının CHAID algoritması ile churn olma durumun değerlendirilmesinde, churn olmayanlar ve churn olanlar için hatalı ve doğru tahmin sayıları hem eğitim veri seti için hem de test veri aşaması için aşağıdaki tablo 17’de karışıklık (confusion) matrisi olarak verilmiştir.

Tablo 5.9.: 1-2 Yaş CHAID Karışıklık Matrisi

CHAID 1-2 Yaş Churn Karışıklık Matrisi				
Eğitim		Değerler	Gerçekleşen	
			0	1
	Tahminlenen	0	21.590	6.482
		1	7.208	11.698
Test		Değerler	Gerçekleşen	
			0	1
	Tahminlenen	0	9.166	2.840
		1	3.114	5.008

Model eğitim veri setinde; 21.590 çocuk hastanın churn olmayacağı doğru şekilde tahminlenmiş, 6.482 çocuk hasta churn olmaması hatalı olarak tahminlenmiş, churn olduğu saptanmıştır. 11.698 çocuk hastanın churn olacağı model tarafından doğru tahminlenmiş, 7.208 çocuk hasta ise, churn olacağı tahminlenmişken, churn olmadığı gözlenmiştir.

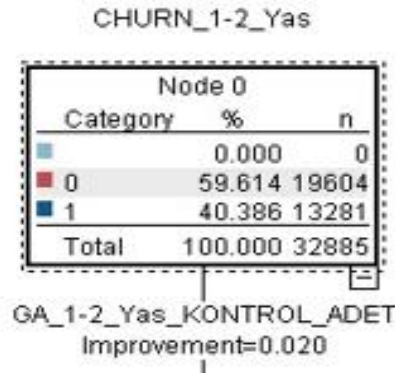
Benzer şekilde, **test veri aşamasında;** 9.166 çocuk hastanın churn olmayacağı modelce doğru şekilde tahminlenirken, 2.840 çocuk hasta churn hatalı olarak tahminlenirken, churn olduğu saptanmıştır. 5.008 çocuk hastanın churn olacağı model tarafından doğru tahminlenmiş, 3.114 çocuk hasta ise, churn olacağı tahminlenmişken, churn olmadığı gözlenmiştir.

Modelin değerlendirilmesinde; **hassasiyet (recall) metriği**, model dahilinde tüm pozitif tahminlenenlerin, yani churn olacağı öngörülenlerin, ne kadarının gerçekte pozitif/kayıp hasta olduğunu incelemektedir. Hassasiyetin eğitim verisinde %64,35,

test aşamasında %63,81 olduğu görülmüştür. Bir başka deyişle, 1-2 yaş çocuk grubunun CHAID algoritması ile tahminlemesinde, modelin eğitim verisi için tüm pozitif/kayıp hasta tahminlenen hastalardan aslında %64,35'inin kayıp hasta olduğunu, test veri setinde ise ilgili oranın %63,81 olduğunu göstermektedir.

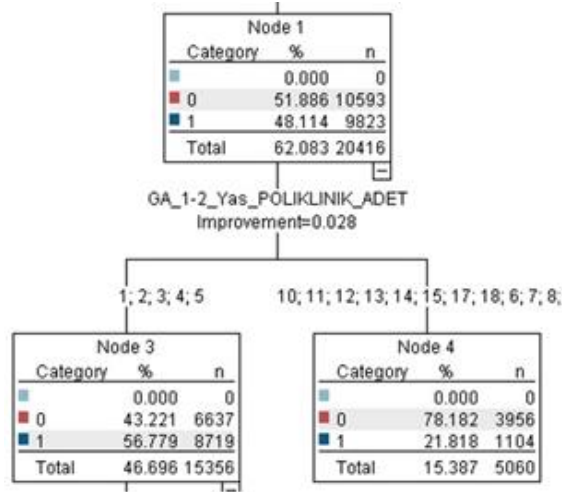
Kesinlik (Precision) metriği, gerçekte pozitif churn olanların aslında kaç tanesinin doğru tahminlendiğinin değerlendirilmesinde; eğitim veri setinde %61,87, test veri aşamasında %61,66 olarak gerçekleştiği saptanmıştır. Bir başka ifade ile, eğitim veri setinde; gerçekte churn olan 100 hastanın 61,87 tanesinin pozitif/kayıp hasta olan olarak tahminlendiği, test veri aşamasında ise gerçekte kayıp hasta olan 100 hastanın 61,66'sının kayıp hasta/pozitif olarak tahminlendiği görülmüştür.

5.2.2.2. CART Algoritması ile Bulgular ve Değerlendirme Metrikleri



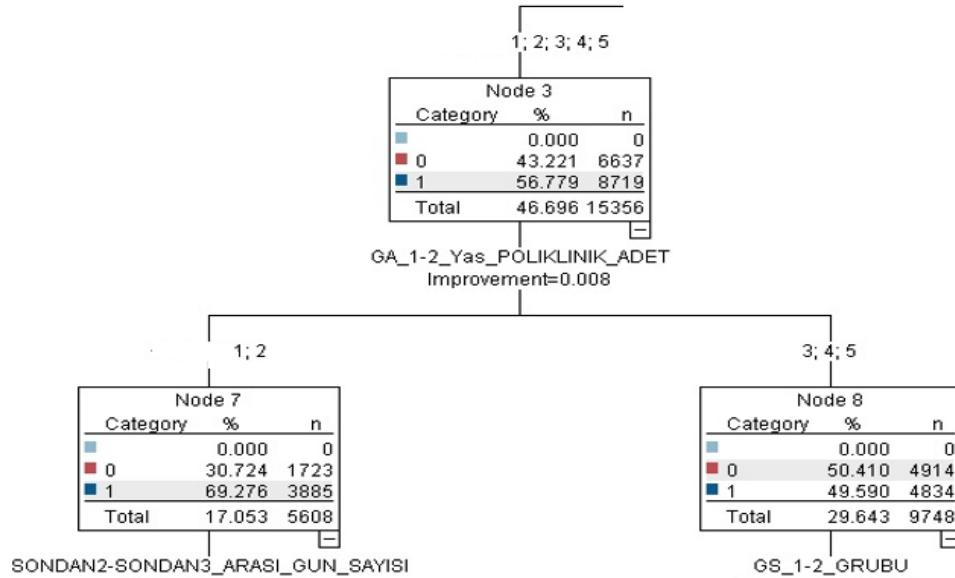
Şekil 5.23.: 1-2 Yaş CART Genel Durumu

1– 2 yaş aralığında kayıp hasta analizine konu olan 32.885 adet çocuk hastadan 13.281 hasta ile %40,4'ünün kayıp hasta olduğu, 19.604 hasta ile 59.6%'nin ise devamlı hasta statüsünde hizmet alımına devam ettiği gözlenmiştir. CART algoritmasında en iyi bölünmeyi veren değer GA_1-2_yas_kontrol_adet yani 1-2 yaş aralığındaki çocuk hastaların kontrol muayene başvuru adetleri olduğu saptanmıştır. Eğer 1 kontrol muayene varsa 20.416 gözlem ile %48,1'inin kayıp hasta olduğu, 2 ve üzeri kontrol muayenesi bulunanlar 12.469 gözlem arasından %27,7'sinin kayıp hasta olduğu görülmüştür.



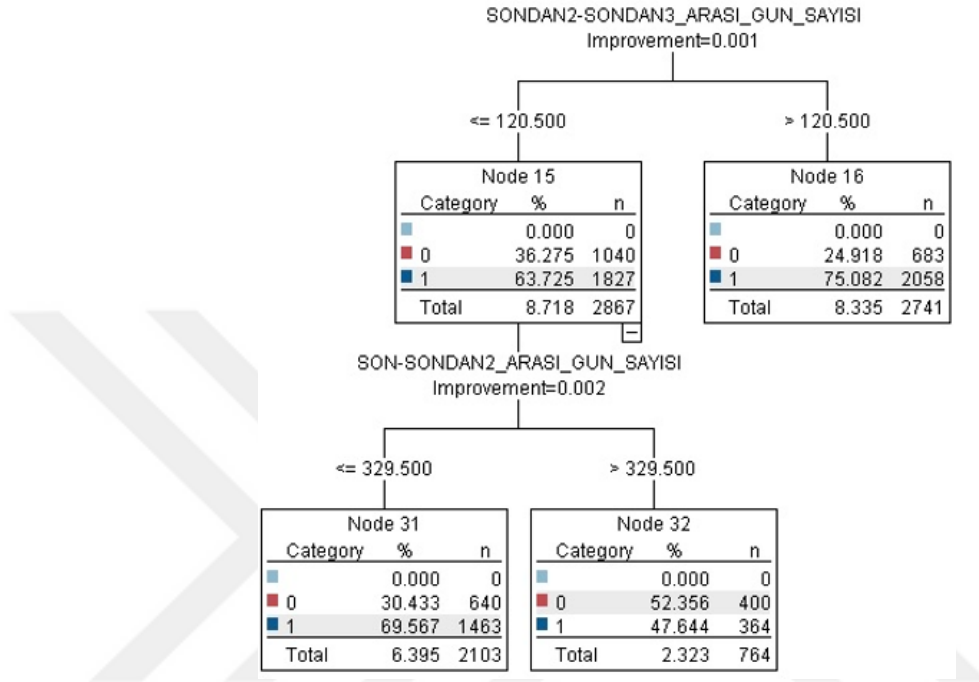
Şekil 5.24.: 1-2 Yaş CART 1. Düğüm ve Dalları

20.416 gözlem içeren birinci düğüm kendi içinde GA_1-2_yas_poliklinik_adet, 1-2 yaş aralığındaki poliklinik muayene başvuru adedince en iyi bölünmüş ve eğer poliklinik muayene başvuru sayısı 1 ve 5 (dahil) arasında ise 15.456 gözlem arasından %56,8'inin, eğer 6 ve üzeri poliklinik muayene başvurusu varsa 5.060 gözlem arasından %21,8'inin kayıp hasta olduğu saptanmıştır.



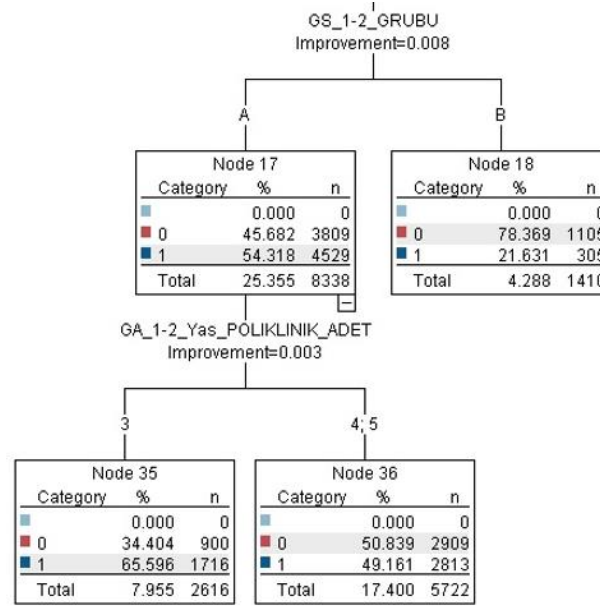
Şekil 5.25.: 1-2 Yaş CART 3. Düğüm ve Dalları

15.356 çocuk hastaya ait gözlem içeren 1-2 yaş aralığında 1-5 adet poliklinik muayene başvuru olan çocuk hastalar, yine kendi içerisinde 1-2 poliklinik muayene başvurusu, 5.608 gözlem, %69,3 kayıp hasta oranı ve 3,4,5 poliklinik muayene başvurusu 9.748 gözlem ve %49,6 kayıp hasta oranı ile bölünmüştür.



Şekil 5.26.: 1-2 Yaş CART 15. Düğüm ve Dalları

5.608 gözlem içeren, 1-2 yaş aralığında 1-2 poliklinik muayene başvurusu bulunan, 7. düğüm de kendi içinde Sondan2-Sondan3_arası_gün_sayısınca en iyi bölünmüş ve eğer başvurular arasında 121 günden az var ise 2.867 gözlem arasından %63,7'sinin, 121 gün üzeri fark var ise, 2.741 gözlem üzerinden %75'inin kayıp hasta olduğu gözlenmiştir. 2.867 gözlem içeren sondan2-sondan3_arası_gün_sayısı 121 gün altı olan sınıfı en iyi bölen kestiricinin de Son_sondan2_arası_gün_sayısı olduğu, eğer başvurular arasında 330'günden az fark var ise 2.103 gözlem arasından %69,6'sının, 330 günden fazla ise 764 gözlem arasından %47,6'sının kayıp hasta olduğu görülmüştür.



Şekil 5.27.: 1-2 Yaş CART 17. Düğüm ve Dalları

9.748 gözlem içeren, 1-2 yaş aralığında 3-4-5 poliklinik muayene başvurusu bulunan, 8. düğüm de kendi içinde GS_1-2_grubu, yani 1-2 yaş aralığındaki geliş sayısınınca belirlenen geliş grubunca en iyi bölünmüş, eğer A (4 ve üzeri poliklinik muayene) grubunda ise 8.338 gözlem üzerinden %54,3'ünün, B grubunda ise (2-3 poliklinik muayene) 1.410 gözlem üzerinden %21,6'sının kayıp hasta olduğu görülmüştür. Ayrıca 8.338 gözlem içeren A grubundaki çocuk hastalar GA_1-2_yas_poliklinik_adedince en iyi bölünmüş ve poliklinik muayene adedi 3 ise, 2.616 gözlem arasından %65'6'sının, poliklinik muayene adedi 4 ve 5 ise, 5.722 gözlem arasından %49,1'inin kayıp hasta olduğu saptanmıştır.

Değerlendirme metriklerinde;

Çalışmada 1-2 yaş arasındaki 67.106 devamlı hasta kabul edilen çocuk hastanın kayıp hasta durumları C & T ile değerlendirilmiştir. Toplam veri setinin 46.978 hasta ile %70'i modelin eğitiminde, 20.128 hasta ile %30'u test için kullanılmıştır.

Tablo 5.10.: 1-2 Yaş CART Değerlendirme Metrikleri

CART 1-2 Yaş Churn Değerlendirme Metrikleri					
	Doğru	Yanlış	Doğruluk	Hassasiyet	Kesinlik
Eğitim	70.262	15.295	68,73%	69,80%	39,32%
Test	29.980	6.554	68,74%	70,07%	42,65%

Modelin eğitim veri setinde, modelin %68,73 oranında doğru tahminlemede, test aşamasında da %68,74'lük doğru tahminde bulunduğu saptanmıştır. 1-2 yaş çocuk hastalarının CART algoritması ile kayıp hasta olma durumun değerlendirmesinde, kayıp hasta olmayanlar ve kayıp hasta olanlar için hatalı ve doğru tahmin sayıları hem eğitim veri seti için hem de test veri aşaması için aşağıdaki tablo 19'da karışıklık (confusion) matrisi olarak verilmiştir.

Tablo 5.11.: 1-2 Yaş CART Karışıklık Matrisi

CART 1-2 Yaş Churn Karşıtlık Matrisi				
Eğitim		Değerler	Gerçekleşen	
			0	1
	Tahminlenen	0	24.856	3.216
		1	11.473	7.433
Test		Değerler	Gerçekleşen	
			0	1
	Tahminlenen	0	10.641	1.365
		1	4.297	3.195

Model eğitim veri setinde; 24.856 çocuk hastanın kayıp hasta olmayacağı doğru şekilde tahminlenmiş, 3.216 çocuk hasta kayıp hasta olmaması hatalı olarak tahminlenmiş, kayıp hasta olduğu saptanmıştır. 7.433 çocuk hastanın kayıp hasta olacağı model tarafından doğru tahminlenmiş, 11.473 çocuk hasta ise, kayıp hasta olacağı tahminlenmişken, kayıp hasta olmadığı gözlenmiştir.

Benzer şekilde, **test veri aşamasında**; 10.641 çocuk hastanın kayıp hasta olmayacağı modelce doğru şekilde tahminlenirken, 1.365 çocuk hasta, hatalı olarak kayıp hasta tahminlenirken, kayıp hasta olduğu saptanmıştır. 3.195 çocuk hastanın kayıp hasta olacağı model tarafından doğru tahminlenmiş, 4.297 çocuk hasta ise, kayıp hasta olacağı tahminlenmişken, kayıp hasta olmadığı gözlenmiştir.

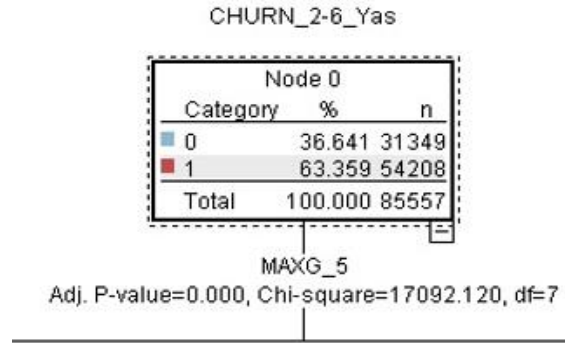
Modelin değerlendirilmesinde; **hassasiyet (recall) metriği**, model dahilinde tüm pozitif tahminlenenlerin, yani kayıp hasta olacağı öngörülenlerin, ne kadarının gerçekte pozitif / kayıp hasta olduğunu incelemektedir. Hassasiyetin eğitim verisinde %69,80, test aşamasında %70,07 olduğu görülmüştür. Bir başka deyişle, 1-2 yaş çocuk grubunun C & T algoritması ile tahminlemesinde, modelin eğitim verisi için tüm pozitif/kayıp hasta tahminlenen hastalardan aslında %69,80'inin kayıp hasta olduğunu, test veri setinde ise ilgili oranın %70,07 olduğunu göstermektedir.

Kesinlik (Precision) metriği, gerçekte pozitif kayıp hasta olanların aslında kaç tanesinin doğru tahminlendiğinin değerlendirilmesinde; eğitim veri setinde %39,32, test veri aşamasında %42,65 olarak gerçekleştiği saptanmıştır. Bir başka ifade ile, eğitim veri setinde; gerçekte kayıp hasta olan 100 hastanın 39,32 tanesinin pozitif/kayıp hasta olan olarak tahminlendiği, test veri aşamasında ise gerçekte kayıp hasta olan 100 hastanın 42,65'inin kayıp hasta/pozitif olarak tahminlendiği görülmüştür.

5.2.3. 2-6 Yaş Grubu Geliş Sıklıkları ve MKA

2-6 yaş aralığında ya da daha önceki tanımlı grup dönemlerinde 3 poliklinik muayene şartını sağlayan 2-6 yaş aralığındaki çocuk hastaların kayıp hasta kabul edilme durumu karar ağaçları algoritmalarından CHAID ve CART ile analiz edilmiştir.

5.2.3.1. CHAID Algoritması ile Bulgular ve Değerlendirme Metrikleri



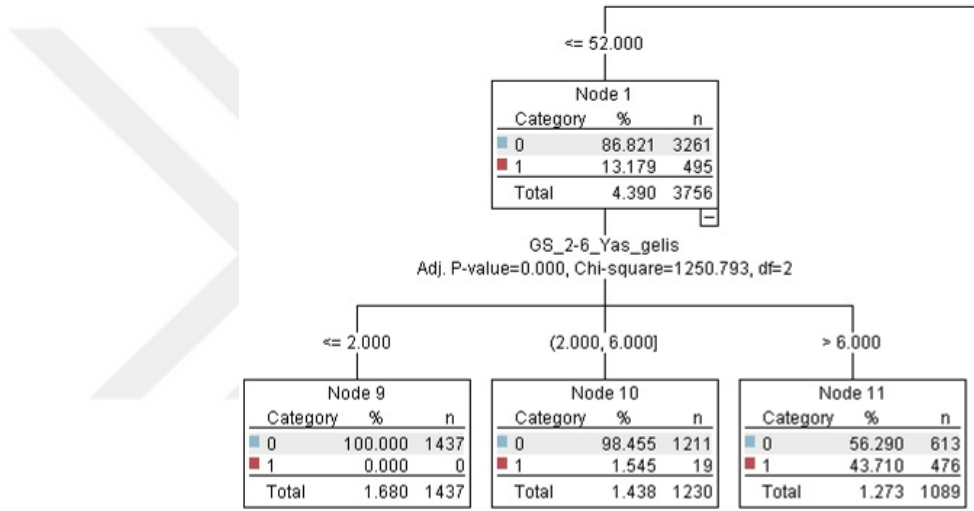
Şekil 5.28.: 2-6 Yaş CHAID Genel Durumu

2– 6 yaş aralığında kayıp hasta analizine konu olan 85.257 adet çocuk hastadan 54.208 hasta ile %63,6'sının kayıp hasta olduğu, 31.349 hasta ile 36.4%'ünün ise devamlı hasta statüsünde hizmet alımına devam ettiği gözlenmiştir. CHAID algoritmasının ki-kare/df kombinasyonunda en iyi bölünmeyi veren değerin MAXG_5, yani çocuk hastalarının 5 ve 6 yaş arasındaki iki başvurusu arasındaki maksimum gün sayısının olduğu saptanmıştır. Burada bölünen 8 dala ilişkin bilgi özetlenerek tablo 20'de sunulmuştur.

Tablo 5.12: 2-6 Yaş CHAID MAXG_5 Sınıfı Başvurular Arası Gün Farkı ve Kayıp hasta Durumu

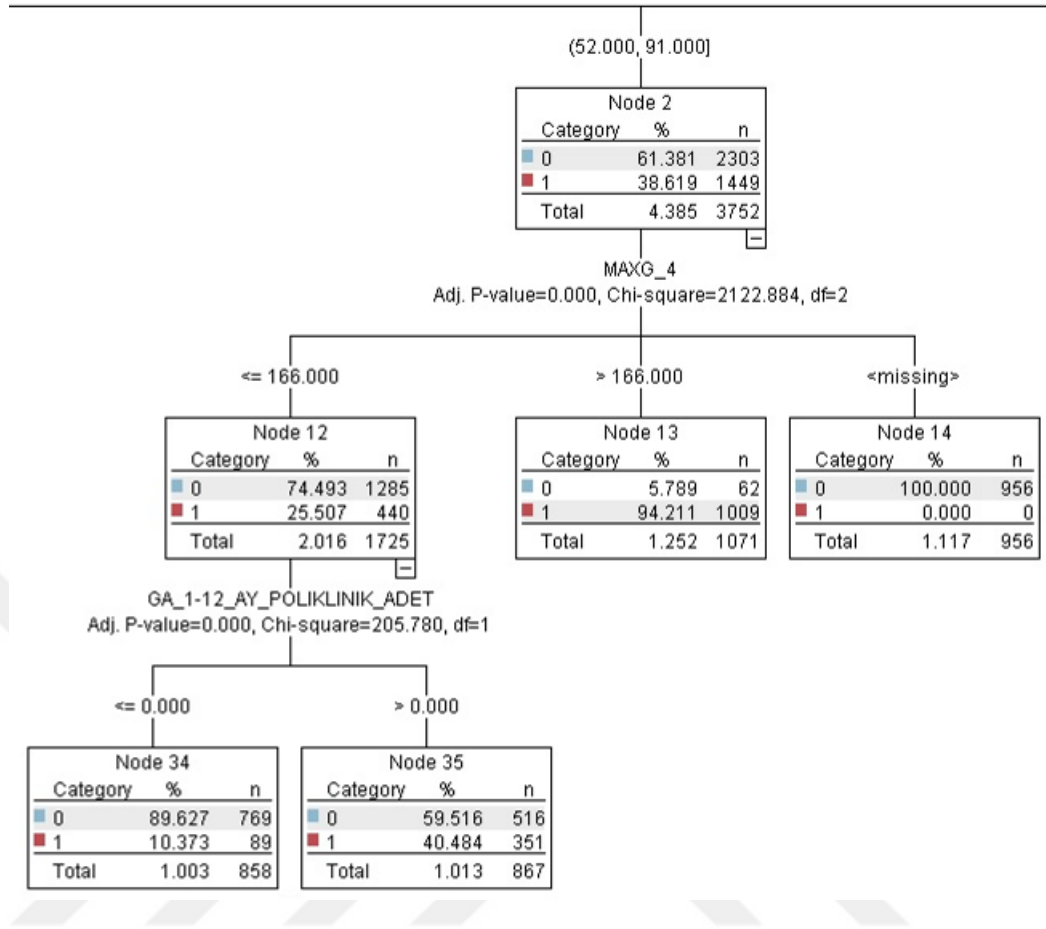
MAXG_5	Devamlı Hasta Adedi	Churn Adedi	Toplam Gözlem Adedi	Devamlı Hasta Oranı	Churn Oranı
52 günden az	3.261	495	3.756	86,8%	13,2%
52-91 gün arası	2.303	1.449	3.752	61,4%	38,6%
91-119 gün arası	2.080	1.651	3.731	55,7%	44,3%
119-146 gün arası	1.930	1.803	3.733	51,7%	48,3%
146-177 gün arası	1.792	2.042	3.834	46,7%	53,3%
177-215 gün arası	134	3.634	3.768	3,6%	96,4%
215 gün sonrası	0	14.995	14.995	0,0%	100,0%
Kayıp değerler	19.849	28.139	47.988	41,4%	58,6%
Toplam	31.349	54.208	85.557	36,6%	63,4%

2-6 yaş arasındaki kayıp hasta oranı %63,4 olarak saptanmış iken, ilk düğümde, çocuk hastaların 5-6 yaş arasındaki gelişleri arasındaki maksimum gün farkı; 52 gün ve altında olduğunda; kayıp hasta oranının %13,2, 52-91 gün arasında %38,6, 91-119 gün arasında %44,3, 119-146 gün arasında %48,3, 146-177 gün arasında %53,3, 177-215 gün arasında %96,4 ve 215 gün üzerinde 100% olduğu saptanmıştır. 2-6 yaş arasında kayıp hasta bilgisi olan fakat MAXG_5 sınıfında 1'den fazla başvurusu olmayanların, 47.988 adet gözlem,%58,6'sının kayıp hasta olduğu saptanmıştır.



Şekil 5.29.: 2-6 Yaş CHAID 1. Düzüm ve Dalları

MAXG_5'in 52 gün ve altında olduğu dalı en iyi bölen kestiricinin ise 2-6 yaş aralığındaki toplam geliş sayısı olduğu saptanmış ve 2-6 yaş arasındaki toplam geliş adedinin 3.756 gözlem olduğu, hastaların 2 ve altında başvurusu varsa hiç kayıp hasta olmadığı, 3 ve 6 arasında başvurusu varsa 1.230 gözlem arasından 19 tanesinin, 6'dan fazla başvurusu olan hastaların ise 1.089 gözlem arasından %43,7 ile 476'sının kayıp hasta olduğu gözlenmiştir.

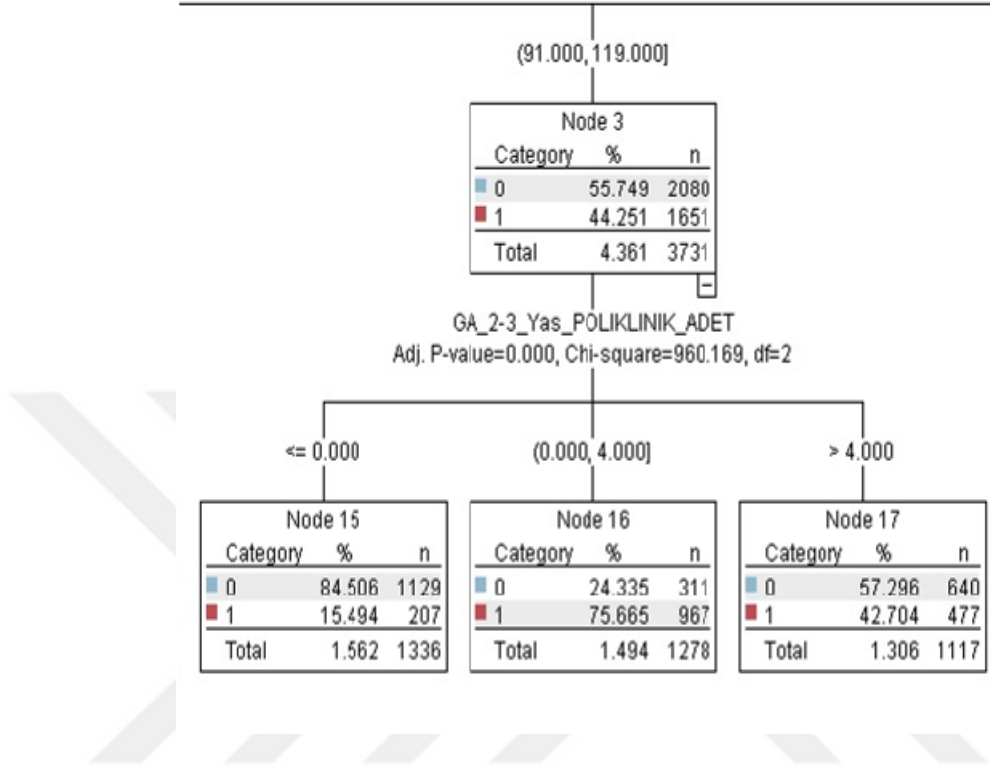


Şekil 5.30.: 2-6 Yaş CHAID 2. Düğüm ve Dalları

İkinci düğümü olan MAXG_5 52-91 günü en iyi bölen kestiricinin MAXG_4 yani 4-5 yaşları arasındaki iki başvuru arasındaki maksimum gün farkı olduğu görülmüştür. MAXG_4 de kendi içinde; 166 gün ve altı, 166 gün üzeri ve kayıp değerler (ilgili yaş aralığında 1 ve 1'den az başvurular) olarak dallanmıştır. MAXG_4'ün 166 gün ve altında olduğu durumlarda 1725 gözlem arasında kayıp hasta oranının %25,5 iken 166 gün üzeri olduğu durumlarda 1071 gözlem arasından %94,2'e yükseldiği saptanmıştır.

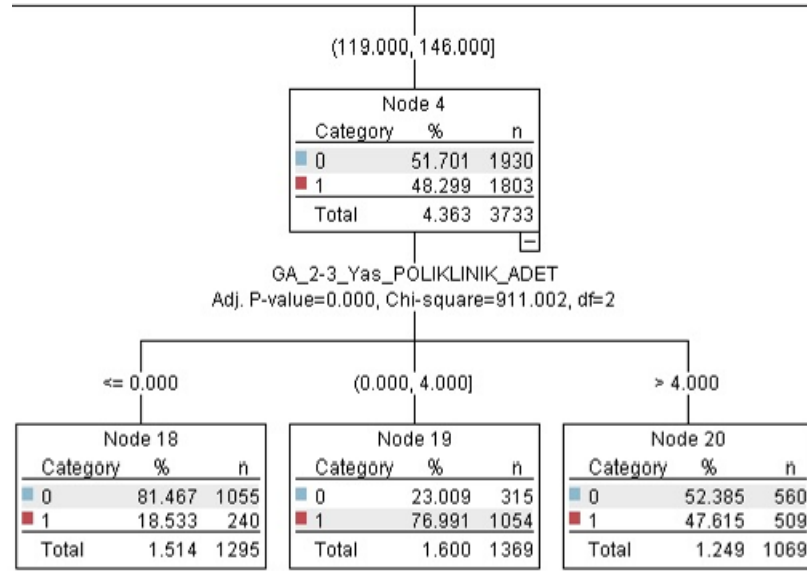
166 günden az olan 1725 gözlemi de kendi içinde; 1-12 ay aralığındaki poliklinik muayene adetlerinin en şekilde anlamlandırdığı görülmüştür. 1-12 ay aralığında herhangi bir poliklinik muayenesi bulunmayan 858 çocuğun 89 tanesinin (10%'nun) kayıp hasta olduğu, 1-12 ay aralığında herhangi bir poliklinik

başvurusu olan çocukların ise 2-6 yaş aralığında 867 gözlem arasından %40,5'nin kayıp hasta olduğu saptanmıştır.



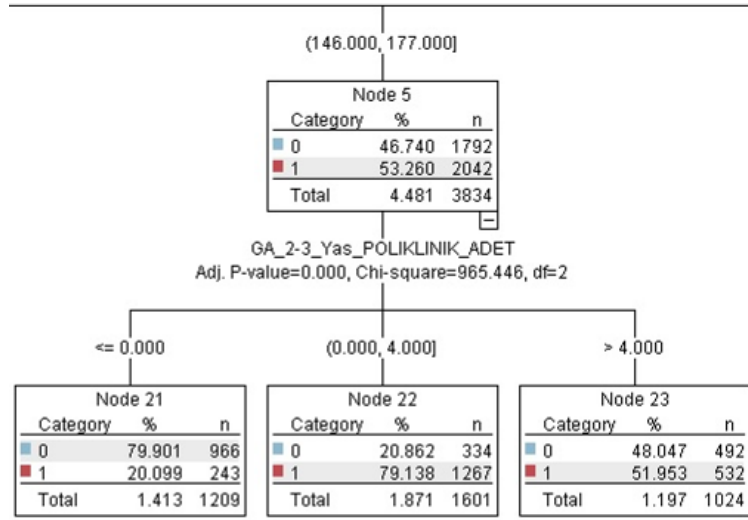
Şekil 5.31.: 2-6 Yaş CHAID 3. Düzüm ve Dalları

Üçüncü düğüm olan MAXG_5 91-119 günü en iyi bölen kestiricinin GA_2-3_Yas_Poliklinik_Adet yani, 2 ve 3 yaş aralığındaki çocuk hastanın poliklinik muayene başvuru adedi olduğu görülmüştür. İlgili grup kendi içinde ise; eğer çocuğun 2-3 yaş aralığında herhangi bir başvurusu yoksa, 1.336 gözlem arasından kayıp hasta oranının %15,5 olduğu, 1-4 arasında poliklinik muayene başvurusu varsa 1.278 gözlem arasından kayıp hasta oranının %75,7 olduğu ve 4'de fazla başvurusu varsa 1.117 gözlem arasından %42,7'sinin kayıp hasta olduğu saptanmıştır.



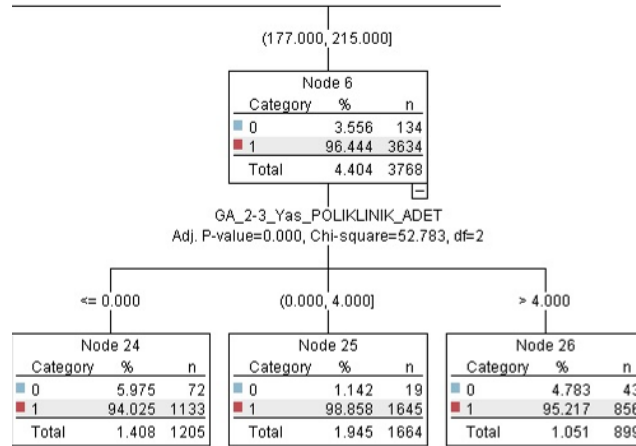
Şekil 5.32: 2-6 Yaş CHAID 4. Düğüm ve Dalları

Dördüncü düğüm olan MAXG_5 119-146 günü en iyi bölen kestiricinin benzer şekilde GA_2-3_Yas_Poliklinik_Adet yani, 2 ve 3 yaş aralığındaki çocuk hastanın poliklinik muayene başvuru adedi olduğu görülmüştür. İlgili grup kendi içinde ise; eğer çocuğun 2-3 yaş aralığında herhangi bir başvurusu yoksa, 1.295 gözlem arasından kayıp hasta oranının %18,5 olduğu, 1-4 arasında poliklinik muayene başvurusu varsa 1.369 gözlem arasından kayıp hasta oranının %76,9 olduğu ve 4'de fazla başvurusu varsa 1.069 gözlem arasından %47,6'sının kayıp hasta olduğu saptanmıştır.



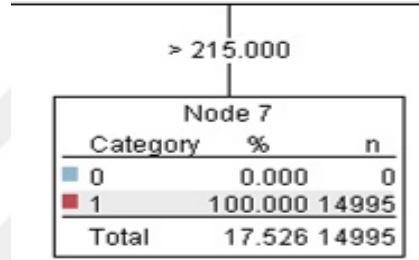
Şekil 5.33: 2-6 Yaş CHAID 5. Düğüm ve Dalları

Beşinci düğüm olan MAXG_5 146-177 günü en iyi bölen kestiricinin benzer şekilde GA_2-3_Yas_Poliklinik_Adet yani, 2 ve 3 yaş aralığındaki çocuk hastanın poliklinik muayene başvuru adedi olduğu görülmüştür. İlgili grup kendi içinde ise; eğer çocuğun 2-3 yaş aralığında herhangi bir başvurusu yoksa, 1.209 gözlem arasından kayıp hasta oranının %20,1, 1-4 arasında poliklinik muayene başvurusu varsa 1.601 gözlem arasından kayıp hasta oranının %76,1 ve 4'de fazla başvurusu varsa 1.024 gözlem arasından %51,96'sının kayıp hasta olduğu saptanmıştır.



Şekil 5.34.: 2-6 Yaş CHAID 6. Düğüm ve Dalları

Altıncı düğüm olan MAXG_5 177-215 günü en iyi bölen kestiricinin benzer şekilde GA_2-3_Yas_Poliklinik_Adet yani, 2 ve 3 yaş aralığındaki çocuk hastanın poliklinik muayene başvuru adedi olduğu görülmüştür. İlgili grup kendi içinde ise; eğer çocuğun 2-3 yaş aralığında herhangi bir başvurusu yoksa, 1.205 gözlem arasından kayıp hasta oranının %94 olduğu, 1-4 arasında poliklinik muayene başvurusu varsa 1.664 gözlem arasından kayıp hasta oranının %98,9 olduğu ve 4'den fazla başvurusu varsa 899 gözlem arasından %95,2'sinin kayıp hasta olduğu saptanmıştır.



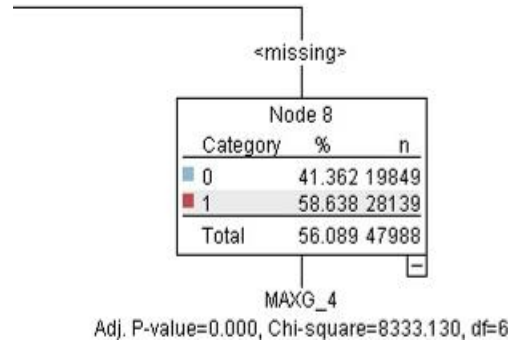
Node 7

Category	%	n
0	0.000	0
1	100.000	14995
Total	17.526	14995

> 215.000

Şekil 5.35: 2-6 Yaş CHAID 7. Düğüm ve Dalları

Yedinci düğüm de MAXG_5 in 215 gün ve üzerinde olduğu 14.995 başvuruyu içermekte ve 100% lük kayıp hasta oranı göstermektedir.



Node 8

Category	%	n
0	41.362	19849
1	58.638	28139
Total	56.089	47988

<missing>

MAXG_4

Adj. P-value=0.000, Chi-square=8333.130, df=6

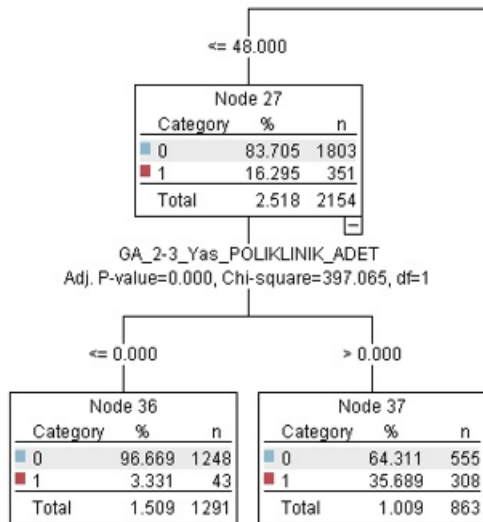
Şekil 5.36.: 2-6 Yaş CHAID 8. Düğüm ve Dalları

8. düğüm ise kayıp değerleri yani, MAXG_5'de en fazla bir başvurunun olduğu sınıfı göstermektedir. 8. düğümü en iyi bölen kestiricinin ise MAXG_4 yani 4-5 yaş

aralığındaki iki başvuru arasındaki en fazla gün sayısı olduğu saptanmıştır. Aşağıdaki tablo 21’ de toplu halde verilen, başvurular arasındaki gün farkı 48 gün ve altında olduğunda 2.154 gözlem arasından; kayıp hasta oranının %16,3, 48-84 gün arasında 1.434 gözlem arasından %39,7, 84-136 gün arasında 2.464 gözlem arasından %48,7, 136-165 gün arasında 1.284 gözlem arasından %52,3, 166-200 gün arasında 1.356 gözlem arasından %79,9, ve 200 gün üzerinde 7.935 gözlem arasından 100% olduğu saptanmıştır.

Tablo 5.13: 2-6 Yaş CHAID MAXG_4 Sınıfı Başvurular Arası Gün Farkı ve Kayıp hasta Durumu

MAXG_4	Devamlı Hasta Adedi	Churn Adedi	Toplam Gözlem Adedi	Devamlı Hasta Oranı	Churn Oranı
48 günden az	1.803	351	2.154	83,7%	16,3%
48-84 gün arası	865	569	1.434	60,3%	39,7%
84-136 gün arası	1.263	1.201	2.464	51,3%	48,7%
136-166 gün arası	613	671	1.284	47,7%	52,3%
166-200 gün arası	273	1.083	1.356	20,1%	79,9%
200 gün sonrası	0	7.935	7.935	0,0%	100,0%
Kayıp değerler	15.032	16.329	31.361	47,9%	52,1%
Toplam	19.849	28.139	47.988	41,4%	58,6%



Şekil 5.37.: 2-6 Yaş CHAID 27. Düğüm ve Dalları

27. düğüm olan MAXG_4 48 günden az dalını en iyi bölen kestiricinin GA_2-3_Yas_Poliklinik_Adet yani, 2 ve 3 yaş aralığındaki çocuk hastanın poliklinik muayene başvuru adedi olduğu görülmüştür. İlgili grup kendi içinde ise; eğer çocuğun 2-3 yaş aralığında herhangi bir poliklinik başvurusu yoksa, 2.154 gözlem arasından kayıp hasta oranının %16,3 olduğu, diğer yandan eğer 1 ve 1'den fazla başvurusu varsa 863 gözlem arasından %35,7'sinin kayıp hasta olduğu saptanmıştır.

(48.000, 84.000]		
Node 28		
Category	%	n
0	60.321	865
1	39.679	569
Total	1.676	1434

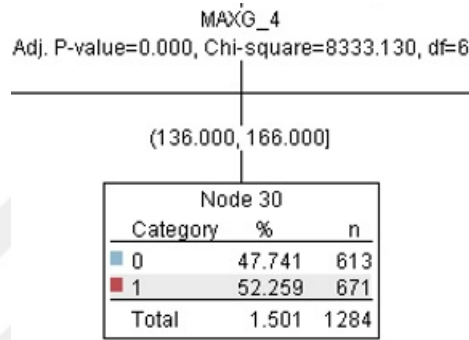
Şekil 5.38.: 2-6 Yaş CHAID 28. Düğüm ve Dalları

28. düğüm olan MAXG_4 48-84 gün dalında, 1.434 gözlem arasından 865'nin devamlı hasta, 569 tanesinin (Gözlemlerin %39,7'si) ise kayıp hasta olduğu saptanmıştır.

(84.000, 136.000]		
Node 29		
Category	%	n
0	51.258	1263
1	48.742	1201
Total	2.880	2464
GA_1-12_AY_POLIKLINIK_ADET		
Adj. P-value=0.000, Chi-square=435.016, df=1		
<= 0.000 > 0.000		
Node 38		
Category	%	n
0	77.523	745
1	22.477	216
Total	1.123	961
Node 39		
Category	%	n
0	34.464	518
1	65.536	985
Total	1.757	1503

Şekil 5.39.: 2-6 Yaş CHAID 27. Düğüm ve Dalları

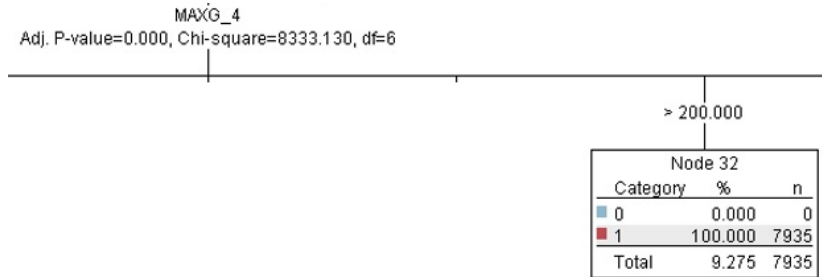
29. düğüm olan MAXG_4 84-136 günü en iyi bölen kestiricinin GA_1-12_Ay_Poliklinik_Adet yani, 1 ve 12 aylar aralığındaki çocuk hastanın poliklinik muayene başvuru adedi olduğu görülmüştür. İlgili grup kendi içinde ise; eğer çocuğun 1-12 ay aralığında herhangi bir başvurusu yoksa, 961 gözlem arasından kayıp hasta oranının %22,5 olduğu, diğer yandan eğer 1'den fazla başvurusu varsa 1.503 gözlem arasından %65,5'inin kayıp hasta olduğu 39. düğümde saptanmıştır.



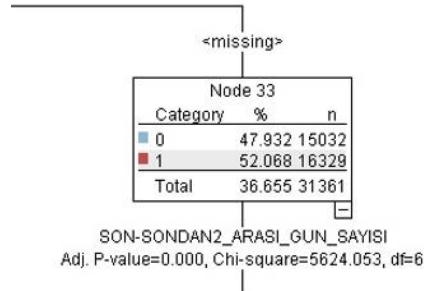
Şekil 5.40.: 2-6 Yaş CHAID 30. Düğüm

30. düğüm olan MAXG_4 136-166 gün dalında, 1.284 gözlem arasından 613'nün devamlı hasta, 671 tanesinin (Gözlemlerin %52,3'ü) ise kayıp hasta olduğu saptanmıştır.

31. düğüm olan MAXG_4 166-200 gün dalında, 1.356 gözlem arasından 273'ünün devamlı hasta, 1083 tanesinin (Gözlemlerin %79,9'u) ise kayıp hasta olduğu saptanmıştır.



32. düğüm olan MAXG_4 200 ve üzeri gün dalında, 7.935 gözlem arasından %100 ile tümünün kayıp hasta olduğu saptanmıştır



Şekil 5.41.: 2-6 Yaş CHAID 32. ve 33. Düğüm

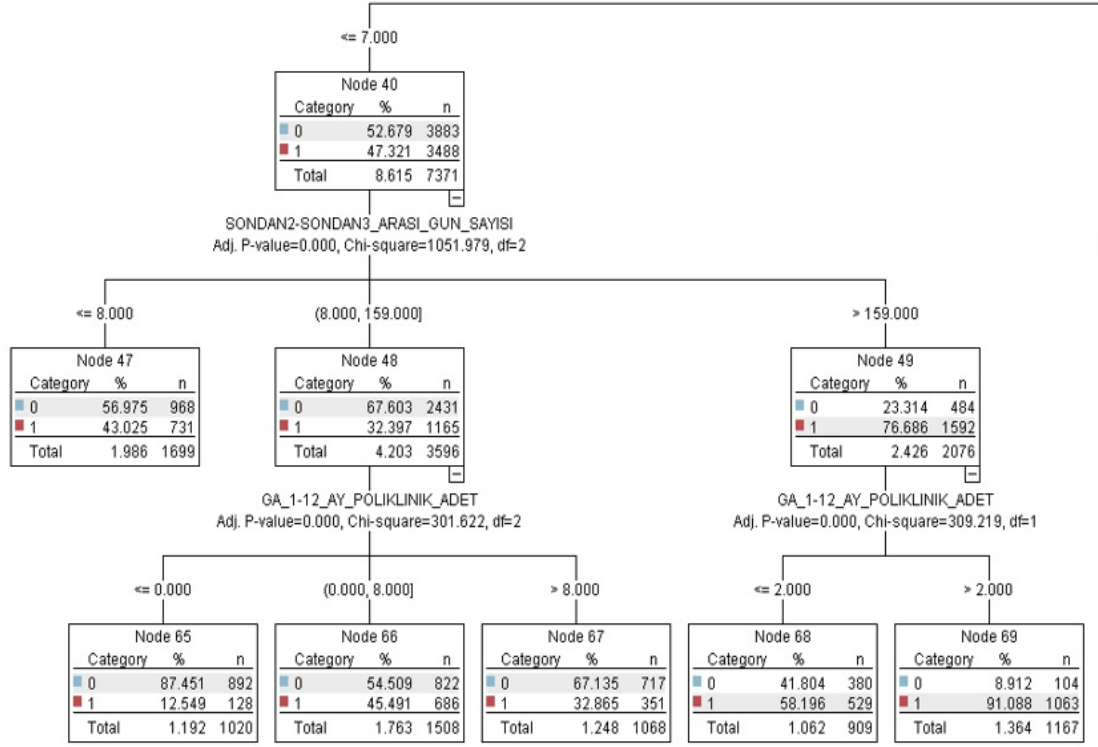
31.361 gözlem barındıran 33.düğüm olan kayıp değerleri en iyi ki-kare/df oranında bölen kestiricinin Son_Sondan2_arası_gün_sayısı olduğu yani, son başvurusu ile ondan bir önceki başvurusu arasındaki gün farkı olduğu saptanmıştır.

Tablo 5.14: 2-6 Yaş CHAID Son ve Sondan2. Başvurular Arası Gün Sayısı ve Kayıp hasta Durumu

Son_Sondan2_Arası_Gün_S ayısı	Devamlı Hasta Adedi	Churn Adedi	Toplam Gözlem Adedi	Devamlı Hasta Oranı	Churn Oranı
7 günden az	3.883	3.488	7.371	52,7%	47,3%
7-16 gün arası	1.814	1.272	3.086	58,8%	41,2%
16-125 gün arası	6.680	3.323	10.003	66,8%	33,2%
125-198 gün arası	1.716	1.517	3.233	53,1%	46,9%
198-312 gün arası	246	2.591	2.837	8,7%	91,3%
312-551 gün arası	294	2.267	2.561	11,5%	88,5%
551 gün sonrası	399	1.871	2.270	17,6%	82,4%
Toplam	15.032	16.329	31.361	47,9%	52,1%

Son başvurusu ve sondan iki başvurular arasındaki gün farkı 7 gün ve altında olduğunda 7.371 gözlem arasından; kayıp hasta oranının %47,3, 7-16 gün arasında 3.086 gözlem arasından %41,2, 16-125 gün arasında 10.003 gözlem arasından %33,2, 125-198 gün arasında 3.323 gözlem arasından %46,9, 198-

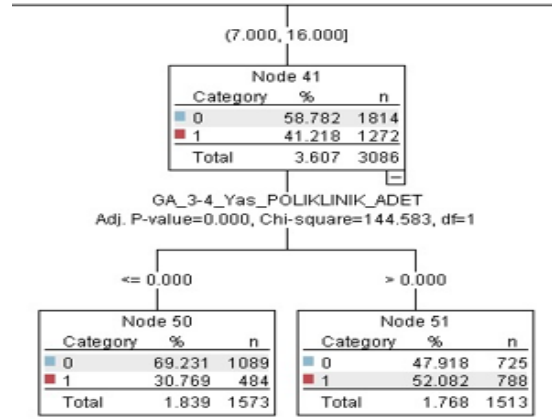
312 gün arasında 2.837 gözlem arasından %91,3, 312-551 gün arasında 2.561 gözlem arasından %88,5, ve 551 gün üzerinde 2.270 gözlem arasından %82,4 olduğu saptanmıştır.



Şekil 5.42.: 2-6 Yaş CHAID 40. Düğüm ve Dalları

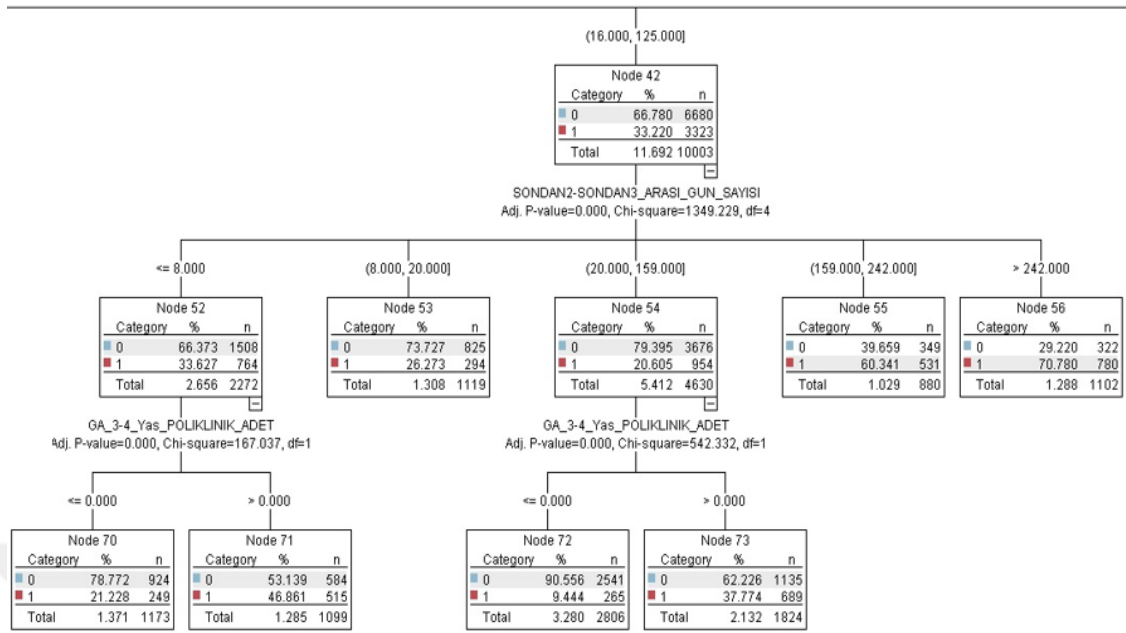
40. düğüm olan ve 7.371 gözlem içeren dalı en iyi bölen kestiricinin Sondan2-Sondan3_arası_gün_sayısı olduğu gözlenmiş ve sondan ikinci ve üçüncü gelişleri arasındaki gün farkı 8 gün ve altındaysa 1.699 gözlem arasından %43'ünün, 8 ile 159 gün arasında ise 3.596 gözlem arasından %32,4'ünün, 159 günden fazla ise 2.076 gözlem arasından %76,7'sinin kayıp hasta olduğu gözlenmiştir. 3.596 gözlem içeren 48. düğüm ise kendi içinde, en iyi GA_1-12_ay_poliklinik_adete bölündüğü ve ilgili dönemde çocuk hastanın poliklinik muayene başvurusu yoksa, 1.020 gözlem arasından %12,55'nin, 1 ile 8 adet poliklinik başvurusu varsa 1.508 gözlem arasından %45'5'inin ve poliklinik başvurusu 8 üzerinde ise, 1.068 gözlem arasından %32,9'unun kayıp hasta olduğu saptanmıştır. Ayrıca, 2.076 gözlem

içeren 49. Düğüm de kendi içinde, en iyi GA_1-12_ay_poliklinik_adete bölündüğü ve ilgili dönemde çocuk hastanın 2 veya daha az poliklinik muayene başvurusu varsa, 909 gözlem arasından %58,2'sinin, 2'den fazla başvurusu varsa 1.167 gözlem arasından %91'inin kayıp hasta olduğu 69. Düğümde saptanmıştır.



Şekil 5.43.: 2-6 Yaş CHAID 41. Düğüm ve Dalları

41. düğüm olan ve 3.086 gözlem içeren dalı en iyi bölen kestiricinin 3-4 yaş aralığındaki poliklinik muayene başvuru adedi olduğu saptanmış ve ilgili dönemde herhangi bir poliklinik muayene başvurusu yoksa 1.573 gözlem arasından %30,7'sinin, 1 ve üzeri poliklinik muayene başvurusu varsa 1.513 gözlem arasından %52'sinin kayıp hasta olduğu 51. düğümde gözlenmiştir.

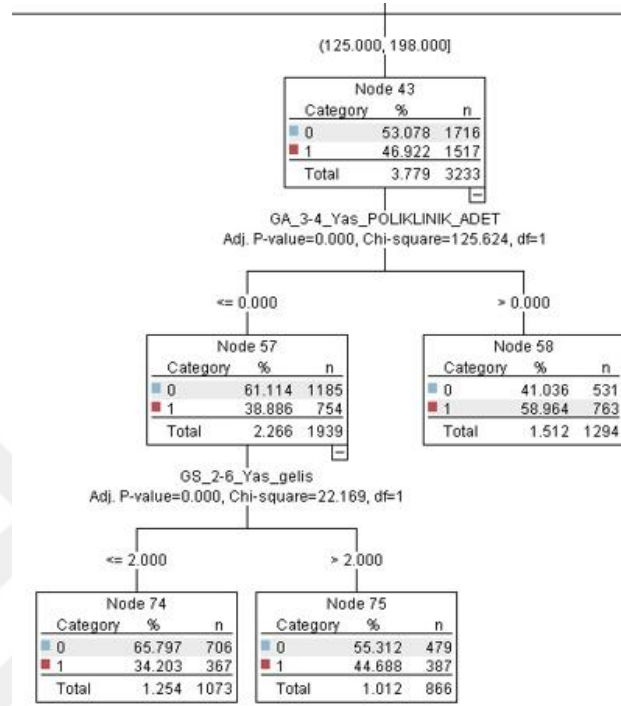


Şekil 5.44.: 2-6 Yaş CHAID 42. Düğüm ve Dalları

42. düğüm olan ve 10.003 gözlem içeren dalı en iyi bölen kestiricinin Sonda2-Sonda3_arası_gün_sayısı olduğu gözlenmiş ve sonda ikinci ve üçüncü gelişleri arasındaki gün farkı 8 günden az ise 2.272 gözlem arasından %33,6'sının kayıp hasta olduğu ve düğümün kendi içinde 3-4 yaş aralığındaki poliklinik başvuru adedince en iyi bölündüğü görülmüş ve eğer ilgili dönemde poliklinik muayene başvurusu yoksa 1.173 gözlem arasından kayıp hasta oranının %21,2 olduğu diğer yandan 1 ve üzeri başvurusu varsa 1.099 gözlem arasından kayıp hasta oranının %46,8'e yükseldiği görülmüştür.

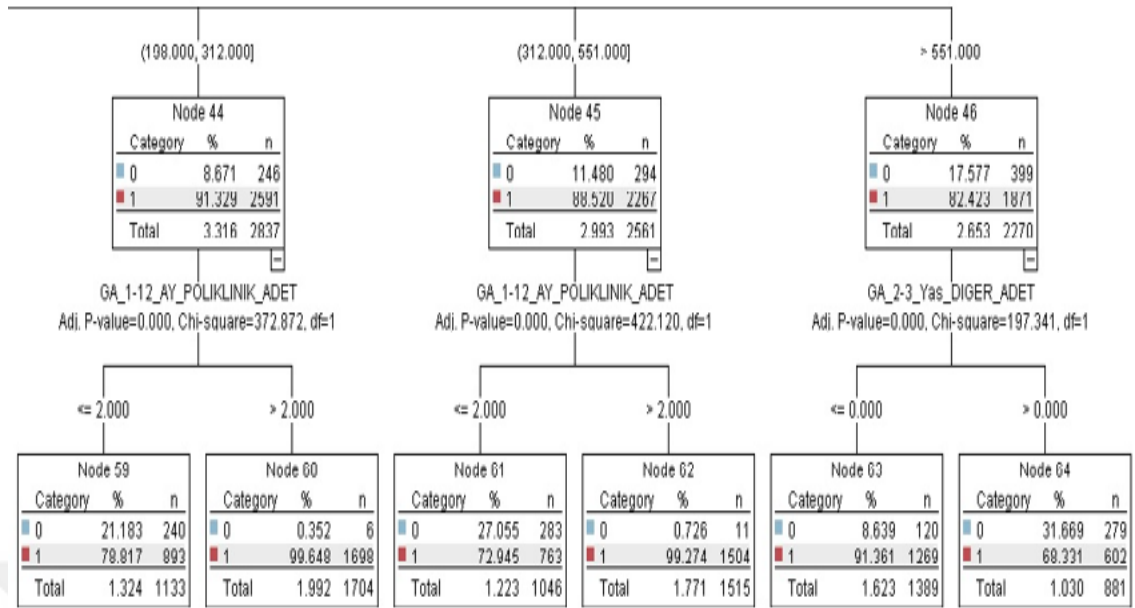
İlgili sınıftaki gelişler arasındaki gün farkının 8 ile 20 gün arasında ise 1.119 gözlem arasından %26,3'ünün, 20 ile 159 gün arasında ise 4.630 gözlem arasından %20,6'sının kayıp hasta olduğu ve düğümün kendi içinde 3-4 yaş aralığındaki poliklinik başvuru adedince en iyi bölündüğü görülmüş ve eğer ilgili dönemde poliklinik muayene başvurusu yoksa 2.806 gözlem arasından kayıp hasta oranının %9,4 olduğu diğer yandan başvuru varsa 1.824 gözlem arasından kayıp hasta oranının %37,8'e yükseldiği 73. düğümde görülmüştür. Gün farkının 159 ile 242 gün arasında ise 880 gözlem arasından %60,3'ünün, gün farkının 242

günün üzerinde olması durumunda 1.102 gözlem arasından %70,8'inin kayıp hasta olduğu 56. düğümde gözlenmiştir.



Şekil 5.45.: 2-6 Yaş CHAID 43. Düğüm ve Dalları

43. düğüm olan son_sondan2 başvuru gün sayısı farkı 125-198 gün olan ve 3.233 gözlem içeren dalı en iyi bölen kestiricinin GA_3-4_Yas_Poliklinik_Adet olduğu, ilgili göndemde poliklinik başvurusu yoksa kayıp hasta oranının %38,9, varsa %59 olduğu görülmüştür. Ayrıca, 3-4 yaş aralığında poliklinik muayene başvurusu olmayan 1.939 gözlemde kendi için 2-6 yaş aralığındaki toplam geliş sayısınınca (poliklinik, kontrol, diğer) bölünmüş ve ilgili dönemde toplam geliş adedi 2 ve altında ise 1.073 gözlem arasından %34,2'sinin, poliklinik başvuru adedi 2'den fazla ise 866 gözlem arasından %44,7'sinin kayıp hasta olduğu gözlemlenmiştir.



Şekil 5.46.: 2-6 Yaş CHAID 44-45-46. Düğüm ve Dalları

44. düğüm olan son_sondan2 başvuru gün sayısı farkı 198-312 gün olan ve 2.837 gözlem içeren dalı en iyi bölen kestiricinin 1-12 ay poliklinik başvuru adedi olduğu, ilgili göndemde poliklinik muayene başvuru adedinin 2 ve altında olması durumunda 1.133 gözlem arasından %79,9'unun, 2 ve üzeri başvuru durumunda ise 1.704 gözlem arasından %99,6'sının kayıp hasta olduğu 60. düğümde görülmüştür.

45. düğümde, olan son_sondan2 başvuru gün sayısı farkı 312-551 gün olan ve 2.561 gözlem içeren dalı en iyi bölen kestiricinin kestiricinin 1-12 ay poliklinik başvuru adedi olduğu, ilgili göndemde poliklinik muayene başvuru adedinin 2 ve altında olması durumunda 1.046 gözlem arasından %72,9'unun, 2 ve üzeri başvuru durumunda ise 1.515 gözlem arasından %99,3'ünün kayıp hasta olduğu görülmüştür.

48. düğüm olan son_sondan2 başvuru gün sayısı farkı 551 üzeri olan ve 2.270 gözlem içeren dalı en iyi bölen kestiricinin GA_2-3_Yas_Diger_adet olduğu, ilgili göndemde herhangi bir başvuru yoksa 1.389 gözlem arasından kayıp hasta oranının %91,4, başvurusu varsa 881 gözlem arasından %69,3 olduğu gözlenmiştir.

Değerlendirme metriklerinde;

2-6 yaş 122.091 çocuk hastanın, CHAID algoritması ile kayıp hasta olma durumun değerlendirmesinde, kayıp hasta olmayanlar ve kayıp hasta olanlar için hatalı ve doğru tahmin sayıları hem eğitim veri seti için hem de test veri aşaması için aşağıda sunulmuştur. Toplam veri setinin 85.557 hasta ile %70'i modelin eğitiminde, 36.534 hasta ile %30'u test için kullanılmıştır.

Tablo 5.15: 2-6 Yaş CHAID Değerlendirme Metrikleri

CHAID 2-6 Yaş Churn Değerlendirme Metrikleri					
	Doğru	Yanlış	Doğruluk	Hassasiyet	Kesinlik
Eğitim	70.262	15.295	82,12%	87,62%	83,59%
Test	29.980	6.554	82,06%	87,24%	83,83%

CHAID algoritması 2-6 yaş kayıp hasta değerlendirme metrikleri de ayrıca tablo 22'de sunulmuştur. Eğitim veri setinde, modelin %82,12 oranında doğru tahminlemede, test aşamasında da %82,06'lık doğru tahminde bulunduğu saptanmıştır. Aşağıdaki tabloda CHAID modeli için karışıklık (confusion) matrisi verilmiştir.

Tablo 5.16: 2-6 Yaş CHAID Karışıklık Matrisi

CHAID 2-6 Yaş Churn Karışıklık Matrisi				
Eğitim		Değerler	Gerçekleşen	
			0	1
	Tahminlenen	0	24.947	6.402
		1	8.893	45.315
Test		Değerler	Gerçekleşen	
			0	1
	Tahminlenen	0	10.656	2.826
		1	3.728	19.324

Model eğitim veri setinde; 24.947 çocuk hastanın kayıp hasta olmayacağı doğru şekilde tahminlenmiş, 6.402 çocuk hasta kayıp hasta olmaması hatalı olarak tahminlenmiş, kayıp hasta olduğu saptanmıştır. 45.315 çocuk hastanın kayıp hasta olacağı model tarafından doğru tahminlenmiş, 8.893 çocuk hasta ise, kayıp hasta olacağı tahminlenmişken, kayıp hasta olmadığı gözlenmiştir.

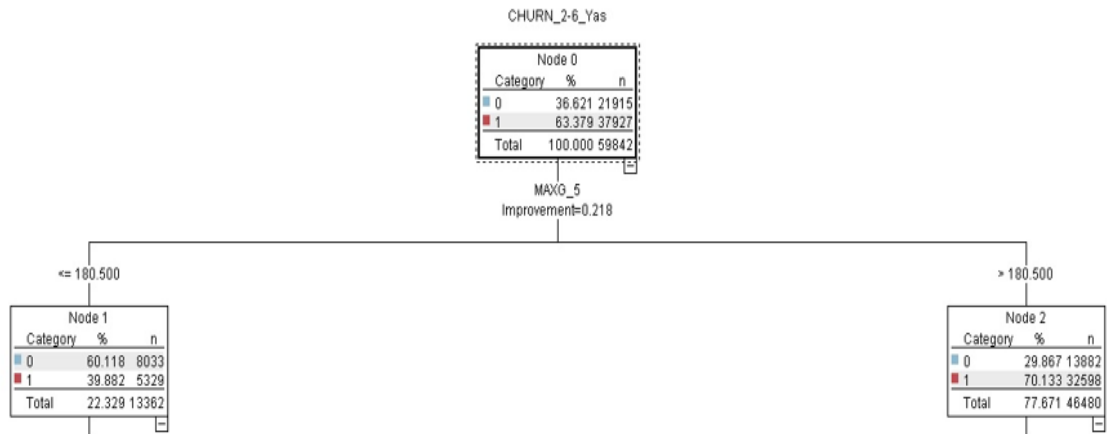
Benzer şekilde, **test veri aşamasında;** 10.656 çocuk hastanın kayıp hasta olmayacağı modelce doğru şekilde tahminlenirken, 2.826 çocuk hasta kayıp hasta hatalı olarak tahminlenirken, kayıp hasta olduğu saptanmıştır. 19.324 çocuk hastanın kayıp hasta olacağı model tarafından doğru tahminlenmiş, 3.728 çocuk hasta ise, kayıp hasta olacağı tahminlenmişken, kayıp hasta olmadığı gözlenmiştir.

Modelin değerlendirilmesinde; **hassasiyet (recall) metriği**, model dahilinde tüm pozitif tahminlenenlerin, yani kayıp hasta olacağı öngörülenlerin, ne kadarının gerçekte pozitif / kayıp hasta olduğunu incelemektedir. Hassasiyetin eğitim verisinde %87,62, test aşamasında %87,24 olduğu görülmüştür. Bir başka deyişle, 2-6 yaş çocuk grubunun CHAID algoritması ile tahminlemesinde, modelin eğitim verisi için tüm pozitif/kayıp hasta tahminlenen hastalardan aslında %87,62'sinin kayıp hasta olduğunu göstermektedir.

Kesinlik (Precision) metriği, gerçekte pozitif kayıp hasta olanların aslında kaç tanesinin doğru tahminlendiğinin değerlendirilmesinde; eğitim veri setinde %83,59, test veri aşamasında %83,83 olarak gerçekleştiği saptanmıştır. Bir başka ifade ile, eğitim veri setinde; gerçekte kayıp hasta olan 100 hastanın 83,59 tanesinin pozitif/kayıp hasta olan olarak tahminlendiği, test veri aşamasında ise gerçekte kayıp hasta olan 100 hastanın 83,83'ünün kayıp hasta/pozitif olarak tahminlendiği görülmüştür.

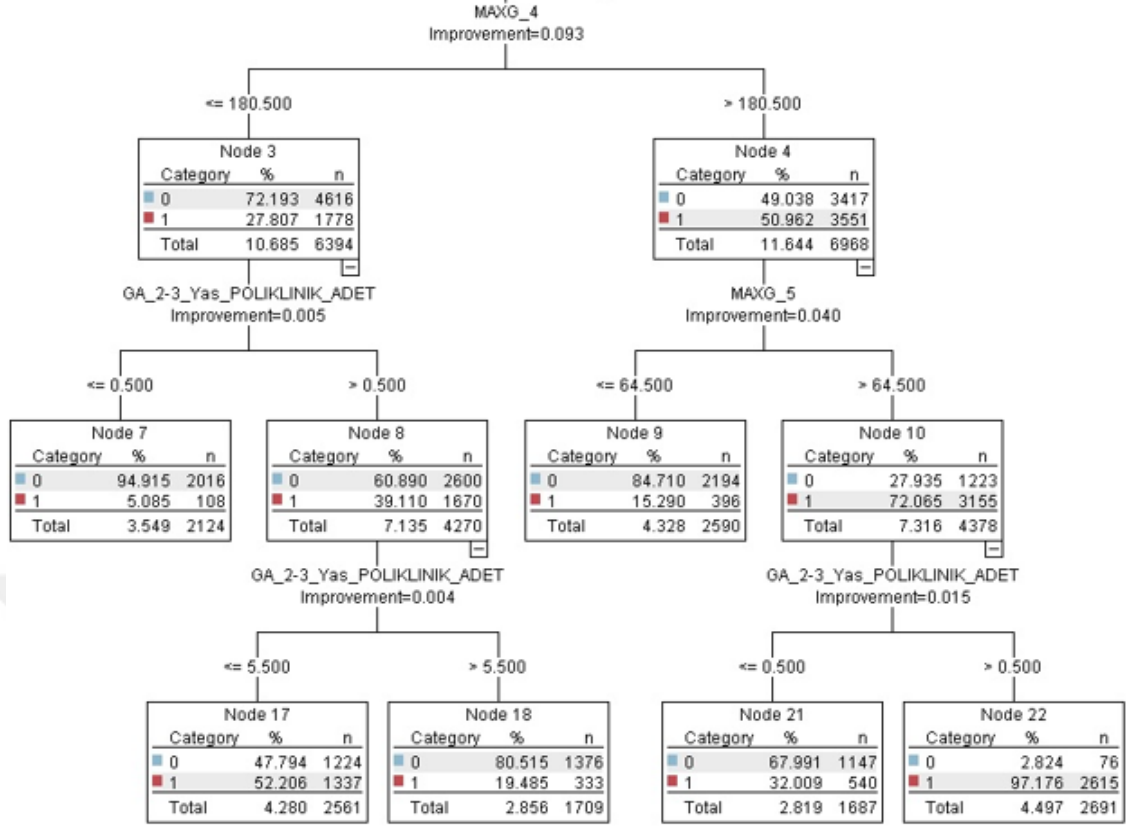
5.2.3.2. CART Algoritması ile Bulgular ve Değerlendirme Metrikleri

2 – 6 yaş aralığında kayıp hasta analizine konu olan 59.842 adet çocuk hastadan 37.927 hasta ile %63,4'ünün kayıp hasta olduğu, 21.915 hasta ile 36.6%'nin ise devamlı hasta statüsünde hizmet alımına devam ettiği gözlenmiştir. CART algoritmasında en iyi bölünmeyi veren değerin MAXG_5, yani çocuk hastalarının 5 ve 6 yaş arasındaki iki başvurusu arasındaki maksimum gün sayısının olduğu saptanmıştır.



Şekil 5.47.: 2-6 Yaş CART Genel Durumu

2-6 yaş arasındaki kayıp hasta oranı %63,4 olarak saptanmış iken, İlk düğümde; çocuk hastaların 5-6 yaş arasındaki gelişleri arasındaki maksimum gün farkı; 180 gün ve altında olduğunda; 13.362 gözlem arasında kayıp hasta oranının %39,9, 180 günden fazla olduğunda 46.480 gözlem arasından %70,1 olduğu gözlenmiştir.

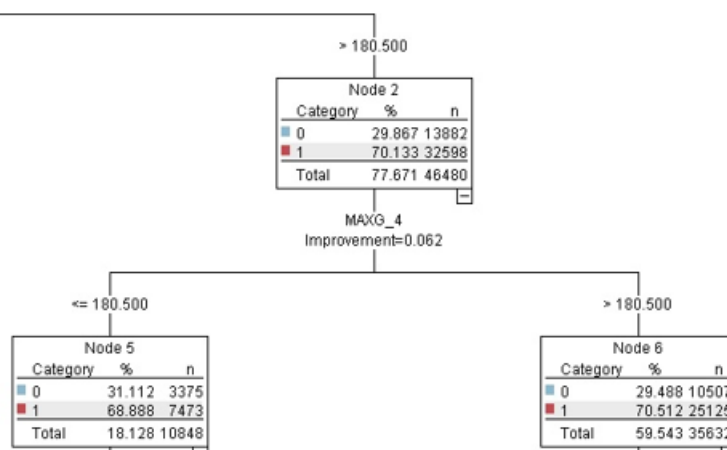


Şekil 5.48.: 2-6 Yaş CART 3 - 4. Düzüm ve Dalları

Birinci düğümün yani 5-6 yaş aralığında iki geliş arasındaki maksimum gün farkının 180 gün ve altında olduğu durumları CART algoritmasına göre en iyi şekilde bölen kestiricinin MAXG_4 olduğu ve eğer MAXG_4 180 günden az ise, 3. düğüm, 6.394 gözlem arasından %27,8'inin, 180 günden fazla ise 6.968 gözlem arasından %51'inin kayıp hasta olduğu görülmüştür. MAXG_4 180 günden az dalının en iyi bölen kestiricinin GA_2-3_Yas_Poliklinik_Adet olduğu görülmektedir. İlgili yaş aralığında çocuk hastanın sağlık sunucusuna eğer herhangi bir poliklinik muayene başvurusu yok ise 2.124 gözlem arasından %5'inin kayıp hasta olduğu diğer yandan eğer başvurusu var ise 4.270 gözlem arasından %39,1'inin kayıp hasta olduğu saptanmıştır. MAXG_5 gün sayısı 180 günden az olan, MAXG_4 gün sayısı 180 günden az olan, 2-3 yaş aralığında poliklinik başvurusu bulunan çocuk hastaların kendi içlerinde ilgili dönemde 5 ve altı poliklinik muayene başvuruları olması durumunda 2.561

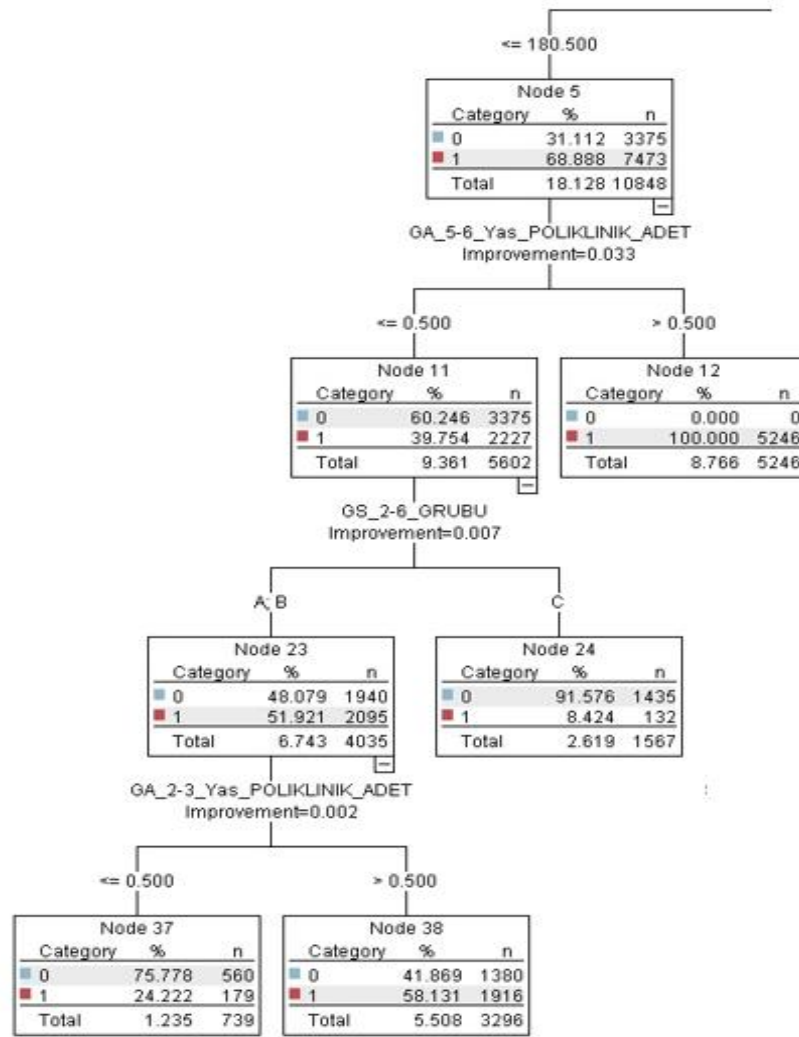
gözlem arasından %52,2'sinin, 5 üzeri başvurusu olanların 1.709 gözlem arasından %19,5'inin kayıp hasta olduğu 19. düğümde saptanmıştır.

MAXG_5'in 180 günün altında ve MAXG_4 gün sayısının 180 günün üzerinde olduğu durumlarda, düğüm 4, gelecekteki başvuruları, MAXG_5'in belirleyici bölen olduğu ve gün sayısı farkı 64 gün altındaysa 2.590 gözlem arasında %15,3 , 64 gün üzerindeyse 4.378 gözlem arasından %72 kadarının kayıp hasta olduğu gözlenmiştir. MAXG_5 gün sayısı 180 günden az olan, MAXG_4 gün sayısı 180 günden fazla olan, MAXG_5 gün sayısı 64'den fazla olanlarda 2-3 yaş aralığında poliklinik başvurusunun belirleyici bölen olduğu ve ilgili dönemde başvurusu bulunmuyorsa 1.687 gözlem arasından %32'sinin, başvurusu olanların 2.691 gözlem arasından %97,2'sinin kayıp hasta olduğu 22. düğümde görülmüştür.



Şekil 5.49.: 2-6 Yaş CART 2. Düğüm ve Dalları

İkinci düğümün yani 5-6 yaş aralığında iki geliş arasındaki maksimum gün farkının 180 günün üzerinde olduğu durumları CART algoritmasına göre en iyi şekilde bölen kestiricinin MAXG_4 olduğu ve eğer MAXG_4 180 günden az ise, 5. düğüm, 10.848 gözlem arasından %68,9'nun, 180 günden fazla ise 35.632 gözlem arasından %70,5'inin kayıp hasta olduğu görülmüştür.

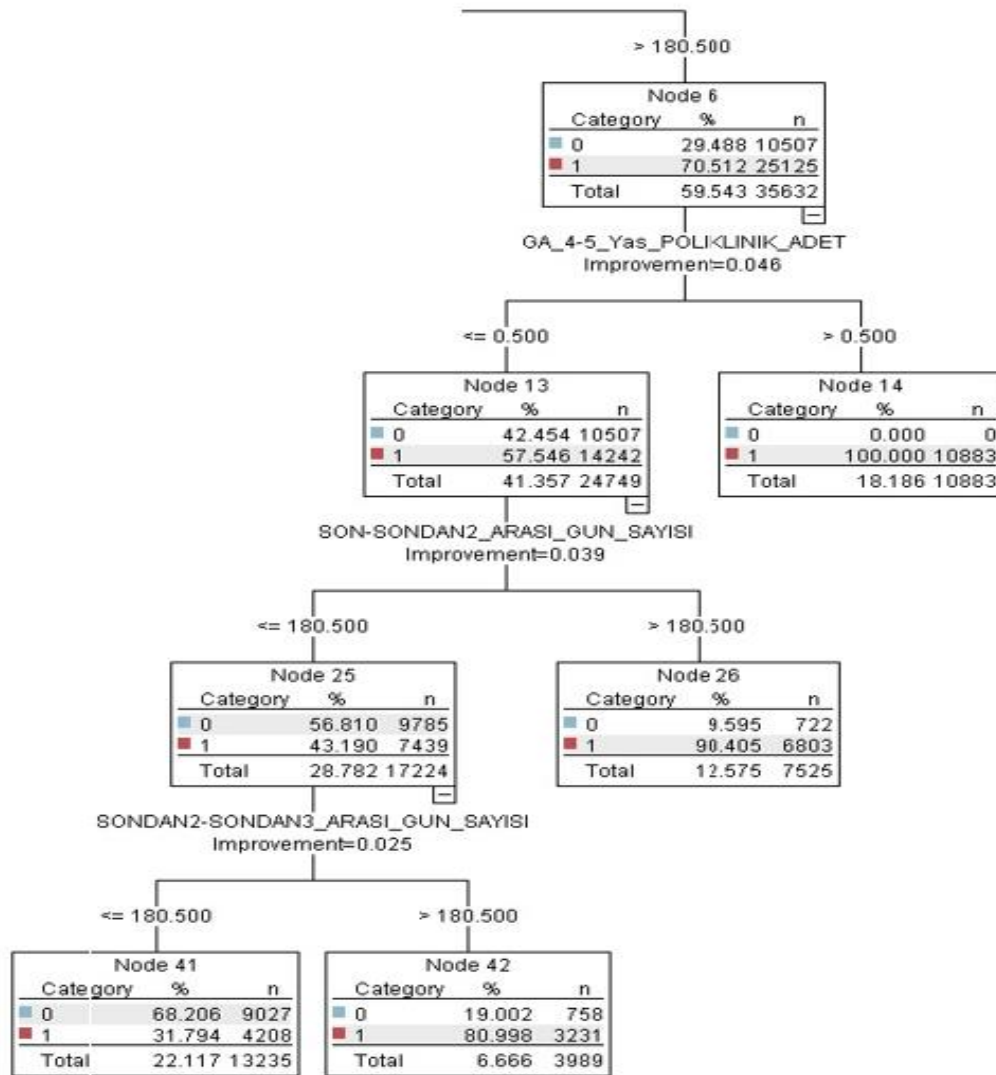


Şekil 5.50.: 2-6 Yaş CART 5. Düğüm ve Dalları

MAXG_5 180 günden fazla, MAXG_4 180 günden az olduğu durumlarda, düğüm 5, en iyi bölünmeyi veren kestiricinin GA_5-6_Yas_Poliklinik_Adet olduğu saptanmıştır. İlgili dönemde herhangi bir başvuru bulunmuyorsa 5.602 gözlem arasından kayıp hasta oranının %39,8, başvuru bulunuyorsa 5.246 gözlem ile %100 kayıp hasta olduğu saptanmıştır.

5-6 yaş aralığında başvurusu bulunmayan 5.602 çocuk hasta grubunda en iyi bölünmeyi sağlayan bağımsız değişkenin GS_2-6_Yaş_ Grubu (2-6 yaş arasındaki kabul edilen hasta grubu) olduğu görülmüş ve eğer ilgili dönemde

çocuk hasta A (10 ve üzeri başvuru), B (6-9 başvuru) grubuna ait olması ve C (3 -5 başvuru) grubuna ait olmasının en belirleyici bölen olduğu görülmüştür. A, B grubuna ait hastalar için 4.035 gözlem arasından %51,9'unun, C grubunda ise 1.567 gözlem arasından %8,4'ünün kayıp hasta olduğu gözlenmiştir. 4.035 gözlem içeren A, B gruplarına ait hastaları ise kendi içinde 2-3 yaş aralığındaki poliklinik muayene adetleri belirleyici olarak böldüğü, ilgili dönemde eğer herhangi bir poliklinik başvurusu yoksa 739 gözlem arasından %24,2'sinin, poliklinik muayene başvurusu varsa 3.296 gözlem arasından %58,1'inin kayıp hasta olduğu saptanmıştır.



Şekil 5.51.: 2-6 Yaş CART 6. Düğüm ve Dalları

MAXG_5 180 günden fazla, MAXG_4 180 den fazla olduğu durumlarda, düğüm 6, en iyi bölünmeyi veren kestiricinin GA_4-5_Yas_Poliklinik_Adet olduğu saptanmıştır. İlgili dönemde herhangi bir başvuru bulunmuyorsa 24.749 gözlem arasından kayıp hasta oranının %57,5, başvuru bulunuyorsa 10.883 gözlem ile %100 kayıp hasta olduğu saptanmıştır. 4-5 yaş aralığında başvurusu bulunmayan 24.749 çocuk hasta grubunda en iyi bölünmeyi sağlayan bağımsız değişkenin Son_Sondan2_arası_gün_farkı olduğu görülmektedir. Eğer gün ilgili başvurular arasındaki gün farkı 180 günden fazla ise 17.224 gözlem arasından kayıp hasta oranının %43,2 olduğu, gün farkının 180 günden fazla olması durumunda ise 7.525 gözlem arasından %90,4'ünün kayıp hasta olduğu belirlenmiştir. 17.224 gözlem içeren 25. düğümü ise Sondan2_Sondan3_arası_gün_farkının en iyi şekilde böldüğü görülmüş ve ilgili başvurular arasındaki gün farkı 180'den az ise 13.235 hasta arasından %31,8'inin, 180 günden fazla ise 3.989 gözlem arasından %81'inin kayıp hasta olduğu 42. düğümde görülmektedir.

Değerlendirme metrikleri;

2-6 yaş 122.091 çocuk hastanın, CART algoritması ile kayıp hasta olma durumun değerlendirmesinde, kayıp hasta olmayanlar ve kayıp hasta olanlar için hatalı ve doğru tahmin sayıları hem eğitim veri seti için hem de test veri aşaması için aşağıda sunulmuştur. Toplam veri setinin 85.557 hasta ile %70'i modelin eğitiminde, 36.534 hasta ile %30'u test için kullanılmıştır.

Tablo 5.17.: 2-6 Yaş CART Değerlendirme Metrikleri

CART 2-6 Yaş Churn Değerlendirme Metrikleri					
	Doğru	Yanlış	Doğruluk	Hassasiyet	Kesinlik
Eğitim	71.195	14.362	83,21%	88,55%	84,42%
Test	30.283	6.251	82,89%	88,04%	84,34%

Modelin eğitim veri setinde, modelin %83,21 oranında doğru tahminlemede, test aşamasında da %82,89'luk doğru tahminde bulunduğu saptanmıştır. Aşağıdaki tablo 26'da CART modeli için karışıklık (confusion) matrisi verilmiştir.

Tablo 5.18: 2-6 Yaş CART Karşıtlık Matrisi

C&T 2-6 Yaş Churn Karşıtlık Matrisi				
Eğitim		Değerler	Gerçekleşen	
			0	1
	Tahminlenen	0	25.430	5.919
		1	8.443	45.765
Test		Değerler	Gerçekleşen	
			0	1
	Tahminlenen	0	10.841	2.641
		1	3.610	19.442

Model eğitim veri setinde; 25.430 çocuk hastanın kayıp hasta olmayacağı doğru şekilde tahminlenmiş, 5.919 çocuk hasta kayıp hasta olmaması hatalı olarak tahminlenmiş, kayıp hasta olduğu saptanmıştır. 45.765 çocuk hastanın kayıp hasta olacağı model tarafından doğru tahminlenmiş, 8.443 çocuk hasta ise, kayıp hasta olacağı tahminlenmişken, kayıp hasta olmadığı gözlenmiştir.

Benzer şekilde, **test veri aşamasında;** 10.841 çocuk hastanın kayıp hasta olmayacağı modelce doğru şekilde tahminlenirken, 2.641 çocuk hasta kayıp hasta hatalı olarak tahminlenirken, kayıp hasta olduğu saptanmıştır. 19.442 çocuk hastanın kayıp hasta olacağı model tarafından doğru tahminlenmiş, 3.610 çocuk hasta ise, kayıp hasta olacağı tahminlenmişken, kayıp hasta olmadığı gözlenmiştir.

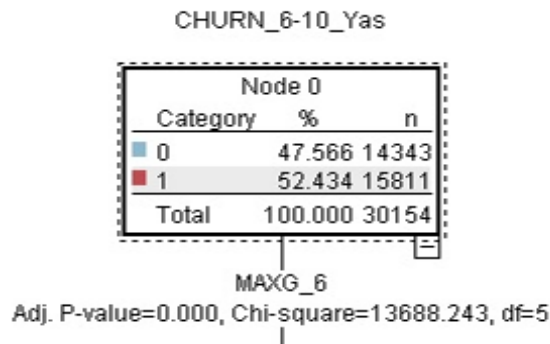
Modelin değerlendirilmesinde; **hassasiyet (recall) metriği**, model dahilinde tüm pozitif tahminlenenlerin, yani kayıp hasta olacağı öngörülenlerin, ne kadarının gerçekte pozitif / kayıp hasta olduğunu incelemektedir. Hassasiyetin eğitim verisinde %88,55, test aşamasında %88,04 olduğu görülmüştür. Bir başka deyişle, 2-6 yaş çocuk grubunun CHAID algoritması ile tahminlemesinde, modelin eğitim verisi için tüm pozitif/kayıp hasta tahminlenen hastalardan aslında %88,55'inin kayıp hasta olduğunu, test veri setinde ise ilgili oranın %88,04 olduğunu göstermektedir.

Kesinlik (Precision) metriği, gerçekte pozitif kayıp hasta olanların aslında kaç tanesinin doğru tahminlendiğinin değerlendirilmesinde; eğitim veri setinde %84,42, test veri aşamasında %84,34 olarak gerçekleştiği saptanmıştır. Bir başka ifade ile, eğitim veri setinde; gerçekte kayıp hasta olan 100 hastanın 84,42 tanesinin pozitif/kayıp hasta olan olarak tahminlendiği, test veri aşamasında ise gerçekte kayıp hasta olan 100 hastanın 84,34'ünün kayıp hasta/pozitif olarak tahminlendiği görülmüştür.

5.2.4. 6-10 Yaş Grubu Geliş Sıklıkları ve MKA

6-10 yaş aralığında ya da daha önceki tanımlı grup dönemlerinde 3 poliklinik muayene şartını sağlayan 6-10 yaş aralığındaki çocuk hastaların kayıp hasta kabul edilme durumu karar ağaçları algoritmalarından CHAID ve CART ile analiz edilmiştir.

5.2.4.1. CHAID Algoritması ile Bulgular ve Değerlendirme Metrikleri



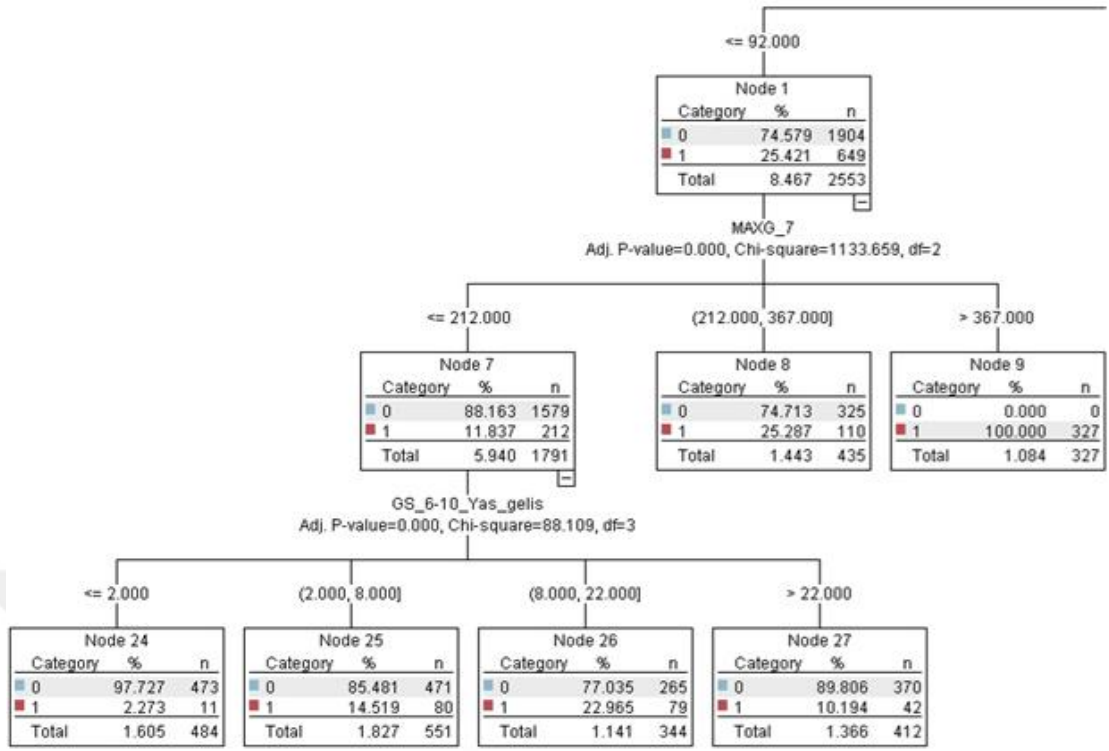
Şekil 5.52.: 6-10 Yaş CHAID Genel Durum

6–10 yaş aralığında kayıp hasta analizine konu olan 30.154 adet çocuk hastadan 15.811 hasta ile %52,4'ünün kayıp hasta olduğu, 14.343 hasta ile %47,6'sının ise devamlı hasta statüsünde hizmet alımına devam ettiği gözlenmiştir. CHAID algoritmasının ki-kare/df kombinasyonunda en iyi bölünmeyi veren değer MAXG_6, yani çocuk hastalarının 6 ve 7 yaş arasındaki iki başvurusu arasındaki maksimum gün sayısının olduğu saptanmıştır. Burada bölünen 6 dala ilişkin bilgi özetlenerek tablo 27'de sunulmuştur.

Tablo 5.19: 6-10 Yaş CHAID MAXG_6 Sınıfı Başvurular Arası Gün Farkı ve Kayıp Hasta Durumu

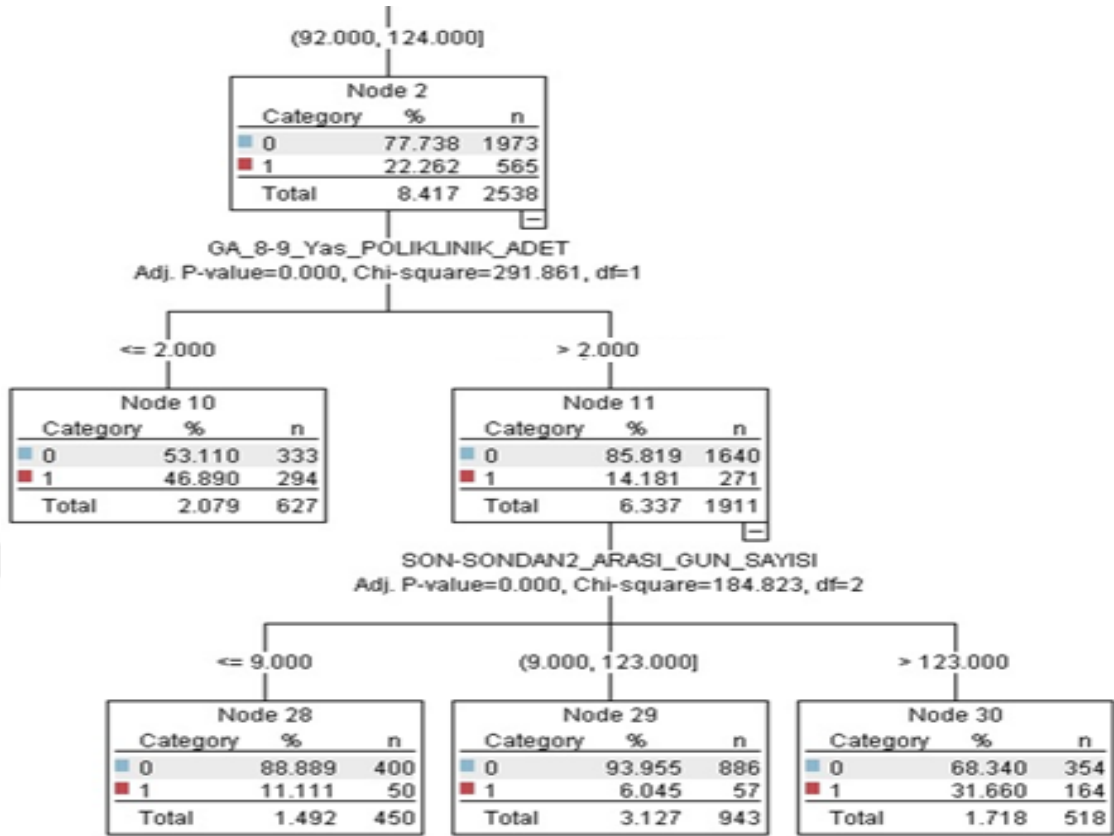
MAXG_6	Devamlı Hasta Adedi	Churn Adedi	Toplam Gözlem Adedi	Devamlı Hasta Oranı	Churn Oranı
92 günden az	1.904	649	2.553	74,6%	25,4%
92-124 gün arası	1.973	565	2.538	77,7%	22,3%
124-217 gün arası	5.586	2.001	7.587	73,6%	26,4%
217-314 gün arası	3.536	1.550	5.086	69,5%	30,5%
314-390 gün arası	1.344	1.195	2.539	52,9%	47,1%
390 üstü ve kayıp değerler	0	9.851	9.851	0,0%	100,0%
Toplam	14.343	15.811	30.154	47,6%	52,4%

6-10 yaş arasındaki kayıp hasta oranı %52,4 olarak saptanmış iken, İlk düğümde; çocuk hastaların 6-7 yaş arasındaki gelişleri arasındaki maksimum gün farkı; 92 gün ve altında olduğunda; kayıp hasta oranının %25,4, 92-124 gün arasında %22,3, 124-217 gün arasında %26,4, 217-314 gün arasında %30,5, 314-390 gün arasında %47,1 ve 390 gün üstü ve 6-10 yaş arasında kayıp hasta bilgisi olan fakat MAXG_6 sınıfında 1'den fazla başvurusu olmayanların kayıp hasta oranının 100% olduğu saptanmıştır.



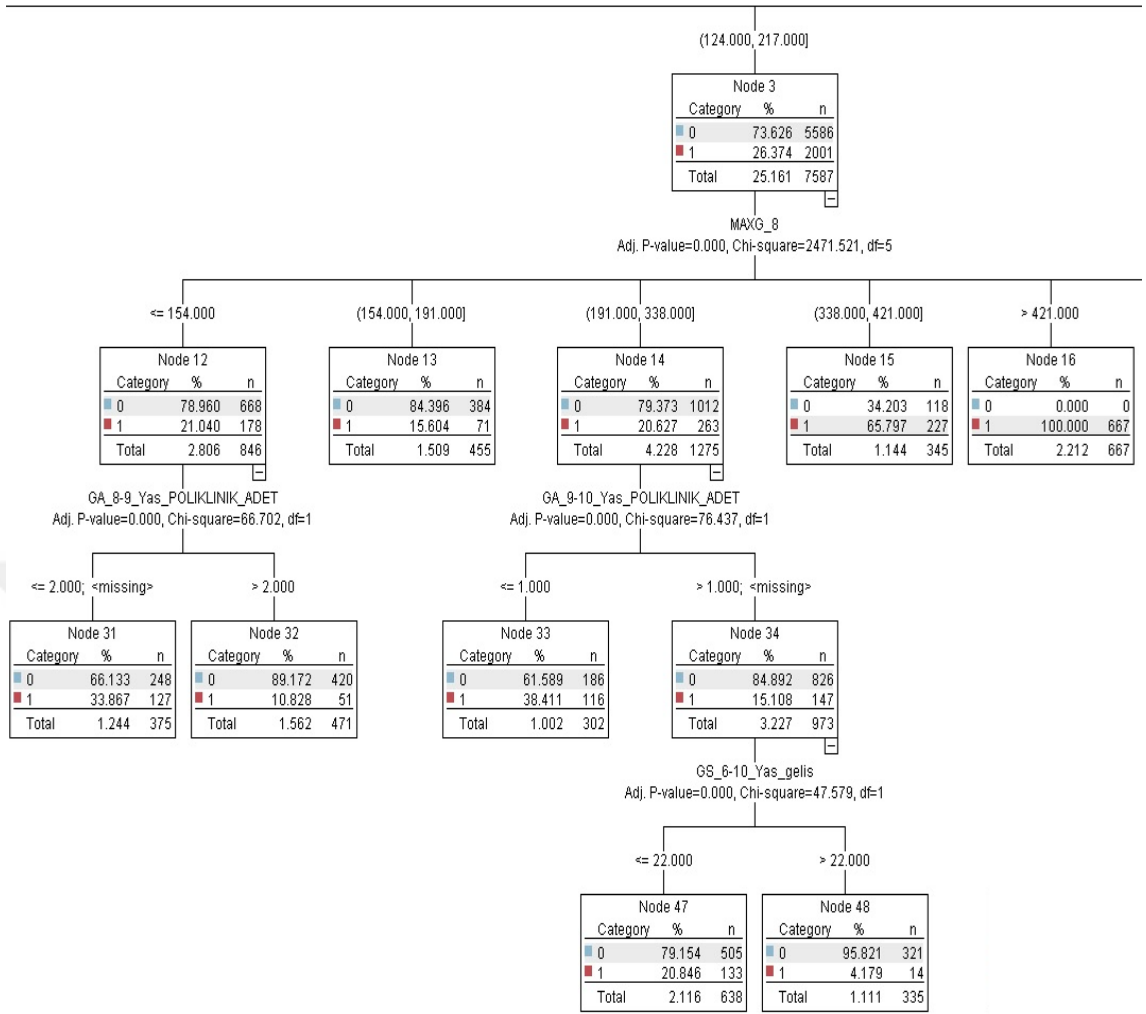
Şekil 5.53.: 6-10 Yaş CHAID 1. D ğ m ve Dalları

MAXG_6'nın 92 g n ve altında olduėu ilk d ğ m - dalı en iyi b len kestiricinin ise MAXG_7 yani 7-8 yaşı arasındaki 2 başvuru arasındaki maksimum g n farkı olduėu saptanmıştır. İlgili grubun kendi i inde maksimum g n farkının 212'den az, 212-367 g n arası ve 367 g n ve  zeri olarak 3 b l nd ė  g r lm şt r. G n farkının 212'den az olduėu 1.791 g zlem arasından %11,8'inin, 212-367 g n arasındaki 435 g zlemden %25,3' n n kayıp hasta olduėu g r lmektedir. 367 g n ve  zeri grubunda 327 g zlem olduėu (7-8 yaşı arasında gelişı olmayan ve daha  nceki /sonraki sene i inde de gelişı kısa s reli bulunmayan) ve bunların %100' n kayıp hasta olduėu saptanmıştır. 1.791 g zlem olduėu, MAXG_7 212 g ne kadar ve kayıp deėerler grubunun kendi i inde en iyi b len kestiricinin GS_6_10_Yaşı_Gelişı yani 6-10 yaşı arasındaki toplam başvuru sayısı olduėu g zlenmiştir. İlgili grup da kendi i inde; 2'den az başvuru, 484 g zlem, %2,3 kayıp hasta oranı, 2-8 başvuru sayısı, 551 g zlem, %14,5 kayıp hasta oranı, 8-22 başvuru adedi, 344 g zlem sayısı, %23 kayıp hasta oranı ve 22  zeri başvuru, 412 g zlem adedi, %10,1 kayıp hasta oranı olarak b l nm şt r.



Şekil 5.54.: 6-10 Yaş CHAID 2. Düğüm ve Dalları

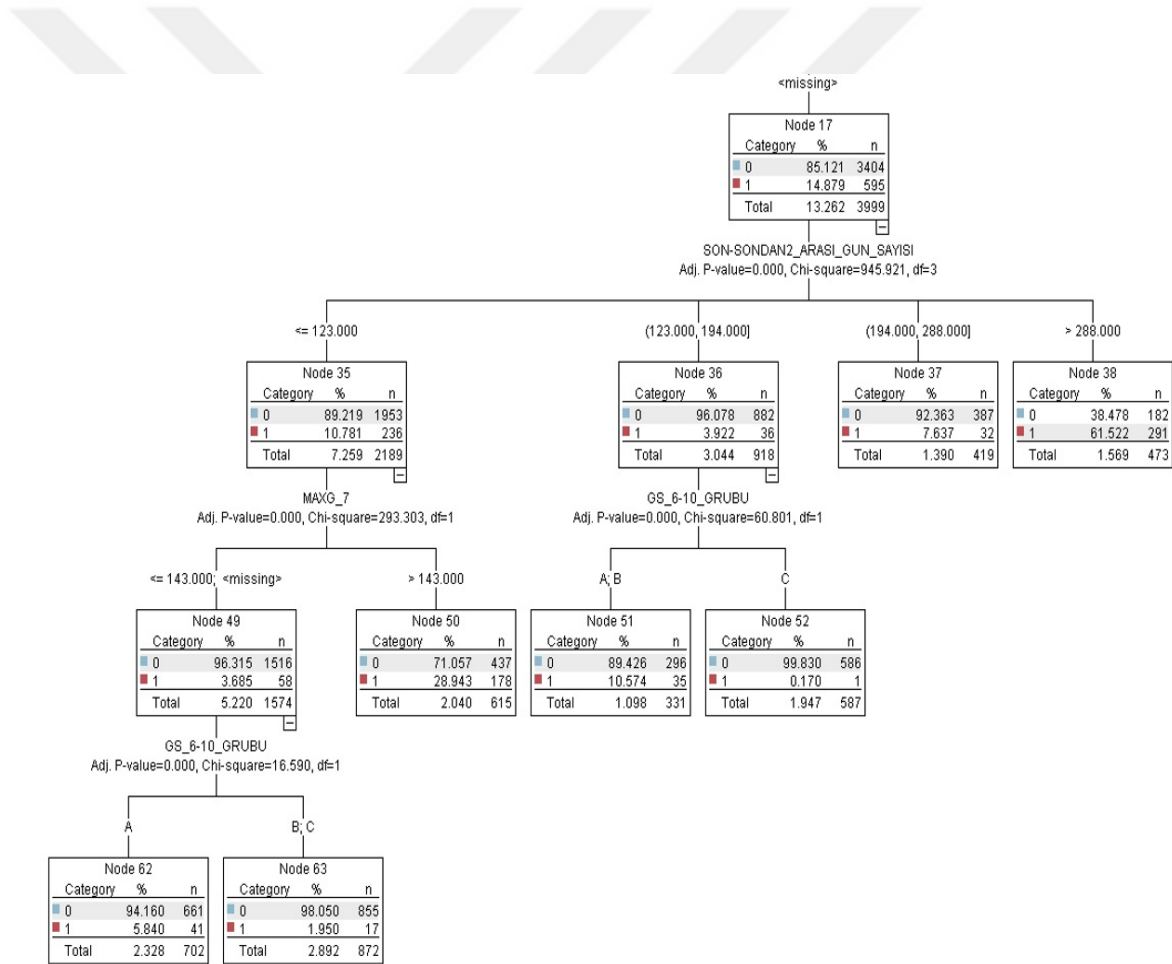
İkinci düğümü olan MAXG_6 92-124 günü en iyi bölen kestiricinin GA_8-9_Poliklinik adet yani 8-9 yaşları arasındaki toplam poliklinik muayene başvuru adedi olduğu görülmüştür. İlgili grup da kendi içinde; 2 başvuru ve altı, 2 başvuru üzeri olarak dallanmıştır. 2 ve altı başvuru durumunda, 627 gözlem arasından %46,9'unun, 2 üzeri başvuru ve kayıp değerler durumu için 1.911 gözlem arasından %14,2'sinin kayıp hasta olduğu saptanmıştır. 1.911 gözlem içeren 2 üzeri başvuru ve kayıp değerler durumunu en iyi bölen kestiricinin son-sondan2_arası_gün_sayısı yani son ve sondan bir önceki başvurular arasındaki gün sayısı olduğu görülmüştür. Eğer ilgili gün farkı 9 ve altında ise 450 gözlem üzerinden %11,1, 9-123 gün arasında ise 943 gözlem arasında %6, 123 gün üzeri ise, 518 gözlem arasından %31,7'sinin kayıp hasta olduğu anlaşılmıştır.



Şekil 5.55.: 6-10 Yaş CHAID 3. Düğüm ve Dalları

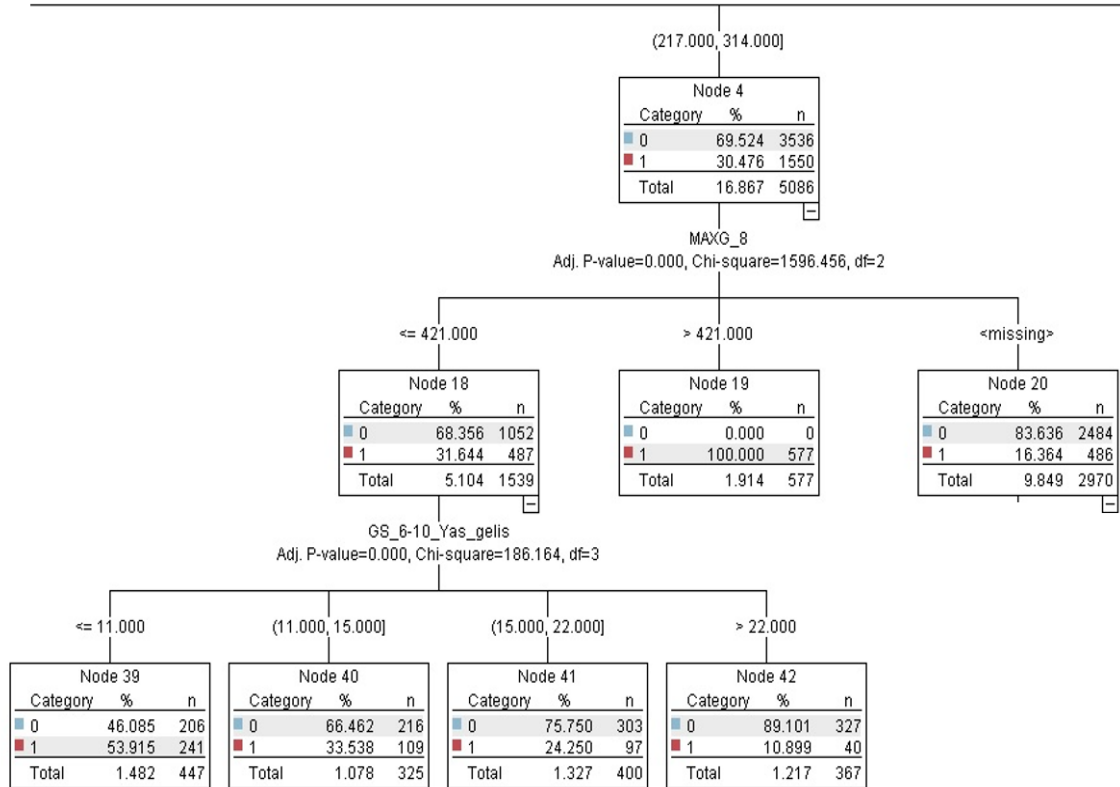
Üçüncü düğümü olan MAXG_6 124-217 günü en iyi bölen kestiricinin MAXG_8 yani 8-9 yaşları arasındaki gelişleri arasındaki maksimum gün farkı olduğu görülmüştür. İlgili grup da kendi içinde; 8-9 yaş aralığındaki ve bir önceki başvuru arasındaki maksimum gün farkı 154 günden az ise, 846 gözlem üzerinden %21'inin kayıp hasta olduğu saptanmıştır. Maksimum gün farkı 154-191 gün ise 455 gözlem arasından %15,6'sının, 191-338 gün ise 1.275 gözlem arasından %20,6'sının, 338-421 gün ise 345 gözlem arasından %65,8'inin, 421 gün üzeri ise 667 gözlem arasından %100'nün kayıp hasta olduğu saptanmıştır. 846 gözlem içeren 12. düğüm, 8-9 yaş aralığındaki poliklinik başvuru adedince

en iyi bölünmüş ve ilgili yaş aralığında poliklinik başvuru adedi 2 ve altında ise 375 gözlem arasından %33,9'unun, 2 üzeri ise 471 gözlem arasından %10,8'inin kayıp hasta olduğu saptanmıştır. MAXG_8, 191-338 gün dalını da benzer şekilde en iyi bölenin GA_9-10_poliklinik_adet olduğu görülmüş, eğer ilgili dönemde herhangi bir poliklinik muayene başvurusu yoksa 302 gözlem arasından %34,4, varsa 973 gözlem arasından %15,1 olduğu görülmüştür. 973 gözlem içeren 34.düğüm de kendi içerisinde 6-10 yaşları arasındaki toplam geliş sayısınınca en iyi bölünmüş ve eğer ilgili dönemde toplam geliş sayısı 22 ve altında ise 638 gözlem arasından %20,9'unun, 22 üzeri ise 335 gözlem arasından %4'ünün kayıp hasta olduğu gözlenmiştir.



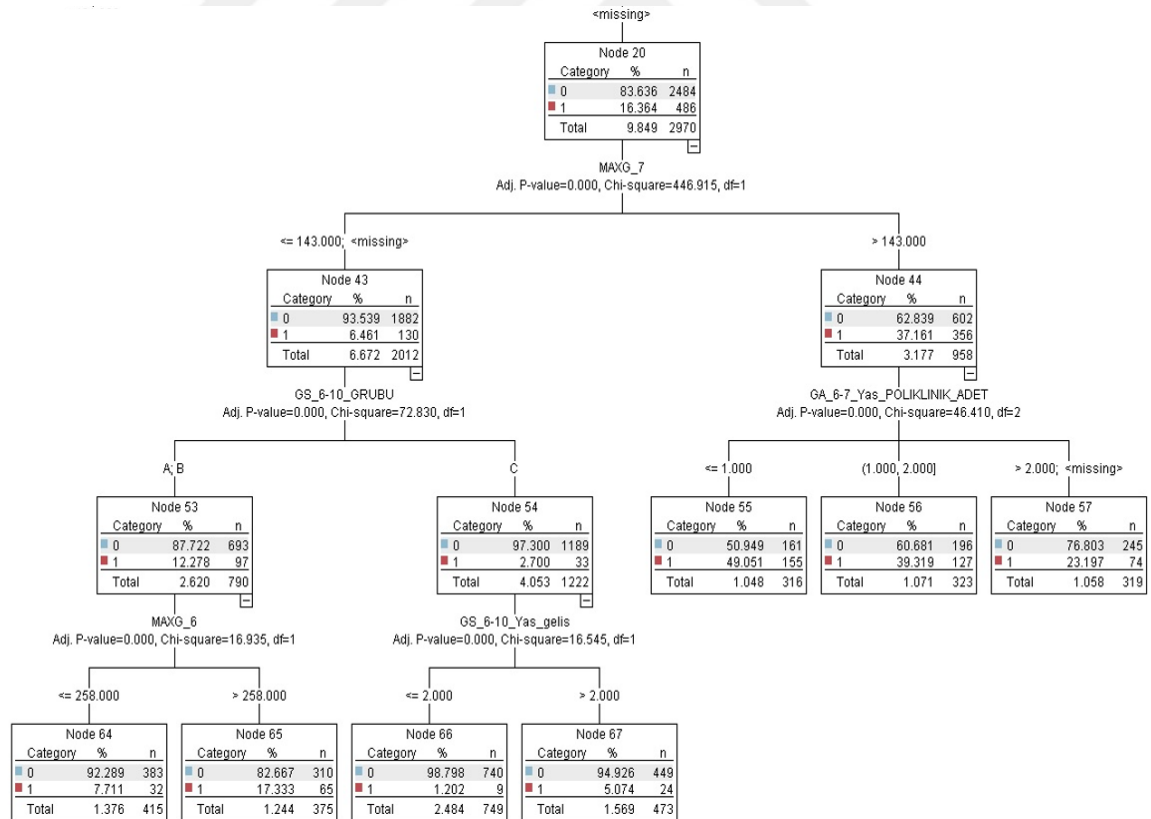
Şekil 5.56.: 6-10 Yaş CHAID 17. Düğüm ve Dalları

8-9 yaş arasında bilgi içermeyen kayıp değerler ise, 3.999 gözlem arasından %14,9'unun kayıp hasta olduğu ve kendi içerisinde Son_sondan2_arası_gün_sayısı, son başvurusu ve ondan bir önceki başvurusu arasındaki gün farkınca en iyi bölünmüş olduğu görülmüştür. İlgili gün sayısı 123'den az ise 2.189 gözlem arasından %10,8'inin, 123-194 arasında ise 918 gözlem arasından %3,9'unun, 194-288 arasında ise 419 gözlem arasından %7,6'sının ve 288 gün üzeri ise 473 gözlem arasından %61,5'inin kayıp hasta olduğu görülmüştür. 35. düğüm olan 123 gün altı gün farkı MAXG_7, 7-8 yaşları arasındaki başvurular maksimum gün farkınca en iyi bölünmüş, eğer ilgili gün farkı 143 ve altında ise 1.574 gözlem arasından %37'sinin, 143'den fazla ise 615 gözlem arasından %28,9'unun kayıp hasta olduğunu görülmüştür. Ayrıca, 143 gün altı grubu da GS_6-10_geliş_grubunca bölünmüş A grubu hastaların, 702 gözlem üzerinden %5,8'nin, B ve C grubu hastaların 872 gözlem üzerinden %2'sinin kayıp hasta olduğu görülmüştür. (A: 8+ başvuru, B: 5-7 başvuru, C: 4 ve 4'den az (3-4 Başvuru) başvurular, eğer daha önceki yaş grubunda bizim hastamız kabul edildi ise 1-4 başvuru).



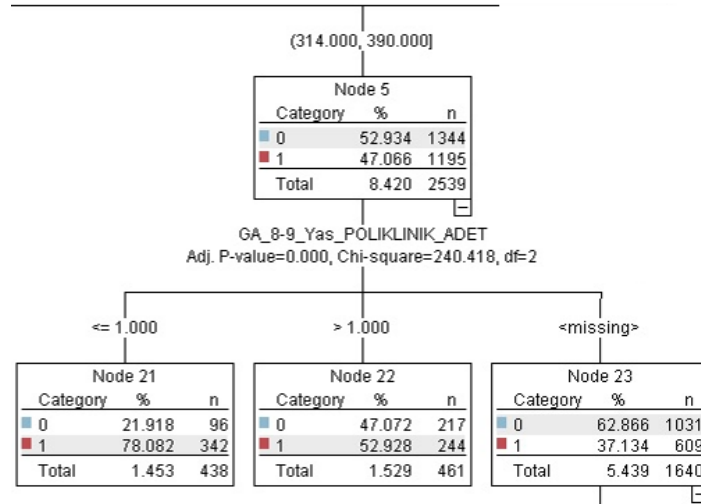
Şekil 5.57.: 6-10 Yaş CHAID 4. Düğüm ve Dalları

Dördüncü düğüm olan MAXG_6 217-314 günü en iyi bölen kestiricinin MAXG_8 yani 8-9 yaşları arasındaki gelişleri arasındaki maksimum gün farkı olduğu görülmüştür. İlgili grup da kendi içinde; 8-9 yaş aralığındaki ve bir önceki başvuru arasındaki maksimum gün farkı 421 günden az ise, 1.539 gözlem üzerinden %31,6'sının, 421 günden fazla ise 577 gözlem arasından %100'ünün kayıp hasta olduğu görülmüştür. Yine 2.970 gözlem içeren MAXG_8 bilgisi içermeyen kayıp değerlerin %16'4'ünün kayıp hasta olduğu görülmüştür. 18. düğüm olan MAXg_8, 421 gün altı dalı kendi içinde GS_6-10_yaş_geliş adedince en iyi bölünmüş ve eğer ilgili dönemdeki toplam başvuru adedi 11 ve altında ise 447 gözlem arasından %53,9'unun kayıp hasta olduğu görülmüştür. Yine 11-15 arasında ise 325 gözlem arasından %33,5'inin, 15-22 ise 400 gözlem arasından %24,3'ünün, 22 üzerinde ise 367 gözlem arasından %10'unun kayıp hasta olduğu saptanmıştır.



Şekil 5.58.: 6-10 Yaş CHAID 20. Düzüm ve Dalları

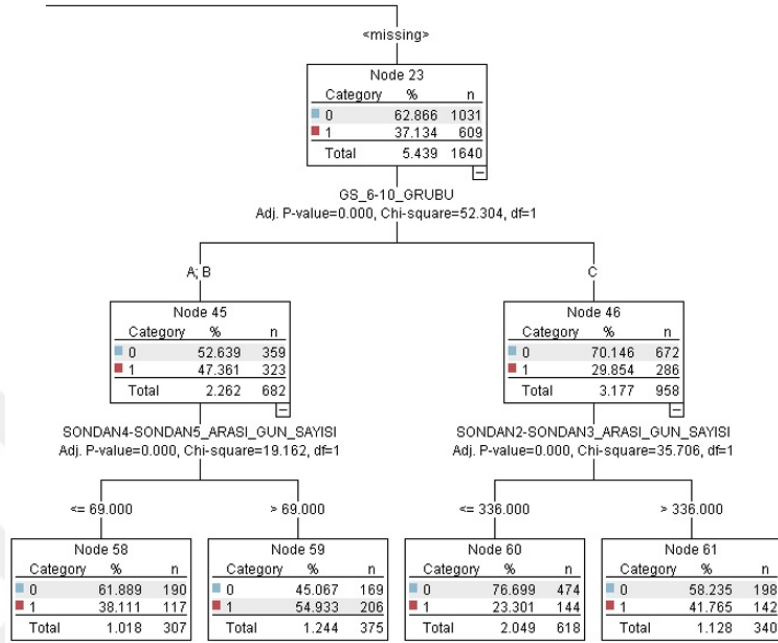
8-9 yaş arasında bilgi içermeyen kayıp değerler ise, 2.970 gözlem arasından %16,4'ünün kayıp hasta olduğu ve kendi içerisinde MAXG_7 gün farkınca en iyi bölünmüş, ilgili gün sayısı 143'den az veya MAXG_7 ait bir bilgi bulundurmeyen kayıp değerler ise 2.012 gözlem arasından %6,5'inin, ilgili gün farkı 143 günden fazla ise 958 gözlem arasından %37,2'sinin kayıp hasta olduğu görülmüştür. 43. düğüm de kendi içerisinde GS_6-10_geliş_grubunca bölünmüş, eğer çocuk hastalar A ve B grubuna ait ise, 790 gözlem arasından %12,3'ünün, C grubuna ait ise 1.222 gözlem arasından %2,7'sinin kayıp hasta olduğu görülmüştür. A, B grubu kendi içinde MAXG_6, gün farkı adedince bölünmüş, eğer ilgili gün farkı 258 ve altında ise 415 gözlem arasından %7,7'sinin, 258 gün üzeri ise 375 gözlem arasından %17,3'ünün kayıp hasta olduğu gözlenmiştir. C grubu çocuk hastalar ise GS_6-10_yaş_geliş adedince bölünmüş ve eğer başvuru adedi 2 ve altında ise 749 gözlem %1,2'sinin, 2 üzeri ise 473 gözlem arasından %5'inin kayıp hasta olduğu görülmüştür.



Şekil 5.59.: 6-10 Yaş CHAID 5. Düğüm ve Dalları

Beşinci düğümü olan MAXG_6314-390 günü en iyi bölen kestiricinin GA_8-9_yaş_poliklinik_adet olduğu, eğer ilgili aralıkta herhangi bir poliklinik başvurusu yok veya 1 tane ise 438 gözlemin %78'inin, 1 ve üzeri ise 461 gözlemin %53'ünün

ve kayıp değerleri içeren 1.640 gözlemin %37,1'inin kayıp hasta olduğu görülmüştür.



Şekil 5.60.: 6-10 Yaş CHAID 23. Düğüm ve Dalları

İlgili daldaki 1.640 gözlem içeren, %37,1 kayıp hasta oranına sahip kayıp değerler ise, kendi içerisinde 6GS_6-10_yaş-grubu, 6-10 yaşları arasındaki toplam gelişleri üzerinden tanımlanan yaş gruplarıncı en iyi bölünmüş ve eğer çocuk hastalar A ve B gruplarına aitse 682 gözlem üzerinden %47,4'ünün, C grubundan ise %29,9'unun kayıp hasta olduğu görülmüştür. A ve B grubu çocuk hastalarda kendi içlerinde Sondan4-Sondan5_arası_gün_sayısıncı bölünmüş ve eğer ilgili gün sayısı farkı 69 ve altındaysa 307 gözlem üzerinden %38,1'inin, 69 gün üzerinde ise 375 gözlem arasından %54,9'unun kayıp hasta olduğu görülmüştür. C grubu çocuk hastalar ise sondan2-sondan3_arası_gün_sayısıncı bölünmüş, ilgili gün sayısı farkı 336 ve altında ise 618 gözlem arasından %23,3'ünün, 336 gün üzerindeyse 340 gözlem arasından %41,8'ünün kayıp hasta olduğu saptanmıştır.

> 390.000; <missing>		
Node 6		
Category	%	n
0	0.000	0
1	100.000	9851
Total	32.669	9851

Şekil 5.61.: 6-10 Yaş CHAID 6. Düğüm

6.düğüm olan MAXG_8 390 gün üzeri ve kayıp değerler (MAXG_8 bilgisi olmayan – 8-9 yaş aralığında başvurusu olmayan çocuklar) 9.851 adet gözlem içermekte ve kayıp hasta oranının %100 olduğu görülmektedir. Bir başka ifadeyle, 6-10 yaş aralığında, 8-9 yaş aralığında herhangi bir gelişi olan ve bir önceki gelişi arasında 390 günden fazla gelişi olmayan çocukların kayıp hasta oldukları görülmüştür.

Değerlendirme metrikleri;

Çalışmada 6-10 yaş arasındaki 43.216 devamlı hasta kabul edilen çocuk hastanın kayıp hasta durumları CHAID ile değerlendirilmiştir. Toplam veri setinin 30.154 hasta ile %70'i modelin eğitiminde, 13.062 hasta ile %30'u testinde kullanılmıştır.

Tablo 5.20: 6-10 Yaş CHAID Değerlendirme Metrikleri

CHAID 6-10 Yaş Churn Değerlendirme Metrikleri					
	Doğru	Yanlış	Doğruluk	Hassasiyet	Kesinlik
Eğitim	26.328	3.826	87,31%	92,92%	82,05%
Test	11.325	1.737	86,70%	92,43%	81,44%

Modelin eğitim veri setinde, modelin %87,31 oranında doğru tahminlemede, test aşamasında da %86,7'lik doğru tahminde bulunduğu saptanmıştır. 6-10 yaş çocuk hastalarının CHAID algoritması ile kayıp hasta olma durumun değerlendirmesinde, kayıp hasta olmayanlar ve kayıp hasta olanlar için hatalı ve doğru tahmin sayıları hem eğitim veri seti için hem de test veri aşaması için aşağıdaki tablo 29'da karışıklık (confusion) matrisi olarak verilmiştir.

Tablo 5.21: 6-10 Yaş CHAID Karşıtlık Matrisi

CHAID 6-10 Yaş Churn Karşıtlık Matrisi				
Eğitim		Değerler	Gerçekleşen	
			0	1
	Tahminlenen	0	13.355	988
		1	2.838	12.973
Test		Değerler	Gerçekleşen	
			0	1
	Tahminlenen	0	5.717	459
		1	1.278	5.608

Model eğitim veri setinde; 13.355 çocuk hastanın kayıp hasta olmayacağı doğru şekilde tahminlenmiş, 988 çocuk hasta kayıp hasta olmaması hatalı olarak tahminlenmiş, kayıp hasta olduğu saptanmıştır. 12.973 çocuk hastanın kayıp hasta olacağı model tarafından doğru tahminlenmiş, 2.838 çocuk hasta ise, kayıp hasta olacağı tahminlenmişken, kayıp hasta olmadığı gözlenmiştir.

Benzer şekilde, **test veri aşamasında;** 5.717 çocuk hastanın kayıp hasta olmayacağı modelce doğru şekilde tahminlenirken, 459 çocuk hasta kayıp hasta hatalı olarak tahminlenirken, kayıp hasta olduğu saptanmıştır. 5.608 çocuk hastanın kayıp hasta olacağı model tarafından doğru tahminlenmiş, 1.278 çocuk hasta ise, kayıp hasta olacağı tahminlenmişken, kayıp hasta olmadığı gözlenmiştir.

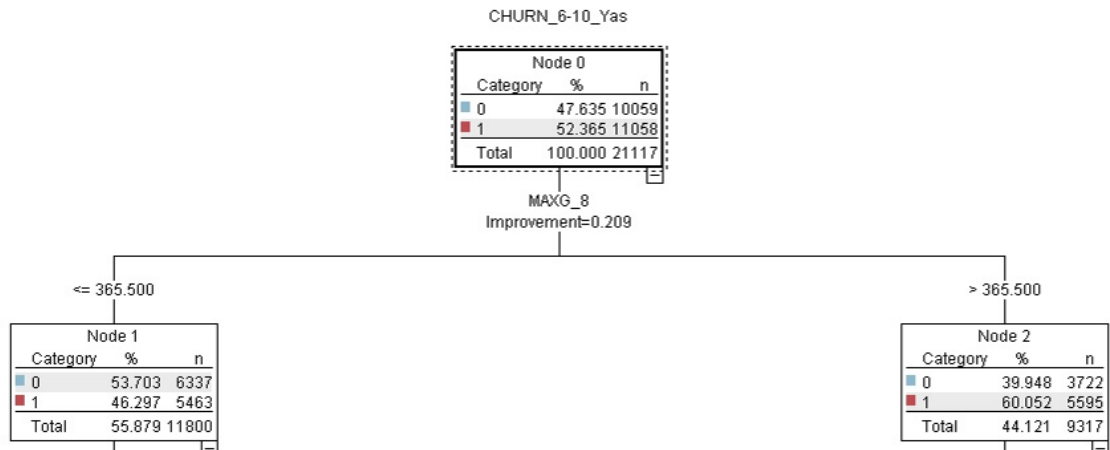
Modelin değerlendirilmesinde; **hassasiyet (recall) metriği**, model dahilinde tüm pozitif tahminlenenlerin, yani kayıp hasta olacağı öngörülenlerin, ne kadarının gerçekte pozitif / kayıp hasta olduğunu incelemektedir. Hassasiyetin eğitim verisinde %92,92, test aşamasında %92,43 olduğu görülmüştür. Bir başka deyişle, 6-10 yaş çocuk grubunun CHAID algoritması ile tahminlemesinde, modelin eğitim verisi için tüm pozitif/kayıp hasta tahminlenen hastalardan

aslında %92,92'sinin kayıp hasta olduğunu, test veri setinde ise ilgili oranın %92,43 olduğunu göstermektedir.

Kesinlik (Precision) metriği, gerçekte pozitif kayıp hasta olanların aslında kaç tanesinin doğru tahminlendiğinin değerlendirilmesinde; eğitim veri setinde %82,05, test veri aşamasında %81,44 olarak gerçekleştiği saptanmıştır. Bir başka ifade ile, eğitim veri setinde; gerçekte kayıp hasta olan 100 hastanın 82,05 tanesinin pozitif/kayıp hasta olan olarak tahminlendiği, test veri aşamasında ise gerçekte kayıp hasta olan 100 hastanın 81,44'ünün kayıp hasta/pozitif olarak tahminlendiği görülmüştür.

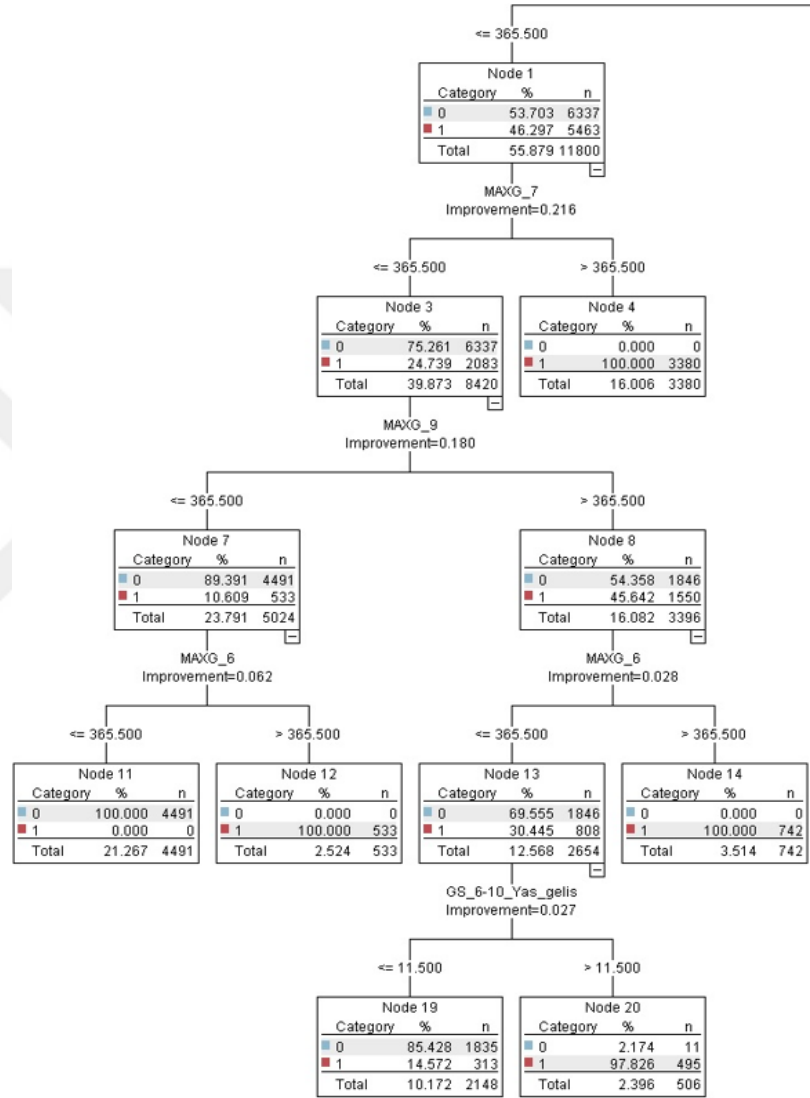
5.2.4.2. CART Algoritması ile Bulgular ve Değerlendirme Metrikleri

6–10 yaş aralığında kayıp hasta analizine konu olan 21.117 adet çocuk hastadan 11.058 hasta ile %52,4'ünün kayıp hasta olduğu, 10.059 hasta ile 47.6%'nin ise devamlı hasta statüsünde hizmet alımına devam ettiği gözlenmiştir. CART algoritmasında en iyi bölünmeyi veren değer $MAXG_8$, yani çocuk hastalarının 8 ve 9 yaş arasındaki iki başvurusu arasındaki maksimum gün sayısının olduğu saptanmıştır.



Şekil 5.62.: 6-10 Yaş CART Genel Durumu

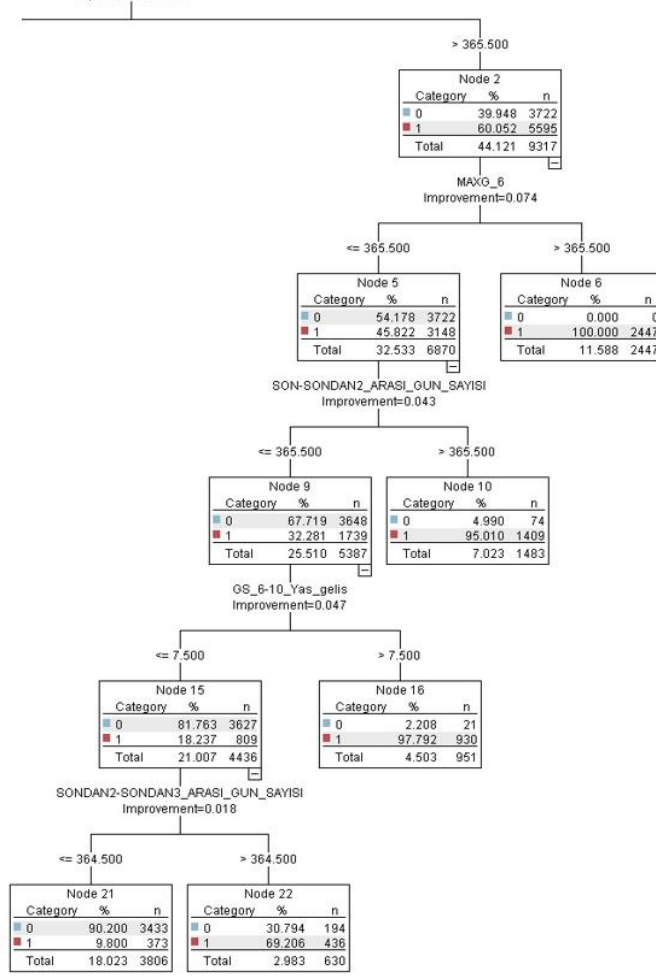
6-10 yaş arasındaki kayıp hasta oranı %52,4 olarak saptanmış iken, İlk düğümde; çocuk hastaların 8-9 yaş arasındaki gelişleri arasındaki maksimum gün farkı; 365 gün ve altında olduğunda; 11.800 gözlem arasında kayıp hasta oranının %46,3, 365 günden fazla olduğunda 9.317 gözlem arasından %60 olduğu gözlenmiştir.



Şekil 5.63.: 6-10 Yaş CART 1. Düğüm ve Dalları

Birinci düğümün yani 8-9 yaş aralığında iki geliş arasındaki maksimum gün farkının 365 gün ve altında olduğu durumları CART algoritmasına göre en iyi şekilde bölen kestiricinin MAXG_7 olduğu ve eğer MAXG_7 365 günden az ise, 3. düğüm, 8.420 gözlem arasından %24,7'sinin, 365 günden fazla ise 3.380

gözlem arasından %100'ünün kayıp hasta olduğu görülmüştür. MAXG_7 365 günden az dalının en iyi bölen kestiricinin MAXG_9 olduğu, 9-10 yaş aralığında eğer bir önceki başvuru ve 9-10 yaş aralığındaki başvuru arasındaki gün farkı 365 gün altında ise, 5.024 gözlem arasından %10,6'sının, 365 günün üzerinde ise, 3.396 gözlem arasından %45,6'sının kayıp hasta olduğu görülmüştür. 5.024 gözlem barındıran 7.düğümü en iyi bölen kestiricinin de MAXG_6 olduğu, 6-7 yaş aralığında eğer bir önceki başvuru ve 6-7 yaş aralığındaki başvuru arasındaki gün farkı 365 gün altında ise, 4.491 gözlem arasından %0'ının, 365 günün üzerinde ise, 533 gözlem arasından %100'ünün kayıp hasta olduğu görülmüştür. 3.396 gözlem barındıran 8.düğümü en iyi bölen kestiricinin de MAXG_6 olduğu, 6-7 yaş aralığında eğer bir önceki başvuru ve 6-7 yaş aralığındaki başvuru arasındaki gün farkı 365 gün altında ise, 2.654 gözlem arasından %30,4'ünün, 365 günün üzerinde ise, 742 gözlem arasından %100'ünün kayıp hasta olduğu görülmüştür. 2.654 gözlem bulunduran 13.düğüm de kendi içinde en iyi bölenin GS_6-10_yaş_gelis, 6-10 yaş arasındaki toplam geliş sayısı olduğu, ilgili geliş sayısı 11 ve altında ise; 2.148 gözlem arasından %14,6'sının, 11 üzerinde ise, 506 gözlem arasından %97,8'inin kayıp hasta olduğu saptanmıştır.



Şekil 5.64.: 6-10 Yaş CART 2. Düğüm ve Dalları

İkinci düğümün yani 8-9 yaş aralığında iki geliş arasındaki maksimum gün farkının 365 gün üzeri olduğu durumları CART algoritmasına göre en iyi şekilde bölen kestiricinin MAXG_6 olduğu ve eğer MAXG_6 365 günden az ise, 5. düğüm, 6.870 gözlem arasından %45,8'inin, 365 günden fazla ise 2.247 gözlem arasından %100'ünün kayıp hasta olduğu görülmüştür. MAXG_6 365 günden az dalının en iyi bölen kestiricinin Son-Sonda2_arası_gün_sayısı olduğu, eğer ilgili aralıktaki gün sayısı farkı 365'den az ise, 35.387 gözlem arasından %32,3'ünün, eğer 365 günden fazla ise 1.483 gözlem arasından %95'inin kayıp hasta olduğu gözlenmiştir. 5.387 gözlem barındıran 9.düğümü en iyi bölen kestiricinin de GS_6-10_geliş_sayısı olduğu, ilgili aralıktaki toplam geliş sayısının 7'den az olduğu durumda, 4.436 gözlem üzerinden kayıp hasta oranının %18,2, 8 ve üzeri durumunda ise 951 gözlem arasından %97,8'inin kayıp hasta olduğu görülmüştür.

4.436 gözlem barındıran 15. düğümü en iyi bölen kestiricinin de Sondan2-Sondan3_arası_gün_sayısı olduğu ve ilgili gün sayısı farkının 365 günden az olması durumunda 3.806 gözlem üzerinden %9,8'inin, 365 gün üzeri olması halinde, 630 gözlem üzerinden %69,2'sinin kayıp hasta olduğu görülmüştür.

Değerlendirme metriklerinde;

Çalışmada 6-10 yaş arasındaki 43.206 devamlı hasta kabul edilen çocuk hastanın kayıp hasta durumları CART ile değerlendirilmiştir. Toplam veri setinin 30.154 hasta ile %70'i modelin eğitiminde, 13.052 hasta ile %30'u test için kullanılmıştır.

Tablo 5.22: 6-10 Yaş CART Değerlendirme Metrikleri

CART 6-10 Yaş Churn Değerlendirme Metrikleri					
	Doğru	Yanlış	Doğruluk	Hassasiyet	Kesinlik
Eğitim	28.764	1.390	95,39%	97,34%	93,77%
Test	12.428	624	95,22%	97,27%	93,57%

Modelin eğitim veri setinde, modelin %95,39 oranında doğru tahminlemede, test aşamasında da %95,22'lik doğru tahminde bulunduğu saptanmıştır. 6-10 yaş çocuk hastalarının CART algoritması ile kayıp hasta olma durumun değerlendirmesinde, kayıp hasta olmayanlar ve kayıp hasta olanlar için hatalı ve doğru tahmin sayıları hem eğitim veri seti için hem de test veri aşaması için aşağıdaki tablo 31'de karışıklık (confusion) matrisi olarak verilmiştir.

Tablo 5.23: 6-10 Yaş CART Karışıklık Matrisi

CART 6-10 Yaş Churn Karışıklık Matrisi				
Eğitim		Değerler	Gerçekleşen	
			0	1
	Tahminlenen	0	13.938	405
		1	985	14.826
Test		Değerler	Gerçekleşen	
			0	1
	Tahminlenen	0	5.995	181
		1	443	6.443

Model eğitim veri setinde; 13.938 çocuk hastanın kayıp hasta olmayacağı doğru şekilde tahminlenmiş, 405 çocuk hasta kayıp hasta olmaması hatalı olarak tahminlenmiş, kayıp hasta olduğu saptanmıştır. 14.826 çocuk hastanın kayıp hasta olacağı model tarafından doğru tahminlenmiş, 985 çocuk hasta ise, kayıp hasta olacağı tahminlenmişken, kayıp hasta olmadığı gözlenmiştir.

Benzer şekilde, **test veri aşamasında;** 5.995 çocuk hastanın kayıp hasta olmayacağı modelce doğru şekilde tahminlenirken, 181 çocuk hasta kayıp hasta hatalı olarak tahminlenirken, kayıp hasta olduğu saptanmıştır. 6.643 çocuk hastanın kayıp hasta olacağı model tarafından doğru tahminlenmiş, 443 çocuk hasta ise, kayıp hasta olacağı tahminlenmişken, kayıp hasta olmadığı gözlenmiştir.

Modelin değerlendirilmesinde; **hassasiyet (recall) metriği**, model dahilinde tüm pozitif tahminlenenlerin, yani kayıp hasta olacağı öngörülenlerin, ne kadarının gerçekte pozitif / kayıp hasta olduğunu incelemektedir. Hassasiyetin eğitim verisinde %97,34, test aşamasında %97,27 olduğu görülmüştür. Bir başka deyişle, 2-6 yaş çocuk grubunun CHAID algoritması ile tahminlemesinde, modelin eğitim verisi için tüm pozitif/kayıp hasta tahminlenen hastalardan aslında %97,34'sinin kayıp hasta olduğunu, test veri setinde ise ilgili oranın %97,27 olduğunu göstermektedir.

Kesinlik (Precision) metriği, gerçekte pozitif kayıp hasta olanların aslında kaç tanesinin doğru tahminlendiğinin değerlendirilmesinde; eğitim veri setinde %93,77, test veri aşamasında %93,57 olarak gerçekleştiği saptanmıştır. Bir başka ifade ile, eğitim veri setinde; gerçekte kayıp hasta olan 100 hastanın 93,77 tanesinin pozitif/kayıp hasta olan olarak tahminlendiği, test veri aşamasında ise gerçekte kayıp hasta olan 100 hastanın 93,37'sinin kayıp hasta/pozitif olarak tahminlendiği görülmüştür.

5.3. Şikâyetler ve MKA

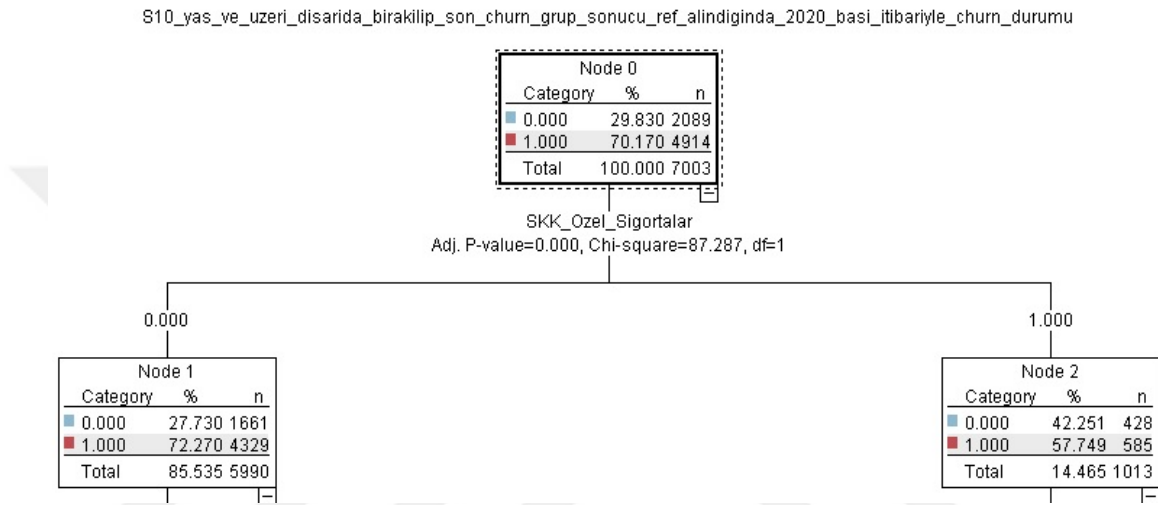
İlgili sağlık kurumunun veri tabanından, MİY yazılımında tanımlı Ş.Ö.T (Şikâyet, Öneri, Teşekkür) standart rapor formatları ile 01.01.2013 ile 31.12.2018 tarihleri arasındaki sisteme kayıtlı 17.672 farklı çocuk hasta ve toplam 23.063 adet şikâyet edinilmiştir. MKA çalışmasına konu olmayan, devamlı hasta kriterini sağlamayan çocuk hastalar kapsam dışına çıkarılmış ve geriye kalan 10.046 hastaya ait 14.233 kayıtlı şikâyet çalışmaya dahil edilmiştir. Modelde incelenen değişkenler aşağıda verilmiştir:

- Hasta numarası
- Birden fazla şikâyeti olma durumu
- İlgili hasta numarasına kayıtlı toplam şikâyet adedi
- Devamlı hasta kriterini sağlama durumu
- Şikâyetin ait olduğu şube
- Şikâyet departmanı
- Şikâyet ana grup (tıbbi, operasyonel, çalışanlar)
- Şikâyet alt grup (İlgi eksikliği, ücret, tıbbi vb.)
- Şikâyet yönetim süreci sonucu (şikâyet yöneticisini görüşme sonrası etiketlediği hastanın memnuniyet durumu; memnun, memnun değil, ulaşılamadı)
- Kayıp hasta /kayıp hasta durum etiketi

Veri temizleme, işleme ve dönüştürme aşamasında; çalışmanın ilk yıllarındaki kayıtlarda görünen, hasta numarası tanımlanmamış çocuk hastalar temizlenmiş, katma değer üretmesi ön görülmeyen şikâyeti kaydeden, şikâyeti yönetip kapatan kişi, şikâyet bildiriminde bulan kişi (hasta mı hasta yakını mı, çocuk hastalarda hasta yakınının bildirimi beklendiği için) veri setinden kaldırılmıştır. İlgili hastalar için, sistemde birden fazla şikâyeti olup olmadığı, kaç şikâyeti olduğu, şikâyetlerinin arasındaki gün farkı, nihai hasta kayıp durum etiketi gibi değişkenler veri işlemesi aşamasında Excel aracılığıyla eklenmiş ve veri seti

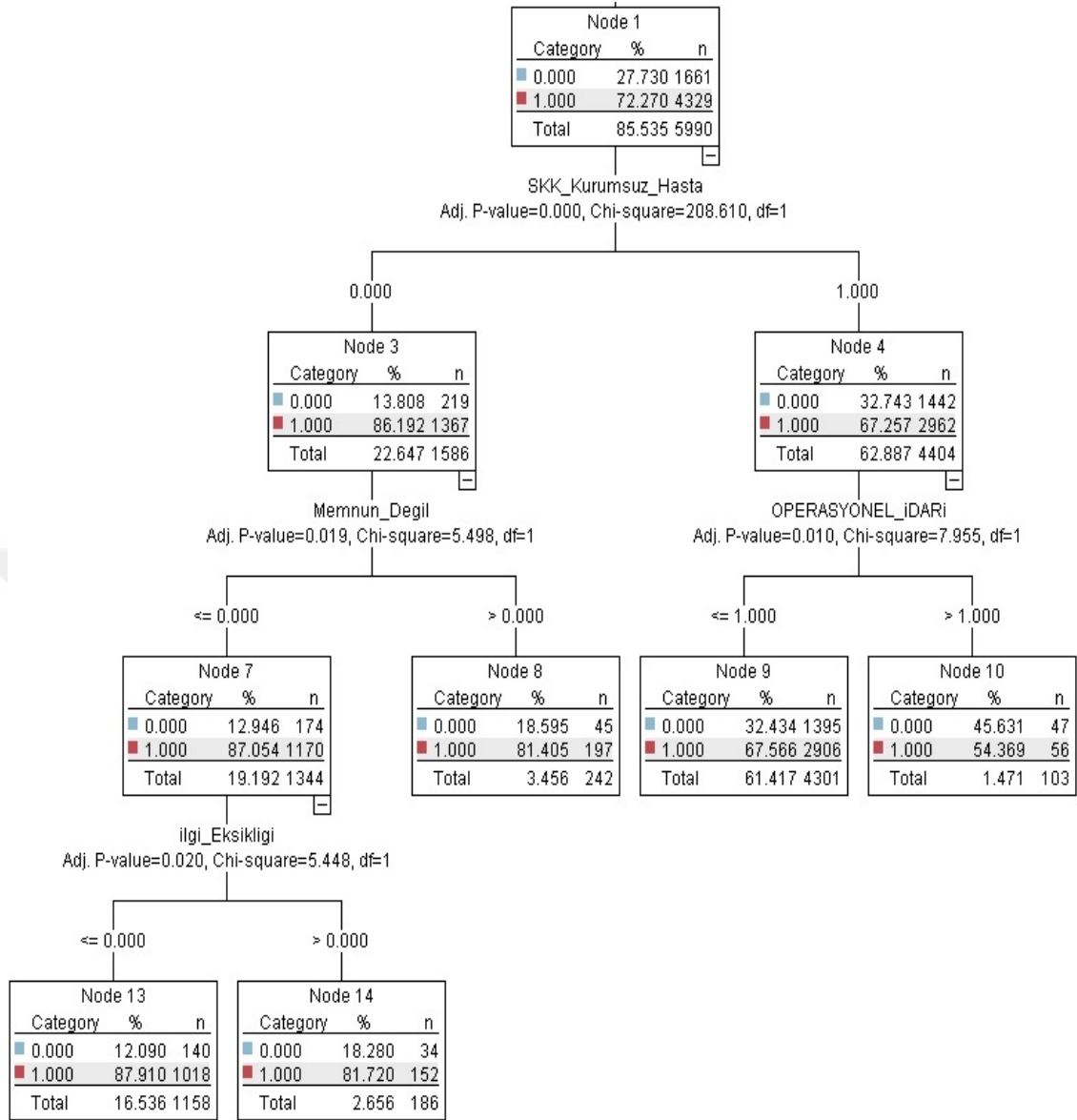
format olarak IBM SPSS Modeler 17.0'a yüklenebilecek hale getirilmiştir. Çalışılan ilgili çocuk hastalar için sisteme kayıt edilmiş şikâyetler, şikâyetlerin adedi, sıklığı ve türü ile kayıp hasta kabul edilme durumu arasındaki örüntüler karar ağacı algoritmalarında CHAID ve CART ile incelenmiş ve tahminlenmiştir.

5.3.1. CHAID Algoritması ile Bulgular ve Değerlendirme Metrikleri



Şekil 5.65.: Şikâyetler ve Kayıp hasta CHAID Uygulama Genel Durumu

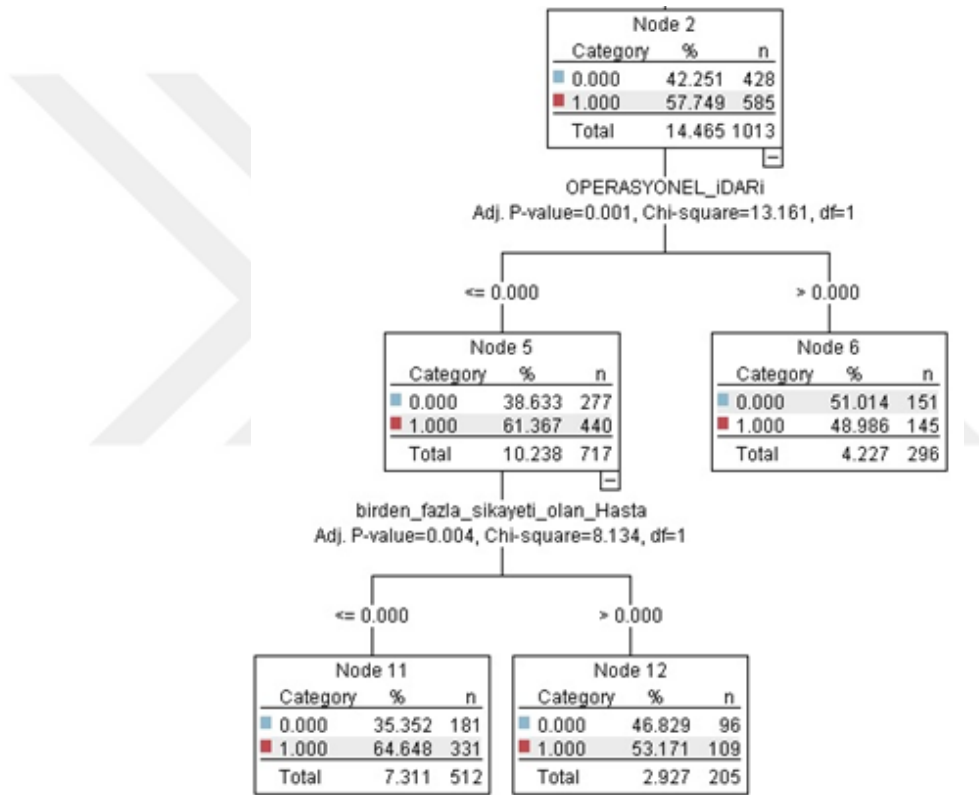
Yapılan CHAID analizine göre öncelikle sistemde şikâyeti bulunan, çalışılan 7.003 hastanın (devamlı hasta kriterini sağlayan hastaların %3,3, ilgili dönem toplu çocuk hastaların %2'si) %70,17'sinin kayıp hasta olarak etiketlendiği görülmüştür. Yapılan Pearson K-Kare testi sonrasında en iyi bölen kestiricinin SKK_Özel_Sigortalar, son kabul kurumunun özel sigortalar durumun olduğu ve hastaların son kabul kurumunun özel sigortalar olduğu durumlarda; 1.013 gözlem arasından %57,7'sinin, özel sigortalar dışında bir kurum olduğunda %72,3'ünün kayıp hasta olduğu gözlenmiştir.



Şekil 5.66.: Şikâyetler CHAID 1. Düğüm ve Dalları

5.990 gözlem içeren birimci düğümü en iyi bölen kestiricinin SKK_Kurumsuz_Hasta, nakit ödemeli hasta olma durumu olduğu ve hastanın kurumsuz hasta sınıfına ait olması durumunda 4.404 gözlem arasından %67,3'ünün, diğer kabul kurumlarına ait olması durumunda ise %86,2'lik kayıp hasta oranına sahip olduğu görülmüştür. 1.586 gözlem içeren 3. düğüm ise Memnun_değil (şikâyet sonrası alınan aksiyonlar doğrultusunda, şikâyeti yöneten kişice atanan hastanın durumu etiketlemesi) etiketince en iyi bölünmüş eğer memnun değil işaretli değilse 1.344 gözlem arasından %87'sinin kayıp hasta olduğu görülmüştür. İlgili düğünde hastaların ilgi ekliliği alt şikâyet

grup başlığında şikâyeti olup olmama durumuna göre bölünmüş, bu grupta şikâyeti olanların 186 gözlem arasından %81,7'sinin kayıp hasta olduğu görülmüştür. Son kabul kurumu özel sigortalar olmayan ve kurumsuz hasta olan sınıfındaki 4. Düğüm 4.404 gözlem içermekte ve hastaların Operasyonel_idari ana grubunda şikâyet i olması durumuna göre iki alt daha homojen gruba ayrılmaktadır. İlgili başlıkta şikâyet i varsa 103 gözlem arasından %5,4'ünün yoksa 4.301 gözlem arasından %67,6'sının kayıp hasta olduğu görülmektedir.



Şekil 5.67.: Şikâyetler CHAID 2. Düğüm ve Dalları

1.013 gözlem içeren düğümün Operasyonel_idari başlıklarda şikâyeti olma durumuna göre en iyi bölündüğü, ilgili grupta şikâyeti varsa 296 gözlem arasından %49'unun, yoksa 717 gözlem arasından %61,4'ünün kayıp hasta olduğu gözlenmiştir. 717 gözlem içeren 5. düğüm de kendi içinde hastanın birden fazla şikâyeti olup olmama durumuna göre bölünmüş ve birden fazla şikâyeti

varsa 205 gözlem üzerinden %53,2'sinin, yoksa 512 gözlem üzerinde %64,6'sının kayıp hasta olduğu görülmüştür.

Değerlendirme Metrikleri

Çalışmada MİY sisteminde şikâyet kaydı bulunan ve devamlı hasta kriterlerini sağlayan 10.046 çocuk hastanın şikâyet ayrıntıları ve kayıp hasta kabul edilme durumu CHAID ile değerlendirilmiştir. Toplam veri setinin 7.003 hasta ile %70'i modelin eğitiminde, 3.043 hasta ile %30'u test için kullanılmıştır.

Tablo 5.24: Şikayetler CHAID Değerlendirme Metrikleri

CHAID Şikayetler - Churn Değerlendirme Metrikleri					
	Doğru	Yanlış	Doğruluk	Hassasiyet	Kesinlik
Eğitim	4.920	2.083	70,26%	71,10%	97,05%
Test	2.125	918	69,83%	71,17%	96,33%

Modelin eğitim veri setinde, modelin %70,26 oranında doğru tahminlemede, test aşamasında da %69,83'lük doğru tahminde bulunduğu saptanmıştır. Çocuk hastalarının CHAID algoritmasıyla şikâyet ve kayıp hasta kabul edilme durumun değerlendirmesinde, kayıp hasta olmayanlar ve kayıp hasta olanlar için hatalı ve doğru tahmin sayıları hem eğitim veri seti için hem de test veri aşaması için aşağıda karışıklık (confusion) matrisi olarak verilmiştir.

Tablo 5.25: Şikâyetler CHAID Karışıklık Matrisi

CHAID Şikayetler - Churn Karışıklık Matrisi				
Eğitim		Değerler	Gerçekleşen	
			0	1
	Tahminlenen	0	151	1.938
		1	145	4.769
Test		Değerler	Gerçekleşen	
			0	1
	Tahminlenen	0	54	839
		1	79	2.071

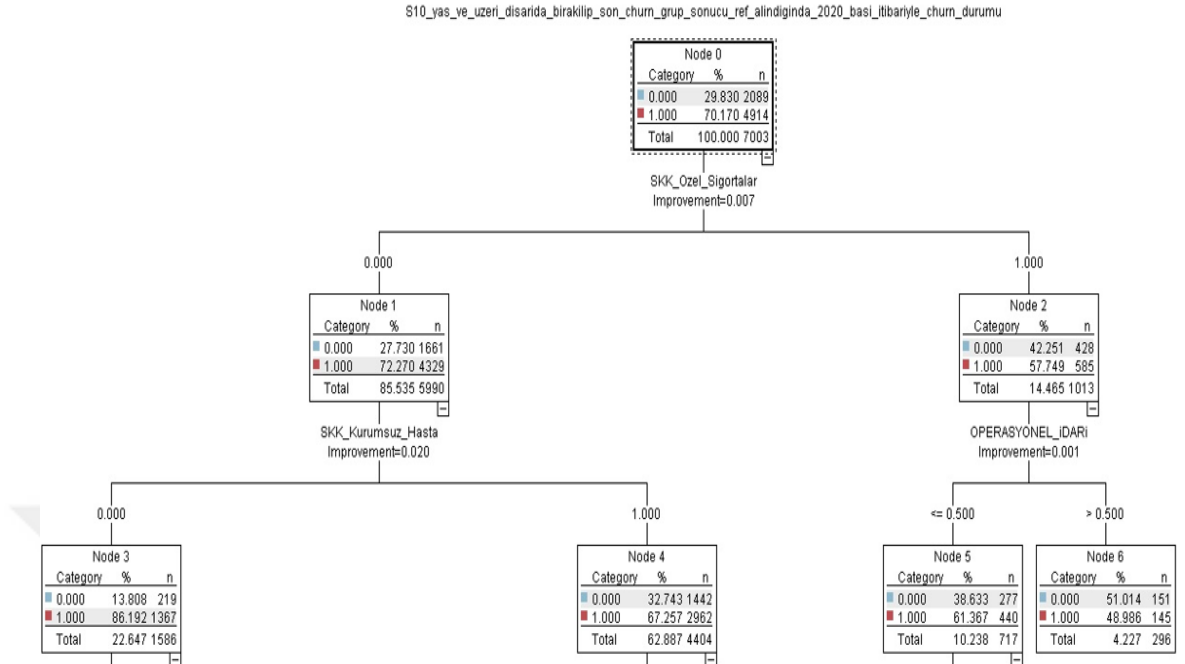
Model eğitim veri setinde; 151 çocuk hastanın kayıp hasta olmayacağı doğru şekilde tahminlenmiş, 1938 çocuk hasta kayıp hasta olmaması hatalı olarak tahminlenmiş, kayıp hasta olduğu saptanmıştır. 4.769 çocuk hastanın kayıp hasta olacağı model tarafından doğru tahmin etmiş, 145 çocuk hasta ise, kayıp hasta olacağı tahminlenmişken, kayıp hasta olmadığı gözlenmiştir.

Benzer şekilde, **test veri aşamasında;** 54 çocuk hastanın kayıp hasta olmayacağı modelce doğru şekilde tahminlenirken, 839 çocuk hasta kayıp hasta hatalı olarak tahminlenmiş, kayıp hasta olduğu saptanmıştır. 2.071 çocuk hastanın kayıp hasta olacağı model tarafından doğru tahminlenmiş, 79 çocuk hasta ise, kayıp hasta olacağı tahminlenmişken, kayıp hasta olmadığı gözlenmiştir.

Modelin değerlendirilmesinde; **hassasiyet (recall) metriği**, model dahilinde tüm pozitif tahminlenenlerin, yani kayıp hasta olacağı öngörülenlerin, ne kadarının gerçekte pozitif / kayıp hasta olduğunu incelemektedir. Hassasiyetin eğitim verisinde %71,10, test aşamasında %71,17 olduğu görülmüştür.

Kesinlik (Precision) metriği, gerçekte pozitif kayıp hasta olanların aslında kaç tanesinin doğru tahminlendiğinin değerlendirilmesinde; eğitim veri setinde %97,05, test veri aşamasında %96,33 olarak gerçekleştiği saptanmıştır. Bir başka ifade ile, eğitim veri setinde; gerçekte kayıp hasta olan 100 hastanın 97,05 tanesinin pozitif/kayıp hasta olan olarak tahminlendiği, test veri aşamasında ise gerçekte kayıp hasta olan 100 hastanın 96,33'ünün kayıp hasta/pozitif olarak tahminlendiği görülmüştür.

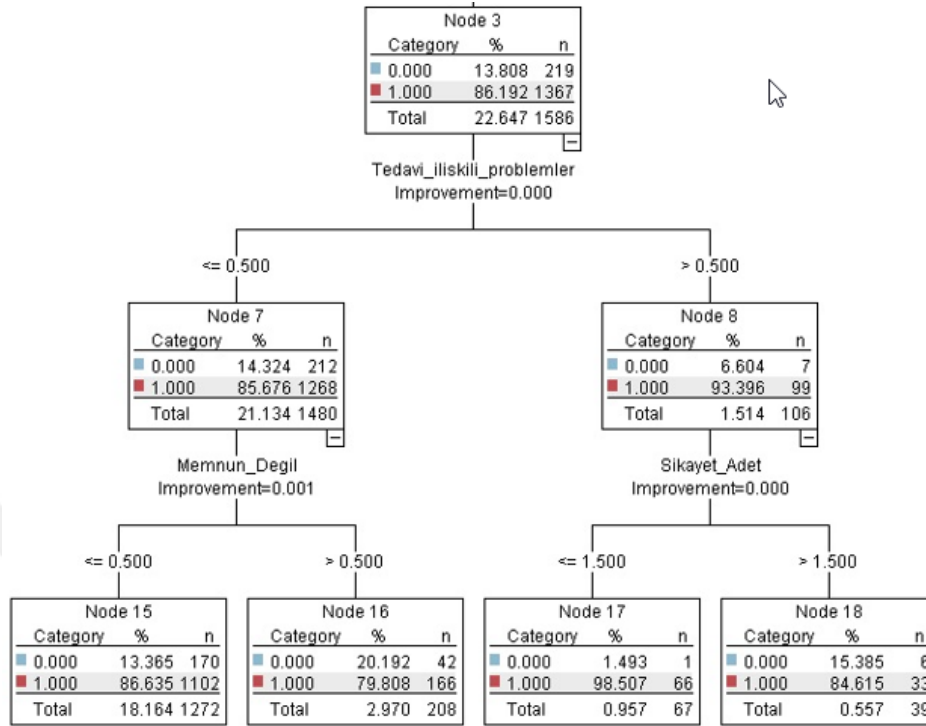
5.3.2. CART Algoritması ile Bulgular ve Değerlendirme Metrikleri



Şekil 5.68.: Şikayetler ve Kayıp hasta CART Uygulaması 1. Düzüm ve 2. Düzüm ve Dalları

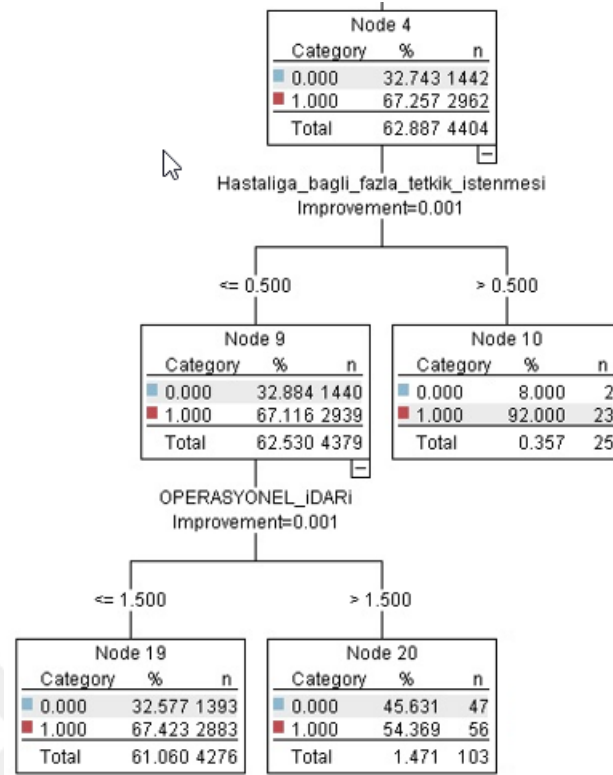
Yapılan CART analizine göre öncelikle sistemde şikâyeti bulunan, çalışılan 7.003 hastanın (devamlı hasta kriterini sağlayan hastaların %3,3, ilgili dönem toplu çocuk hastaların %2'si) %70,17'sinin kayıp hasta olarak etiketlendiği görülmüştür. Yapılan Gini İndeksi sonrasında en iyi bölen kestiricinin SKK_Özel_Sigortalar, son kabul kurumunun özel sigortalar durumun olduğu ve hastaların son kabul kurumunun özel sigortalar olduğu durumlarda; 1.013 gözlem arasından %57,7'sinin, özel sigortalar dışında bir kurum olduğunda %72,3'ünün kayıp hasta olduğu gözlenmiştir. 5.990 gözlem barındıran birinci düğümün ise SKK_Kurumsuz_Hasta, son kabul kurumunun Kurumsuz – nakit ödemeli hastalar olduğu durumunca en iyi bölündüğü ve hastanın kurumsuz hasta etiketli olması durumunda 4.404 gözlem arasından %67,3'ünün, diğer kabul kurumlarına ait olduğu durumlarda %86,2'sini kayıp hasta olduğunu göstermektedir. Son kabul kurumunun özel sigortalar olduğu 1.013 gözlem içeren ikinci düğümü en iyi bölen kestiricinin Operasyonel_Idari, hastaların operasyonel ve idari başlıklar altında

yer alan şikâyet leri olduğu görülmüş ve eğer hastanın ilgili grupta şikâyet i varsa 296 gözlem arasından %48'inin, yoksa %61'inin kayıp hasta olduğu görülmüştür.



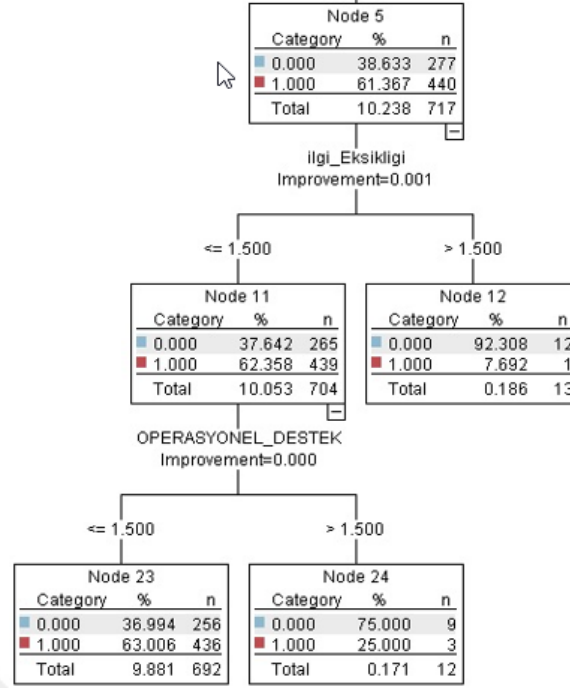
Şekil 5.69.: Şikayetler CART 3. Düzüm ve Dalları

1.586 gözlem içeren 3. Düzüm ise, kendi içinde Tedavi ilişili_problemler değişkenince en iyi bölünmüş, hastaların tedavi ile ilişkili şikâyet leri varsa 106 gözlem arasından %93'ünün, yoksa 1.480 gözlem arasından %85,7'sinin kayıp hasta olduğu görülmüştür. 1.480 gözlemlili 7. düğüm de kendi içinde Memnun_Değil, (şikâyet sonrası alınan aksiyonlar doğrultusunda, şikâyeti yöneten kişice atanan hastanın durumu etiketlemesi) etiketi olduğu görülmüştür. 8. düğüm ise kendi içinde Sikayet_Adet, sistemde bulunan şikâyet adedince en iyi bölünmüş ve hastanın sistemdeki şikâyet i 2 ve altında ise 67 gözlem ile kayıp hasta oranının %98,5 olduğu, 2 üzerinde ise %84,6 olduğu görülmüştür.



Şekil 5.70.: Şikayetler CART 4. Düğüm ve Dalları

4.404 gözlem içeren 4. düğüm, hastalığa bağlı fazla tetkik istenmesi etiketince en iyi bölünmüş, eğer sistemde böyle bir şikâyet i olan hasta varsa 25 gözlem üzerinden kayıp hasta oranının %92, eğer yoksa 4.379 gözlem arasından %67.1'inin kayıp hasta olduğu görülmüştür. 9. düğüm de kendi içinde operasyonel ve idari başlıklarda şikâyeti olup olmama durumuna göre bölünmüştür.



Şekil 5.71.: Şikayetler CART 5. Düğüm ve Dalları

717 gözlem içeren 5. düğüm kendi için ilgi eksikliği başlıklı şikâyet adedinin 2 üstü ve 2 ve altı olup olmama durumuna göre bölündüğü ve eğer bu gruptaki hastaların ilgi eksikliği etiketli şikâyetleri 2 ve altında ise %62,4'ünün kayıp hasta olduğu gözlenmiştir. Bu grupta kendi içinde Operasyonel_Destek yani destek hizmetleri ana grubundaki şikâyetin adedine göre dallanmıştır.

Değerlendirme Metriği

Çalışmada MİY sisteminde şikâyet kaydı bulunan ve devamlı hasta kriterlerini sağlayan 10.046 çocuk hastanın şikâyet ayrıntıları ve kayıp hasta kabul edilme durumu CART ile değerlendirilmiştir. Toplam veri setinin 7.003 hasta ile %70'i modelin eğitiminde, 3.043 hasta ile %30'u test için kullanılmıştır.

Tablo 5.26: Şikayetler CART Değerlendirme Metrikleri

CART Şikayetler - Churn Değerlendirme Metrikleri					
	Doğru	Yanlış	Doğruluk	Hassasiyet	Kesinlik
Eğitim	3.258	3.745	46,52%	86,28%	29,93%
Test	1.394	1.649	45,81%	87,90%	27,02%

Modelin eğitim veri setinde, modelin %76,52 oranında doğru tahminlemede, test aşamasında da %45,81'lik doğru tahminde bulunduğu saptanmıştır. Modelin doğruluk oranının düşük olması sebebiyle karışıklık matrisi ayrıca paylaşılmamıştır. Modelin değerlendirilmesinde; **hassasiyet (recall) metriği**, model dahilinde tüm pozitif tahminlenenlerin, yani kayıp hasta olacağı öngörülenlerin, ne kadarının gerçekte pozitif / kayıp hasta olduğunu incelemektedir. Hassasiyetin eğitim verisinde %86,28, test aşamasında %87,90 olduğu görülmüştür.

Kesinlik (Precision) metriği, gerçekte pozitif kayıp hasta olanların aslında kaç tanesinin doğru tahminlendiğinin değerlendirilmesinde; eğitim veri setinde %29,93, test veri aşamasında %27,02 olarak gerçekleştiği saptanmıştır. Bir başka ifade ile, eğitim veri setinde; gerçekte kayıp hasta olan 100 hastanın 29,93 tanesinin pozitif/kayıp hasta olan olarak tahminlendiği, test veri aşamasında ise gerçekte kayıp hasta olan 100 hastanın 27,02'sinin kayıp hasta/pozitif olarak tahminlendiği görülmüştür.

5.4. Hasta Memnuniyet Anketleri ve MKA

Anket sorularından NPS skorunun belirleyicisi olan “Hastanemizi bir yakınınıza veya arkadaşınıza tavsiye eder misiniz / etme ihtimaliniz nedir? (0-10 arası değerlendirme) ile kayıp hasta kabul kriteri arasındaki çalışmada; ilgili sağlık kurumunun veri tabanından MİY yazılımında tanımlı anketler raporu ile 01.01.2016 ile 31.12.2018 tarihleri arasındaki sisteme kayıtlı 25.885 adet çocuk hastalar için doldurulmuş anket edinilmiştir. Devamlı hasta kriterlerini sağlayan ve MKA kapsamında değerlendirilen çocuk hastalar ile eşleştirilerek 13.545 adet ankete indirilmiştir. Veri temizleme, işleme ve dönüştürme aşamasında, NPS ile ilişkili sorunun boş bırakıldığı durumlar veri setinden çıkarılmış ve 13.307 adet hasta NPS sorusu ve MKA değerlendirmesine dahil edilmiştir. NPS sorusuna 0-6 arası puan verenler “Kötüleyenler”, 7-8 verenler “Pasifler” ve 9-10 verenler “Destekleyenler” olarak etiketlenmiştir. Birden fazla anket cevabı olan hastalarda en son dolduran anket baz alınmıştır. MKA çalışmasına konu olan çocuk hastalardan kaç tanesine anket gönderildiğinin bilgisi olmamakla birlikte,

çalışmaya konu olan çocuk hastaların sadece yaklaşık %9,5'lik bir kısmı için sistemde anket bulunmaktadır.

Soruya verilen cevaplar 0-10 ve grupları ile kayıp müşteri durumu IBM SPSS Modeler 17.0 programında Ki-Kare testi ile incelendiğinde;

Tablo 5.27: Anketler ve Kayıp hasta Ki-Kare Testi Sonuçları

Grup		0	1	Total
K1	Adet	422	988	1.410
	Satır %	29,9%	70,1%	100,0%
K2	Adet	493	1.089	1.582
	Satır %	31,2%	68,8%	100,0%
P	Adet	1.114	2.146	3.260
	Satır %	34,2%	65,8%	100,0%
D	Adet	2.600	4.455	7.055
	Satır %	36,9%	63,1%	100,0%

Ki-Kare = 37.648, df = 3, p = 0,000

K1: Kötüleyenler Grup 1: 0,1,2,3 puan verenler

K2: Kötüleyenler Grup 2: 4,5 ve 6 puan verenler

P: Pasifler: 7 ve 8 puan verenler

D: Destekleyenler: 9 ve 10 puan verenler

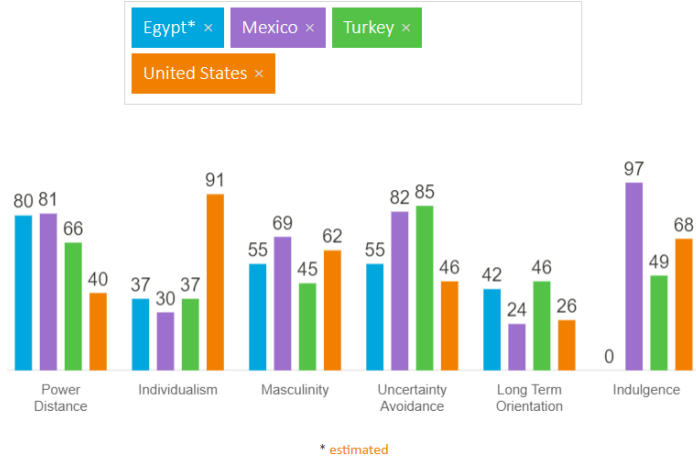
Ki-Kare analizi sonucunda gruplara göre kayıp hasta olup olmama düzeyi arasında istatistiki olarak anlamlı ($p < 0,05$) bir fark bulunmaktadır. Çocuk hasta için doldurulan ankette Destekleyen grubunda olan bir kişinin kayıp hasta oranı %63 ile diğer gruplardan daha düşüktür. Bu oranın, Pasifler grubunda %65 olduğu görülmüştür. Kötüleyenler grubu (0 ve 6 arasında puan verenler), 0-3 verenler K1, 3-6 verenler K2 olarak bölünmüştür. K1 grubunda olan hastaların % 70,1'inin, K2 grubunda olanların ise %68,8'inin kayıp hasta olduğu gözlenmiştir.

Ki-Kare testinden hareketle, kayıp hasta kabul edilen hastalardan sistemde anketi bulunan 13.307 hastanın anketteki “Hastanemizi bir yakınınıza veya arkadaşınıza tavsiye eder misiniz / etme ihtimaliniz nedir” sorusuna cevabı, 0 ve 3 arasında ise %70,1’inin, 3 ve 6 arasında ise %68,8’inin kayıp hasta statüsüne geçme potansiyeli olduğu tahminlenebilmektedir.

Memnuniyetsiz müşterilerin %96’sının şikâyetlerini ilgili mal/hizmet sunucusuna aktarmadığı (Beal, 2014), müşterilerin %92’sinin ağızdan ağıza pazarlamayı diğer pazarlama kanallarına göre daha güvenilir bulduğu kabul edildiğinde (Nielsen,2012); NPS daha da önem kazanmaktadır. Burada NPS’ye nasıl ulaşıldığı, anketlerin hangi müşterilere, hangi aralıkla, hangi formatta ve nasıl sorulduğu NPS sonucunu etkileyebilmektedir. Aynı zamanda, müşterilerin anketlere dönüş oranı, kültürel konjonktür de değerlendirmede önem arz etmektedir.

Hofstede’nin 72 farklı ülkedeki IBM çalışanlarına uyguladığı organizasyonel değerler çalışması ile başlayan ve bugün kültürel değerlendirmelerde referans olan “Cultural Dimensions” Kültürel Boyut Ölçütü’ndeki (Hofstede ve Minkov, 2010; Hofstede, 2001) “individualism” bireysellik ölçütünde; Türkiye’deki insanların bireysellikten çok gruplar halinde hareket ettikleri paylaşılmıştır. Bu doğrultuda, mutsuz hastaların deneyimlerini daha bireyselliğin yüksek olduğu ülkelere nispeten daha fazla kişi ile paylaşacakları düşünülmüştür. Örneğin, Amerika’da mutsuz olan bir hastanın şikâyet bildiriminde bulunma ihtimali Türkiye’dekine göre daha fazla iken, Türkiye’de mutsuz ama sisteme kayıtlı bir şikâyeti bulunmayan bir hastanın, Amerika’daki örneğine kıyasla daha fazla insanla olumsuz deneyimini paylaşacakları düşünülmektedir.

Yapılan bir çalışmada; Amerika, Meksika, Mısır ve Türkiye üzerine bireyselliğin “individualism” ve toplumsallığın “collectivism” etkin olduğu toplumlarda tüketici davranışları ve negatif ağızdan ağıza pazarlama eğilimleri incelenmiştir (Chapa, vd., 2014). İlgili ülkelerin bireysellik, toplumsallık karşılaştırma Şekil 5.72’de verilmiştir.



Şekil 5.72.: Kültürel Boyut Ölçütünde Ülke Karşılaştırması

Kaynak: <https://www.hofstede-insights.com/fi/product/compare-countries/>

Her ne kadar toplumsallığın güçlü olduğu Türkiye, Mısır ve Meksika'nın kendi içindeki davranış paternleri değişse de, toplumsallığın yaygın olduğu Türkiye ve bireyselliğin yaygın olduğu Amerika değerlendirmesinde; Amerika'da mutsuz bir müşterinin görüşlerini ilgili kurum ile paylaşma eğilimi Türkiye'dekilerden fazla iken, Türkiye'deki müşterilerin olumsuz deneyimleri yakınları ve çevresi ile Amerika'dakilere nispeten daha fazla paylaştığı, benzer oranda mal/hizmet sunucusu değiştirdikleri ve daha fazla kişiyi ilgili hizmet sunucunu değiştirmeye davet ettiğini göstermiştir.

SONUÇ VE ÖNERİLER

İşletmeler için büyümeyi ve sürdürülebilirliği sağlayabilmenin en temel faktörlerinden birisi, sistemlerine yeni müşterileri dahil etmek ve/veya mevcut müşterilerinin mal/hizmet alımlarının devamlılığını sağlayabilmektir. Özellikle son yıllarda yaşanan krizler doğrultusunda artan maliyetler ve kızışan rekabet ortamı mevcut müşterilerin, işletmelerin kendi ekosistemi içinde tutundurulmasının önemini arttırmıştır. Yapılan çalışmalarda yeni müşteri kazanmanın mevcut müşteriyi mutlu edip elde tutmaya göre beş ile on kat daha maliyetli olduğuna yer verilmektedir (Kotler ve Armstrong 2017). Yine çalışmalar, mal/hizmet alımı sonrası memnuniyetsiz ayrılan müşterilerin %96'sının şikâyet başvurusunun olmadığını fakat deneyimlerini en az 15 kişi ile paylaştıklarını göstermektedir (Beal, 2014). Nilsen (2012) de çalışmasında müşterilerinin %92'sinin ağızdan ağıza pazarlamayı diğer pazarlama-iletişim kanallarını kullanmaya tercih ettiklerini, tanıdıklarından duydukları ya da müşteri bloglarından okumuş oldukları müşteri deneyimlerini daha güvenilir bulduklarını göstermiştir. Bu bilgiler ışığında, memnuniyetsizlik sonucu kaybedilen müşterinin sadece ilgili dönemdeki katkıları değil, YBMD düşünüldüğünde ve güçlenen sosyal medya desteği ile ulaşabileceği potansiyel müşteri etki alanı değerlendirildiğinde, bu soruna çözüm olarak geliştirilen MKA daha da kritik bir hal almaktadır.

Konu sağlık olunca, hastalar hizmet alım sonrası memnuniyetsizliklerini diğer sektörlere nispeten daha fazla kişi ile paylaşmaktadırlar. Ayrıca, artan maliyet ve rekabet baskısı da sağlık sektöründe MKA uygulamalarının önemini daha da arttırmaktadır. Müşteri sürekliliğinin bulunduğu ve genellikle abonelik esasına dayalı bankacılık, sigortacılık ve telekomünikasyon gibi sektörlerde yaygın olarak kullanılan MKA, sağlık alanında son birkaç yılda görülmeye başlanmıştır. Yazında özellikle hastane sektöründe dünya genelinde çok nadir çalışmalar gözlenirken, Türkiye'de ise yapılan yazın taramasında ilgili çalışmalara rastlanılmamıştır.

Sağlık hizmet sunucuları hastalarının tercihlerini, ne zaman ve hangi durumlarda hizmet sunucusu değiştirdiklerini tam analiz edemedikleri için her sağlık kuruluşu,

mevcut hastalarını elde tutmak, memnuniyetlerini ölçmek ve sağlamak için farklı politikalar izliyorlar olsalar da hasta geribildirimleri sonrası alınan aksiyonların ötesine geçememektedir. Hastalarının tercih örüntülerini saptayabilmek için hastaneler kimine memnuniyet anketi göndererek, kimini telefonla arayarak, kiminden de basılı formları doldurmasını isteyerek reaktif uygulamalar yapmaktadır. Halbuki, hasta profil ve davranışlarına göre hizmet alım yaşam döngülerinin ve MKA'nın gerçek zamanlı olarak takip edilmesi ise, hastanelere kayıp hastaların önceden tahmin edilebilmesi, proaktif stratejilerin belirlenerek kaybı önleyici zamanında aksiyonların alınabilmesi gibi avantajlar sağlayacaktır.

Yapılan bu çalışmada, sağlık sektöründe MKA'nın uygulanabilirliğini araştırmak, kayıp hastaları tanımlayabilecek ve tahmin edebilecek bir modeli tasarlamak ve hasta kayıplarını önlemeye çalışacak aksiyon planlarına ışık tutmak amaçlanmıştır. Çalışma kapsamında, sağlık sektöründe hasta ve müşteri kavramlarından hangisinin daha kabul edilebilir olduğu, hastaların yaş grupları, sosyal güvenceleri ve geliş sıklıkları ile kayıp hasta durumları, sisteme kayıtlı şikayetleri ile kayıp hasta durumları ve NPS'yi etkileyen anket sorusu ve kayıp hasta durumları arasındaki ilişkileri tahminleme modeli doğrultusundaki araştırmalar yer almıştır. Ayrıca, kullanılan VM algoritmalarından CHAID ve CART'dan hangisinin daha iyi sonuç vereceğinin tespiti çalışmaları yapılmıştır. Aşağıda hipotez sorularının sonuçları yazın, sektörel bilgi ve kültürel çerçeve doğrultusunda paylaşılmıştır.

Yapılan yazın ve sektörel pratik uygulamalar incelemesi doğrultusunda, sağlık alanında öncelikle medikal hizmet alıcılarının müşteri yerine hasta olarak anılmasının hem etik hem de kültürel değerler doğrultusunda daha doğru olduğu, medikal alan dışındaki diğer hizmetlerdeki memnuniyetlerinin sağlanmasında müşteri yaklaşımının benimsenmesinin etkin olacağına kanaat getirilmiştir.

MKA'nın sıklıkla uygulandığı bankacılık, telekomünikasyon ve sigortacılık gibi abonelik esasına dayalı müşteri sürekliliğinin bulunduğu sektörlerin yanı sıra sağlık alanında, özel hastane sektöründe de MKA'nın uygulanabileceği saptanmıştır. Sağlık alanında MKA'nın yapılabilmesi için öncelikle, devamlı hasta kabul ve kayıp hasta kabul ölçütlerinin tanımlanması gerekmektedir. Her bir

hastalık farklı bir akış içerdiği için öncelikle çalışmanın hedef grubu belirlenmiştir. Çalışmada örneklem olarak çocuk hastalar alınmıştır. Bu grubun seçiminde, anne-babaların çocuklarının sağlıklarına (hastalık hali ve planlı takiplerde) kendilerine olduğundan daha fazla duyarlı olmaları, çocukların hizmet alımının devamlılık arz etmesi, yaş grupları ve hizmet alım sıklıklarına göre gruplandırılabilmesi, anne-babaların konu çocukları olduğunda aldıkları hizmet sonrası gözlem, şikâyet ve önerilerini daha sık dile getirmeleri önemli rol oynamıştır.

Devamlı-kayıp hasta tanımlamaları için yarı yapılandırılmış mülakat tekniği kullanılmıştır. Mülakatlar, sağlık kuruluşunun üç farklı şehirdeki beş farklı hastanesinde görev yapan 15 farklı çocuk hekimi ile medikal açıdan ve altı üst düzey yöneticisiyle yönetim açısından gerçekleştirilmiştir. Yapılan çalışma sonunda devamlı hasta kabul kriterinin üste üste 3 poliklinik (ödemeli) muayene olduğu saptanmıştır. Yaş gruplarına göre değerlendirildiğinde; 0-1,1-2 yaş gruplarında üste üste üç kez poliklinik muayenenin 1 yıllık dönem içinde, 2-6 ve 6-10 yaş gruplarında ise 2 yıllık dönemde sağlanabileceğine yönetici görüşleri sonucunda belirlenmiştir. Üst düzey yönetici mülakatlarından, iki poliklinik (ödemeli) muayene sonrası kendi hastalarımız kabul edebileceği görüşü hâkim olsa da bazen bir hastanın aynı hastane içinde başka doktorları sıra ile denemesi de göz önünde bulundurulduğunda, bir hastaya bizim hastamızdır denilebilmesi için ilgili zaman diliminde en az üç poliklinik muayene başvurusunun olması gerektiği ortak fikrine varılmıştır.

Kayıp hasta kabul kriterlerinin belirlenmesinde de aynı yöntem kullanılmış, 0-1 yaş grubunda 2 ay, 1-2 yaş grubunda 4 ay, 2-6 yaş grubunda 6 ay, 6-10 yaş aralığında 1 yıl ve üzeri çocuk hastanın başvurusunun bulunmaması durumu olarak tanımlanmıştır. Yapılan bu tanımlamalar, MKA için zemin hazırlamakta ve diğer hastalıklar için de yaklaşım örneği teşkil etmektedir.

Devamlı hasta kabul ölçütleri sağlandığında; Büyük Veri üzerinde VM teknikleri ile hastaların davranış kalıpları saptanabilir ve gelecekteki hareketlerini tahmin etmede kullanılabilir. Kim hangi durumda hastamızdır ve hangi durumda kayıp müşteri/hasta kabul edilebilir sorusunun cevabı netleştikten ve süreklilik arz eden

durumlar tanımlandıktan sonra; bu ayrımın müşterilerin davranışları, sosyal güvenceleri, geliş sıklıkları, sistemde bulunan şikâyet adet ve tipleri, varsa yapılan anketlere verdikleri cevaplar ile ilişkilendirilebilir. Böylelikle, kuruluşlar bir sonraki aşamaya, tahminleme kısmına geçebilir ve alacakları önlemlerle hastaları kayıp hasta/müşteri konumuna yaklaştırmadan gerekli önlemlerini alabilirler.

MKA çalışmalarının her bir branşta değil konjenital, kalıtsal, kronik rahatsızlıklar ya da hamilelik-doğum dönemi gibi düzenli takip gerektiren birçok tıbbi branş için, zaman dilimi tanımlaması, devamlı hasta kabul ve kayıp hasta kabul ölçütlerinin tanımlanması ile yapılabileceği saptanmıştır. Dolayısıyla, hastaneler MKA ile hastalarının hizmet sunucusunu değiştirme eğilimlerini saptayabilir ve geliştirilebilecek erken uyarı sistemleri aracılığıyla kaybı önleyici faaliyetlere odaklanabilirler. Örneğin, gebelik tanısı almış bir anne adayının, gebelik takibini ilgili hastanede yapma durumu ve/veya doğumunu ilgili hastanede gerçekleştirme durumu ile anne adayının sosyodemografik yapısı, geliş sıklıkları, sosyal güvencesi, şikâyet bildirim durumu ya da anket yanıtları arasındaki örüntüler incelenebilir ve tahmin modelleri kurulabilir. Böylelikle, hastaneler anne adaylarının davranış ve eğilimleri üzerinden, kayıp hasta durumu önleyici faaliyetlerde bulunabilir ve hastalarını tutundurabilirler.

Yapılan çalışma doğrultusunda çocuk hastaların yaş gruplarına göre kayıp hasta oranlarının farklılık gösterdiği görülmüştür. Devamlı hasta kabul edilen tüm çocuklarda, 0-1 yaş grubunda kayıp hasta oranının %38, 1-2 yaş grubunda %40, 2-6 yaş grubunda %63 ve 6-10 yaş grubunda %44 olduğu görülmüştür. Yazında kayıp hastalar üzerine yapılmış çalışmalar yaygın olmasa da Meltzer (2022), Amerika pazarı için yaptığı bir çalışmada hastaların %30-40'ının ilk başvuru sonrası tekrar bir gelişinin olmayacağını ve memnuniyetsiz ayrılan hastaların %67'sinin hizmet sunucusunu değiştireceğini paylaşmıştır.

Araştırılan örüntülerden ilki olan geliş sıklıkları ve kayıp hasta kabul edilme durumu arasında anlamlı bir bağ olduğu ve tahminlemede kullanılabileceği saptanmıştır. Yapılan çalışmada, 1-2 yaş grubunda %39,98'lik genel kayıp hasta oranına sahip modele alınan çocuk hastaların değerlendirmesinde, karar ağacını en iyi bölen kestiricinin 1-2 yaş arasındaki toplam poliklinik muayene (ödemeli)

başvuru adedi olduğu, ilgili dönemde devamlı hasta kriterinin sağlanması sonrasında eğer 1 poliklinik muayene başvurusu varsa %70,67'sinin, 2 başvurusu varsa %67,3'ünün ve 7 başvurusu varsa %20,9'unun kayıp hasta olduğu saptanmıştır. Bir diğer örnekte, CART algoritmasında, 6-10 yaş grubu için karar ağacını en iyi bölen açıklayıcı değişkenin, çocuk hastaların 8-9 yaş başvuruları arasındaki maksimum gün farkı olduğu görülmüş, eğer ilgili gün farkı 365 günün altındaysa kayıp hasta oranının %46,29, 365 günün üzerinde ise %60,05 olduğu saptanmıştır. Ayrıca, 8-9 yaş başvuruları arasındaki gün farkı 365 günün altında ve 7-8 yaş başvuruları arasındaki gün farkı 365 günün altındaysa 8.420 çocuk hasta arasından kayıp hasta oranının %24,74 olduğu, diğer yandan eğer ilgili gün farkı 365 günün üzerinde ise 3.380 çocuk hastadan %100'ünün kayıp hasta olduğu saptanmıştır. 6-10 yaş grubunun CHAID ile değerlendirilmesinde, CART uygulamasında olduğu gibi karar ağacını en iyi bölen değişkenin, çocuk hastaların 8-9 yaş başvuruları arasındaki maksimum gün farkı olduğu görülmüştür. CHAID'e göre, eğer 8-9 yaş aralığındaki başvurular arasındaki fark 92 günden az ise kayıp hasta oranının %25,4, 92-124 gün ise %22,3, 124-217 gün ise %26,4, 217-314 gün ise %30,5 ve 314-390 gün ise kayıp hasta oranının %47,1'e yükseldiği görülmüştür. Her iki algoritmanın tahmin başarı performansları karşılaştırıldığında, CHAID'in test veri setinde %86,70'lik bir doğruluk oranı olduğu, CART algoritmasının ise 95,22% doğruluk oranı ile daha başarılı bir tahminlemede bulunduğu saptanmıştır.

En yüksek kayıp hasta oranı ile dikkat çeken 2-6 yaş grubuna ayrıca odaklanması önerilmektedir. Yapılan VM doğrultusunda CART algoritmasına göre 2-6 yaş grubu çocuk hastaların kayıp hasta olma durumunda, en belirgin etmenin çocuk hastaların 5-6 yaş başvuruları arasındaki maksimum gün farkı olduğu, ilgili gün farkının 180'den az olması durumunda kayıp hasta oranının %39,88, 180 günden fazla olması durumunda ise %70,13 olduğu saptanmıştır. 2-6 yaş grubu CHAID ile değerlendirildiğinde, CART algoritmasında olduğu gibi karar ağacını en iyi bölen değişkenin, çocuk hastaların 5-6 yaş başvuruları arasındaki maksimum gün farkı olduğu görülmüştür. CHAID'e göre, eğer 5-6 yaş aralığındaki başvurular arasındaki fark 52 günden az ise kayıp hasta oranının %13,2, 52-91 gün

ise %38,6, 91-119 gün ise %44,3, 119-146 gün ise %48,3, 146-177 gün ise %53,3 ve 177-215 gün ise kayıp hasta oranının %96,4 olduğu görülmüştür. Her iki algoritmanın tahmin başarı performansı karşılaştırıldığında, CHAID'in test veri setinde %82,06'lık bir doğruluk oranı olduğu, CART algoritmasının ise 82,89% doğruluk oranı ile daha başarılı bir tahminlemede bulunduğu saptanmış diğer yandan modeller arasındaki farkın minimal olduğu gözlenmiştir. CART ve CHAID bulguları doğrultusunda ilgili sağlık kurumunun, 2-6 yaş grubu hastalarının ilgili dönemdeki başvuruları arasındaki gün farkı 177'ye yaklaştığında, hastaları takibe almaları, çağrı merkezi, ilgili bölüm asistanı veya doktoru tarafından memnuniyet aramaları yapmaları ya da iletişimi kuvvetlendirmek ve sağlık sunumunu kişiselleştirmek amacıyla farklı hizmetleri sunarak tutundurma faaliyetlerinde bulunmaları önerilmektedir.

Hangi Karar Ağacı algoritmasının geliş sıklıkları ve kayıp hasta tahminlemede daha başarılı sonuçlar vereceğinin karşılaştırmasında, CHAID'in test veri setinde 0-1 yaş için %69,98, 1-2 yaş için %70,42, 2-6 yaş için %82,06 ve 6-10 yaş grubu için %86,70 oranında doğru tahminlemede bulunduğu görülmüştür. Diğer yandan CART algoritmasında, 0-1 yaş için %69,75, 1-2 yaş için %68,74, 2-6 yaş için %82,89 ve 6-10 yaş grubu için %95,22 oranında doğru tahminlemede bulunduğu görülmüştür. Özetle, 0-1 ve 1-2 yaş gruplarında tahmin performansları yakın olmakla birlikte CHAID'in, 2-6 ve 6-10 yaş gruplarında CART algoritmasının kayıp hasta tahminini en başarılı şekilde yaptığı saptanmıştır. Ayrıca, karar ağaçları algoritmalarının değerlendirme metriklerinde, doğruluk, hassasiyet ve kesinlik değerlerinde, modellerin hasta yaşı arttıkça daha başarılı tahminlerde bulunduğu görülmüştür. Bu durumun, ebeveynlerin ilk yaşlarda ve özellikle ilk çocuklarda devamlı takibinde kalacakları, bizim doktorumuz diyecekleri doktor arayışının olması, ilerleyen yaşlarda ise hem beklenen geliş sıklıklarının azalması hem de doktor hasta bağının artık belirlenmiş olmasından kaynaklandığı düşünülmektedir.

Şikayetler ve kayıp hasta değerlendirmesinde, kayıp hasta olan hastaların sadece %7'si (10.046 çocuk) için sistemde bir şikâyet kaydı bulunduğu, aynı oranın devamlı hasta kriteri sağlayan ve çalışmaya konu olan çocuk hastalar

için %4,8 ve ilgili dönemdeki toplam çocuk hastalar üzerinden ise %2'nin altında olduğu görülmüştür. Yapılan çalışmada, devamlı hasta kriterini sağlayan çocuklardan sisteme kayıtlı şikâyeti bulunanlarının %70'inin kayıp hasta olduğu görülmüştür. Bu doğrultuda, müşterilerin %96'sının herhangi bir şikâyetle bulunmadığı bilgisiyle (yapılan çalışma da bu bulguyu desteklemektedir) birlikte, şikâyeti olan hastaların yakın takibe alınması, iletişimin ve ilgili süreçlerin iyileştirilmesi ve memnuniyetlerinin sağlanması önem arz etmektedir. Hastaların sadece %4'lük kısmından gelen şikâyetlerin, kayıp hasta olan fakat herhangi bir şikâyeti bulunmayan büyük grubu da temsil edebileceği düşüncesiyle, hastanelerin her bir şikâyete ekstra önem vermeleri ve yapılan iyileştirmeleri hastalarla da paylaşımları önerilmektedir. Uygulanan VM bulguları doğrultusunda, şikâyeti olan hastalarda, sosyal güvencenin özel sigorta olduğu durumlarda kayıp hasta oranının %57,75, Kurumsuz hastalarda ise %67,26 olduğu görülmüştür. Buradan hareketle, sağlık kurumunun şikâyet yönetiminde Kurumsuz hastalarda daha hassas olması gerektiği görülmüştür.

Kullanılan veri madenciliği modellerinden CHAID'in %70,26 ile doğru tahminleme de CART'tan daha başarılı olduğu görülmüştür. Diğer yandan bu oranın, geliş sıklıkları çalışmasındaki tahmin doğruluğu oranına göre daha düşük olmasının, hastaların sadece %4,8'inin şikâyetlerini dile getirmiş olması yani birçok hastanın şikâyetle bulunmadan kayıp hasta olmasından kaynaklandığı düşünülmüştür.

Çocuk hastalar için kağıt anket, e-posta ve cep telefonuna SMS yollarıyla toplanılan anket formlarının içindeki NPS belirleyici sorusu "Hastanemizi bir yakınınıza veya arkadaşınıza tavsiye eder misiniz / etme ihtimaliniz nedir? (0-10 arası değerlendirme)" ile kayıp hasta durumu arasındaki ilişki üzerine Ki-Kare aracılığıyla yapılan analizde, gruplara göre kayıp hasta olup olmama düzeyi arasında istatistiki olarak anlamlı ($p < 0,05$) bir fark bulunduğu saptanmıştır. Çocuk hasta için doldurulan ankette, Destekleyen (9 ve 10 puan verenler) grubunda olan bir kişinin kayıp hasta oranı %63 ile diğer gruplardan daha düşüktür. Bu oranın, Pasifler grubunda %65 olduğu görülmüştür. Kötüleyenler grubu (0 ve 6 arasında puan verenler), 0-3 verenler K1, 3-6 verenler K2 olarak bölünmüştür. K1 grubunda olan hastaların %70,1'inin, K2 grubunda olanların ise %68,8'inin kayıp

hasta olduđu gözlenmiştir. Çalışmaya konu olan çocuk hastaların sadece %9,5'i için sistemde bir anket bulunmakta ve Kötüleyenler, Pasifler ve Destekleyenler grupları arasındaki kayıp hasta oranı farkları çok yüksek değildir. Diğer yandan hastanelere, memnuniyetsiz bir hastanın deneyimini en az 15 kişi ile paylaşması, müşterilerin %92'sinin ağızdan ağıza pazarlamayı diğer pazarlama kanallarına göre daha güvenilir bulduđu (Nielsen, 2012) bilinciyle, ilgili anket sorusuna düşük puan veren hastaları takibe almaları ve yıpratılan iletişimin tekrar inşa edilmesi için iletişime geçmeleri önerilmektedir.

Yapılan bu çalışmalar, içinde bulunulan toplumun kültürel yapısı doğrultusunda değerlendirilmelidir. Hosftede'nin Kültürel Boyut Ölçütündeki kişilerin bireysellik ve toplumsallık endekslerindeki değişiklikler, hastaların resmi kanallarla şikâyet etme, deneyimlerini başkalarıyla paylaşma ve tanıdıklarını memnun kalmadığı hastaneden hizmet alımından engelleme çabasında farklılık göstermektedir. Bu doğrultuda, Türkiye, her geçen yıl bireysellik seviyesinde artış olsa da, bireylerin gruplar halinde hareket ettiği, Hosftede'nin çalışmasında toplumsallığın yüksek olduđu bir ülke olarak yer almaktadır. Chapa ve arkadaşlarının 2014 yılında Amerika, Mısır, Meksika ve Türkiye üzerine yaptıkları çalışmada da görüldüğü gibi toplumsal yaklaşımın belirgin olduđu Türkiye'de bir müşterinin görüşlerini ilgili kurum ile paylaşma eğilimi bireyselliğin hakim olduđu Amerika'daki bir müşteriden az iken, Türkiye'deki müşterilerin olumsuz deneyimleri yakınları ve çevresiyle Amerika'dakilere nispeten daha fazla paylaştığı, benzer oranda mal/hizmet sunucusu değiştirdikleri ve daha fazla kişiyi ilgili hizmet sunucusunu değiştirmeye davet ettiğini göstermiştir. Ayrıca, sektörel boyutun da kendi içinde değerlendirilmesi gerekmektedir, zira, hizmet devamlılığı olan çocuk hastalar, kanser hastaları, doğum takibindeki hastalar, içerisinde bulundukları kültür ve işletmenin kültürü nedeniyle; memnuniyetsizliklerini paylaşmaları durumunda hizmet verenin uyarılabileceği, cezalandırılabilceği ve bir dahaki hizmet alımında daha kötü hizmet alma kaygısı nedeniyle veya paylaştıklarında dinlenmeyeceklerini ya da herhangi bir işlem yapılmayacağını düşündüğünde gözlemlerini paylaşmayabilmektedirler. Diğer yandan, aynı hasta grupları resmi olmayan ortamlarda deneyimlerini paylaşmaktadırlar. Dolayısıyla, kurumlar öncelikle gelen şikâyetleri çözmek için (%4,8'lik kısım paylaşıyor) 20 kat daha

fazla önem vermelidirler ve müşterilerinin gözlemlerini daha farklı kanallarla, daha sık aralıkla toplamalıdır.

Gelecek Çalışmalar İçin Önermeler

Çocuk hastalar üzerine yapılan bu çalışma, bir dönem içerisinde hizmet alım sürekliliği içeren diğer hastalıklar için de MKA'nın uygulanabileceğini göstermiştir. Bu doğrultuda kronik, kalıtsal, kanser, diyabet takibi, gebelik takip ve doğum süreçleri gibi durumlar için de uygulanabilir. Başarılı uygulanması durumunda hem hastanın tedavi bütünlüğü açısından hem kurumsal-sektörel fayda açısından, hem de ülkenin sağlık sistemi genelinde katkılar sağlayacağı düşünülmektedir. MKA uygulamaları hastalar açısından değerlendirildiğinde; hastaların memnuniyetsizlik ve takibinde hizmet sunucusu değiştirme nedenleri kuruluşlarca tespit edildiğinde, hastalardaki memnuniyetsizlik büyük ölçüde giderilebilecektir. Hastalar güvenebilecekleri yeni bir hizmet sunucusu arayışına girmek zorunda kalmayacaklardır. Tedavi bütünlüğü açısından hastaların tedavilerini aynı hizmet sunucusunda (veri bütünlüğü, iletişim kolaylığı vs.) tamamlayıcı olması daha iyi medikal sonuçlar getirecektir. Planlı takip modülleri ile hastaların sağlıkları kendileri için yakından takip edilecektir ve bu da hastaların bilinç düzeyinin artmasına ve/veya sağlıklı takiplerinin yaptırılmasına olanak tanıyacaktır.

MKA kamu açısından değerlendirildiğinde ise; hasta tercihlerinin rekabetin artması sonrasında şekillendiği düşünülürse, vatandaşlar için daha kaliteli ve tamamlayıcı (bütüncül) bir medikal hizmet sunumunun yapılmasına olanak sağlanabilecektir. Özellikle hasta tutundurma faaliyetleri kapsamında, planlı takip modülleri ile (eğer devlet hastanelerince de benimsenirse) toplum sağlığına ciddi katkılar sağlayacağı öngörülmektedir. Günümüzde, toplum sağlığı merkezleri yeni doğan aşıları, gebelik takipleri gibi konularda bir takip yapıyor olsa, bu hizmet çerçevesi ve kalitesi bu vb. modeller üzerinden oldukça arttırılabileceği düşünülmektedir.

Yukarıdaki gelişimlerin sağlık sunucuları tarafından sahiplenilmesi için de sağlık hizmetlerinin ödenmesi mantığında bazı gelişmelere gerek duyulmaktadır.

Günümüzde, Türkiye’de ve birçok diğer ülkelerde, sağlık hizmetleri verilen hizmet/paket başı fiyat uygulaması ile ödenmektedir. Hastaya sunulan hizmetin değer üretilip üretilmediğine bakılmaksızın ödemesi hizmet/paket başına ilgilî birimler (SGK, özel sigorta, hasta vs.) tarafından yapılmaktadır. İlgili ödeme yöntemi ise hastanın, sağlık hizmet sunum sisteminin odağında olmasını desteklememektedir. Ödemelerin hizmet adedini göre değil de hizmet sunumunun amaçlanan çıktıları (medikal sonuçlar, hasta memnuniyeti) ve amaçlanan çıktılara ulaşırken harcanan maddi kaynaklar dengesini göze alan Değer Bazlı Sağlık (Value Based Healthcare) sistemine göre yapılmasının bu tip analizleri daha zarurî hale getireceği ve genel hizmet sunumunun kalitesinde artışa olanak sağlayacağı öngörülmektedir.

Çalışılan, anketlerdeki NPS skorunu belirleyen soru ve kayıp hasta ilişkisi doğrultusunda, anketlerdeki hangi soruların NPS skorunu daha çok etkilediği ve dolayısıyla hastanenin hangi departman ve süreçlerinde (hasta hizmetleri, hemşirelik, otelcilik, idari vb.) yaşanan aksaklıkların kayıp hasta durumunu etkilediği de incelenebilir. Bu durum kurumlara, yapılan anketlerin daha iyi analiz edilebilme ve anket sorularına verilen cevaplar ile müşterilerinin kurumun lehine mi aleyhine mi deneyimlerini paylaşacaklarını ve kayıp hasta olma durumunu tahmin ve değerlendirme imkânı tanıyabilecektir. Yapılacak tahminler doğrultusunda da kurumlar MİY süreçlerine bir ayrı boyut kazandırıp, iletişim ve pazarlama faaliyetlerini stratejik olarak yeniden şekillendirebileceklerdir.

KAYNAKÇA

- Aaker, David A.(1991). Managing Brand Equity: Capitalizing On The Value Of Brand Name.*The Free Press*.[https://doi.org/10.1016/0148-2963\(94\)90009-4](https://doi.org/10.1016/0148-2963(94)90009-4)
- Adak, M. F. ve Yurtay, N. (2014). Gini Algoritmasını Kullanarak Karar Ağacı Oluşturmayı Sağlayan Bir Yazılımın Geliştirilmesi.*Bilişim Teknolojileri Dergisi*, 6 (3), 1-6.
- Adwan O., Faris H., Jaradat H., Harfoushi O., Ghatasheh N. (2014)“Predicting Customer Churn in Telecom Industry using Multilayer Preceptron Neural Networks: Modeling and Analysis”, *Life Science Journal*.11(3).
- Aggarwal, C. C., (2015). *Data classification: algorithms and applications*. CRC Press. Boca Raton
- Agrawal, R. ve Srikant, R. (1995). Mining sequential patterns. *Proceedings of the Eleventh International Conference on Data Engineering*, 3-14. <https://doi.org/10.1109/ICDE.1995.380415>
- Agrawal, R. ve Srikant, R. (1994) Fast Algorithms for Mining Association Rules in Large Databases. *Proceedings of the 20th International Conference on Very Large Data Bases*, 12-15 September 1994, 487-499.
- Akar, Ö. ve Güngör, O. (2012). Rastgele Orman algoritması kullanılarak çok bantlı görüntülerin sınıflandırılması . *Jeodezi ve Jeoinformasyon Dergisi*, (106) , 139-146. <https://doi.org/10.9733/jgg.241212.1t>
- Akyiğit, H. E. Ve Taşcı, T. (2022). Sigortacılık Sektöründe Makine Öğrenilmesi ile Müşteri Kaybı Analizi. *Tasarım Mimarlık ve Mühendislik Dergisi*, 2 (1), 66-79. <https://dergipark.org.tr/tr/pub/dae/issue/68555/1073934> adresinden alınmıştır.
- Alan, M.A.(2019).“Association Rules Mining on Medical Data”, *C.Ü. İktisadi ve İdari Bilimler Dergisi*, 20(1),410-419.
- Albayrak,. Y. S. ve Yılmaz, Ö. K. (2009). Veri Madenciliği: Karar Ağacı Algoritmaları ve İmkb Verileri Üzerine Bir Uygulama . *Süleyman Demirel Üniversitesi İktisadi Ve İdari Bilimler Fakültesi Dergisi* , 14 (1) , 31-52 . <https://Dergipark.Org.Tr/Tr/Pub/Sduibfd/Issue/20831/223135> adresinden alınmıştır.
- Aliyeva, A. (2023) Introduction to SEMMA, Retrieved from, <https://www.linkedin.com/pulse/introduction-semma-data-mining-using-sas-enterprise-abbasaliyeva/>
- Alshamsi, A., (2022), "Customer Churn prediction in ECommerce Sector", [Master's thesis].Rochester Institute of Technology.

- Anand, A. (2003). *A Study and Comparison of Data Clustering Techniques*. [Master Thesis], Faculty of The Graduate School of The University of Texas, El Paso, Texas, USA.
- Atalay, M., Çelik, E. (2017). Büyük Veri Analizinde Yapay Zekâ ve Makine Öğrenmesi Uygulamaları, *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, Cilt: 9, Sayı: 22, ss 155-172.
- Atılğan, E. (2011). *Karayollarında Meydana Gelen Trafik Kazalarının Karar Ağaçları ve Birlikte Analizi İle İncelenmesi*. [Yayınlanmamış Yüksek Lisans Tezi], Hacettepe Üniversitesi, FBE.
- Atınç, Y. ve Ayan, K., (2013) "Cancer risk analysis by fuzzy logic approach and performance status of the model," *Turkish Journal of Electrical Engineering and Computer Sciences*: Vol. 21, No. 3, Article 20. <https://doi.org/10.3906/elk-1108-22>
- Aydan, S. (2020), 'Hasta mı, Müşteri mi?' Tartışmasında Farklı Yaklaşımlar, *Türkiye Klinikleri J Health Sci.* ;5(2):375-8.
- Aydın, Ş.E. (2017), *Yapay Zeka Teknolojisi: Yapay Zekaların Dünü, Bugünü ve Yarını*, Çukurova Üniversitesi, Sosyal Bilimler Enstitüsü, Dönem Projesi.
- Bakır, Y. (2006). *Sağlık Kuruluşlarında Pazarlama Faaliyetleri, Müşteri Memnuniyeti*. [Yayınlanmamış Yüksek Lisans Tezi]. Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü, Hatay.
- Bardi, Ş. ve Can, A. V. (2021). Diskriminant analizi ve C5.0 algoritması ile finansal başarısızlığın tahmini: BİST Kobi Sanayi Endeksi'ndeki işletmeler örneği. *Ömer Halisdemir Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi* , 14 (3) ,1071-1090. <https://doi.org/10.25287/ohuiibf.925344>
- Beal, A. (2014). *Repped: 30 Days to a Better Online Reputation*. Marketing Pilgrim LLC.
- Becker's Hospital Review (2022). How thinking of patients as customers can improve healthcare. Retrieved from <https://www.beckershospitalreview.com/patient-experience/how-thinking-of-patients-as-customers-can-improve-healthcare.html>
- Berry, M. J. A. ve Linoff, G. S. (2004). *Data Mining Techniques for Marketing, Sales, and Customer Relationship Management* (Second Edition). Wiley Publishing Inc.
- Bilgin, M. (2018) *Makine Öğrenmesi*. 2. Basım. İstanbul, Papatya Yayıncılık.

- Blattberg, Kim ve Neslin, 2008. *Database marketing: Analyzing and managing customers*. Springer, s. 612.
- Bleuel, W. (2019). CSAT or CES: Does It Matter?. *Graziadio Business Review*. Volume 22 Issue 1, (1).
- Booms, B., ve Bitner, M. (1981). Marketing Strategies and Organisation Structures for Service Firms', In Marketing of Services: 1981 Special Educators 'Conference Proceedings, James Donnelly and William George, (Eds), Chicago, *American Marketing Association*, 46-51.
- Bounsaythip, C. ve Esa, R. R. (2001). "Overview of Data Mining For Customer Behavior Modeling", *VTT Information Technology Research Report*, Version:1, s. 1-53.
- Braun, M. ve Schweidel, D. A. (2011). Modeling customer lifetimes with multiple causes of churn. *Marketing Science*, 30(5), 881-902.
- Bravante, J. J. A. ve Robielos, R. A. C. (2022). Game Over: An Application of Customer Churn Prediction using Survival Analysis Modelling in Automobile Insurance, *IEOM Society International*.
- Breiman, L. (2001), Random forests. *Machine Learning*, 45, 5-32.
<https://doi.org/10.1023/A:1010933404324>
- Bruyckere, P. D., (2018), The difference between data, information, knowledge and wisdom. Retrieved from <https://theeconomyofmeaning.com/2018/06/29/the-difference-between-data-information-knowledge-and-wisdom/>
- Burez, J. ve Van den Poel, D. (2008). Separating financial from commercial customer churn: A modeling step towards resolving the conflict between the sales and credit department. *Expert Systems with Applications*, 35(1-2), 497-514.
- Buttle, F., Maklan, S. (2015), *Customer Relationship Management*. Routledge.
- Buttle, F. (2004). *Customer relationship management. Concepts and tools*. Oxford Elsevier.
- Büber, R., Başer, H. (2012). Sağlık İşletmelerinde Müşteri Memnuniyeti: Vakıf Üniversitesi Hastanesinde Bir Uygulama. *Sosyal ve Beşerî Bilimler Dergisi*, 1, 265-274. <https://dergipark.org.tr/tr/pub/sobiadsbd/issue/11353/135685>
- Büyüktanır, B., (2022), Du-Cba: Veriden Habersiz Ve Artırımlı Sınıflandırmaya Dayalı Birlikte Kuralları Çıkarma Mimarisi. [Yüksek Lisans Tezi]. Marmara Üniversitesi, Fen Bilimleri Enstitüsü.

- Capri, H. L. (2015). *Data Mining: Principles, Applications and Emerging Challenges*. Nova Science Publishers.
- Cengizci, A. D., (2020). *Otel İşletmelerinde Kayıp Müşteri Tahminlemesi*. [Doktora Tezi]. Akdeniz Üniversitesi.
- Chapa, O., Hernandez, M.D., Wang, Y.J. ve Skalski, C. (2014) Do Individualists Complain More than Collectivists? A Four-Country Analysis on Consumer Complaint Behavior, *Journal of International Consumer Marketing*, 26:5, 373-390, <https://doi.org/1080/08961530.2014.933089>
- Chapman, P., Clinton, J., Keber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R. (2000). *CRISP-DM 1.0 Step by step DM guide*. Edited by SPSS
- Chauhan, S., Saini, S., Bathla, R., Rana, A. (2020). *Application of Machine Learning to Predict Hospital Churning, 8th International Conference on Reliability*. Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO) Amity University, Noida, India.
- Chen, Z. Y., Fan, Z. P. ve Sun, M. (2012). A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data. *European Journal of Operational Research*. 223(2), 461-472. <https://doi.org/10.1016/j.ejor.2012.06.040>
- Chen, I. J., ve Popovich, K. (2003). Understanding customer relationship management (CRM) People, process and technology. *Business Process Management Journal*, 9(5), 672–688.
- Cissé, I. T. ve Yılmaz, Ö. (2022). Yaşlıları Günlük Yaşamlarında Destekleyici Mobil Sağlık Uygulaması Geliştirilmesi . *Avrupa Bilim ve Teknoloji Dergisi*. *Ejosat Özel Sayı 2022 (ICAENS-1)* , 601-609 . <https://doi.org/10.31590/ejosat.108364>
- Clifton, C., (2022), Data Mining, *Encyclopedia Britannica*, Retrieved from <https://www.britannica.com/technology/data-mining>
- Costa D., Mercieca-Bebber R., Tesson S., Seidler Z., Lopez A., (2019), Patient, client, consumer, survivor or other alternatives? A scoping review of preferred terms for labelling individuals who access healthcare across settings, <https://doi.org/10.1136/bmjopen-2018-025166>
- Çalış Boyacı, A., Durmaz, K. İ. ve Gencer, C. (2018). Uçak Seferlerindeki Rotaları Etkileyen Faktörlerin Analizi, *International Journal of Economic and Administrative Studies*, 18. EYİ Special Issue, 179-190.

- Çalış, A., Kayapınar, S. ve Çetinyokuş, T. (2014). Veri Madenciliğinde Karar Ağacı Algoritmaları İle Bilgisayar Ve İnternet Güvenliği Üzerine Bir Uygulama Endüstri Mühendisliği, 25 (3),2-19.Retrieved from <https://dergipark.org.tr/tr/pub/endustrimuhendisligi/issue/46771/586362> adresinden alınmıştır.
- Çalışkan, S.K. ve Soğukpınar, İ., (2008). "Kxknn: K-Means ve K-En Yakın Komşu Yöntemleri İle Ağlarda Nüfuz Tespiti", 2. Ağ ve Bilgi Güvenliği Sempozyumu kitabı içinde, Girne, 120–124.
- Çallı, L. ve Kasım, S.,(2022). "Using Machine Learning Algorithms to Analyze Customer Churn in the Software as a Service (SaaS) Industry",*Academic Platform Journal of Engineering and Smart Systems.vol. 10, no. 3*, pp. 115-123. <https://doi.org/10.21541/apjess.1139862>
- Çelik, U., Akçetin ve E., Gök, M., (2017).*Rapidminer ile Uygulamalı Veri Madenciliği*, Pusula Yayıncılık.
- Çeltikçi, E. (2018), A systematic review on machine learning in neurosurgery: The future of decision-making in patient care. *Turk Neurosurg.* 28(2). 67-173. <https://doi.org/10.5137/1019-5149.JTN.20059-17.1>
- Çınar, A. ve Silahtaroglu, G. (2015). Veri Madenciliği Teknikleri ile Müşteri Memnuniyetine Etki Eden Gizli Nedenlerin Keşfi . *Marmara Üniversitesi İktisadi ve İdari Bilimler Dergisi* , 33 (2) , 309-330 . Retrieved from <https://dergipark.org.tr/tr/pub/muiibd/issue/499/4516>.
- Çiçek, A., Arslan, Y. (2020). Müşteri Kayıp Analizi İçin Sınıflandırma Algoritmalarının Karşılaştırılması. *İleri Mühendislik Çalışmaları ve Teknolojileri Dergisi*.1(1),13–19. <https://dergipark.org.tr/en/pub/imctd/issue/56439/765347>
- Çiftçi, S., (2006), *Uzaktan Eğitimde Öğrencilerin Ders Çalışma Etkinliklerinin Log Verilerinin Analiz Edilerek İncelenmesi* .[Yüksek Lisans Tezi]. Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Çoğun, Y. H, Karacaçıl, Z., Kızılyıldırım S. (2022), *Doğa ve Mühendislik Bilimlerinde Güncel Yaklaşımlar*, sayfa 363-373.
- Çokluk, Ö. (2010). Lojistik regresyon analizi: Kavram ve uygulama. (Turkish). *Logistic Regression: Concept and Application.* 10(3), 1357-1407.
- Demir, F. O. ve Kırdar, Y. (2007). Müşteri İlişkileri Yönetimi: Crm. *Review Of Social,Economic & Business Studies*.(7/8), 293-308
- Demir, M. (2021). *Özellik Seçim Yöntemleri Kullanılarak Sınıflandırma Algoritmalarının Performanslarının Karşılaştırılması*.[Yüksek Lisans Tezi]. Afyon Kocatepe Üniversitesi,Fen Bilimleri Enstitüsü.

- Deng, Z. , Lu, Y., Wei, K. K. ve Zhang, J. (2010). Understanding customer satisfaction and loyalty: An empirical study of mobile instant messages in China. *International Journal of Information Management*, Elsevier, vol. 30(4), pages 289-300. <https://doi.org/10.1016/j.ijinfomgt.2009.10.001>
- Dick, A.S. ve Basu, K. (1994) Customer Loyalty: Toward an Integrated Conceptual Framework. *Journal of the Academy of Marketing Science*. 22, 99-113. <https://doi.org/10.1177/0092070394222001>
- Dolgun, M.Ö., (2014). *Veri Madenciliği Sınıflama Yöntemlerinin Başarılarının Bağımlı Değişken Prevelansı Örneklem Büyüklüğü ve Bağımsız Değişkenler Arası İlişki Yapısına Göre Karşılaştırılması*. [Doktora Tezi]. Hacettepe Üniversitesi Sağlık Bilimleri Enstitüsü, Ankara.
- Donat, H. (2021), *Öz Nitelik Seçimi İle Telekomünikasyon Sektöründe Kayıp Müşteri Analizi*. [Yüksek Lisans Tezi]. İstanbul Üniversitesi, Cerrahpaşa Lisansüstü Eğitim Enstitüsü.
- Donskoy, V. I. (2013). "Splitting criteria, binary decision tree synthesis, and algorithm LISTBB". *Taurida Journal of Computer Science Theory and Mathematics*, 1(22):11- 34.
- Dunham, M. H. (2003). *Data mining: Introductory and advanced topics*. Pearson Education.
- Dur, R. , Koçer, S. ve Dünder, Ö. (2022). Evaluation of Customer Loss Analysis for Marketing Campaigns in the Banking Sector. *Politeknik Dergisi*. 1-1. <https://doi.org/10.2339/politeknik.1036034>
- Duran, O., (2022). "Karar Ağacı Algoritmaları Nelerdir? <https://www.matematiksel.org/karar-agaci-algoritmaları-nelerdir/>
- Durmuş, B., (2020). Veri, Enformasyon (Malumat), Bilgi ve Bilgelik Kavramları. <https://webdeyazilim.com/veri-enformasyon-malumat-bilgi-bilgelik-kavramlari.html>
- DYCHE, Jill (2002). The CRM handbook: a business guide to customer relationship management. Addison Wesley Technology Series.
- Avuçlu, E. (2019). *Geleceğin Dünyasında Bilimsel ve Mesleki Çalışmalar*. Ekin Basım Yayın Dağıtım, Syf. 1-18.
- Emel, G. G., Taşkın, Ç. (2002). Genetik Algoritmalar ve Uygulama Alanları, Uludağ Üniversitesi, İktisadi ve İdari Bilim Fakültesi Dergisi, vol. 21, no. 1, pp. 129–152, 2002.

- Emel, G. G., Taşkın, Ç. (2005) .Veri Madenciliğinde Karar Ağaçları ve Bir Satış Analizi Uygulaması, *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi*, Cilt: 6 Sayı: 2.
- Ercan, U. (2016). *Veri Madenciliği ile Hane halkı Yağ Tüketiminin Modellenmesi*. [Doktora Tezi]. Akdeniz Üniversitesi, Sosyal Bilimler Enstitüsü.
- Ersöz, F. (2015). *Veri madenciliği teknikleri ve uygulamaları*. Seçkin Yayıncılık.
- Eşmekaya, E., Şeker, Ş. E. (2017). ETL Süreçleri (ETL Process), *YBS Ansiklopedi*, v. 4, is. 3, pp. 18- 34
- Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., (1996).Knowledge Discovery and Data Mining:Towards a Unifying Framework, *KDD-96 Proceedings*.AAAI. Syf.41
- Gaddam, L. ve Kadali, S.L.H. (2022). *Comparison of Machine learning algorithms on Predicting Churn within Music streaming service*. [Thesis in Computer Science]. Faculty of Computing, Blekinge Institute of Technology.
- Gallo, A. (2014). The Value of Keeping The Right Customers, *Harvard Business Review*. <https://hbr.org/2014/10/the-value-of-keeping-the-right-customers>
- Ganesh, J., Arnold M. J., ve Reynolds K. E. (2000). Understanding the Customer Base of Service Providers: An Examination of the Differences Between Switchers and Stayers. *Journal of Marketing*, 64 (July), 65-87.
- Gedleç, Ş., Yılmaz ve H. B., (2020) Karar Ağaçlarında Algoritma Seçimi, <https://www.datascienceearth.com/karar-agaclarinda-algoritma-secimi/>
- Genoveva, (2015), Anaylzing customer satisfaction and customer loyalty based on brand image and percieved service quality, *Journal of US-China Public Administration*, June 2015, Vol. 12, No. 6, 497-508
- Gorunescu, F. (2011). Data mining: Concepts, models and techniques. *Intelligent Systems Reference Library*.Springer.
- Gour, R. (2019). Unique Real-Life Applications of SVM. Retrieved from 06 January, 2023, from <https://medium.com/@rinu.gour123/8-unique-real-life-applications-of-svm-8a96ca43313>
- Gupta S, Lehmannand Dr. (2004). Valuing customers. *Journal of Marketing Research*. 41(1). 7-18.
- Gülpinar, V. (2013). Yapay Sinir Ağları Ve Sosyal Ağ Analizi Yardimi İle Türk Telekomünikasyon Piyasasında Müşteri Kaybı Analizi. *Marmara Üniversitesi*

Gülsevin, G. ve Türkan, A.H., (2012) "Evaluation of Efficiencies of Hospitals in Afyonkarahisar Using Data Envelopment Analysis", *Afyon Kocatepe University Journal of Sciences*, 12, 1-8.

Gürsoy, U. Ş. (2009), Customer churn analysis in telecommunication sector. *İstanbul Üniversitesi İşletme Fakültesi Dergisi. Cilt 39, Sayı 1, 35 - 49.*

H, Y. (2013). *Veri madenciliği yöntemleri*. Papatya Yayıncılık.

Hadden, J. (2008). *A customer profiling methodology for churn prediction*. [Yayımlanmamış Doktora Tezi]. School of Applied Sciences, University of Cranfield, England.

Han, J. ve Kamber, M. (2000) *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers.

Han, J., Kamber ve M., Pei, J. (2012), *Data mining: concepts and techniques*, Third Edition, Morgan Kaufmann Publishers.

Hansemark, O.C. ve Albinsson, M.E. (2004). Customer satisfaction and retention: the experiences of individual employees. *Managing Service Quality*, 14, 40-57.

Hastie, T., Tibshirani, R. ve Friedman, J. (2001). *The Elements of Statistical Learning*. Springer-Verlag. <https://doi.org/10.1007/978-0-387-21606-5>

Hidber, C., (1999). *Online Association Rule Mining*. ACM SIGMOD International Conference on Management of Data, s. (145-156).

Hofstede, G. (2001). *Culture's Consequences: Comparing Values, Behaviors. Institutions and Organizations Across Nations*. 2nd Edition, Sage Publications.

Hofstede, G. ve Minkov, M. (2010). *Cultures and Organizations: Software of the Mind*. 3rd Edition, McGraw.

Holland, J., (1992). *Adaptation in Natural and Artificial Systems*. University of Michigan Press, second edition.

Hung S.-Y., Hung W.-H., Tsai C.-A. ve Jiang S.-C. (2010). Critical factors of hospital adoption on CRM system: Organizational and information system perspectives. *Decis. Supp. Syst.* 48(4):592–603. retrieved from IBM.com (2023) Karar Ağaçları, <https://www.ibm.com/docs/tr/cloud-paks/cp-data/4.5.x?topic=palette-decision-tree-nodes>

- Iriana, R. ve Buttle, F. (2007). Strategic, operational, and analytical customer relationship management: attributes and measures. *Journal of Relationship Marketing*, (4), 23-42.
- Irmak, S., Köksal, C. D. ve Asilkan, Ö. (2012). Hastanelerin Gelecekteki Hasta Yoğunluklarının Veri Madenciliği Yöntemleri İle Tahmin Edilmesi . *Uluslararası Alanya İşletme Fakültesi Dergisi* , 4 (1) ,101-114 .
<https://dergipark.org.tr/tr/pub/uaifd/issue/21593/231863> adresinden alınmıştır.
- Jain, Dipak ve Siddhartha, S. S. (2002). Customer lifetime value research in marketing: A review and future directions. *Journal of interactive marketing* 16.2. 34-46
- Jill, A. (2014). *Harvard Business School*. Retrieved from
<https://hbr.org/2014/10/the-value-of-keeping-the-right-customers>
- Kaba, A, (2021). *Makine öğrenmesi Yöntemleri İle Faizsiz Finansman Sektöründe Müşteri Kayıp Tahmini: Churn Analizi*. [Yüksek Lisans Tezi]. Mimar Sinan Güzel Sanatlar Üniversitesi, Fen Bilimleri Enstitüsü.
- Karabatak, M., (2008). *Özellik Seçimi, Sınıflama ve Öngörü Uygulamalarına Yönelik Birliktelik Kuralı Çıkarımı ve Yazılım Geliştirilmesi*. [Doktora Tezi]. Fırat Üniversitesi Fen Bilimleri Enstitüsü.
- Karabulut, D. (2021). *Hastane Bilgi Yönetim Sistemlerinde Veri Madenciliği: Hasta Profil Tahmini*. [Yüksek Lisans Tezi]. Karabük Üniversitesi.
- Kasım, S. (2022). *Veri Madenciliği Yöntemleriyle Müşteri Kaybı Analizi: Yazılım Sektörü*. [Yüksek Lisans Tezi]. Sakarya Üniversitesi, Fen Bilimleri Enstitüsü.
- Kavuncubaşı, Ş., (2000). *Hastane ve Sağlık Kurumları Yönetimi*. 1. Baskı, Siyasal Yayınevi, sayfa: 18-56.
- Kılıç Depren, S., Aşkın, Ö. E., ve Öz, E., (2017). *Comparing Data Mining Methods for Classification of Students Science Success* . END 2017. (pp.47-51).
- Kırım, A., (2001). *Strateji ve Birebir Pazarlama CRM*. Sistem Yayıncılık.
- Kısa, A., ve Tokgöz, N., (2007). *Sağlık Kurumları Yönetimi*, Anadolu Üniversitesi Açıköğretim Yayınları. 323 s.
- Kilimci Z. H., (2022). The Effectiveness of Homogeneous Classifier Ensembles on Customer Churn Prediction in Banking, Insurance and Telecommunication Sectors. *International Journal of Computational and Experimental Science and Engineering(IJCESEN)* Vol. 8-No.3, pp. 77-84.

- King, S. F. ve Burgess, T. F. (2007). Understanding success and failure in customer relationship management. *Industrial Marketing Management*, 37(4), 421–431.
- Kişioğlu, Pınar ve Y. İlker Topçu (2011). Applying Bayesian Belief Network Approach to Customer Churn Analysis: A Case Study on the Telecom Industry of Turkey. *Expert Systems with Applications*, 38, s.7150-7157. <https://doi.org/10.1016/j.eswa.2010.12.045>
- Koçdar, S. (2018). Bilgi Toplumu ve e-Devlet.Y. Güney & M. R. Okur (Ed.). *Bilgi Toplumu* (2-26) içinde . Anadolu Üniversitesi
- Kotler, P. (2000). *Marketing Management: The Millennium Edition*. Person Prentice Hall.
- Kotler, P. ve Armstrong, G. (2017). *Principles of marketing. (14th Edition)*, ISBN-13: 978-0-13-216712-3, Pearson.
- Kotler, P., Armstrong, G., Saunders, J. ve Wong, V. (2002). *Principles of Marketing*. 3rd European Edition, Prentice Hall.
- Kotler, P. ve Keller, K. L.(2009). *Marketing Management*. Pearson Prentice Hall, 182 syf.
- Kotler, P. ve Keller, K.L. (2005).*Marketing Management*. 12th Edition, Prentice-Hall of India Private Limited.
- Koyuncugil, A. ve Özgülbaş, N. (2010). Veri Madenciliği: Tıp ve Sağlık Hizmetlerinde Kullanımı ve Uygulamaları . *Bilişim Teknolojileri Dergisi* , 2 (2). <https://dergipark.org.tr/tr/pub/gazibtd/issue/6616/87893>
- Kumar, V. ve Reinartz, W. (2018). *Customer relationship management: Concept, strategy, and tools*. Springer. s. 17
- Kutlu, B. ve Badur, B. (2009). Yapay sinir ağları ile borsa endeksi tahmini. *Yönetim Dergisi*, 20 (63), 45- 40.
- Kuzey, C.,(2012). *Veri Madenciliğinde Destek Vektör Makinaları ve Karar Ağaçları Yöntemlerini Kullanarak Bilgi Çalışanlarının Kurum Performansı Üzerine Etkisinin Ölçülmesi ve Bir Uygulama*. [Doktora Tezi]. İstanbul Üniversitesi Sosyal Bilimler Enstitüsü.
- Kwoon H., Kim H. H., An J., Lee J. ve Park Y. R., (2021). Lifelog Data-Based Prediction Model of Digital Health Care App Customer Churn: Retrospective Observational Study. *J Med Internet Res*; 23(1):e22184
- Larose, D. T. (2006). *Data mining methods and models*.Wiley-IEEE Press.

Lezki, Ş. (2019). Temel Kavramlar. İçinde F. Er (Ed.), *Veri Madenciliği*, (2-31): Anadolu Üniversitesi.

Majava, J., Nuottila, J., Haapasalo, H. ve Kris M. Y. Law, (2013). Customer Definition and Representation in Market-Driven Product Development. Diversity, Technology, and Innovation for Operational Competitiveness: Proceedings of the 2013 International Conference on Technology Innovation and Industrial Management.

Martinez-Plumed F., Contreras-Ochando, L., Ferri, C., Hernandez-Orallo, J., Kull, M., Lachiche N., Mar'ia Jose Ramirez-Quintana, M. ve J., Flach, P. (2019) CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories, *Ieee Transactions On Knowledge And Data Engineering*. <https://doi.org/10.1109/TKDE.2019.2962680>.

McCarthy, E. J. (1964). *Basic marketing: A managerial approach*. Irwin Inc.

McDonald M. (2012), *Market Segmentation: How to do it and How to profit from it?* 4th Ed. John Wiley & Sons Ltd.

Meltzer, E. (2022). 45+ Patient Retention and Churn Rate Statistics. Retrieved from <https://etactics.com/blog/patient-retention-statistics>

Minsky, M., & Papert, S. (1969). *Perceptrons*. M.I.T. Press.

Mohammadzadeh, M., Hoseini, Z. Z. ve Derafshi, H. (2017). A data mining approach for modeling churn behavior via RFM model in specialized clinics Case study: A public sector hospital in Tehran. *Procedia computer science*, 120, 23–30. <https://doi.org/10.1016/j.procs.2017.11.206>.

Morgan, N. A. ve Rigo, L. L., (2006). The value of different customer satisfaction and loyalty metrics in predicting business performance. *Marketing Science*, 25(5), 426-439.

Müller, A. C. ve Guido, S. (2017). *Introduction to machine learning with Python: A guide for data scientists*. Published by O'Reilly Media.

Nettleton D. (2014). Credit card churn forecasting by logistic regression and decision tree. *Expert Systems Applications*. 38(12). 15273–15285. <https://doi.org/10.1016/j.eswa.2011.06.028>

Nielsen, (2012). Nielsen Global Trust in Advertising Survey. Retrieved from <https://www.nielsen.com/us/en/insights/article/2012/consumer-trust-in-online-social-and-mobile-advertising-grows/> .

Nisbet, R., Elder, J. ve Miner, G. (2009). *Handbook of statistical analysis and data mining applications*. Academic Press Elsevier.

Odabaşı, Y. (2000). *Satışta ve Pazarlamada Müşteri İlişkileri Yönetimi*. Sistem Yayıncılık.

Odabaşı, Y. ve Barış, G. (2002). *Tüketici Davranışı*. Mediacat.

Oliveira, V. L. M. (2012). *Analytical customer relationship management in retailing supported by data mining techniques*. [Yayınlanmamış Doktora Tezi]. University of Porto, Portekiz.

Olson, D.L. ve Delen, D. (2008). Advanced Data Mining Techniques. *Computer Science*. <https://doi.org/10.5860/choice.45-6838>

Özcan, C., (2014). *Veri Madenciliğinin Güvenlik Uygulama Alanları ve Veri Madenciliği ile Sahtekârlık Analizi*. [Yüksek Lisans Tezi]. İstanbul Bilgi Üniversitesi.

Özdemir, A. , Yalçın Aslay, F. ve Çam, H. (2010). Veri Tabanında Bilgi Keşfi Süreci: Gümüşhane Devlet Hastanesi Uygulaması . *Sosyal Ekonomik Araştırmalar Dergisi* , 10 (20) , 347-366 . <https://Dergipark.Org.Tr/Tr/Pub/Susead/Issue/28415/302475> .

Özekes, S. (2003). Veri Madenciliği Modelleri ve Uygulama Alanları . *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi* , 2 (3) , 65-82 . <https://dergipark.org.tr/tr/pub/ticaretfbid/issue/21343/22895>.

Öztürk, K. ve Şahin, M. E. (2018). Yapay Sinir Ağları ve Yapay Zekâ'ya Genel Bir Bakış. *Takvim-i Vekayi* , 6 (2) , 25-36.

Öztürk, Ü. (2014). *Lojistikte Fiyatlandırmayı İyileştirme Amaçlı Olarak Veri Madenciliği Teknikleri İle Bir Öneri*. [Yayınlanmamış Yüksek Lisans Tezi]. İstanbul Beykent Üniversitesi, Fen Bilimleri Enstitüsü.

Özyazar, Ö. (2019). *Sürdürülebilir Tıbbi Sistemler İçin Veri Madenciliği Uygulamaları: Diyabet Üzerine Bir Çalışma*. [Yüksek Lisans Tezi]. Marmara Üniversitesi.

Palmer, A., Gervilla, E. ve Jiménez, R. (2011). *Data mining: machine learning and statistical techniques*. INTECH Open Access Publisher.

Payne, (1993). *The Essence of Service Marketing*. Prentice Hall.

Pehlivan, G. (2006). *Chaid Analizi Ve Bir Uygulama*. [Yüksek Lisans Tezi]. Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü.

Pejić B. M., Pivar, J. ve Jaković, B. (2021). Churn Management in Telecommunications: Hybrid Approach Using Cluster Analysis and Decision Trees. *Journal of Risk and Financial Management*, 14(11), 544.

Pekmezci, T. Demireli, C. ve Batman, G. (2008). İç Müşteri Memnuniyeti: Konya Un Fabrikalarında Bir Uygulama. *Dumlupınar Üniversitesi Sosyal*

Bilimler Dergisi, 22,1-16.

Peppers, D. ve Rogers, M. (2016). *Managing Customer Relationships: A Strategic Framework*. Third Edition. Wiley.

Porter M. ve Kaplan R. (2016). How to Pay for Health Care, Value Based Healthcare, HBR, (accessed of date April 10th, 2022). Retrieved from <https://www.isc.hbs.edu/health-care/value-based-health-care/Pages/publications.aspx>.

Prasad, U. D. (2012). Prediction Of Churn Behavior Of Bank Customers. *Business Intelligence Journal*, 5(1), 96-101.

Provost, F. ve Fawcett, T. (2013). *Data Science for Business: What you need to know about data mining and data-analytic thinking*. O'Reilly Media, Inc., syf.23-25.

Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers Inc.

Reddy, M. G. A., Raghavaraju, S. ve Lashyry, P. (2022), "Ensemble Approach on the Online Shopping Churn Prediction," *2022 6th International Conference on Trends in Electronics and Informatics (ICOEI)*, pp. 01-08, <https://doi.org/10.1109/COEI53556.2022.9776921>.

Reichheld, F. (2001). *Harvard Business School*. Publishing PQDT. Retrieved from https://media.bain.com/Images/BB_Prescription_cutting_costs.pdf

Reichheld, F. F, Darnell, D. ve Burns, M. (2021) Net Promoter 3.0: A better system for understanding the real value of happy customers. *Harvard Business Review*, November-December.

Reichheld, F. F. (2003). The One Number You Need To Grow. *Harvard Business Review*, 81(12), 46-55

Reichheld, F. F. ve Teal, T. (2001). *The loyalty effect: The hidden force behind growth, profits, and lasting value*. Massachusetts: Harvard Business School Press. s. 88

Reichheld, F. F. ve Teal, T. (1996). *The Loyalty Effect the Hidden Force behind Growth, Profits, and Lasting Value*. MA Harvard Business School.

Reichheld, F.F. ve Sasser, E. (1990). Zero Defections: Quality Comes to Services. *Harvard Business Review*, 68, 105-111

Rokach, L. ve Maimon, O. (2014). *Data mining with decision trees: theory and applications*. (2nd edition). World Scientific Publishing Co. Pte. Ltd.

Rokach, L. ve Maimon, O. (2008). *Data Mining with Decision Trees. Theory and Applications*. World Scientific.

Ruby Newell-Legner (2015), "Understanding Customers" The Future of Marketing 2016: New Roles, and Trends. *Harvard Business School*. Nov. 30.

Ruggieri, S. (2002) Efficient c4.5 [classification algorithm]. *IEEE Transactions on Knowledge and Data Engineering*, vol. 14, no. 2, pp. 438–444.

Russell, L. ve Ackoff, R. L., (1989) From Data to Wisdom. *Journal of Applied Systems Analysis* 16. 3-9.

Saputra, D. A., Handayani, P. W. ve Satria, R. (2019). Customer relationship management (CRM) implementation evaluation using maturity assessment in telecommunication industry: Case study of an Indonesian company. 2019 International Conference on Advanced Computer Science and Information Systems (ICACSIS). Institute of Electrical and Electronics Engineers Inc. Oct 2019. 513-518.

Sarıman, G. (2014). Veri Madenciliğinde Kümeleme Teknikleri Üzerine Bir Çalışma: K-Means ve K-Medoids Kümeleme Algoritmalarının Karşılaştırılması. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 15 (3, 192-202.
<https://dergipark.org.tr/tr/pub/sdufenbed/issue/20796/222134>
adresinden alınmıştır.

Scholkopf, B. (2018). *Learning With Kernels: Support Vector Machines, Regularization, Optimization, And Beyond*. MIT Press.
ISBN 0-262-53657-9. Ocllc 1039411838.

Schmid, H. (2013). Probabilistic part-of-speech tagging using decision trees. In *New methods in language processing*. Routledge, page 154

Sebik, N. B. ve Bülbül, H. I., (2018). Veri Madenciliği Modellerinin Akciğer Kanseri Veri Seti Üzerinden Başarılarının İncelenmesi. *Tübav Bilim Dergisi*, Cilt:11, Sayı:3, Sayfa: 1-7.

Seid, M.H. ve Woldeyohannis, M.M. (2022). Customer Churn Prediction Using Machine Learning: Commercial Bank of Ethiopia. 2022 International Conference on Information and Communication Technology for Development for Africa (ICT4DA). pp. 1-6,
<https://doi.org/10.1109/ICT4DA56482.2022.9971224>.

Sernovitz, A. (2012). Word of Mouth Marketing: How Smart Companies Get People Talking. 3rd ed. Greenleaf Book Group Press.

Seymen, Ö., F., (2022). Yeni Bir Müşteri Kayıp Yönetimi Modeli: Perakende Sektörü Uygulaması. [Doktora Tezi]. Sakarya Üniversitesi, Syf 32.

Sezer, E. A., Bozkır, A. S., Yağız, S. ve Gökçeoğlu C. (2010). Karar Ağacı Derinliğinin CART Algoritmasında Kestirim Kapasitesine Etkisi: Bir Tünel Açma Makinesinin İlerleme Hızı Üzerinde Uygulama. Akıllı Sistemlerde Yenilikler ve Uygulamaları Sempozyumu, Kayseri.

Shah. S. ve Kursak. A., (2004). Data mining and genetic algorithms based gene/SNP selection. Artificial Intelligence in Medicine, 31. 183 -196.

Simmons, R (2014). Choosing the right customer. Harvard Business Review Dergisi, Mart Sayısı.

Singh, K.N., Mantri, J.K. ve Kakulpati, V. (2023). Churn Prediction of Clinical Decision Support Recommender System. In: Swarnkar, T., Patnaik, S., Mitra, P., Misra, S., Mishra, M. (eds) Ambient Intelligence in Health Care. Smart Innovation, Systems and Technologies, vol 317. https://doi.org/10.1007/978-981-19-6068-0_36

Soysal, A. , Doğan, S. ve Baynal, T. (2017). Özel Sağlık Kurumlarında Müşteri İlişkileri Yönetimi. Kahramanmaraş Sütçü İmam Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi , 7 (1) , 39-66 .

Suner, A. ve Demirarslan, M. (2021). Sağlık Veri Setlerinde Öz nitelik Seçiminin Sınıflandırma Performansına Etkisi. Journal of Artificial Intelligence in Health Sciences, 1(1), 6–11. <https://doi.org/10.52309/jai.2021.2>

Sülekli, H.E., (2019). Yoğun Bakım Ünitelerinde Yatan Hastalara İlişkin Mortalite ve Yatış Süresine Etki Eden Faktörlerin Veri Madenciliği Yöntemleri ile İncelenmesi. [Yüksek Lisans Tezi]. Hacettepe Üniversitesi, Sağlık Bilimleri Enstitüsü, Ankara.

Şeker, Ş. E., (2016), Müşteri Kayıp Analizi (Customer Churn Analysis) http://ybsansiklopedi.com/wp-content/uploads/2016/06/musteri_kayip.pdf.

Şeker, Ş.E, (2018) CRISP-DM: Endüstriler Arası Standart İşleme – Veri Madenciliği için (Cross Industry Standard Processing – Data Mining), YBS Ansiklopedi, Cilt 5, Sayı 2.

Şenyürek, A. (2019). Telekomünikasyon Sektöründe Aboneliklerini İptal Edecek Müşterilerin Yapay Öğrenme Yöntemleri İle Tahmin Edilmesi. [Yüksek Lisans Tezi]. Yıldız Teknik Üniversitesi, İstanbul.

Tarcan, E. (2001). Hizmet Yönetiminde Kalite ve Müşteri Tatmini Ölçümü.

İstanbul Üniversitesi Yayınları, No:4288, 2001

Taya, Ö. (2022). Destek Vektör Makineleri ve Lojistik Regresyon Yöntemi İle Sınıflandırma: Türkiye'in İthalat ve İhracat Örneği. [Yüksek Lisans Tezi]. İnönü Üniversitesi Sosyal Bilimler Enstitüsü.

Tokay, G. (2019). Müşteri İlişkileri Yönetiminde Veri Tabanlı Pazarlamaya Dair Bir Araştırma. [Yüksek Lisans Tezi]. Bahçeşehir Üniversitesi SBE, İstanbul.

Tolon, M. (2007). Tüketici tatmininin yapay sinir ağları yöntemiyle ölçülmesi ve Ankara'daki perakendeci mağazaların müşterileri üzerinde bir uygulama. [Yayınlanmamış Doktora Tezi]. Gazi Üniversitesi Sosyal Bilimler Enstitüsü İşletme Anabilim Dalı: Ankara

Tontuş, H. (2020). Sarf Malzeme Kullanımından Veri Madenciliği Birliktelik Kurallarının Elde Edilmesi, Kuralların Analizi Ve Sınıflandırılması. [Yüksek Lisans Tezi]. Gazi Üniversitesi, Fen Bilimleri Enstitüsü.

Türk, Y. (2005). Perakendeci İşletmelerde Personelin Davranışsal Özellikleri İle Müşteri Memnuniyeti Arasındaki İlişki . Süleyman Demirel Üniversitesi İktisadi Ve İdari Bilimler Fakültesi Dergisi , 10 (1) , 195-219.

Türker, A. K., Göleç, A., Aktepe, A., Ersöz, S., İpek, M. ve Çağıl, G. (2019). Siparişlerin gecikme tahmini için veri madenciliğine dayalı gerçek zamanlı bir sistem tasarımı ve uygulaması. Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi, 35 (2), 709-724.<https://doi.org/10.17341/gazimmfd.478648>.

UKessays, (2018). What is churn? An overview. [online]. Retrieved from: <https://www.ukessays.com/essays/statistics/churn.php?vref=1>. [Accessed 10 April 2022].

Uncuoğlu, E. (2018). Destek Vektör Makinaları Kullanarak Kişisel Termal Konfor Modellemesi. [Yüksek Lisans Tezi]. Pamukkale Üniversitesi Fen Bilimleri Enstitüsü.

Uzun, M. ve Özgöz, A. (2022). Müşteri Şikayetleri Yönetiminin Müşteri Memnuniyeti Üzerine Etkisi . 19 Mayıs Sosyal Bilimler Dergisi, 3 (3), 231-249.<https://doi.org/10.52835/19maysbd.1067740>

Verma, A. K., Pal, S. ve Kumar, S. (2019). Classification of Skin Disease using Ensemble Data Mining Techniques. Asian Pacific journal of cancer prevention:APJCP, 20(6),1887–1894. <https://doi.org/10.31557/APJCP.2019.20.6.1887>

Yakut, E. (2012). Veri Madenciliği Tekniklerinden C5.0 Algoritması Ve Destek Vektör Makineleri İle Yapay Sinir Ağlarının Sınıflandırma Başarılarının Karşılaştırılması: İmalat Sektöründe Bir Uygulama. [Doktora Tezi]. Atatürk Üniversitesi Sosyal Bilimler Enstitüsü.

- Yetiz, F. , Terzioğlu, M. ve Kayakuş, M. (2021). Makina Öğrenmesi Yöntemleri ile Türk Mevduat Bankalarının Müşteri Tahminine Yönelik Bir Uygulama. Sosyoekonomi , 29(50) , 413-432
- Yıldız, M. ve Şeker, Ş., E., (2016), Veri Madenciliği Araçları (Data Mining Tools), YBS Ansiklopedi, Cilt 3, Sayı: 4. Sayfa 12-19.
- Yohannes Y. ve Hoddinott J., (2018). Classification and Regression Trees: An Introduction Technical Guide. International Food Policy Research Institute, Mart 1999, S.2, Classification and Regression Trees an Introduction.
- Yörük, R. ve Eşmekaya, E. (2018). Müşteri Yaşam Boyu Değeri. YBS ansiklopedi, cilt 5, sayı 2.
- Zeleny, M. (1987). Management Support Systems: Towards Integrated Knowledge Management. Human Systems Management, 7, 59-70.
- Zygiaris, S. (2018). Database Management Systems: A Business – Oriented Approach Using Oracle, MYSQL, and Ms Access. (First Edit). Emerald Pulishing Limited.

EKLER

EK 1 Çocuk Hastalıkları Hekimleri İçin Görüşme Formu

Müşteri Kayıp Analizi, genellikle telekomünikasyon, bankacılık ve sigortacılık gibi müşteri sürekliliği bulunan sektörlerde uygulanmaktadır. Temelde, müşterilerin davranışlarının analizi ile müşteri kaybını tahmin etmeye dayalı bir yaklaşımdır. Müşteri Kayıp Analizi uygulanan işletmelerde kaybı muhtemel olan müşteri, yapılan analiz ile saptanarak çeşitli tutundurma faaliyetleri ile hizmet alımının devamı amaçlanmaktadır.

Bu doğrultuda, çocuk hastaların kuruluşunuzun hastası kabul edilme ölçütleri, yaşlarına göre geliş sıklıklarının değerlendirilmesi ve kayıp hasta olma ölçütleri büyük veri içinden yapılacak analiz ile birlikte hekimlerin görüşleri ile karşılaştırılacak ve yaşlara göre geliş sıklıkları ile kayıp hasta olma durumu incelenecektir. Çalışmanın devamında çocuk hastaların gelişlerine ait sistemde şikâyet olup olmaması, tekrar eden şikâyetin olup olmaması, şikâyet alt başlıkları ve varsa uygulanan anket sorularına verilen cevaplar ile kayıp hasta olma durumu (hastaların sosyal güvencelerine göre) analiz edilecektir.

Aşağıdaki soruların Covid-19 etkisi ile değişen ebeveyn taleplerinden (online görüşme, whatsapp mesajlaşma vs) bağımsız olarak değerlendirilmesi gerekmektedir.

Sorular

- 1) Kliniğe başvuran bir hastayı kaç geliş sonrası kendi hastamız olarak kabul edebiliriz?
 - a) Bir poliklinik muayene (1 geliş)
 - b) İki poliklinik bir kontrol muayene (2 geliş)
 - c) Üç poliklinik muayene (3 geliş)
 - d) Diğer (Açıklayınız).....
- 2) Çocuk hastaları geliş sıklıkları açısından yaş gruplarına göre nasıl sınıflandırırsınız?
 - a) 0-1 yaş, 1-3 yaş, 3-7 yaş, 7-12 yaş, 12 yaş ve üzeri
 - b) 0-2 yaş, 2-5 yaş, 5- 10 yaş, 10 yaş üzeri
 - c) 0-1 yaş, 1-3 yaş, 3-6 yaş, 6-10 yaş, 10 yaş ve üzeri
 - d) Diğer (Açıklayınız)
- 3) Belirlediğiniz gruplarıdaki çocukların yaş aralıklarına göre tahmini yıllık geliş sıklıklarını nasıl değerlendirirsiniz? (Geliş adedi: Poliklinik muayene + kontrol muayene)
 - a) 0-1 yaş: kez, 1-3 yaş: kez, 3-7 yaş: kez, 7-12 yaş: kez, 12 yaş ve üzeri: kez,
 - b) 0-2 yaş: kez, 2-5 yaş: kez, 5-10 yaş: kez, 10 yaş ve üzeri: kez,
 - c) 0-1 yaş: kez, 1-3 yaş: kez, 3-6 yaş: kez, 6-10 yaş: kez, 10 yaş ve üzeri: kez,
 - d) Diğer (Açıklayınız).....
- 4) Belirlenen yaş aralıklarına göre ne kadar süre sonra bir hastamızı kayıp/hizmet sunucusu değiştirmiş olarak değerlendirebiliriz?
 - a) 0-1 yaş: .. ay, 1-3 yaş: ... ay, 3- 7 yaş: .. ay, 7-12 yaş:.... ay, 12 yaş ve üzeri:.. yıl
 - b) 0-2 yaş: ... ay, 2-5 yaş: ay, 5-10 yaş: ay, 10 yaş ve üzeri: yıl
 - c) 0-1 yaş: ...ay, 1-3 yaş: ay, 3- 6 yaş: . ay, 6-10 yaş: .. ay, 10 yaş ve üzeri: ..yıl
 - d) Diğer (Açıklayınız).....

EK 2 Geliş Sıklıkları ve MKA'da Kullanılan Değişkenler

Başlık	Anlamı
Hastano	Benzersiz Tanımlayıcı / Anahtar
Dogum Tarihi	Hastanın Doğum Tarihi
Son_Gelis_Sube	Başvurunun Olduğu Son Şube
Son_Gelis_Sube_Istanbul	Başvurunun Olduğu Son Şubenin Kategorisi-Istanbul
Son_Gelis_Sube_Anadolu	Başvurunun Olduğu Son Şubenin Kategorisi-Anadolu
Son_Kabul_Kurumu_Grubu	Son Kullanılan Sosyal Güvence Grubu (Nakit, Özel Sigorta, Sgk Vs.)
Son_Gelis_Tipi	Son Başvurunun Tipi (Ayaktan, Yatan, Günübirlik)
Uyruk	Uyruk Belirteci
Son_Adres_Ili	Son Başvurudaki Hastanın Adresinin Kayıtlı Olduğu İl
Pol_Mua_Basvuru_Tarihi	Poliklinik Muayene Başvuru Tarihi
Kontrol_Mua_Basvuru_Tarihi	Kontrol Muayene Başvuru Tarihi
Diger_Basvuru_Tarihi	Poliklinik Muayene Ve Kontrol Muayene Dışındaki Diğer Ayaktan Ya Da Yatan Başvurular
SKK_Bankalar	Sosyal Güvence Grubu Belirteci (Bankalar)
SKK_Ozel_Sigortalar	Sosyal Güvence Grubu Belirteci (Özel Sigortalar)
SKK_Kurumsuz_Hasta	Sosyal Güvence Grubu Belirteci (Kurumsuz/Nakit/Bireysel Hastalar)
SKK_Yurtdisi_Bireysel_Hastalar	Sosyal Güvence Grubu Belirteci (Yabancı Bireysel Hastalar)
SKK_Buyukelcilik_Ve_Konsolosluklar	Sosyal Güvence Grubu Belirteci (Büyükelçilik Ve Konsolosluklar)
SKK_Diger_Kurumlar	Sosyal Güvence Grubu Belirteci (Diğer Anlaşmalı Kurumlar/Firmalar)
SKK_Personel_Yakınları	Sosyal Güvence Grubu Belirteci (Personel Yakınları)
SKK_Yurtdisi_Sigortalar	Sosyal Güvence Grubu Belirteci (Yurtdışı Özel Sigortalar)
SKK_Sosyal_Guvenlik_Kurumu	Sosyal Güvence Grubu Belirteci (Sosyal Güvenlik Kurumu)
Uyruk_Turk	Türk Uyruklular
Uyruk_Yabancı	Yabancı Uyruklular
GA_0-1_Ay_Diger_Adet	0-1 Ay Aralığındaki Çocuk Hastanın Poliklinik Ve Kontrol Muayene Dışındaki Başvuru Adetleri
GA_0-1_Ay_Kontrol_Adet	0-1 Ay Aralığındaki Çocuk Hastanın Kontrol Muayene Adedi
GA_0-1_Ay_Poliklinik_Adet	0-1 Ay Aralığındaki Çocuk Hastanın Poliklinik Muayene Adedi

GA_1-12_Ay_Diger_Adet	1-12 Ay Aralığındaki Çocuk Hastanın Poliklinik Ve Kontrol Muayene Dışındaki Başvuru Adetleri
GA_1-12_Ay_Kontrol_Adet	1-12 Ay Aralığındaki Çocuk Hastanın Kontrol Muayene Adedi
GA_1-12_Ay_Poliklinik_Adet	1- 12 Ay Aralığındaki Çocuk Hastanın Poliklinik Muayene Adedi
GA_1-2_Yas_Diger_Adet	1-2 Yaş Aralığındaki Çocuk Hastanın Poliklinik Ve Kontrol Muayene Dışındaki Başvuru Adetleri
GA_1-2_Yas_Kontrol_Adet	1-2 Yaş Aralığındaki Çocuk Hastanın Kontrol Muayene Adedi
GA_1-2_Yas_Poliklinik_Adet	1- 2 Yaş Aralığındaki Çocuk Hastanın Poliklinik Muayene Adedi
GA_2-3_Yas_Diger_Adet	2-3 Yaş Aralığındaki Çocuk Hastanın Poliklinik Ve Kontrol Muayene Dışındaki Başvuru Adetleri
GA_2-3_Yas_Kontrol_Adet	2-3 Yaş Aralığındaki Çocuk Hastanın Kontrol Muayene Adedi
GA_2-3_Yas_Poliklinik_Adet	2- 3 Yaş Aralığındaki Çocuk Hastanın Poliklinik Muayene Adedi
GA_3-4_Yas_Diger_Adet	3-4 Yaş Aralığındaki Çocuk Hastanın Poliklinik Ve Kontrol Muayene Dışındaki Başvuru Adetleri
GA_3-4_Yas_Kontrol_Adet	3-4 Yaş Aralığındaki Çocuk Hastanın Kontrol Muayene Adedi
GA_3-4_Yas_Poliklinik_Adet	3-4 Yaş Aralığındaki Çocuk Hastanın Poliklinik Muayene Adedi
GA_4-5_Yas_Diger_Adet	4-5 Yaş Aralığındaki Çocuk Hastanın Poliklinik Ve Kontrol Muayene Dışındaki Başvuru Adetleri
GA_4-5_Yas_Kontrol_Adet	4-5 Yaş Aralığındaki Çocuk Hastanın Kontrol Muayene Adedi
GA_4-5_Yas_Poliklinik_Adet	4-5 Yaş Aralığındaki Çocuk Hastanın Poliklinik Muayene Adedi
GA_5-6_Yas_Diger_Adet	5-6 Yaş Aralığındaki Çocuk Hastanın Poliklinik Ve Kontrol Muayene Dışındaki Başvuru Adetleri
GA_5-6_Yas_Kontrol_Adet	5-6 Yaş Aralığındaki Çocuk Hastanın Kontrol Muayene Adedi
GA_5-6_Yas_Poliklinik_Adet	5-6 Yaş Aralığındaki Çocuk Hastanın Poliklinik Muayene Adedi
GA_6-7_Yas_Diger_Adet	6-7 Yaş Aralığındaki Çocuk Hastanın Poliklinik Ve Kontrol Muayene Dışındaki Başvuru Adetleri
GA_6-7_Yas_Kontrol_Adet	6-7 Yaş Aralığındaki Çocuk Hastanın Kontrol Muayene Adedi
GA_6-7_Yas_Poliklinik_Adet	6-7 Yaş Aralığındaki Çocuk Hastanın Poliklinik Muayene Adedi
GA_7-8_Yas_Diger_Adet	7-8 Yaş Aralığındaki Çocuk Hastanın Poliklinik Ve Kontrol Muayene Dışındaki Başvuru Adetleri

GA_7-8_Yas_Kontrol_Adet	7-8 Yaş Aralığındaki Çocuk Hastanın Kontrol Muayene Adedi
GA_7-8_Yas_Poliklinik_Adet	7-8 Yaş Aralığındaki Çocuk Hastanın Poliklinik Muayene Adedi
GA_8-9_Yas_Diger_Adet	8-9 Yaş Aralığındaki Çocuk Hastanın Poliklinik Ve Kontrol Muayene Dışındaki Başvuru Adetleri
GA_8-9_Yas_Kontrol_Adet	8-9 Yaş Aralığındaki Çocuk Hastanın Kontrol Muayene Adedi
GA_8-9_Yas_Poliklinik_Adet	8-9 Yaş Aralığındaki Çocuk Hastanın Poliklinik Muayene Adedi
GA_9-10_Yas_Diger_Adet	9-10 Yaş Aralığındaki Çocuk Hastanın Poliklinik Ve Kontrol Muayene Dışındaki Başvuru Adetleri
GA_9-10_Yas_Kontrol_Adet	9-10 Yaş Aralığındaki Çocuk Hastanın Kontrol Muayene Adedi
GA_9-10_Yas_Poliklinik_Adet	9-10 Yaş Aralığındaki Çocuk Hastanın Poliklinik Muayene Adedi
GA_Toplam_Gelis_Adedi	Toplam Geliş Adedi (Poliklinik, Kontrol Ve Diğer)
Son-Sondan2_Arasi_Gun_Sayisi	Son Geliş - 1 Önceki Geliş Gün Farkı
Sondan2-Sondan3_Arasi_Gun_Sayisi	Bir Önceki - 2 Önceki Geliş Gün Farkı
Sondan3-Sondan4_Arasi_Gun_Sayisi	Sondan 2 Önce Geliş - 3 Önceki Geliş Gün Farkı
Sondan4-Sondan5_Arasi_Gun_Sayisi	Sondan 3 Önce Geliş - 4 Önceki Geliş Gün Farkı
Sondan5-Sondan6_Arasi_Gun_Sayisi	Sondan 4 Önce Geliş - 5 Önceki Geliş Gün Farkı
Sondan6-Sondan7_Arasi_Gun_Sayisi	Sondan 5 Önce Geliş - 6 Önceki Geliş Gün Farkı
Sondan7-Sondan8_Arasi_Gun_Sayisi	Sondan 6 Önce Geliş - 7 Önceki Geliş Gün Farkı
Sondan8-Sondan9_Arasi_Gun_Sayisi	Sondan 7 Önce Geliş - 8 Önceki Geliş Gün Farkı
Sondan9-Sondan10_Arasi_Gun_Sayisi	Sondan 8 Önce Geliş - 9 Önceki Geliş Gün Farkı
GS_0-1_Yas_Gelis	0-1 Yaş Toplam Geliş Sayısı
GS_0-1_Grubu	Geliş Sayısına Göre Segmenti (A,B,C, N/A)
GS_1-2_Yas_Gelis	1-2 Yaş Toplam Geliş Sayısı
GS_1-2_Grubu	Geliş Sayısına Göre Segmenti (A,B,C, N/A)
GS_2-6_Yas_Gelis	2-6 Yaş Toplam Geliş Sayısı
GS_2-6_Grubu	Geliş Sayısına Göre Segmenti (A,B,C, N/A)
GS_6-10_Yas_Gelis	6-10 Yaş Toplam Geliş Sayısı
GS_6-10_Grubu	Geliş Sayısına Göre Segmenti (A,B,C, N/A)
MAXG_0	0-1 Yaş Arasında İki Başvuru Arasındaki Max Gün Sayısı
MAXG_1	1-2 Yaş Arasında İki Başvuru Arasındaki Max Gün Sayısı

MAXG_2	2-3 Yaş Arasında İki Başvuru Arasındaki Max Gün Sayısı
MAXG_3	3-4 Yaş Arasında İki Başvuru Arasındaki Max Gün Sayısı
MAXG_4	4-5 Yaş Arasında İki Başvuru Arasındaki Max Gün Sayısı
MAXG_5	5-6 Yaş Arasında İki Başvuru Arasındaki Max Gün Sayısı
MAXG_2-6_Yas	2-6 Yaş Arasında İki Başvuru Arasındaki Max Gün Sayısı
MAXG_6	6-7 Yaş Arasında İki Başvuru Arasındaki Max Gün Sayısı
MAXG_7	7-8 Yaş Arasında İki Başvuru Arasındaki Max Gün Sayısı
MAXG_8	8-9 Yaş Arasında İki Başvuru Arasındaki Max Gün Sayısı
MAXG_9	9-10 Yaş Arasında İki Başvuru Arasındaki Max Gün Sayısı
MAXG_6-10_Yas	6-10 Yaş Arasında İki Başvuru Arasındaki Max Gün Sayısı
Churn_0-1_Yas	0-1 Yaş Aralığındaki Churn Durumu (1 Evet/ 0 Hayır)
Churn_1-2_Yas	1-2 Yaş Aralığındaki Churn Durumu (1 Evet/ 0 Hayır)
Churn_2-6_Yas	2-6 Yaş Aralığındaki Churn Durumu (1 Evet/ 0 Hayır)
Churn_6-10_Yas	6-10 Yaş Aralığındaki Churn Durumu (1 Evet/ 0 Hayır)
2020_Churn_Durumu	01.01.2020 İtibariyle Churn Durumu (1 Evet/ 0 Hayır)