

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

TELEKOMÜNİKASYON SEKTÖRÜNDE İPTAL ANALİZİ

**YÜKSEK LİSANS TEZİ
Pınar KİŞİOĞLU**

Anabilim Dalı : Endüstri Mühendisliği

Programı : Endüstri Mühendisliği

HAZİRAN 2009

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

TELEKOMÜNİKASYON SEKTÖRÜNDE İPTAL ANALİZİ

YÜKSEK LİSANS TEZİ
Pınar KİŞİOĞLU
(507071120)

Tezin Enstitüye Verildiği Tarih : 04 Mayıs 2009
Tezin Savunulduğu Tarih : 01 Haziran 2009

Tez Danışmanı : Doç. Dr. Y. İlker TOPÇU (İTÜ)
Diğer Jüri Üyeleri : Prof. Dr. Ramazan EVREN (İTÜ)
Doç. Dr. Şebnem BURNAZ (İTÜ)

HAZİRAN 2009

ÖNSÖZ

Tez çalışması süresince her konuda yol gösterici ve sabırlı yaklaşımlarından ötürü danışmanım Sayın Doç. Dr. Y. İlker Topçu'ya, her konuda desteğini yanımda hissettiğim ve yüksek lisans eğitimim boyunca moral ve motivasyonumu yüksek tutan Sayın Özgür Aykanat'a, her konuda manevi desteklerini esirgemeyen aileme ve arkadaşlarıma sonsuz teşekkürlerimi sunarım.

Yüksek lisans eğitimimi devam ettirmemi sağlayan, verdiği maddi desteklerinden dolayı Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK) na teşekkürü bir borç bilirim.

Mayıs 2009

Pınar Kışioğlu

İÇİNDEKİLER

Sayfa

ÖNSÖZ.....	iii
KISALTMALAR	vii
ÇİZELGE LİSTESİ.....	ix
ŞEKİL LİSTESİ.....	xi
ÖZET.....	xiii
SUMMARY	xv
1. GİRİŞ	1
1.1. Çalışmanın Amacı	1
2. TELEKOMÜNİKASYON SEKTÖRÜNDEKİ DEĞİŞİMLER.....	3
2.1 İptal Analizi.....	4
2.2 Yazın Taraması	7
3. BAYES AĞLARI	15
3.1 Nedensel Haritalar	19
3.1.1 Seri bağlantılar	20
3.1.2 Iraksayan bağlantılar.....	20
3.1.3 Yakınsayan bağlantılar	21
3.2 Matematiksel İfadeler	22
3.2.1 Bayes teoremi.....	26
3.3 Bayes Ağlarının Genel Yapısı.....	29
3.3.1 Bayes ağları için zincir kuralı	30
3.3.2. Bayes ağlarının Netica’da modellenmesi	31
4. KULLANILAN YÖNTEMLER	39
4.1 Karar Ağaçları.....	39
4.1.1 SPSS ile Karar Ağacı Uygulaması	41
4.2 Korelasyon Analizi.....	48
4.2.1 SPSS ile korelasyon analizi uygulaması.....	49
4.3 Eşdoğrusallık analizi	51
4.3.1 SPSS ile eşdoğrusallık analizi uygulaması	52
5. UYGULAMA	55
5.1 Verilerin Toplanması.....	55
5.2 Verilerin Hazırlanması.....	56
5.3 Sürekli Değişkenlerin Kesikli Hale Getirilmesi.....	57
5.4. Değişkenler Arasındaki İlişkinin Gösterilmesi	61
5.5 Problemin Bayes Ağları İle Modellenmesi.....	62
5.6 Senaryo Analizi	64
5.6.1 Senaryo 1	65
5.6.2 Senaryo 2	69
5.6.3 Senaryo 3	75
6. SONUÇ VE ÖNERİLER	79
KAYNAKLAR.....	83
EKLER.....	85

KISALTMALAR

DET	: Detection Error Tradeoff
DBSCAN	: Density-Based Spatial Clustering of Applications with Noise
SQL	: Structured Query Language
CHAID	: Chi-squared Automatic Interaction Detection
CRT	: Classification and Regression Trees
QUEST	: Quick, Unbiased, Efficient Statistical Tree
SPSS	: Statistical Package for the Social Sciences
VIF	: Variance Inflation Factor

ÇİZELGE LİSTESİ

Sayfa

Çizelge 3.1 : P(YktG Ykt) koşullu olasılık çizelgesi.....	30
Çizelge 3.2 : P(AÇ Ykt, TB) koşullu olasılık çizelgesi.....	30
Çizelge 5.1 : 9-15 TL faturası olan aboneler için beklenen kazanç analizi.....	68
Çizelge 5.2 : 15-30 TL faturası olan aboneler için beklenen kazanç analizi.....	69
Çizelge 5.3 : 0-15 TL faturası olan aboneler için beklenen kazanç analizi.....	74
Çizelge 5.4 : 15-30 TL faturası olan aboneler için beklenen kazanç analizi.....	75
Çizelge 5.5 : 0-30 TL arası faturalar için tarife değişiminden beklenen kazanç analizi.....	77
Çizelge 5.6 : 30-400 TL arası faturalar için tarife değişiminden beklenen kazanç analizi.....	78

ŞEKİL LİSTESİ

Sayfa

Şekil 3.1 : Araba çalıştırma problemi için nedensel harita.	19
Şekil 3.2 : Seri bağlantı.	20
Şekil 3.3 : Iraksayan bağlantı.	20
Şekil 3.4 : Cinsiyet, saç ve boy uzunlukları arasındaki nedensel ilişkiler.	21
Şekil 3.5 : Yakınsayan bağlantı.	21
Şekil 3.6 : Soğuk algınlığı, boğaz iltihabı ve ateş çıkması arasındaki nedensel ilişkiler.	22
Şekil 3.7 : Ayrık olaylar.	23
Şekil 3.8 : Bileşik olaylar.	25
Şekil 3.9 : İndirgenmiş uzay.	26
Şekil 3.10 : U uzayının parçalanışı.	27
Şekil 3.11 : Örnek bir ağ yapısı.	29
Şekil 3.12 : Araba çalıştırma problemi için Bayes Ağı.	29
Şekil 3.13 : Netica Programı.	31
Şekil 3.14 : Netica’da menü gösterilimi.	32
Şekil 3.15 : Netica’da kısa yolların gösterilimi.	32
Şekil 3.16 : Düğüm özelliklerinin değiştirildiği pencere.	33
Şekil 3.17 : Netica’da basit bir BA örneği.	34
Şekil 3.18 : C düğümüne ait koşullu olasılık çizelgesi.	35
Şekil 3.19 : A düğümüne ait koşullu olasılık çizelgesi.	35
Şekil 3.20 : Verilerin farklı bir veri kümesinden alınması.	36
Şekil 3.21 : BA’nın derlenmesi.	37
Şekil 3.22 : Fayda düğümünün olasılık çizelgesi.	37
Şekil 3.23 : Havanın yağmurlu olması durumu.	38
Şekil 4.1 : Karar Ağacı.	40
Şekil 4.2 : Karar ağacının seçilmesi.	42
Şekil 4.3 : Değişkenlerin seçilmesi.	42
Şekil 4.4 : Hedef bölümlerin seçilmesi.	43
Şekil 4.5 : Düğüm sayısının belirlenmesi.	43
Şekil 4.6 : Model çıktısının özelleştirilmesi.	44
Şekil 4.7 : Tahmin değerlerinin kaydedilmesi.	44
Şekil 4.8 : Model sonucu ile ilgili genel bilgi.	45
Şekil 4.9 : Örnek karar ağacı.	45
Şekil 4.10 : Örnek karar ağacına ait çizelge gösterimi.	46
Şekil 4.11 : Açıklayıcı değişkenlerin değerleri.	46
Şekil 4.12 : Açıklayıcı değişkenlerin değerleri.	47
Şekil 4.13 : Risk ve sınıflandırma çizelgeleri.	48
Şekil 4.14 : Değişkenlerin tahmini olasılıkları.	48
Şekil 4.15 : Korelasyon analizi için programın açılması.	50
Şekil 4.16 : Korelasyon analizi için değişkenlerin seçilmesi.	50
Şekil 4.17 : Satış ve yakıt verimliliğine ait korelasyon matrisi.	51
Şekil 4.18 : Satış ve beygir gücüne ait korelasyon matrisi.	51

Şekil 4.19 : Eşdoğrusallık analizinde kullanılan değişkenler.....	52
Şekil 4.20 : Eşdoğrusallık analizinin sonuç çizelgesi.	53
Şekil 5.1 : Yaş değişkeninin karar ağacında sınıflandırılması.....	58
Şekil 5.2 : Abonelik süresi değişkeninin karar ağacında sınıflandırılması.	59
Şekil 5.3 : Ortalama fatura tutarı değişkeninin karar ağacında sınıflandırılması.	59
Şekil 5.4 : Ortalama konuşma süresi değişkeninin karar ağacında sınıflandırılması.	60
Şekil 5.5 : Ortalama arama sıklığı değişkeninin karar ağacında sınıflandırılması. ...	60
Şekil 5.6 : Değişkenler arasındaki korelasyon matrisi.	61
Şekil 5.7 : Eş doğrusallık analizi sonucu.....	62
Şekil 5.8 : Nedensel Harita.	62
Şekil 5.9 : İptal analizi için Bayes ağı.	63
Şekil 5.10 : Derlenmiş Bayes ağı.	64
Şekil 5.11 : Senaryo 1 için olasılık çizelgesi.	65
Şekil 5.12 : Senaryo 1 için Bayes ağı.....	66
Şekil 5.13 : Senaryo 1 için ayrıntılı Bayes ağı.....	67
Şekil 5.14 : Senaryo 1 için iptal yüzdelerini gösteren grafik.	67
Şekil 5.15 : Senaryo 2 için derlenmiş Bayes ağı.....	70
Şekil 5.16 : Senaryo 2 için ayrıntılı Bayes ağı.....	71
Şekil 5.17 : Farklı operatörleri aramayan aboneler için Bayes ağı.	72
Şekil 5.18 : Farklı operatörleri arayan aboneler için Bayes ağı.....	72
Şekil 5.19 : Senaryo 2 için iptal yüzdelerini gösteren grafik.	73
Şekil 5.20 : 15-30 TL arası faturası olan aboneler için iptal yüzdelerini gösteren grafik.	74
Şekil 5.21 : Senaryo 3 için derlenmiş Bayes ağı.....	76
Şekil 5.22 : Tarife bazında ortalama fatura tutarı ve iptal yüzdeleri.	76

TELEKOMÜNİKASYON SEKTÖRÜNDE İPTAL ANALİZİ

ÖZET

Telekomünikasyon sektöründe özelleştirmeler ve piyasaya yeni kurumların katılımıyla şirketler açısından pazarda tutunabilmek önemli bir unsur haline gelmiştir. Artan rekabetle birlikte telekomünikasyon hizmeti sağlayıcıları için abone kaynaklı iptal yönetimi önem kazanmıştır.

Etkili bir iptal yönetimi için şirketler yeni müşteri kazanmak yerine eldeki abonelerini tutma yaklaşımına yönelmiştir. Bugüne kadar yapılan araştırmalarda telekomünikasyon sektöründe iptal etmeye eğilimli aboneler tahmin edilmeye çalışılmıştır.

Bu çalışmada iptal etme eğiliminde olan abonelerin davranış özellikleri Bayes ağları ile modellenmiştir. Bayes ağları yönteminin kullanılmasının sebebi faktörler arasındaki nedensel ilişki temeline dayanmasıdır. Çalışmada kullanılan veriler Türkiye’de faaliyet gösteren telekomünikasyon şirketlerinden birinden alınmıştır. İlk olarak, Bayes ağlarında sadece kesikli değişkenler kullanıldığı için sürekli değişkenler CHAID karar ağacı algoritması ile kesikli hale getirilmiştir. Ardından değişkenler arasındaki ilişkiyi göstermek için korelasyon analizi ve eşdoğrusallık testi yapılmıştır. Elde edilen sonuçlarla uzman görüşleri birleştirilerek Bayes ağının temelini oluşturan bir nedensel harita oluşturulmuştur. Bayes ağının sonuçlarına göre abonelerin ortalama konuşma süreleri, ortalama fatura tutarları, farklı operatörleri ortalama arama sıklığı ve bağlı oldukları tarife türünün iptali açıklayan önemli değişkenler olduğu görülmüştür. Çalışmanın sonunda üç farklı senaryoyla iptal eden abonelerin özellikleri ortaya konmuş ve bu abonelerin iptal oranını azaltacak kampanya önerileri sunulmuştur. Daha önce yapılan çalışmalarda telekomünikasyon sektöründe iptal analizi için Bayes ağlarına başvurulmamıştır. Bu çalışmayla iptal yönetimine yeni bir yöntemle farklı bir bakış açısı getirilmiştir.

CHURN ANALYSIS IN TELECOMMUNICATION INDUSTRY

SUMMARY

In telecommunication industry, it is really important to take place in the market for many organizations since the deregulations and attendance of new firms to the market. As competition increases, customer churn becomes a great issue to deal with by the telecommunication providers.

For an effective churn management companies try to retain their existing customers, instead of acquiring new ones. Previous researches focus on predicting the customers with a propensity to churn in telecommunication industry.

In this study, a model is constructed by Bayesian Belief Network to identify the behaviors of customers with a propensity to churn. Bayesian Belief Network is used because it investigates the casual relationship between factors. The data used is collected from one of the telecommunication providers in Turkey. First, as only discrete variables are used in Bayesian Belief Networks, CHAID algorithm is applied to discretize continuous variables. Next, correlation analysis and multicollinearity test are performed to show the relationship between variables. Then a causal map as a base of Bayesian Belief Network is brought out via the results of correlation analysis, multicollinearity test and experts' opinions. According to the results of Bayesian Belief Network, average minutes of calls, average billing amount, the frequency of calls to people from different providers and tariff type are the most important variables that explain customer churn. At the end of the study, three different scenarios that examine the characteristics of the churners are analyzed and promotions are suggested to reduce the churn rate. Previous researches does not deal with Bayesian Belief Network for churn analysis in telecommunication industry. This study shows a new perspective to churn management.

1. GİRİŞ

Günümüzde firmalar ağır rekabet ortamında ayakta kalabilmek için müşteri odaklı bir yaklaşımı benimsemek zorundadır. Bir çok şirket müşteri kaybını önlemek ve müşteri sadakatini arttırmak için yollar aramaktadır. Bu nedenle müşterilerini kaybetmek istemeyen kuruluşlar çeşitli araştırmalar yapmakta ve bu çalışmaların sonuçlarına göre kampanyalar düzenlemektedir.

Türkiye’de müşteri kaybıyla en çok ilgilenen sektörlerden biri müşteri kaybının çok kolay olduğu telekomünikasyondur. Özellikle 2008 yılında çıkan numara taşınabilirliği kararıyla telekomünikasyon operatörleri arasında abone geçişlerinin daha rahat yapılmasına olanak sağlanmıştır. Bu nedenle telekom operatörleri ellerindeki aboneleri memnun edecek kampanyalara başlamıştır.

1.1. Çalışmanın Amacı

Bu çalışmanın amacı telekomünikasyon sektöründe abone iptalini etkileyen en önemli faktörleri belirlemektir. Bunu yaparken faktörler arasındaki nedensel ilişki temeline dayandığı için Bayes ağları yöntemi kullanılmıştır. İptal eden abonelerin özellikleri incelenerek farklı müşteri segmentleri oluşturulmuş ve bu müşteri segmentlerinde iptal oranını azaltacak önerilerde bulunulmuştur.

2. TELEKOMÜNİKASYON SEKTÖRÜNDEKİ DEĞİŞİMLER

Dünya genelinde telekomünikasyon sektöründe süregelen bir değişim mevcuttur. 1996'da Amerika Birleşik Devletleri'nde kısa (şehiriçi) ve uzun mesafeli (şehirlerearası, uluslararası) aramalar, kablolu TV, radyo-televizyon yayını ve kablosuz hizmetlerdeki engelleri koruyan yasalar kaldırılmıştır. Avrupa'da, devlet kısıtlarının ortadan kalkmasıyla İngiltere, İsveç ve Finlandiya'da yeni pazarlar oluşmuştur. 1998'de 15 Avrupa ülkesi kendilerini serbestleştirmiştir [1].

Gelişmekte olan ülkeler Şili, Malezya ve Peru'da hükümete ait telefon tekelleri yürürlükten kaldırılmıştır. Bu yeniden yapılanmalar yeni pazar fırsatları yaratmakta ve zorunlu tekellerin müşterilere sağladığı herkese açık olmayan erişimi sonlandırmaktadır [2].

Rekabet arttıkça ve iletişim ağı başlıca satış ürünü haline geldikçe hizmet sağlayıcılar maliyetleri düşürmek, hizmetlerine değer katmak ve farklılığı garanti etmekle yüz yüze gelmişlerdir. Müşteri farklı taşıyıcılar içinden seçim yapabildiğinde müşteri kaybetme kaygısı ve pazarlama zekası, pazar payının korunması ve arttırılmasında stratejik fonksiyonlar haline gelmiştir [1].

Rekabetçi koşullar altında müşteriler taşıyıcı hareketlerinin odak noktası haline gelmişlerdir. Müşteri gereksinimleri sadece hizmet tekliflerine karar vermekle kalmıyor, belli müşteri tiplerine odaklanmak isteyen firmaların iletişim ağını şekillendirmekte ve organizasyonel yapısını etkilemektedir [2].

Telekomünikasyon sektöründe rekabet iki farklı eğilime yol açmıştır:

- Sektör konsolidasyonu: Rekabet içindeki firmalar yeni pazarlara girmek ve yeni hizmetler sunmak için yollar aramaya başlamışlardır. Bunun için en mantıklı ve uygun yol birleşme ve diğerlerini kontrol altına almaktan geçmektedir.
- Yakınsama: Telekomünikasyon hizmetlerinin (video, ses, veri vb.) çoklu bantlarının entegrasyonu ekonomik bir gerçek haline gelmiştir ve firmalar tarifeler ve ürün paketleri şeklinde çoklu hizmetler sunmaya başlamışlardır.

- Müşteri stratejisi: Telekom endüstrisinin ‘büyük ağlar kur, müşteri kazan’ söylemleri ‘yeni hizmetler sun, müşterilerini memnun et’ olarak değişmiştir. Kısaca sektör müşteri kazanma yöneliminden müşteriye elde tutma yönelimine geçmiştir [1].

Bütün bu değişimler bir şirketin destekleyici altyapısının duyarlı ve esnek olmasını gerektirmiştir. Karar destek bilgilerini çalışana erişebilir kılmak kurumu hızlı, rekabetçi, işi kontrol edebilir hale getirir. Doğru bir şekilde uygulandığında karar destek uygulamaları karlılığı daha iyi anlamayı ve pazarlama amacına odaklanmayı sağlar. Devletin kısıtlayıcı etkilerinin kaldırıldığı telekomünikasyon sektöründe gelişmekte olan kurum için gerekli veriler işlenmemiş olarak bol miktarda mevcuttur. Müşteri verileri çok olduğu kadar yeterli derecede kullanılmayan bir kaynak olmuştur. Fakat tek başına veri, şirketin kurumsal amaçları ve bilgi teknolojisi olmadan anlamlı bilgiye dönüşmez [1].

Yeni stratejilerin kurulumunu sağlayan teknoloji öncelikle şirketin kurumsal fonksiyonlarıyla entegre olmalıdır. Müşterileri kazanma, kontrol etme ve elde tutma kabiliyeti kurumun kendi mimari altyapısıyla birlikte hız ve etkinliğine bağlıdır. Telekomünikasyon şirketlerinin gelecekteki başarısı veya başarısızlığı kendi pazarlarına ve müşterilerine ait bilgilerden geçmektedir [1].

2.1 İptal Analizi

Müşteriler, abonmanlıklarını sürdürdükleri şirketten ayrılarak rakip bir firmayla anlaştıklarında ‘churner’ (iptal eden) olurlar. İptal etme müşterinin geçiş sürecidir. Bu durum çok sayıda müşterisi olan şirketler için önemlidir. Kredi kartı sağlayıcıları, sigorta şirketleri ve telekomünikasyon şirketleri bunun en önemli örneklerindendir [3].

Telekomünikasyon sektörü geçtiğimiz yıllarda birçok değişim yaşamıştır. Piyasaların serbestleştirilmesiyle birlikte hemen hemen bütün Avrupa ülkeleri kısa (şehiriçi) ve uzun (şehirlerarası, uluslararası) mesafeli servisleri yabancı taşıyıcılara açmıştır. Yeni rekabetçi ortamın aniden ortaya çıkmasıyla ‘churn’ sorunu tırmanmıştır [3].

Rekabetin sertleştiği kablosuz telekomünikasyon sektörü için aylık iptal oranının %2.2 civarında olduğu tahmin edilmektedir [4].

Asya ve Avrupa pazarlarında piyasaya yeni giren her bir firma iptal sorununa biraz daha katkısı olmaktadır. 1998’de Avrupa’da 30 yeni telekom firması piyasaya girmiştir ve uzmanların analizlerine göre firmaların pazarda tutunabilmeleri için %15 lik paya sahip olmaları gerekmektedir. Geçmişte abone sayısının artmasıyla bu sorun yaşanmazken piyasanın durağanlaşması ve kişi başına ortalama gelirin düşmesiyle öldürücü rekabet artacaktır [5].

İptallerin telekomünikasyon sektörüne maliyeti oldukça büyüktür. Avrupa ve Amerika piyasaları için her yıl 4 milyar \$, sebep olmaktadır. Yıllık iptal oranları %25 ile %30 arasında değişmekte ve bu oranın üst seviyelerinde sayılara ulaşan firmalar yeni müşteriler kazanmak için gerekli yatırımı yapamamaktadırlar. Nedeni ise kaybedilen her bir müşterinin zararının yeni bir müşteriyle kapatılması için üç yıllık uzun bir süre gerekmektedir [5]. Bununla beraber yeni müşteriyi kazanmak, mevcut müşteriyi elde tutmaktan daha pahalıya mal olmaktadır. Ayrıca bir şirket yeni müşteriyi kazanım masraflarını tamamen karşılamadan müşteri iptal edebilir. Mevcut müşteriler üzerinde harcama yapmak yeni müşterileri kazanmaktan daha verimli olduğu açıktır [3]. Bir çok firmanın telekomünikasyon sektörüne girmesiyle rekabetin yönü yeni aboneler edinmekten eldeki aboneyi tutmaya ve aboneleri rakip firmalardan korumaya doğru değişmiştir. İletişimde rakip şirketlerin sayısı arttıkça iptal etmenin nasıl engelleneceği ve müşteri sadakatinin ne şekilde sağlanabileceği tartışılan konular haline gelmiştir. Aslında bir firma bunları tartışırken yüksek orandaki abone iptali pazarın aktif bir şekilde çalıştığının bir göstergesidir [6].

Zamanla iptal yönetimi çok önemli bir rekabet aracı olarak ortaya çıkmış ve müşteri odaklı pazarlama çabalarının temelini oluşturmuştur [3]. Mattersion’un işaret ettiği gibi bir çok telekom uzmanı açısından iptal ile nasıl başedilebileceğini araştırmak şirketlerinin ayakta kalmasını sağlayacak bir anahtar haline gelecektir [7]. Telekom firmalarının yönetiminin karşı karşıya kaldığı bu sorun nedeniyle iptal edecek abonelerin tahmini ve kalmalarına yönelik teşvik edici bir plan hazırlamak gereklidir [5].

Etkili iptal yönetimi ile bir şirket hangi tür müşterilerin iptal etmeye yakın ve hangilerinin sadık olacağına karar verebilir. Sürecin bir parçası olarak müşteri değerine karar verilebilir. Örneğin; hiç kar kazandırmayan ya da çok az kazandıran müşterilerin aboneliklerini sona erdirmeleri istenebilir. Bir şirket bu çeşit bilgiye sahip olduğunda yöneticiler iptalleri azaltmak için stratejik kararlar alabilir,

ayrılanlardan değerli olanları geri kazanabilir ve gelecekte doğru müşteri türüne göre maliyet etkili davranabilir.

Fakat bir çok müşteri iptal etmeye eğilimli olduğunu aboneliklerini sonlandırmak için aradıkları zaman gösterirler. Bu durumda bir yönetici kendine ‘acaba doğru abonelere mi sahibim yoksa yaptığım yatırıma karşılık alamadan iptal edecek abonelere mi sahibim?’, ‘abonelerimize ait doğru bir bakışa sahip miyiz ve böylece iptal etmeye eğilimli olanların profilini çıkartabilir miyiz?’ sorularını sormalıdır. Bu soruların yanıtları da büyük ölçüde doğru bilgiye ulaşmayı sağlayacak yollara sahip olmaya dayanmaktadır [5].

Telekomünikasyon sektöründe iptal ile mücadele etmenin bir çok yolu vardır. Bu yöntemlerin ortak amacı:

- Bir müşterinin iptal edip etmeyeceğini ve ne zaman iptal edeceğini tahmin etmek,
- Müşterilerin neden iptal ettiklerini anlamaktır.

Hangi müşterilerin iptal etmeye yatkın olduğunu tahmin ederek telekomünikasyon şirketi, müşterilerini kalmaya teşvik edici öneriler geliştirerek iptallerin oranını düşürebilir. Müşterilerin neden iptal ettiklerini anlayarak telekomünikasyon şirketi müşterilerini memnun edebilmek için servis hizmetlerini (tarife paketleri vb.) değiştirmek üzerinde çalışabilir [3].

Tahmin etme veya müşteri hareketlerini anlamadan önce veriler bir araya getirilmelidir. Bu, genellikle içinde açık, değişken olmayan, tarihsel verilerin olduğu bir veri ambarı ile yapılmaktadır. Bir iptal analizi için veri ambarında saklanacak verilerin çeşitleri şunlardır:

- Müşteriye ait istatistiksel karakteristikler. Yaş, cinsiyet, medeni durum, ikamet vb.
- Arama istatistikleri. Bir gün içinde farklı zamanlardaki görüşmelerin uzunlukları, şehiriçi ve şehirlerarası görüşmelerin sayısı vb.
- Her bir müşteriye ait fatura bilgileri. Müşterinin şehiriçi ve uzun mesafeli görüşmeler için ödediği miktar.
- Ekstra hizmet bilgisi. Müşterinin uzun mesafeli konuşmalar için özel bir tarifeyle bağlanıp bağlanmadığına ilişkin bilgi.

- Müşteri tarafından satın alınan ses ve veri ürün ve hizmetleri. Yayın hizmeti, özel iletişim ağları, hizmete sunulmuş veri aktarma bağlantıları vb.
- Şikayet bilgileri. Çağrı merkezlerine şüpheli fatura bilgileri, düşmeyen telefonlar, çalışmayan özel servis hizmetleri gibi bırakılan şikayetler
- Güven (sadakat) tarihi

Kaba bir hesapla aylık iptal miktarı şu şekilde hesaplanabilir:

$$\text{Aylık iptal oranı} = (C_0 + A_1 - C_1) / C_0$$

Burada C_0 ayın başındaki abone sayısını, C_1 ayın sonundaki abone sayısını, A_1 ise ay boyunca yeni müşteri sayısındaki artışı göstermektedir. Örnek olarak bir firmanın ay başında 100 tane müşterisi olsun, ay boyunca 20 yeni abone elde etsin ve ay sonunda 110 aboneye sahip olsun. Firma bir ay boyunca 10 abone kaybetmiştir yani %10 luk bir abone kesimi iptal etmiştir [8].

Abone iptali şirkete bağlı iptal ve aboneye bağlı iptal olarak ikiye ayrılabilir. Şirkete bağlı iptaller telekomünikasyon sağlayıcısı şirket tarafından abonenin isteğine bakılmaksızın yapılan iptaldir. Çoğu kez üst üste ödenmemiş faturalar yüzünden şirket aboneliğin iptalini yapmaktadır. Aboneye bağlı iptal ise müşterinin isteğine bağlı olarak yapılan iptallerdir. Her ne kadar karlılık sağlamayan aboneleri elemek önemli bir durum olsa da aboneye bağlı iptal ele alınması gereken önemli bir konudur [8].

İptal yönetimi, ayrılmayı düşünen potansiyel abonelerin tahminini yapan ve onları iptalden uzaklaştırmanın yollarını arayan bir süreç olarak tanımlanabilir. İptal yönetiminin amacı abone kaynaklı kayıpların enküçüklenmesi ve karlılık sağlayan aboneleri elde tutarak karlılığın enbüyüklenmesidir [8].

2.2 Yazın Taraması

Dünya çapında telekomünikasyon sektöründe iptal sorununu ele alan bir çok çalışma yapılmaktadır. Örneğin, Tayvan'da iletişim sektöründe meydana gelen abone kaynaklı iptal oranı aylık %2 dir. Hung, Yen ve Wang yaptıkları çalışmada iptal yönetimi için önce hangi abonelerin iptal edeceği tahmin edildikten sonra etkili bir elde tutma yönteminin belirlenmesi gerektiğinden bahsetmişlerdir. Kümeleme, karar ağaçları ve sinir ağları tekniklerini kullanmışlardır. Makalede öncelikle karar ağacı tekniği kullanılmıştır. Karar ağacı modellerinden biri bütün aboneler üzerinde

uygulanırken, diğer model kümeleme yöntemiyle ayrılan abone segmentlerine ayrı ayrı uygulanmıştır. Kümeleme yönteminde K-means tekniği ile abone 5 farklı kümede toplanmıştır. Abone segmentasyonu yapılmadan uygulanan karar ağacı algoritması olarak C5.0 seçilmiştir. Bu modelde iptal oranı 0.71 olarak bulunmuştur. Daha önce ayrılan kümelerin her birine ayrı ayrı karar ağacı modeli uygulanmış ve iptal oranları 0.50, 1.19, 0.32, 0.30, 1.37 olarak bulunmuştur. Makalede elde edilen sonuçlara göre ilk 6 ay tutarlı bir yapı varken yeni yıllar birlikte abonelerin davranışları değişiklik göstermiştir. İlk 6 ayda sinir ağları karar ağaçlarına göre; segmentasyon yapılmayan karar ağacı, segmentasyon yapılan karar ağacına göre daha iyi sonuç vermiştir [7].

Modeller arasında bir farklılık olup olmadığına bakmak için t-testi uygulanmış. Buna göre

- Segmentasyon yapılmayan karar ağacı modeli segmentasyon yapılanına göre daha iyi sonuç vermiştir.
- Sinir ağları segmentasyon yapılmayan karar ağacı modeline göre daha iyi sonuç vermiştir [7].

Tayvan’da yapılan bir diğer çalışmada Wei ve Chiu tarafından abone iptaline arama davranışı tabanlı iptal tahmini daha tatmin edici sonuçlar verdiği ortaya konulmuştur. Burada arama desenlerinde meydana gelen farklılıkların abonenin iptal edip etmeyeceğine dair sinyaller verdiği gösterilmiştir. Kullanılacak değişkenler seçilirken abonelerin telefonla görüşme süreleri, kaç farklı telefon numarasını aradıkları ve abonelik süreleri gözönüne alınmıştır. Çalışmada kullanılan veri, Tayvan’daki büyük bir mobil telekomünikasyon şirketinden alınmıştır. Ekim 2000-Ocak 2001 tarihleri arasında aktif olan 11400 abone ile 4500 iptal etmiş abone seçilmiştir. Makalede tahmin süresi ve elde tutma çalışmalarının yapıldığı süreler gözönüne alınmıştır. Veri madenciliği tekniği olarak multi-classifier class-combiner kullanılmıştır. Bu tekniğe göre verilen bir S kümesi belirlenen bir orana bölünmektedir. Örneğin 1:10 ile S kümesini 10 parçaya bölündüğü kabul edilir. Her bir parça büyük kümenin özelliklerini taşır. Modelin performansı kayıp oran (miss rate) ve yanlış oran (false rate) ile belirlenmiştir. Kayıp oranla iptal ettiği halde model tarafından belirlenemeyen aboneler, yanlış oranla iptal etmediği halde model tarafından iptal ettiği belirlenen abonelerin sayısı verilmiştir. Bu kayıp oran ve yanlış

oran deęerleri DET (detection error tradeoff) eęrisinde toplanmıřtır. DET eęrisinin orijine yakın olması tahminlerin ok iyi yapıldıęını gsterir [9].

alıřma sırasında ncelikle veri kmesi 2, 4, 8 ve 16 kmeye ayrılıp iptal tahmini yapıldıęında 1:2 oranının daha iyi sonu verdięi grlmřtr. Gzlem yapılan 30 gnlk sre 2,3,5 ve 10 olmakzere alt periyodlara blnmřtr. Alt periyod 2 olarak alındıęında sonuların daha iyi olduęu gzlemlenmiřtir. İptal tahminini etkileyen faktrlerden bir dięeri olan elde tutma sresi 7, 14 ve 21 olarak seilmiřtir. Sonuta elde tutma sresi arttıka DET eęrisi orijinden uzaklařmakta ve sonuların mkemellięi azalmaktadır [9].

Telekomnikasyon sektrnde hizmet veren bir řirket iin iptal eden her mřteri nemli bir kayıptır. Bunların iinden karlılıęı yksek olanların iptal etmesi ok daha nemli bir kayıptır. Bu nedenle bir ok firma elde tutma yntemlerini deęerli abonelerizerinde yoęunlařtırmaktadır. Ferreira, Vellasco, Pacheco, Barbosa yayınladıkları makalede mobil iletiřim sektrnde deęerli mřterilerin iptal etmesiyle ilgilenmiřlerdir. Ferreira ve dięerlerine gre deęerli mřterilerin iptal edip etmeyeceęi nceden tahmin edilerek byk bir rekabet gc ve finansal kaynak saęlanabilir. Arařtırmada kullanılan veriler Brezilya’da faaliyet gsteren 100,000 aboneye sahip bir mobil iletiřim operatrnden alınmıřtır. Ekim 2000-Haziran 2001 aralıęında ncelikle 1245 abonenin verisi alınmıř fakat aylık iptal oranı olan %1.245 ok dřk olduęu iin veri kmesi 3500 abone ve %28 iptal oranıyla belirlenmiřtir. Veri yıęının sadece Haziran 2001’de iptal eden mřteriler alınmıř, Haziran 2001’den nce iptal edenler bu analizden ıkartılmıřtır. Veritabanından toplam 37 farklı deęiřken alınmıřtır. Bunlar, fatura bilgileri (aylık fatura geliri vb.), kullanım bilgileri (kullanılan hizmetlerin eřidi vb.), demografik veri (abonenin yařı, cinsiyeti vb.), mřteri iliřkileri verisi (abonelik sresi vb.), piyasa verisi (reklam giderleri vb.) bařlıkları altında toplanmıřtır. Verilerin gsterilimi iki farklı aıdan yapılmıřtır. Bunlardan biri basit gsterimdir, veriler sadece normalize edilmiřtir. Dięer gsterim ise geliřtirilmiř gsterimdir. Burada veriler normalize edilmesinin yanı sıra verilerzerinde bařka yntemlerde (En kk kareler yntemi vb.) uygulanmıřtır. Bu yntemler, bir girdinin amalanan durumun (Haziran ayındaki iptal durumu) deęiřkenlięini ne kadar aıkladıęını gstermesi aısından uygulanmıřtır. İptal analizi iin sinir aęları, karar aęaları, genetik algoritmalar ve neuro-fuzzy tekniklerinden faydalanılmıřtır. Kullanılan tekniklerle yapılan analizlerin sonucunda iptal

edebilecek müşterilerin sayısı tahmin edilmiştir. En iyi tahmin modelinin sinir ağlarına ait olduğu görülmüştür. Eğer etkili elde tutma yöntemleri uygulanırsa bu aboneler iptal etmek yerine mevcut iletişim operatörlerini kullanmak isteyecekleri açıktır. Bu durumda firmanın elde edeceği aylık ortalama kar her bir abone başına 80 dolar, bir senede elde edilebilecek toplam kar ise 70 milyon dolar civarında olacağı hesaplanmıştır. Böylece iyi bir abone elde tutma stratejisiyle gerçekten önemli gelir elde edileceği bu çalışmayla gösterilmiştir [10].

Ferreira ve diğerlerinin çalışmasına benzer bir çalışmada Amerika’da yapılmıştır. Mozer, Wolniewicz, Grimes, Johnson ve Kaushansky tarafından yapılan çalışmada aboneyi elde tutmak ve karlılığı artırmak amacıyla öncelikle iptal etme olasılığı olan abonelerin tahmin edilmesi için modellerin kurulması, modelleri kurduktan sonra olası tasarrufları hesaplayarak tahminlerin karar verme sürecine etkisi hesaplanmıştır. Çalışmada kullanılan veriler 46,744 aboneden oluşan veri tabanına sahip bir Amerikan telekomünikasyon firmasına aittir. Ekim-Aralık 1998 aralığında seçilen abonelerin Ocak yada Şubat 1999 da iptal edip etmediğini tahmin etmeye yönelik araştırma yapılmıştır. Bu aboneler en az hizmeti (sesli mesajlaşma vs) kullanan aylık ortalama geliri 234 dolar olan müşterilerdir. Ekim ayında (1998) firmanın bütün abonelerinden elde ettiği toplam gelir 14 milyon dolardır [11].

Abonelerle ilgili veriler; ayrıntılı arama kayıtları, finansal veriler (fatura bilgileri vb.), sunulan hizmetler, pazar araştırması (reklam maliyetleri, rakiplerin durumu vb.), aboneye ait demografik verilerden elde edilmektedir. Analiz edilecek değişkenlerin belirlenmesinde müşteri memnuniyetini etkileyen faktörler gözönüne alınmıştır. Önem sıralarına göre arama kalitesi, fiyat seçenekleri, operatörün kapasitesi, müşteri hizmeti, kapsama alanı vb. etkili faktörlerden bazılarıdır. Bunların yanında araştırma için toplam 134 değişken kullanılmıştır. Lojistik regresyon, karar ağaçları ve sinir ağları kullanılan tekniklerdendir. Modellerin değerlendirilmesi için Lift eğrisinden yararlanılmıştır. Lift eğrisini çizebilmek için xy-düzleminde koordinatlardan biri toplam abone sayısını diğeri ise iptal edeceği tahmin edilen abone sayısını göstermektedir. Ocak-Şubat 1999 tarihlerinde iptal ettiği belirlenen abonelere ait Lift eğrisi çizilmiştir. Bundan sonra kullanılan tekniklerle oluşan modellerin performansı bu eğriye yakınlıklarıyla ölçülmektedir. En doğru sonucu gösteren Lift eğrisine en yakın olan Lift eğrisine sahip olan model en iyi sonucu

vermiş demektir. Elde edilen sonuçlara göre sinir ağlarının lojistik regresyon ve karar ağacı tekniklerine göre daha iyi sonuç verdiği anlaşılmaktadır [11].

Duke Teradata'nın yayınladığı rapora göre Amerika'da yılı itibariyle telefon görüşmelerinin %25'inin kablosuz yolla yapıldığı ve 2006 yılı itibariyle kablosuz iletişim pazarının 189 milyon aboneye ulaştığını belirtilmiştir [12]. Pazarın büyüklüğü bir çok araştırmayı da beraberinde getirmektedir. Nath ve Behara tarafından ABD'deki bir telekomünikasyon firmasında yapılan çalışmada abone iptallerinin tahmini için Naive Bayes algoritması kullanılmıştır. Araştırma için kullanılan veri 150'den fazla değişkene sahip 100,000 aboneden oluşan bir veri tabanından alınmıştır. Veri kümesi 3 gruba ayrılmıştır. Birinci grup olan eğitilen grupta iptal (bağımlı değişken) değişkeni diğer bağımsız değişkenlerle birlikte vardır. Bu çalışmada eğitilen veri üzerinde oluşturulan modeller şimdiki test verisi ve gelecekteki test verisi üzerinde tahmin için kullanılmıştır [13].

Temmuz-Aralık 2001 tarihleri arasında en az 6 aydır abone olan aboneler seçilmiştir. Analiz için seçilen değişkenler abonelerin demografik verilerini, kullanım sıklığını, servis kalitesini gösteren değişkenlerdir. Naive-Bayes algoritması kullanılarak oluşturulan model, %68 oranında doğruluğa sahiptir [13].

Türkiye'de diğer bir çok sektörde olduğu gibi GSM sektörü de gelişmeye devam etmektedir. Karahoca ve Kara yaptıkları çalışmada iptal yönetimi için farklı bir kümeleme tekniğinden yararlanmışlardır. Bunun nedeni olarak etkili bir iptal yönetimi için abonelerin karlılıklarına göre segmentlere ayrılması gerektiğini öne sürmektedirler. Böylece değerli aboneleri elde tutabilmek için oluşan her bir segmente uygun promosyon ve kampanyalar yapılabilir. Bu çalışmada Türkiye'deki bir GSM operatörüne ait 250 abonenin ayrıntılı arama kayıtları, demografik veri ve fatura verileri alınmıştır. Bağımlı değişken olarak abonenin iptal değeri (iptal eden abone için 1, iptal etmeyen abone için 0 değeri atanır), bağımsız değişkenler olarak abonelik süresi, abonenin aylık harcama ve aylık gelir, ortalama arama süresi (1 ay, 3 ay, 4 ve 6 ay boyunca) ve ortalama SMS miktarı (1 ay, 3 ay, 4 ve 6 ay boyunca) alınmıştır. Çalışmada Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algoritması kullanılmış ve K-means, Expectation Maximization and Farthest-First kümeleme yöntemleri ile karşılaştırma yapılmıştır. DBSCAN yöntemi ile iptal analizi ve yeni abonenin hangi kümeye ait olduğunu saptamak amaçlanmıştır [14].

Verilerin hazırlanması aşamasında her bir aboneye 0-250 arasında bir puan verilmiş. Bunun nedeni müşterileri gruplara ayırabilmek için öncesinde grupların genel niteliklerini belirlemenin işleri kolaylaştırmasıdır. Bunun sonucunda 3 tane grup elde edilmiştir: iptal etmeye eğilimi yüksek, orta ve düşük olan grup. DBSCAN yöntemine göre bir kümenin her bir elemanı için verilen yarıçap komşuluğunda belirlenen minimum sayıda eleman olmalıdır. Son aşamada kümeler belirlenen yarıçap ve minimum eleman sayısına göre belirlenir. Yeni üyenin hangi kümeye ait olduğu böylelikle belirlenir [14].

Diğer yöntemler ile karşılaştırıldığında DBSCAN yöntemi en iyi sonucu verdiği gözlenmiştir. Diğer yöntemlerle bulunan kümelerin aksine DBSCAN yöntemiyle bulunan kümelerin ortalamaları birbirinden uzak, standart sapmaları ise küçük değerler almıştır. Bu kümelerin birbirinden iyi derecede ayrıldığını göstermektedir [14].

Abonelerin iptale eğilimli olup olmadıklarını anlamının bir diğer yolu da anket yöntemidir. Anketle elde edilen sonuçlar çeşitli istatistik tekniklerle açıklanabilir. Buna örnek olacak bir çalışma Pakistan’da Jahanzeb ve Jabeen tarafından yapılmıştır. İki telekomünikasyon firmasında ait abonelere sorulan sorularla müşteri memnuniyeti karşılaştırılmıştır. Araştırma, Pakistan’da faaliyet gösteren Ufone ve Telenor’a ait veriler alınarak yapılmıştır. Ufone 2001’de Pakistan Telecommunication Mobile Limited şirketi altında kurulmuş ve 2006 itibarıyla 7,5 milyon aboneye sahip bir firmadır. 2004’de Pakistan pazarına giren bir Norveç firması olan Telenor ise pazarın %6,7 sine sahip, 2006 itibarıyla 3,5 milyon abonesi olan bir firmadır [15].

Her bir firmadan 50’şer tane abonenin bilgisi alınarak örnek hazırlanmıştır. Aboneler faturalı ödeme yapan, en az 6 aydır abone ve Pak. Rs. 40,000-60,000 aylık geliri olan 30-45 yaş arasındaki erkek abonelerden oluşmaktadır.

Bu çalışmada yer alan hipotez:

H_0 : Ufone, Telenor’ a göre müşteri memnuniyetinde daha yüksek puana sahip olacaktır.

H_1 : Ufone, Telenor’ a göre müşteri memnuniyetinde daha düşük puana sahip olacaktır.

Seilen bu iki gruba nce bir anket uygulanmıřtır. Anket sonularından biri Ufone kullananları %64', Telenor kullananların %84' daha nce farklı bir operatrden řu andaki operatrne gemiřtir. Demek ki Telenor, Ufone'a gre daha etkili bir mřteri kazanma politikası izlemiřtir [15].

Ankette abonelere nceki hatlarını neden iptal ettikleri sorulmuřtur. Bu soruya alınan cevaplarda dřk ses kalitesi, sınırlı kapsama alanı, aė sorunları, yksek fiyatlar iptal sebepleri arasında gsterilmiřtir. Anketin bir diėer sonucu ise abonelerin oėu hatlarını deėiřtirmeye karar verdiklerinde operatrlerine bildirmemektedir. Ufone aboneleri dolaylı yoldan haber verdiėinden Ufone bu tarz aboneler iin farklı paketler uygulayabilmektedir. Telenor'da benzer bir durum sz konusu deėildir [15].

Anketin bir bařka sonucu oėu abone bir operatre baėlı kalmaya karar verdiėinde diėer operatrlerle ilgilenmezler. nk mevcut telefon numaralarını korumak isterler. Genel olarak bir hattı kapatıp yeni bir hat almak zor ve zaman kaybettiren bir durum olarak grlyor. 2006 yılının sonunda ise mobil telefon numarasının tařınabilirliėiyle ilgili dzenleme yapılmıřtır. Buna gre aboneler hatlarını deėiřtirdiklerinde numaraları aynı kalmaktadır. Sonulara bakılırsa Ufone aboneleri operatrlerinin etkili bir mřteri elde tutma politikası uyguladıklarına inanıyor. Telenor abonelerine gre operatrleri uzun vadede etkili bir mřteri elde tutma stratejisi belirleyememiřtir [15].

alıřmanın sonunda anket verilere gre iki firma arasında mřteri memnuniyeti aısından bir fark olup olmadıėını belirlemek iin t-testi yapılmıřtır. T-testinin sonucuna gre Ufone, Telenor'a gre mřteri memnuniyeti konusunda daha yksek puan almıřtır [15].

Tm dnyada telekomnikasyon sektrnde gzlenen rekabet firmaları iletiřimin yanında mřterilerine bazı ek hizmetleri de sunmalarını gerektirmiřtir. Bu konuyla ilgili olarak Sohn ve Kim yaptıkları alıřmada mobil iletiřimde saėlanan ek hizmetlerin kullanımıyla ilgili desenlerin ortaya ıkarılmasını ele almıřlardır. Bunun iin Kore'deki faaliyet gsteren bir firmanın verilerinden yararlanarak kmeleme ve iliřkilendirme metotları kullanılmıř ve hangi abonelerin hangi tr ek hizmeti kullandıklarına dair desenler ortaya ıkarılmıřtır. Kore'deki bir mobil iletiřim řirketinden Mart-Mayıs 2001 tarihleri arasında aboneliėi devam eden 17000 aboneye ait veriler kullanılmıřtır. Toplam 115 baėımsız deėiřkenden 21 tanesi alınarak faktr

analizi uygulanmıřtır. Buna gre 5 faktr oluřmuřtur. 5 faktre baėlı olarak 17000 abone kmelere ayrılmıřtır. 7 farklı algoritma arasından, en iyi sonu veren K-means (Newton) kmeleme tekniėi olmuřtur. alıřma doėrudan iptal analiziyle ilgili olmasa da firmanın hangi tarz abonelere sahip olduėunu bilmesi aısından nemlidir [16].

3. BAYES AĞLARI

Bayes ağları değişkenler arasındaki olasılıklı ilişkileri gösteren bir grafik modeldir. Değişkenler arasındaki ilişkileri birbirleri arasındaki nedensel bağlantılardan yola çıkarak belirlenir [17].

Bir Bayes ağı neden-sonuç diyagramı oluşturduğu için sistematik ve bütünlük süreçlere sahip bir çok alanda uygulama olanağına sahiptir [18].

Alternatif yöntemlerin bazı dezavantajları bulunmaktadır. Sezgisel tabanlı analizler ve analitik yöntemler karmaşık problemler için yetersizdir. İstatistiksel yaklaşımlar veri toplama konusunda büyük bir çaba gerektirirken, simülasyon ise öncesinde matematiksel bir modelin kurulmasını gerektirir. Karar ağaçları ise aralarında korelasyon olan değişkenleri sınıflandırmada yetersiz kalmaktadır. Çok kullanılan bir diğer yöntem olarak yapay sinir ağları ise veri kümesinden çok fazla etkilenmektedir [18].

Bayes ağları kalitatif ve kantitatif kısımlardan oluşmaktadır. Kalitatif kısımda (aynı zamanda ‘yapısal öğrenme’ (structural learning) denmektedir) değişkenler arası bağımsızlıkların grafiksel şekilde gösterimidir. Bu grafik gösterim yönlü döngüsüz haritalardan oluşmaktadır. İki çeşit yapısal öğrenme yöntemi vardır. Bunlardan biri skor hesabı, diğeri ise bağımsızlık testine bağlı olarak kısıt tabanlı yaklaşımdır. Kullanıcının ön bilgilerine dayanarak bir Bayes ağı kurulduktan sonra Bayes yaklaşımı, veriden optimal bir model yapısı oluşturur. Kısıt tabanlı yaklaşım ise optimal model yapısını her bir veri çifti için koşullu bağımlılıklara dayanarak oluşturur. Bayes yaklaşımıyla karşılaştırıldığında kısıt tabanlı yaklaşım hesaplama kolaylığı nedeniyle daha çok tercih edilir [18].

Bayes ağları bir çok konuda avantaj sağlamasına karşılık bazı dezavantajlara da sahiptir. Bayes ağlarıyla yapılan bir analizde sadece kesikli değişkenler kullanılmalıdır. Bu yüzden sürekli değişkenlerin olduğu bir analizde bu değişkenler çeşitli yöntemlerle kesikli hale getirilmelidir. Fakat sürekli değişkenleri kesikli hale getirme süreci veri kaybına sebep olabilmektedir ve verinin hassasiyeti azalmaktadır [18].

Yapay sinir ağıları ve karar ağaçları tahmin yapmak için uygun modeller olarak düşünülebilir. Yapay sinir ağıları, girdilerle saklı düğümler arasındaki bağıl ağırlığı hesaplayarak girdi düğümleri arasındaki etkileşimi ele almaktadır. Bu yüzden yapay sinir ağıları sorunu çözmek için ideal görünebilir. Fakat 'kara kutu' yapısından dolayı açık ilişkileri belirlemede sınırları vardır. Yani yapay sinir ağılarıyla değişkenler arasındaki nedensel ilişkiler ortaya çıkartılamaz. Diğer bir taraftan karar ağaçlarında değişkenler birbirlerine doğrudan bağlıdır. Her ne kadar bu tarz bağlantılar değişkenler arasında doğrudan etkileri varsa da yapay sinir ağılarıyla karşılaştırıldığında tahmin etmede daha düşük bir performansa sahiptir [19].

Bayes ağıları basit bir şekilde neden ve sonuç ilişkisini ele aldığı için gittikçe önemli bir veri madenciliği aracı olmaya başlamıştır. Bayes ağıları bu nedensel ilişkileri yönlü döngüsüz haritalar aracılığıyla yapmaktadır [19].

Kalitatif yapıda yönlü döngüsüz haritalardan yararlanılmaktadır. Burada düğümler değişkenleri, yönlü oklar ise koşullu bağımsızlıkları göstermektedir. İki çeşit yapısal öğrenme algoritması vardır. Bunlardan ilki arama ve skor atama tabanlı bir algoritmadır. Bu algoritma istatistiksel olarak elde edilen skor fonksiyonuna bağlıdır. Bu algoritma daha sonra skor fonksiyonunu enbüyükleyen bir yapı seçer. İkinci yapısal öğrenme algoritması ise koşullu bağımsızlık testine bağlı olarak oluşturulan algoritmadır. Daha çok tercih edilen bu algoritma değişkenler arasındaki koşullu bağımsızlığı tahmin eder [20].

Bayes ağlarının bir diğer yapısı da kantitatif yani parametre öğrenme yapısıdır. Bu yapıda her bir değişken için bağımlılık ilişkileri bileşik koşullu olasılık dağılımları ile ifade edilmektedir [20].

Bayes ağlarının avantajları şu şekilde sıralanabilir. İlk olarak Bayes ağları eksik veriyle uygun bir şekilde çalışır. Kayıp değerli örneklerde olasılıkları toplayarak veya birleştirerek sorun giderilebilir. İkinci bir avantaj ise Bayes ağları nedensel ilişkileri öğrenerek için kullanılabilir. Böylece problemin daha iyi kavranmasını sağlamak ve sonuçları tahmin etmeye yarar. Bunların dışında bir ağ oluşturmak zaman kaybına sebep olabilir ve çok fazla çaba gerektirebilir. Fakat Bayes ağlarında öncelikle ağın yapısı belirlenmekte ve daha sonra modele yeni bir değişken kolaylıkla eklenebilmektedir. Son olarak veri kullanıcının ön bilgilerine dayanarak

olasılıksal olarak birleştirildiği için bu yöntem sağlam bir modelin oluşturulmasını sağlar [20].

Kore’de yapılan bir çalışmada Bayes Ağları kullanılarak büyük mühendislik projelerinin risk yönetimi için bir model oluşturulmuştur. Uygulama Kore’de gemi inşaat sektöründe yapılmıştır. Uzman görüşü ve literatür araştırmasına göre 26 farklı risk faktörü belirlenmiştir. Nisan 2007’de 11 gemi inşaat şirketinden 252 uzmana bir anket uygulanmıştır. Bayes ağları proje risk yönetimine uygulanmış ve bir sebep sonuç diyagramı kolayca oluşturulmuştur. Anket sonucu elde edilen verilerle riskler arasında bileşik koşullu olasılıklar hesaplanarak risk olasılıkları elde edilmiştir. Çalışmanın sonucunda büyük ölçekli ve orta ölçekli şirketlerde en yüksek risk faktörleri tasarım değişikliği, işgücü, hammadde tedariği olarak belirlenmiştir. Büyük ölçekli ve orta ölçekli şirketlerde farklı proje performans riskleri belirlenmiştir. Büyük ölçekli şirketlerde zaman çizelgesinde gecikme ve şartname memnuniyetsizliği, küçük ölçekli şirketlerde ise bütçe yetersizliği ve zaman çizelgesi gecikmeleri önemli risk faktörleri olarak belirlenmiştir [18].

Kore’de yapılan diğer bir çalışmada sinema sektörü için Bayes ağları uygulaması yapılmıştır. Çalışmada gişe sonuçlarının tahmini ve belirsizliği azaltmak amaçlanmaktadır. Bu makalede önceki çalışmalardan farklı olarak değişkenler arasındaki nedensel ilişkiler üzerine yoğunlaşmıştır. Gişe başarısının tahmini öncelikle değişkenler belirlenmiştir. Ocak 2005 ile Ekim 2006 tarihleri arasında vizyona girmiş 100 adet filme ait veriler farklı kaynaklardan toplanmıştır. Her bir sürekli değişken sınıflandırma ağaçları (CART) algoritmalarıyla kesikli hale getirildikten sonra yapısal öğrenme süreciyle Bayes ağı oluşturulmuştur. Duyarlılık analiziyle gişe performansını için kritik değişkenler belirlenip başarı olasılığını en çok arttıran değişkenler bulunmuştur. Son olarak Bayes ağları modeli gişe başarısını tahmin etmek için kullanılmış ve yapay sinir ağları ve karar ağaçları algoritmalarının sonuçlarıyla karşılaştırılmıştır. Çalışmanın sonunda Bayes ağlarıyla oluşturulan modelin gişe başarısını daha doğru bir şekilde tahmin ettiği görülmüştür [19].

Bayes ağları bir çok farklı sektörde uygulama alanı bulmuştur. Türkiye’deki ulaşım sektörü için farklı ulaşım şekillerinin kullanımının paylaşımıyla ilgili bir karar destek sistemi oluşturmayı amaçlayan bir çalışma gerçekleştirilmiştir. Nedensel haritanın düğümlerini oluşturacak temel değişkenlerin bulunması için ulaşım sektörüyle uğraşan uzmanların görüşleri alınmıştır. Değişkenler arasındaki nedensel ilişkilerin

belirlenmesi için geleneksel ekonometrik teknikler, yapay sinir ağıları ve uzman görüşünden yararlanılmıştır. Bayes ağlarının kullanımıyla nedensel ilişkiler ebeveyn-çocuk yapısıyla gösterilmiştir. Her bir değişkenin olası durumlarına ait olasılıklar belirlendikten sonra koşullu olasılıklar hesaplanmıştır. Duyarlılık analizinin sonucunda çıkan sonuçlarla ilgili olarak farklı senaryolar geliştirilerek analizler yapılmıştır. Makalede, Türkiye'nin ulaştırma problemi bütün olarak ele alınması ve uzmanların gelecekte vereceği kararlara rehberlik etmesi açısından önemlidir [21].

Sağlık sektöründe uygulanan bir çalışmada ise belirsizlik altında en iyi sonucu verdiği için Bayes ağları yaklaşımı seçilmiştir. Şizofreni hastalığının oluşmasında etiyolojik, genetik ve çevresel faktörlerin etkisi olduğu bilinmektedir. 223 şizofreni hastasından alınan verilerle bir veri seti oluşturulmuştur. Bu çalışmada Bayes ağlarıyla genotiplerin fenotipler üzerindeki etkileri araştırılmıştır. Daha açık olarak hedef değişkenlerin (hastalığa yakalanma yaşı ve sinirsel belirtiler) en küçük ve en büyük değerleri arasında kesikli hale getirebilmektir. Sonuçlara göre seçilen özelliklerin hastalığa yakalanma yaşı ve sinirsel belirtilerin sınıflandırılmasında etkileri olduğu ortaya çıkarılmıştır. Buna göre hastalığa yakalanma yaşı hastaları alt gruplara ayırmak için en önemli klinik özellik olduğu belirlenmiştir [22].

Şirketlerin ARGE bölümünün inovasyon performansı üzerindeki etkisini belirlemeyi amaçlayan bir çalışma Kore'de yapılmıştır. Bunun için 2003 ve 2004 yıllarına ait birtakım veriler kullanılmıştır. Bağımlı değişken ise şirketin inovasyonda başarılı ise 1 ve başarısız ise 0 değerini alır. Öncelikle değişkenlerin teknolojik inovasyon üzerindeki etkisinin olup olmadığı probit model (lojistik regresyon modeline alternatif bir yöntemdir. Bağımlı değişken kategorik olduğunda kullanılır.) ile araştırılmıştır. Burada bağımlı değişken ürün inovasyonu ve süreç inovasyonu olarak iki gruba ayrılmıştır. Bayes ağlarını kullanabilmek için sürekli değişkenler tek seviye karar ağacıyla kesikli hale getirilmiştir. Parametre öğrenme ile bayes ağları koşullu olasılık dağılımları üretilmiştir. Yapısal öğrenmenin sonuçlarına göre ürün inovasyonu için devlet desteği ve süreç inovasyonu için diğer şirketlerle işbirliği içinde olma en önemli faktörler olmuştur. Ayrıca entropi azaltma ile ürün inovasyonu devlet desteğine oldukça duyarlıdır. Ancak diğer şirketlerle işbirliğinin süreç inovasyonu üzerinde önemli etkisi varken süreç inovasyonunun işbirliğine duyarlı olmadığı görülmüştür. Bayes ağlarının inovasyon araştırmasında ilk defa uygulandığı bu makalede Bayes ağlarıyla elde edilen sonuçlar yapay sinir ağları ve

probit model ile karşılaştırılmıştır. Sonuç olarak Bayes ağlarının analiz ve tahmin için en kullanışlı model olduğu anlaşılmıştır [20].

Bir diğer çalışmada ise Türkiye'nin enflasyonun karmaşık yapısının analizi için Bayes ağlarından yararlanılmıştır. İlk olarak enflasyon modeli uzman görüşleriyle bilişsel haritalar yardımıyla yapılandırılmıştır. Ocak 1994 ile Aralık 2001 aralığındaki verilerin ayrıntılı incelenmesiyle bağımlı olasılıklar hesaplanmıştır. Bayes nedensel ağlarıyla Türkiye'nin enflasyon modeli değerlendirilmiştir. Bunlara ek olarak çalışmanın sonunda sistemde oluşacak bir değişikliğin sonuçlarının anlaşılması amaçlanmıştır [23].

İzleyen bölümde Bayes ağlarının temel yapısını oluşturan nedensel haritalar ayrıntılı bir şekilde açıklanmıştır.

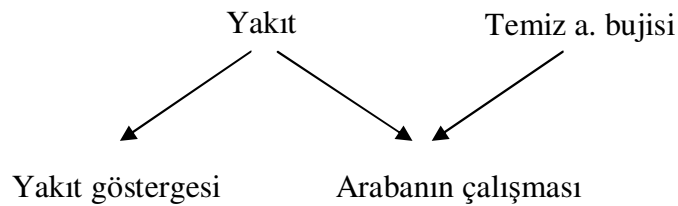
3.1 Nedensel Haritalar

Nedensel haritalar, bir değişkenler kümesi ve yönlendirilmiş bağlantılardan oluşan haritalara denir. Eğer bir haritada A dan B ye bir bağlantı varsa, B ye A nın çocuğu ve A ya da B nin atası denir.

Nedensel haritalarda değişkenler olayları temsil eder. Bir değişken birden fazla durum içerebilir. Örneğin; bir değişken arabanın rengini temsil edebilir. Bu değişken için sarı, kırmızı, mavi, beyaz, siyah gibi farklı durumlar sözkonusudur. Ayrıca değişkenler sayılabilir ya da sürekli durumdan oluşabilir.

Örnek: (Araba Çalıştırma Problemi)

Bir arabanın çalışmamasının birçok sebebi olabilir. Bunlardan bazıları aracın yakıtının kalmaması ve ateşleme bujilerinin kirli olmasıdır. Yakıtın olup olmadığını anlamak için yakıt göstergesine bakmak gerekir. Problemi daha iyi ifade etmek için neden-sonuç içinde ilişkileri içeren bir nedensel harita oluşturalım.

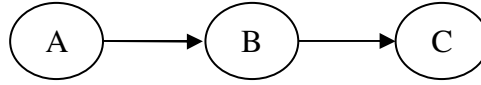


Şekil 3.1 : Araba çalıştırma problemi için nedensel harita.

Yakıt değişkeni için {var, yok}, temiz ateşleme bujisi için {evet, hayır}, yakıt göstergesi için {dolü, 1/2, boş}, arabanın çalışması için {evet,hayır} durumları mevcuttur. Değişkenlerin durumlarını belirledikten sonra arabanın çalışmasını hangi yönde etkilediğini belirleyebiliriz. Şekil 3.1'deki bütün etkilerin pozitif yöndedir. Yani yakıtın olması arabanın çalışmasını pozitif yönde etkiler.

3.1.1 Seri bağlantılar

A, B ve C olaylarından A olayının B olayına etkisi, B olayının da C olayı üzerinde etkisi varsa bu üç olay seri bağlantılıdır denir.

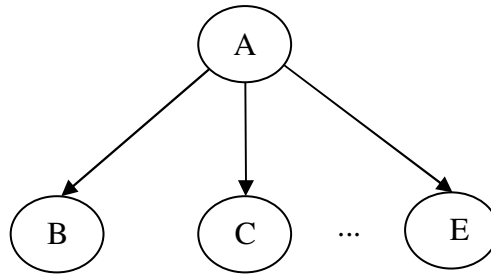


Şekil 3.2 : Seri bağlantı.

Başka bir deyişle A olayı B olayının kesinliğini ve sonuçta C olayının kesinliğini etkiler. C olayı bilindiği takdirde B olayı ve dolayısıyla A olayının bilinmesi sağlanır. Diğer bir yandan B olayı biliniyorsa A ve C olayları arasındaki bağ kopar ve A ile C olayları birbirinden bağımsız hale gelirler.

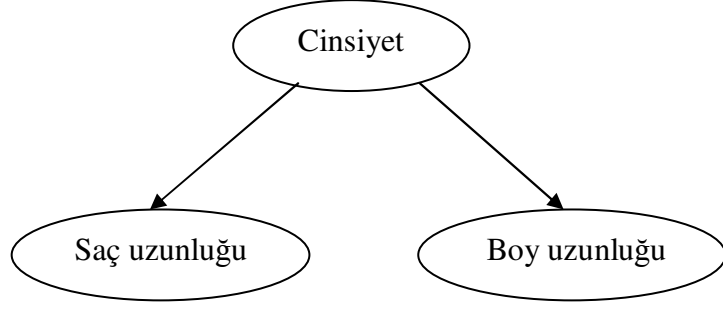
3.1.2 İraksayan bağlantılar

Şekil 3.3'de görülen yapı ıraksayan bağlantıya bir örnektir. A olayı bilinmiyorsa bu olayın B, C,... çocuklarının kesinliği üzerine etkisi vardır. Eğer A olayı biliniyorsa B, C ve diğer olaylar birbirinden bağımsız hale gelir.



Şekil 3.3 : İraksayan bağlantı.

Örnek: Şekil 3.4'de cinsiyet ile saç uzunluğu ve boy uzunluğu arasındaki nedensel ilişki gösterilmektedir.



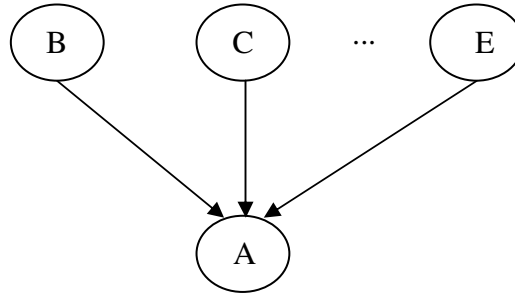
Şekil 3.4 : Cinsiyet, saç ve boy uzunlukları arasındaki nedensel ilişkiler.

Cinsiyet {erkek, kadın}, saç uzunluğu {uzun, kısa}, boy uzunluğu {<1.65, >1.65} durumlara sahip olsun.

Eğer bir insanın cinsiyeti hakkında fikrimiz yoksa saç uzunluğuna bakarak bir varsayımda bulunabiliriz ve bu varsayım aracılığıyla kişinin boy uzunlu hakkında yorum yapabiliriz. Ancak gözlemlenen kişinin erkek olduğunu biliniyorsa saç uzunluğu kişinin boy uzunluğu hakkında fikir vermez.

3.1.3 Yakınsayan bağlantılar

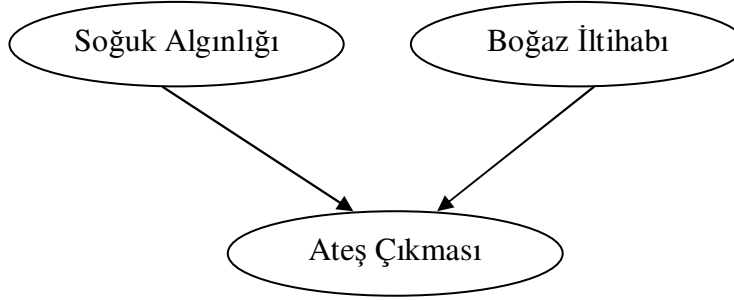
Yakınsayan bağlantılar daha önce bahsedilen bağlantı türlerinden farklıdır. Eğer A olayının haricinde B, C,... olayları bilinmiyorsa bu durumda B, C,... olayları birbirinden bağımsızdır. Yani A olayı gözlemlenebilir ise A'nın ebeveyni olan olayların birbiri üzerine etkisi yoktur.



Şekil 3.5 : Yakınsayan bağlantı.

Bir olayın olası bir nedeni hakkındaki bilgi bu olayın diğer olası nedenleri hakkında bilgi edinmemizi sağlayamaz. Fakat sonuç olan olay bilinmiyorsa bu olayın olası nedenleri hakkında yorum yapılabilir. Örneğin; B ve C olayları A olayının sebepleri olsun. C olayı gerçekleştiğinde B olayının kesinliği azalır. Aynı şekilde C olayı gerçekleşmezse B olayının gerçekleşme olasılığı artar.

Örnek: Şekil 3.6’da soğuk algınlığı ve boğaz iltihabı ile ateş çıkması arasındaki nedensel ilişki gösterilmiştir.



Şekil 3.6 : Soğuk algınlığı, boğaz iltihabı ve ateş çıkması arasındaki nedensel ilişkiler.

Eğer ateş çıkması hakkında bir şey bilmiyorsak bir kişinin soğuk algınlığı geçiriyor olması onun boğaz iltihabı geçirip geçirmediği hakkında bilgi vermez. Ancak bu kişinin ateşinin çıktığı gözlemleniyor ise soğuk algınlığı geçirmediği biliniyorsa boğaz iltihabı olma olasılığı yüksektir.

Nedensel haritalar gibi koşullu bağımsızlık da Bayes ağlarının temelini oluşturur. Bir a olayının olasılığı $P(a)$ verildiğinde aslında bu olasılık bilinen başka faktörlerle koşullandırılmıştır. Örneğin; bir zarın atıldığında 5 gelme olasılığı $1/6$ ’ dır. Aslında bu olasılık zarın hilesiz olduğu koşulu altında doğrudur. Zar hakkında hiçbir şey bildiğimizde zarın hilesiz olduğunu varsayabiliriz. Dolayısıyla ifade ‘zarın hilesiz olduğu bilindiğinde zarın 5 gelme olasılığı $1/6$ ’ dır’ olmalıdır. Bu durumda bir olasılık ifadesi bilinen durumlar üzerine koşullandırılmış ifadelerdir.

Bayes ağlarının ayrıntılarına geçmeden önce matematiksel altyapısı değinmek gereklidir.

3.2 Matematiksel İfadeler

Örnek uzay: bir deneyin gerçekleşmesi mümkün olan tüm sonuçlarına denir. Örnek; bir zar havaya atıldığında 1,2,3,4,5 ve 6 sayıları elde edilecektir. Bu durumda oluşacak örnek uzay $U=\{1,2,3,4,5,6\}$ olacaktır.

Örnek olay: herhangi bir deney için örnek uzayın bir alt kümesine denir. Bir paranın iki kez atılması deneyinde iki yazı elde edilmesi bir olaydır. Aynı şekilde iki tura gelmesi veya bir yazı ve bir tura gelmesi de ayrı olaylardır.

Rasgele olay: gerçekleşmesi tamamen rastlantıya bağlı olan olaya denir. Örnek; bir zarın havaya atıldığında altı gelmesi rastlantıya bağlıdır.

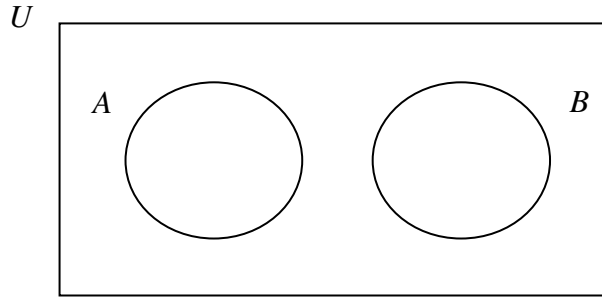
A ve B , U örnek uzayında iki olay olsun.

Birleşim: A veya B olaylarının en az birinin elde edilmesiyle oluşan olaydır. Birleşim $A \cup B$ ile gösterilir.

Kesişim: A ve B olaylarının aynı anda elde edilmesiyle oluşan olaydır. Kesişim $A \cap B$ ile gösterilir.

Tümleyen: A olayının elde edilmemesiyle oluşan olaydır. A olayının tümleyeni $\bar{A}$ ile gösterilir.

Ayrık olay: A ve B olayları aynı anda gerçekleşemiyorsa bu iki olay ayrıktır denir. Bir diğer ifadeyle $A \cap B = \emptyset$ ise A ve B ayrık (mutually exclusive) olaylardır. Örneğin bir paranın atılması olayını ele alalım. $U = \{Y, T\}$ kümesi bu deney için örnek uzaydır. A olayına ‘paranın yazı gelmesi’, B olayına ise ‘paranın tura gelmesi’ diyelim. $A = \{Y\}$, $B = \{T\}$ olaylarının hiçbir ortak noktası olmadığı için $A \cap B = \emptyset$ dir. Yani A ve B olayları ayrık olaylardır.



Şekil 3.7 : Ayrık olaylar.

İki olayın ayrık olması sonlu sayıdaki olaylar için genişletilebilir:

$A_i \cap A_j = \emptyset$ ($i \neq j$) ise $A_1, A_2, \dots, A_n$ olayları ikişerli ayrıktırlar.

Bağımsız olaylar: A ve B olayları U örnek uzayında iki olay olsun.

$$P(A \cap B) = P(A) \times P(B) \quad (3.1)$$

ise A ve B olaylarına bağımsız olaylar denir. Burada

$P(A \cap B)$: A ve B olaylarının birlikte gerçekleşme olasılığı

$P(A)$: A olayının gerçekleşme olasılığı

$P(B)$: B olayının gerçekleşme olasılığıdır.

Örneğin bir para iki kez atılsın. İki kez yazı gelme olasılığını bulmaya çalışalım.

$A = \{ \text{ilk atışta yazı gelmesi} \}$

$B = \{ \text{ikinci atışta yazı gelmesi} \}$

A ve B olayları birbirinden bağımsız oldukları için (3.1)'den

$$P(A \cap B) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

olarak elde edilir.

Bağımlı olay: A ve B olayları U örnek uzayında iki olay olsun.

$P(A \cap B) \neq P(A) \times P(B)$ ise A ve B olaylarına bağımlı olaylar denir.

Bir deney için örnek uzay U olsun. O halde U örnek uzayındaki bir A olayının $P(A)$ olasılığıyla ilgili aşağıdaki aksiyomlar kabul edilir:

1. $P(A) \geq 0$

2. $P(U) = 1$

3. A ve B ayrık olaylar ise $P(A \cup B) = P(A) + P(B)$.

3. aksiyom sonlu yada sonsuz sayıdaki olaylar için genelleştirilebilir:

$A_1, A_2, \dots, A_n, \dots$ bir U uzayında sonlu yada sonsuz sayıda ikişerli ayrık olaylar dizisi olsun. Bu durumda

$$P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots \text{ dir.}$$

Bazı olasılık kuralları ise bu aksiyomlardan yararlanılarak türetilir.

Teorem 1. A ve B bir U örnek uzayına ait olaylar olsun. Eğer A olayı B olayının alt kümesi ise $P(A) \leq P(B)$ dir.

Teorem 2. Herhangi bir A olayı için $P(A) \leq 1$ dir.

Teorem 3. A, U örnek uzayında bir olay olsun. $P(\bar{A}) = 1 - P(A)$ dir.

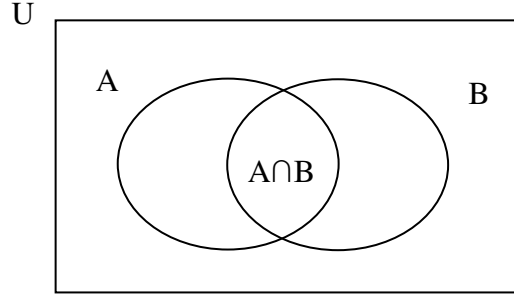
Teorem 4. Boş kümenin olasılığı sıfırdır. $P(\emptyset) = 0$.

Teorem 5. A ve B bir U örnek uzayına ait olaylar olsun.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (3.2)$$

Teorem 6. A_i ler ($j = 1, 2, \dots, n$) bir U uzayına ait olaylarsa $P(U_{j=1}^n A_j) \leq \sum_{j=1}^n P(A_j)$ eşitsizliği mevcuttur.

Bileşik Olasılık : A ve B bir U örnek uzayına ait olaylar olsun. Bileşik olasılık A ve B olaylarının aynı anda gerçekleşmesinin olasılığıdır. Bir başka deyişle A ve B olaylarının bileşik olasılığı $P(A \cap B)$ olasılığına eşittir.



Şekil 3.8 : Bileşik olaylar.

Marjinal Olasılık : A olayı bir U örnek uzayına ait olsun. A olayının marjinal olasılığı ayrık kısımlarının olasılıklarının toplamına eşittir. A olayı aşağıdaki şekilde yazılabilir:

$$A = (A \cap B) \cup (A \cap \bar{B}) \quad (3.3)$$

$A \cap \bar{B}$ ve $A \cap B$ ayrık olaylardır. Bu sebeple olasılığın üçüncü aksiyomundan

$$P(A) = P(A \cap B) + P(A \cap \bar{B}) \quad (3.4)$$

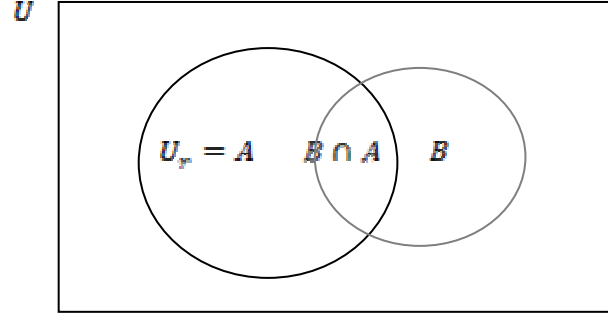
ifadesi A olayının marjinal olasılığını verir.

Koşullu Olasılık: Bir U örnek uzayında A olayının gerçekleştiğini kabul edelim.

A olayının haricinde hiçbir olayın gerçekleşmesin. Bu durumda U örnek uzayı $U_r = A$ uzayına indirgenmiş olur. U örnek uzayındaki B olayı için sadece A olayı ile bağlantılı olan kısmı yani $B \cap A$ söz konusudur.

A olayı bilindiğinde B olayının koşullu olasılığı:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad (3.5)$$



Şekil 3.9 : İndirgenmiş uzay.

3.5 denklemine benzer şekilde

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (3.6)$$

yazabiliriz. Buradan

$$P(A \cap B) = P(B) \times P(A|B) \quad (3.7)$$

elde ederiz. Buna aynı zamanda çarpım kuralı denmektedir.

Eğer A ve B olayları bağımsız olaylar ise

$$P(A|B) = \frac{P(A) \times P(B)}{P(B)} = P(A) \quad (3.8)$$

ve aynı şekilde $P(B|A) = P(B)$ olduğu açıktır.

3.2.1 Bayes teoremi

$P(A \cap B) = P(B) \times P(A|B)$ olduğundan bu ifadeyi denklem 3.6'da yerine koyarsak

$$P(B|A) = \frac{P(B) \times P(A|B)}{P(A)} \quad (3.9)$$

elde ederiz. 3.9 denklemi tek bir olayın olduğu durum için Bayes teoremini verir.

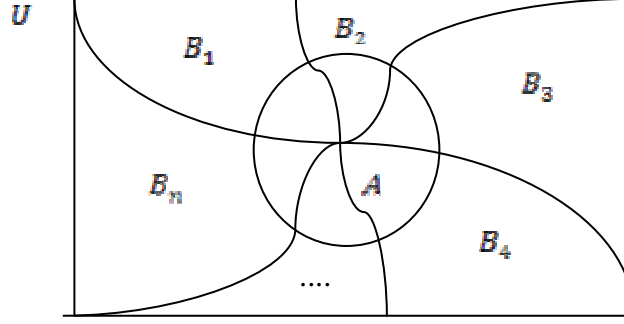
Bir örnek uzayın parçalanışı:

Verilen bir U örnek uzayında n tane olay $B_1, B_2, \dots, B_n$ olduğunu varsayalım. Öyleki

$$\bullet \bigcup_{i=1}^n B_i = U$$

$$\bullet i = 1, 2, \dots, n, j = 1, 2, \dots, n \text{ ve } i \neq j \text{ için } B_i \cap B_j = \emptyset$$

ise $B_1, B_2, \dots, B_n$ olayları U uzayının bir parçalanışını gösterir.



Şekil 3.10 : U uzayının parçalanışı.

U uzayında gözlemlenebilir bir A olayı şu şekilde yazılabilir:

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \dots \cup (A \cap B_n)$$

$i \neq j$ için B_i, B_j ayrık olaylar olduğundan $(A \cap B_i)$ ve $(A \cap B_j)$ ayrıktır. Bu durumda

$$P(A) = \sum_{j=1}^n P(A \cap B_j) \quad (3.10)$$

Bu aynı zamanda toplam kuralı olarak bilinmektedir. A olayının olma olasılığının ayrık parçalarının olasılıklarının toplamına eşit olduğu esasına dayanmaktadır. 3.10 denkleminde çarpım kuralını uygularsak

$$P(A) = \sum_{j=1}^n P(A|B_j) \times P(B_j) \quad (3.11)$$

elde ederiz.

$i = 1, 2, \dots, n$ için $P(B_i|A)$ koşullu olasılığı

$$P(B_i|A) = \frac{P(A \cap B_i)}{P(A)} \quad (3.12)$$

şeklinde. Çarpım kuralını pay kısmına, toplam kuralını da paydaya uygularsak

$$P(B_i|A) = \frac{P(A|B_i) \times P(B_i)}{\sum_{j=1}^n P(A|B_j) \times P(B_j)} \quad (3.13)$$

elde edilir. 3.13 denklemi Bayes Teoreminin uzayın birden fazla parçalanış altında olduğu durumu göstermektedir. Burada U uzayında A olayı gözlemlenebilir, B_i olayları ise gözlemlenemeyen olaylardır.

Örnek: (Monty Hall Problemi)¹

¹ Kaynak: http://en.wikipedia.org/wiki/Monty_Hall_problem.

Bir televizyon kanalında yayınlanan yarışmada yarışmacıya kırmızı, yeşil ve mavi renklerde üç kapalı kapı gösterilmektedir. Kapılardan birinin arkasında araba bulunmaktadır ve yarışmacıdan bir kapıyı seçmesi istenmektedir. Yarışmacının kırmızı kapıyı seçtiğini varsayalım. Hangi kapının arkasında araba olduğunu bilen sunucu yeşil renkli kapıyı açtırır ve kapının arkası boş çıkar. Bu durumda sunucu yarışmacıya kırmızı kapıda hala ısrar edip etmediğini sorar. Buradaki asıl sorun kapıyı değiştirmek araba kazanma şansını arttırır mı?

Açılmamış iki kapı olduğundan şansınızın %50-%50 devam ettiğini düşünebilirsiniz. Ancak durum bundan ibaret değildir.

Kapıların arkasında araba bulunma durumlarını A_k , A_y , A_m ile gösterelim. Başlangıçta bu kapıların arkasında araba bulunma olasılığı $P(A_k) = P(A_y) = P(A_m) = 1/3$ ve kırmızı kapıyı seçtiğimizi varsayalım. Sunucunun yeşil kapıyı açtırmasına B olayı diyelim.

Eğer araba kırmızı kapının arkasında ise sunucu yeşil ya da mavi kapıyı açmakta serbest olacaktır. O halde $P(B|A_k) = 1/2$.

Eğer araba yeşil kapının arkasında ise sunucu mavi kapıyı açacaktır. $P(B|A_y) = 0$.

Eğer araba mavi kapının arkasında ise sunucu yeşil kapıyı açacaktır. $P(B|A_m) = 1$.

Bu durumda (3.10 denkleminde)

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{P(B)} = \frac{\frac{1}{2} \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}$$

$$P(A_y|B) = \frac{P(B|A_y)P(A_y)}{P(B)} = \frac{0 \frac{1}{3}}{\frac{1}{2}} = 0$$

$$P(A_m|B) = \frac{P(B|A_m)P(A_m)}{P(B)} = \frac{1 \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$

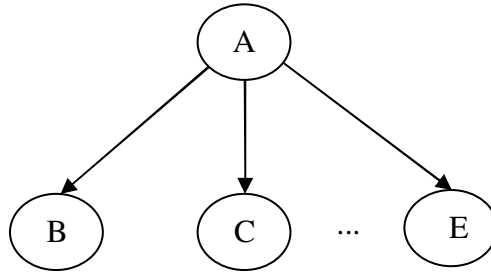
Görüldüğü gibi arabanın mavi kapı arkasında olma olasılığı daha yüksek çıkmıştır. Bu yüzden mavi kapıyı seçilmelidir.

3.3 Bayes Ağlarının Genel Yapısı

- Değişkenlerden oluşan bir küme ve değişkenler arasındaki yönlü doğrultuların oluşturduğu bir küme mevcut,
- Her bir değişken sonlu sayıda birbirinden ayrık durumlardan oluşuyor,
- Değişkenler ve yönlü doğrultuların oluşturduğu döngüsüz bir harita oluşturuyor,
- $B_1, B_2, \dots, B_n$ ebeveynlerine sahip her bir A düğümü için koşullu olasılık tabloları $P(A|B_1, B_2, \dots, B_n)$ oluşturuluyor

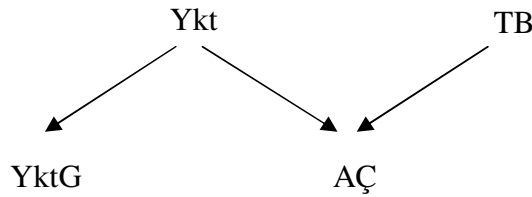
ise bu yapıya Bayes Ağı denir.

Eğer A'nın ebeveyni yoksa tablo koşullu olmayan olasılık tablosuna indirgenir. Örneğin Şekil 3.11'deki gibi bir ağda A düğümüne ait ön olasılıklar belirlenmelidir. Bu sadece matematiksel açıdan değil insanların kesinlik için nedenselliği ortaya koymasını sağlayan bir gerekliliktir.



Şekil 3.11 : Örnek bir ağ yapısı.

Araba çalıştırma problemi için Bayes Ağı oluşturmaya çalışalım.



Şekil 3.12 : Araba çalıştırma problemi için Bayes Ağı.

Bu modelde $P(Ykt)$, $P(TB)$, $P(AÇ|Ykt, TB)$, $P(YktG|Ykt)$ olasılıklarına ihtiyaç vardır. Başlangıçta $P(Ykt) = (0.98, 0.02)$ ve $P(TB) = (0.96, 0.04)$ olarak belirlensin. Diğer koşullu olasılık çizelgeleri aşağıda verilmiştir.

Çizelge 3.1 : $P(YktG|Ykt)$ koşullu olasılık çizelgesi.

	Ykt = Var	Ykt = Yok
YktG = Dolu	0.39	0.001
YktG = 1/2	0.60	0.001
YktG = Boş	0.01	0.998

Çizelge 3.1’de araçta yakıt olması veya olmaması durumuna göre yakıt göstergesinin durumu gösterilmektedir. Örneğin, yakıt varsa yakıt göstergesinin dolu gösterme olasılığı %39, yarısını dolu gösterme olasılığı %60 ve boş gösterme olasılığı 0.01’dir.

Çizelge 3.2 : $P(AÇ|Ykt, TB)$ koşullu olasılık çizelgesi.

	Ykt = Var	Ykt = Yok
TB = evet	(0.99,0.01)	(0,1)
TB = hayır	(0.01, 0.99)	(0,1)

Çizelge 3.2’de gösterilen parantez içindeki ikili sayılar AÇ değişkeni için (evet, hayır) olasılıklarını temsil etmektedir. Örneğin; aracın ateşleme bujileri temiz ve yakıtı varken aracın çalışma olasılığı 0.99, çalışmama olasılığı ise 0.01 dir. Aracın çalışması ve çalışmaması birbirinden ayrık durumlar olduğundan ikili sayıların toplamı olan 1’e eşit olacaktır.

3.3.1 Bayes ağları için zincir kuralı

$U = \{A_1, A_2, \dots, A_n\}$ değişkenler uzayı ve $P(U) = P(A_1, A_2, \dots, A_n)$ birleşik olasılıkları mevcut olsun. Bu durumda $P(A_i)$ ve $P(A_i|e)$ (e, A hakkında elde bulunan bilgi) olasılıkları hesaplanabilir. Fakat değişken sayısının artmasıyla $P(U)$ eksponansiyel olarak artacak ve U kümesinin eleman sayısı çok artmadan dahi olasılıklar tablosu izlenemez ölçüde büyüyecektir. Bu yüzden $P(U)$ nun daha küçük bir gösterimine ihtiyaç vardır.

Bayes Ağı, U üzerinde bu çeşit bir gösterimdir. Eğer Bayes Ağındaki koşullu olasılıklar U için geçerli ise $P(U)$ bu olasılıklardan hesaplanabilir.

Teorem: BA, $U = \{A_1, A_2, \dots, A_n\}$ üzerinde tanımlı bir Bayes Ağı olsun. Bu durumda $P(U)$ birleşik olasılık dağılımı BA’da belirtilen tüm olasılıkların çarpımıdır. Yani, $pa(A_i)$ A_i ’nin ebeveyn kümesi olmak üzere

$$P(U) = \prod_i P(A_i|pa(A_i)) \quad (3.14)$$

İspat: (teoremin ispatı matematiksel tümevarım yöntemiyle yapılacaktır.)

$n = 1$ olduğu durum: U 'daki tek eleman A olsun. Bu durumda $P(U) = P(A)$ olacağı açıktır.

Zincir kuralının $n - 1$ değişkenden oluşan bütün ağlar için doru olduğunu varsayalım ve U, n değişkenli bir yönlü döngüsüz harita için uzay olsun. Ağ döngüsüz olduğundan çocuğu olmayan en az bir A değişkeni mevcuttur.

A değişkeninin ağdan çıkartıldığını varsayalım. Tümevarım hipotezinden $P(U \setminus \{A\})$, $P(A|pa(A))$ dışında tanımlı bütün olasılıkların çarpımına eşittir.

$$P(U) = P(A|U \setminus \{A\}) \cdot P(U \setminus \{A\})$$

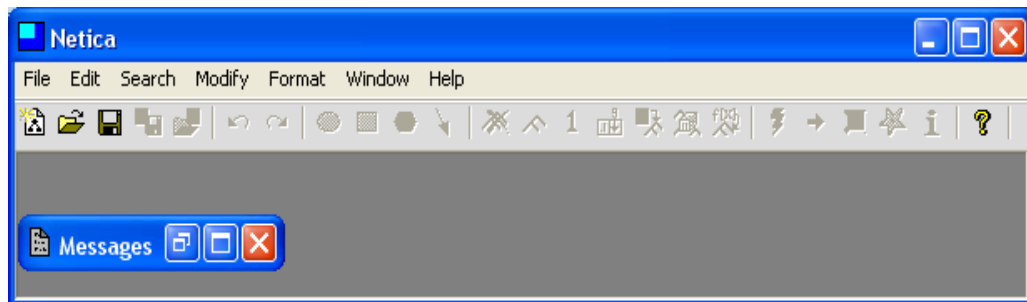
$U \setminus \{A\}$, U uzayından A değişkeni çıkarıldığında elde edilen tüm değişkenleri göstermektedir. $pa(A)$ verildiğinde

$$P(U) = P(A|U \setminus \{A\}) \cdot P(U \setminus \{A\}) = P(A|pa(A)) \cdot P(U \setminus \{A\}) \quad (3.15)$$

Bu eşitlik değişkenlerin birbirinden bağımsız olması özelliğine dayanarak elde edilmiştir. $P(A|U \setminus \{A\})$, A ebeveynleri dışındaki diğer değişkenlerden bağımsız olduğundan $P(A|pa(A))$ 'ya eşittir. 3.15 denkleminin sağ tarafı tanımlı tüm olasılıkların çarpımıdır.

3.3.2. Bayes ağlarının Netica'da modellenmesi

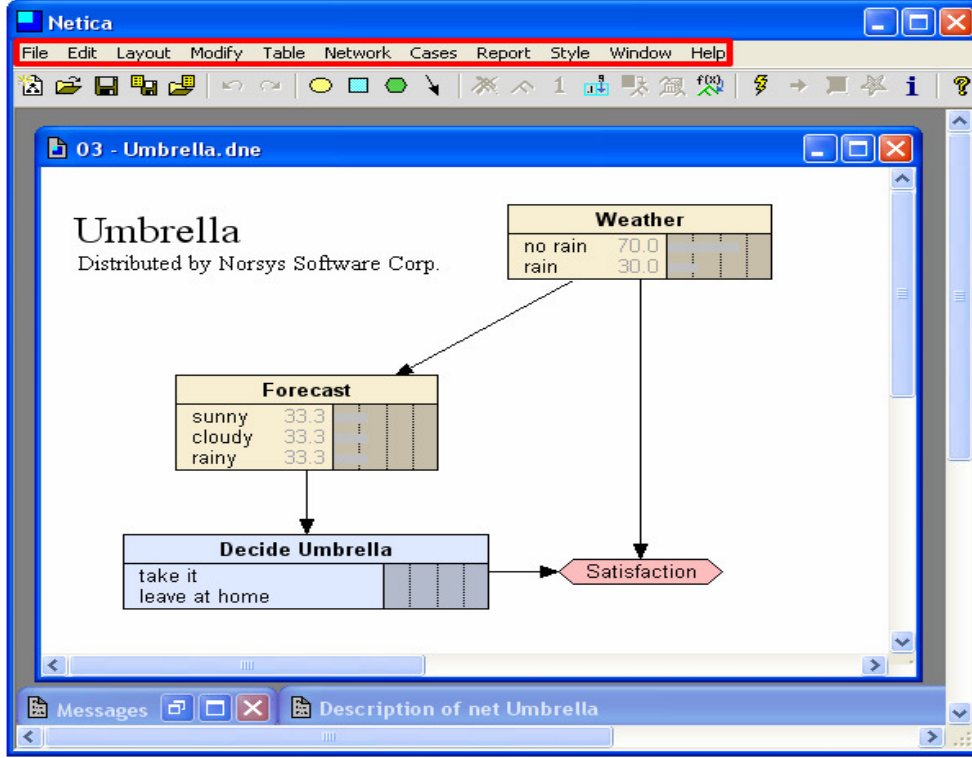
Netica; Bayes ağları, karar ağları ve nedensel haritaların uygulamasını yapmak için yapılmış bir yazılım paketidir. Norsys Software Corp. tarafından piyasaya sürülmektedir. Programı indirebilmek için www.norsys.com adresine girildikten sonra Downloads menüsünün altında programın en yeni sürümü indirilebilir. Netica yazılımı aslında ücretli olup ücretsiz edinmek istenildiğinde sınırlı versiyona sahiptir. Sınırlı versiyon indirildiği takdirde program normal işleyişinde çalışacak ancak çok fazla düğüme (15 düğüm ve üzeri) sahip modelleri çalıştırmayacaktır. Netica programı kurulduktan sonra açılan pencere Şekil 3.13'deki gibidir.



Şekil 3.13 : Netica Programı.

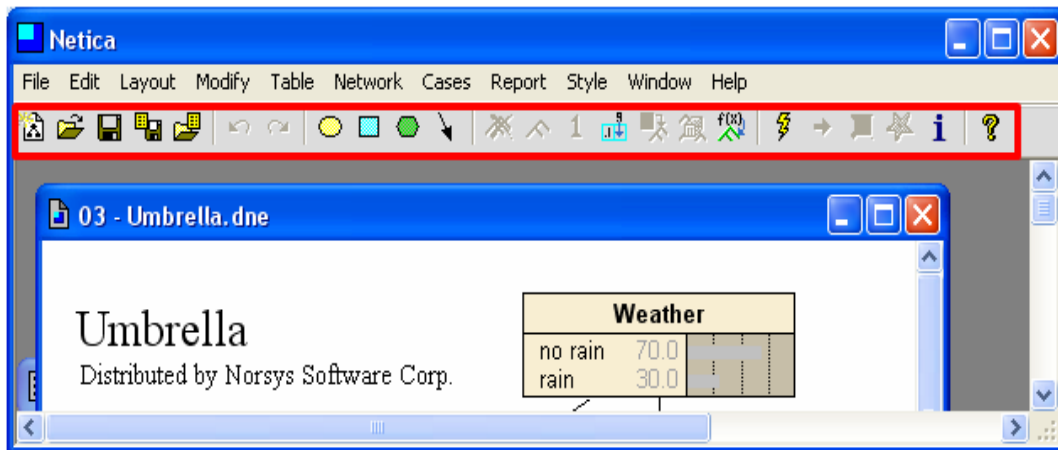
Messages kısmında kullanıcının program çalışırken yaptığı her adım not edilmektedir. Ayrıca programın verdiği hataları da buradan görülebilir.

Netica'da bir ağ açıldığında zaman zaman menü genişler. (Şekil 3.14)



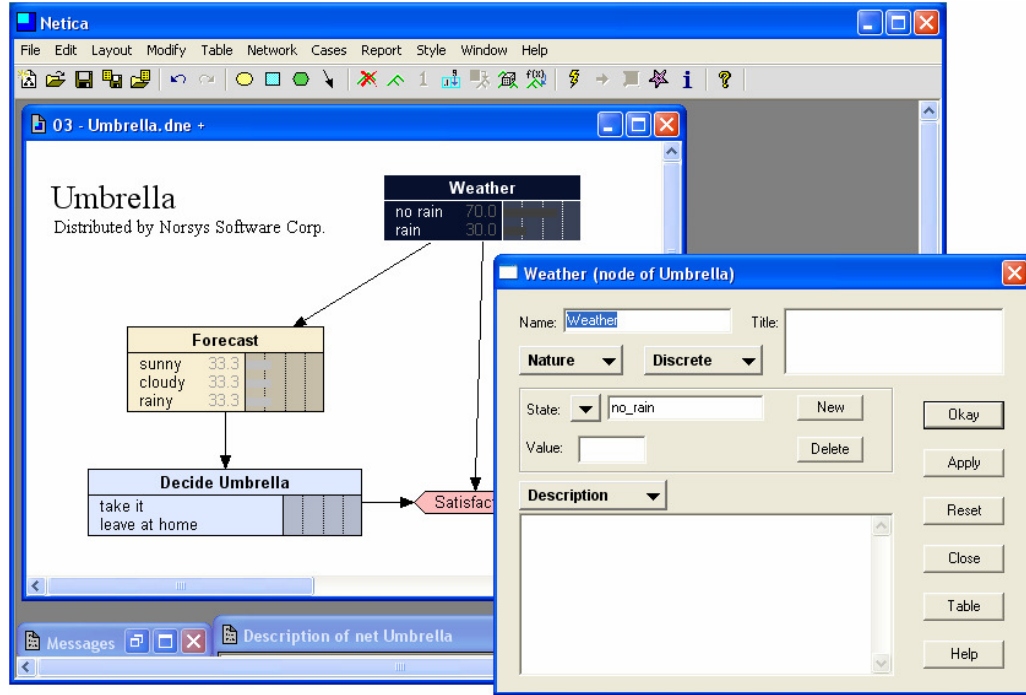
Şekil 3.14 : Netica'da menü gösterilimi.

Menünün altında yer alan araç çubuğunda ise bazı fonksiyonlara ait kısa yollar bulunmaktadır. (Şekil 3.15)



Şekil 3.15 : Netica'da kısa yolların gösterilimi.

Açılan ağ modelinde bir düğümün üzerine çift tıklandığında düğümün özelliklerini gösteren bir pencere açılacaktır. (Şekil 3.16)



Şekil 3.16 : Düğüm özelliklerinin değiştirildiği pencere.

Name kısmına düğümün adı girilir. Bu kısımda Türkçe karakterler ve boşluk kullanılamaz. Title kısmına eğer bir başlık girilirse düğüm başlığı ağ modelinde gösterilecektir. Burada boşluk kullanılabilir. Eğer girilmediyse düğümün sadece adı gösterilecektir. Name kısmının altında yer alan butona basıldığında düğümün çeşidi (nedensel, karar, fayda düğümü) belirlenir. Hemen yanındaki butonda ise düğümün kesikli veya sürekli olduğu gösterilir. Bayes ağlarındaki düğümlerin hepsi kesikli olmak durumundadır. Bu yüzden düğüm sürekli olarak seçilirse value kısmında değerler aralık verilerek girilir. Böylece sürekli bir düğüm kesikli hale getirilmiş olunur. State yazan kısımda ise düğüme ait durumlar belirlenir. Description yazan yerde ise modele ait açıklamalar yazılabilir.

Bir ağdaki düğümler modelin değişkenlerini veya faktörlerini temsil eder. Dört çeşit düğüm vardır:

- Nedensel düğüm: standart değişkenlerdir.
- Karar düğümü: belli kararların faydasını temsil eden değerleri içeren düğümdür.

Bir karar düğümü fayda düğümü olmadan kullanılamaz. Karar düğümü kesikli olmak zorundadır.

- Fayda düğümü: belli çıktıların fayda değerlerini içeren düğümdür. Fayda düğümü karar düğümü olmadan kullanılamaz. Fayda düğümü sürekli olmak zorundadır.

- Sabit

Verilen şemsiye örneğinde 2 tane nedensel düğüm, bir tane karar ve bir tane fayda düğümü vardır.

Yeni bir nedensel düğüm oluşturmak için

- araç çubuğundaki  üzerine tıklanır.
- sayfada boş alana tıklanır.

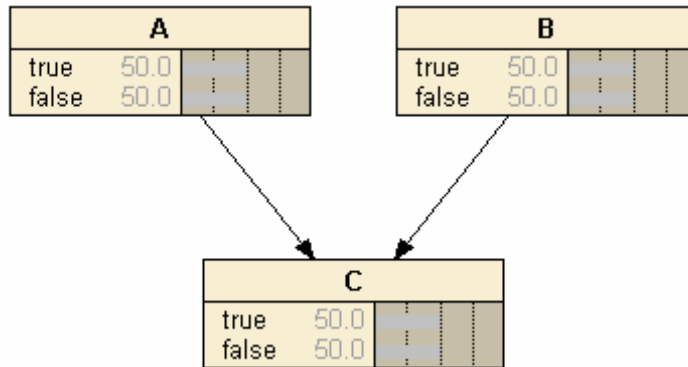
Eğer hızlı bir şekilde birden fazla nedensel düğüm eklemek için

- araç çubuğundaki  üzerine çift tıklanır.
- sayfa üzerine tıkladıkça yeni bir düğüm eklenir.
- Düğüm eklemeyi sonlandırmak için tekrar  üzerine tıklanır.

Düğüm üzerine çift tıklayarak açılan pencerede düğümün özellikleri belirlenir.

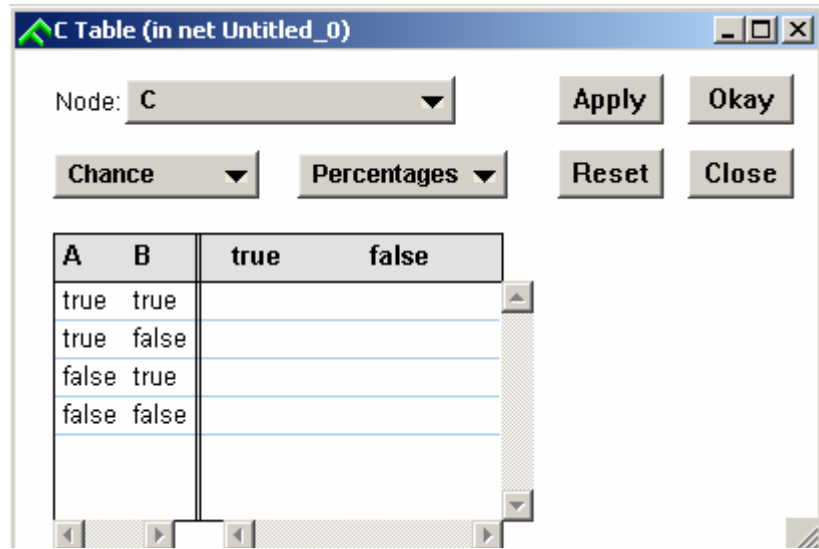
İki düğüm arasındaki ilişkiyi göstermek için araç çubuğunda  üzerine tıklanır. Düğümlerin arasındaki neden-sonuç ilişkisine göre sürükle bırak tekniğiyle bağlantı oluşturulur. Burada dikkat edilmesi gereken husus bağlantıların yönlü bir döngü oluşturmamasıdır.

Koşullu olasılık tabloları aralarında bağlantı oluşturulan düğümler arasındaki ilişkileri açıklayan tablolardır. Aşağıdaki tabloda C düğümü A ve B düğümlerinin çocuğudur. A, B ve C düğümlerinin üçü de true, false durumlarına sahiptir.



Şekil 3.17 : Netica’da basit bir BA örneği.

O halde C düğümüne ait koşullu olasılık çizelgesi aşağıdaki gibidir.

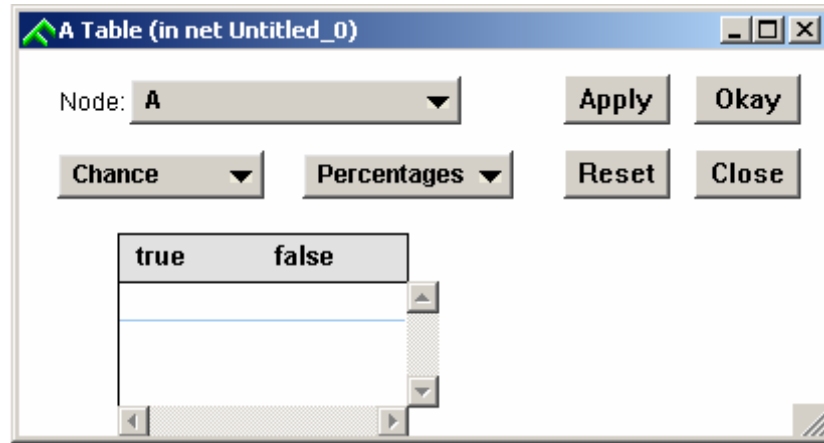


A	B	true	false
true	true		
true	false		
false	true		
false	false		

Şekil 3.18 : C düğümüne ait koşullu olasılık çizelgesi.

İlk satıra bakılacak olunursa A ve B doğruyken C nin doğru olma ve yanlış olma şansı yüzde kaçtır sorusuna ait durum belirtilmiştir.

A ve B düğümüne ait olasılık çizelgeleri de vardır. Bu düğümün ebeveyn düğümleri olmadığı için sadece kendi durumlarının olma olasılıkları yer alacaktır.



	true	false

Şekil 3.19 : A düğümüne ait koşullu olasılık çizelgesi.

Bir düğümüne ait olasılık tabloları iki şekilde görüntülenebilir:

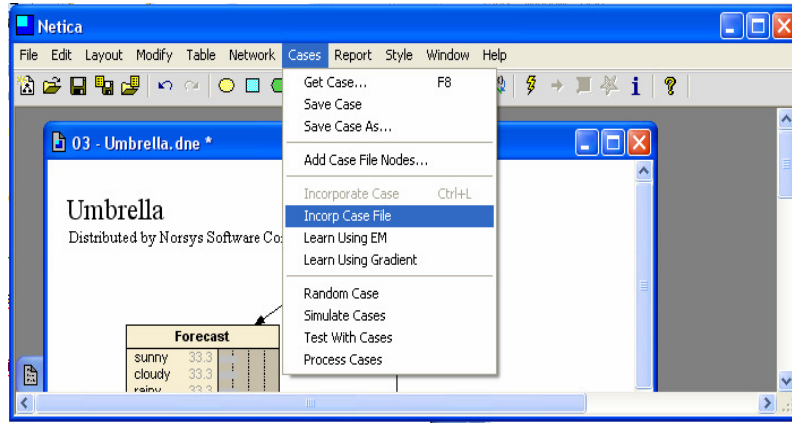
- Düğümün üzerine iki kez tıklanır ve açılan pencerede Table tıklanır.
- Düğüm seçildikten sonra araç çubuğundaki  üzerine tıklanır.

Olasılık tabloları kullanıcı tarafından elle girilebilir. Bunun için olasılık çizelgesi açıldıktan sonra olasılıklar yazılabilir.

Burada dikkat edilmesi gereken nokta her bir satırın toplamı 1 (yüzdesel olarak %100) e eşit olmalıdır. Eğer eşitlik olmazsa Okay butonuna basıldıktan sonra program hata mesajı verecektir.

Netica olasılık çizelgelerindeki olasılıkları bir veri kümesinden alabilir. Bunun için Excel’de bir çalışma sayfası açılır. İlk satıra (A1 hücresine) // ~->[CASE-1]->~ yazılır. İkinci satıra ise ağdaki düğümlerin adları yazılır. A2 hücresine ağın ilk düğüm, B2 hücresine ikinci düğüm ve benzer şekilde diğer düğümlerin adları yazılır. Üçüncü satırdan itibaren veriler hücrelere girilir. Eğer kayıp veri varsa asterisk * kullanılmalıdır. Netica’nın verileri alabilmesi için veri dosyasının uzantısı ‘Text (Tab delimited) (*.txt)’ uzantılı olarak kaydedilmesi gerekir.

Netica’da ağ oluşturulduktan sonra verilerin alınması için menüde Cases seçeneğinin altında Incorp Case File seçilir.



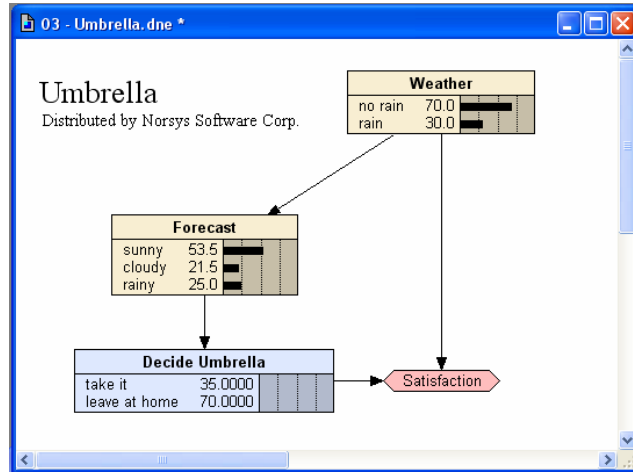
Şekil 3.20 : Verilerin farklı bir veri kümesinden alınması.

Verilerin olduğu dosya seçildikten sonra Enter Degree başlıklı bir pencere açılır. Buraya mümkün olduğunca büyük sayılar (bir milyon gibi) girilmelidir. Bu Netica’nın dosyadaki her bir veriyi girilen sayı kadar okumasını sağlar.

Ağdaki bütün olasılık tabloları tamamlandıktan sonra olasılık dağılımlarını görebilmek için program derlenir. Bunun için iki yol vardır:

- Network menüsünden Compile seçilir.
- Araç çubuğunda  tıklanır.

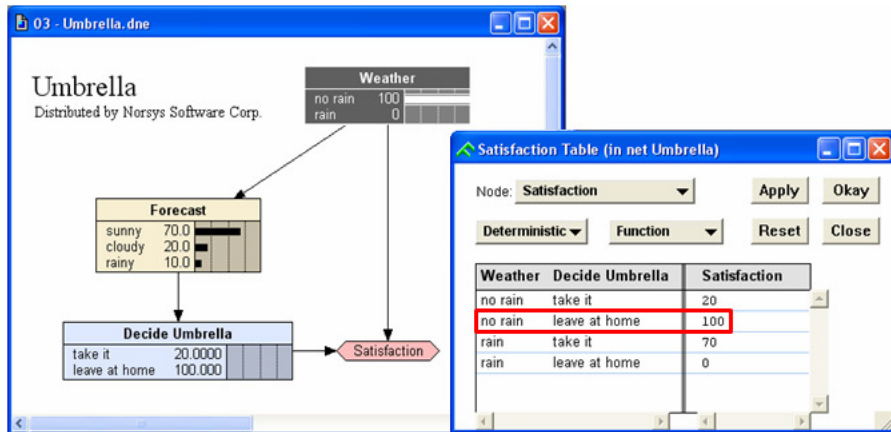
Her bir düğüm için olasılık dağılımları aşağıdaki şekilde elde edilir.



Şekil 3.21 : BA'nın derlenmesi.

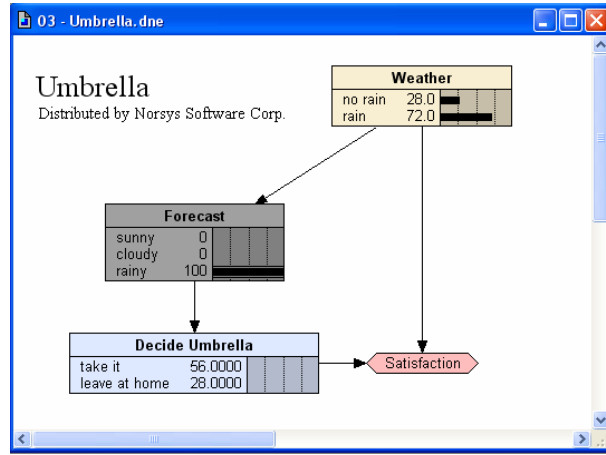
Program derlenip sonuçlar alındıktan sonra düğümlerdeki durumlara göre senaryo analizi ve tahmin yapılabilir. Bunun için bir düğümün üzerine bir kez tıklanır ve istenen durumu seçmek için bu durumun üzerine tıklanır. Örneğin 'weather' düğümünde 'no rain' durumu seçilsin. Bu durumda sadece 'no rain' olduğu durum için ağ güncellenir.

Havanın yağmurlu olmadığı bilindiğinde havanın güneşli olma olasılığı 0.70, bulutlu olma olasılığı 0.20 ve yağmurlu olma olasılığı 0.10 dur. Buna göre şemsiyeyi evde bırakma kararı alınmalıdır. Bu seçimden sağlanacak fayda (satisfaction) Şekil 3.22'den görüldüğü gibi %100 dür.



Şekil 3.22 : Fayda düğümünün olasılık çizelgesi.

Aynı şekilde çocuk düğüm için bir durum seçilirse ebeveyn düğümler sayesinde nedenler ortaya çıkartılabilir. Örneğin hava durumu yağmurlu olarak tahmin ediliyorsa %72 ihtimalle hava yağmurludur.



Şekil 3.23 : Havanın yağmurlu olması durumu.

4. KULLANILAN YÖNTEMLER

Tez çalışması boyunca Bayes ağının oluşturulması için karar ağaçları, korelasyon analizi ve eşdoğrusallık analizinden yararlanılmıştır. Bu bölümde yararlanılan yöntemler ayrıntılı şekilde anlatılmıştır.

4.1 Karar Ağaçları

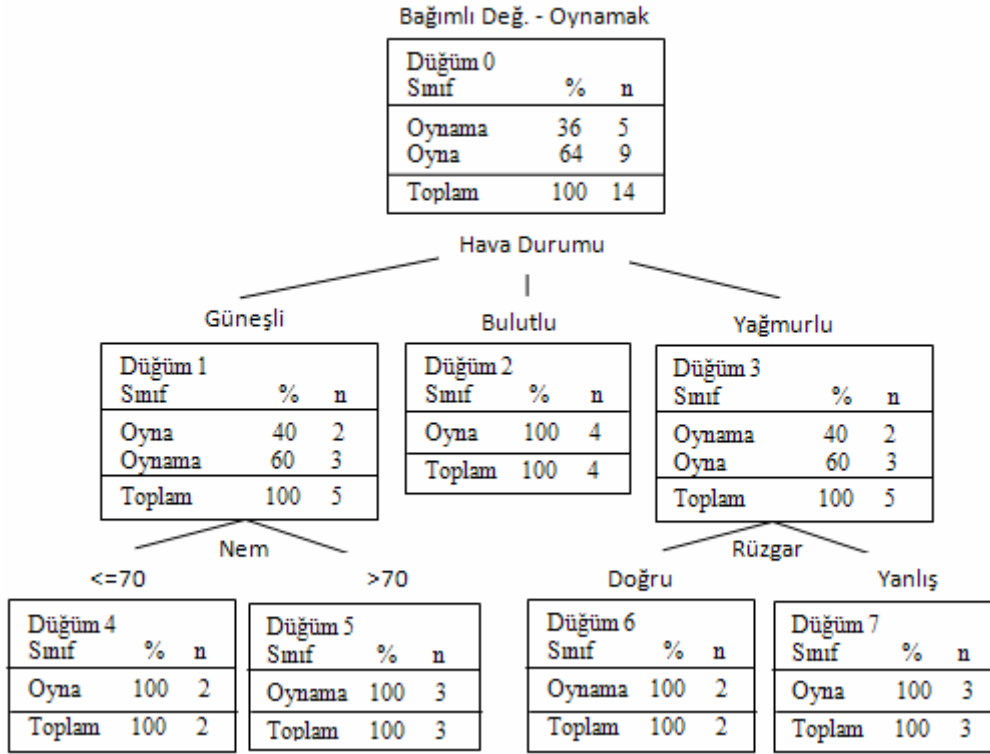
Karar ağaçları yönteminde ağaç yapısında bir sınıflandırma modeli oluşturulur. Karar ağaçları ile durumlar gruplara ayrılır veya bağımsız değişkenlerin değerlerine dayanarak bağımlı değişkenin değerini tahmin edilir.

İki tip karar ağacı mevcuttur:

- Sınıflandıma ağaçları: bütün kayıtları alır ve her birini bir sınıfa atar. Sınıflandırmanın doğru olup olmadığını güvenilirliği hesaplayarak verir.
- Regresyon ağaçları: sayısal bir değer alan değişkenin değerini tahmin eder.

Bütün karar ağaçları aslında aynı yapıdadır. Verilere ağaç modeli uygulandığında her bir kayıt ağacın yönünde bir dizi testle ilerler. Bu süreç her bir kayıt bir yaprağa yerleşene kadar devam eder.

Şekil 4.1’de basit bir karar ağacı görülmektedir. Bu karar ağacı tenis oynayıp oynamamaya karar vermek için kullanılacaktır.



Şekil 4.1 : Karar Ağacı.

Örneğin; hava güneşli ve nem oranı yüksek ise tenis oynanmayacak, hava yağmurlu ve rüzgar hafif ise tenis oynanacaktır.

Karar ağaçları aşağıda sözü edilen amaçlar için kullanılır:

Segmentasyon: insanların hangi grubun üyesi olacağını belirler.

Katmanlara ayırma: durumları birkaç kategoriye atamayı sağlar. Örneğin yüksek, orta ve düşük risk grupları.

Tahmin: kurallar üreterek bunları geleceği tahmin etmede kullanır. Bir arabanın veya evin ikinci el fiyatının tahmini.

İlişkinin belirlenmesi: sadece belli alt gruplarda olabilecek ilişkilerin ortaya çıkarılmasını sağlar.

Sürekli değişkenleri kesikli hale getirme: sürekli değişkenler minimum veri kaybıyla kesikli değişken haline getirilir.

Her bir kaydın bir kategoriye atanması için genellikle karar ağaçları kullanılır. Daha rahat anlaşılması, açıklanması ve SQL gibi dillere çevrilmesinden dolayı karar ağaçları iyi bir seçimdir.

Karar ağaçları sayısal girdilerle kullanıldığında veri hazırlanması sırasında veri daha az zarar görür. Diğer tekniklerde (kümeleme, sinir ağları) kayıp daha fazladır. Bir diğer faydası ise önemli değişkenlerin ağacın üst kısımlarında yer almasıdır.

Karar ağaçlarında kategorik verilerin kullanılması bazen sorun yaratabilir. Kategorik veriler çalı gibi bir ağaca sebep olabilirler. Diğer algoritmalar sınıfları birleştirerek çok sayıda sınıftan daha az sayıda sınıfın oluşmasını sağlayabilirler. Sınıf sayısı arttıkça olası gruplar çok hızlı büyüdüğünden araştırma pratik olmayan bir hal alır.

Yaygın olarak kullanılan karar ağacı algoritmalarından bazıları şunlardır:

CHAID (Chi-squared Automatic Interaction Detection): bu algorithmada her bir adımda bağımlı değişken üzerinde en çok etkisi olan bağımsız değişken seçilir. Bu algoritmanın etkin olabilmesi için büyük veri kümelerine ihtiyaç vardır. Çoklu regresyon analizine bir alternatif olarak özellikle veri kümesi regresyon analizine uygun olmadığında kullanılır.

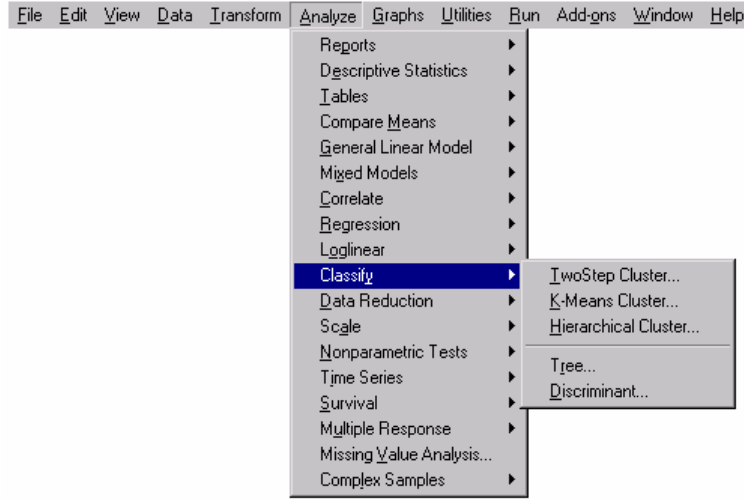
CRT (Classification and Regression Trees): bağımlı değişkenin kategorik veya sayısal olmasına bağlı olarak sırasıyla sınıflandırma veya regresyon ağacı oluşturulur. Bu algoritma, bağımlı değişkene göre veriyi olabildiğince homojen bir şekilde gruplara ayırır. Bir kural oluştuktan sonra çocuk düğümlerde aynı şekilde uygulanır. Herbir gözlem sadece ve sadece bir düğüme atanır.

QUEST (Quick, Unbiased, Efficient Statistical Tree): diğer algoritmalara göre daha hızlı bir şekilde çalışır. Sadece bağımlı değişkenin nominal olduğu durumda kullanılır.

4.1.1 SPSS ile Karar Ağacı Uygulaması

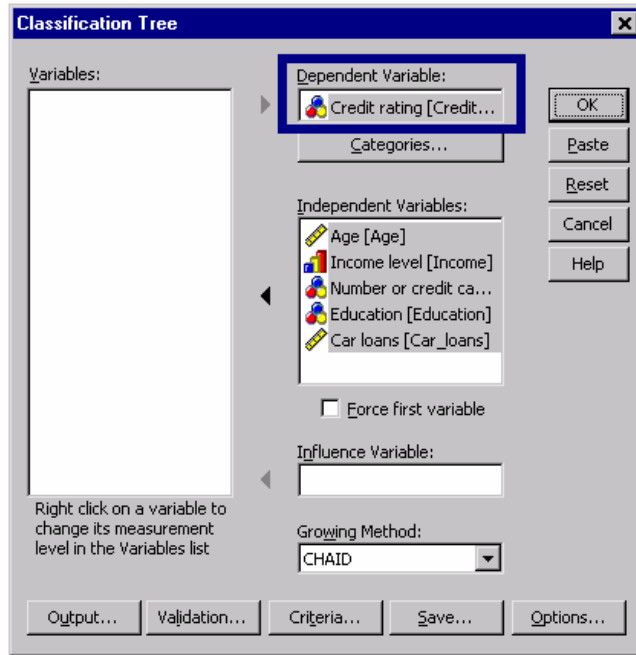
Bir bankada müşterilerin kredi alıp ödeyip ödemeyenlere ait geçmiş verilerin tutulduğu bir veritabanı bulunmaktadır. Sınıflandırma ağaçları kullanarak bu iki tip müşteri grubuna ait özellikler belirlenebilir ve böylece oluşturulan model ile kredi talep edenlerin aldıkları krediyi ödeyip ödemeyecekleri tahmin edilebilir.

Bu örnek için CHAID algoritması kullanılacaktır. SPSS programı açıldıktan sonra menüden ‘Analyze’ komutundan ‘Classify’ ve ardından ‘Tree’ komutu seçilebilir.



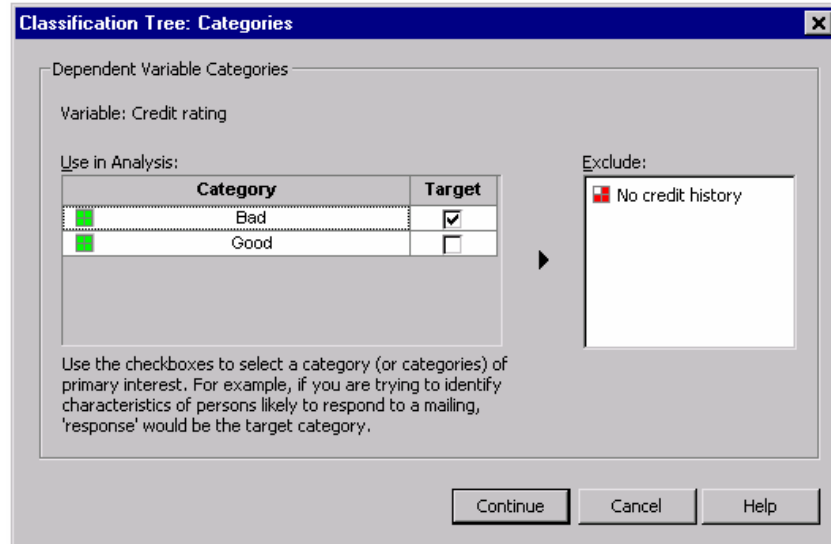
Şekil 4.2 : Karar ağacının seçilmesi.

‘Credit rating’ değişkeni bağımlı değişken olarak seçilmiştir. Diğer değişkenler ise bağımsız değişken olarak seçilmiştir.



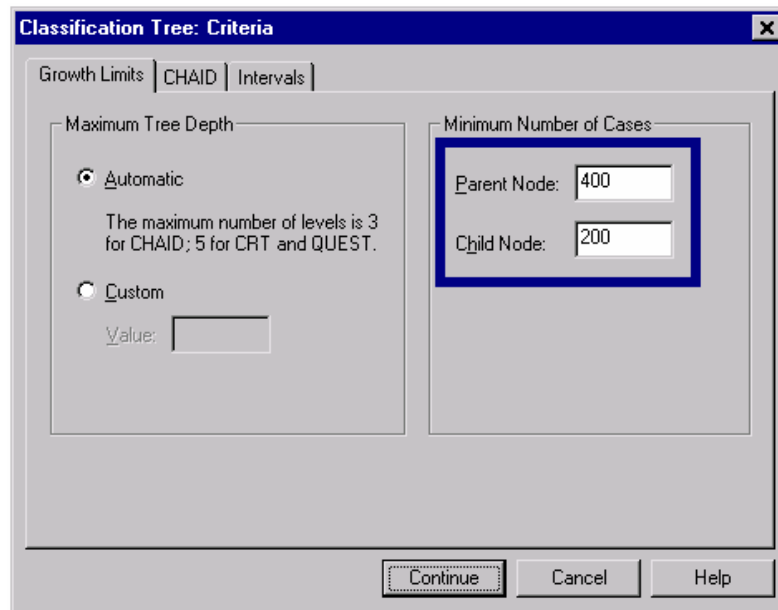
Şekil 4.3 : Değişkenlerin seçilmesi.

Seçilen bağımlı değişkenin altındaki ‘Categories’ butonuna basıldığında bağımlı değişkenin gösterilmesi istenen hedef bölümleri belirlenir. Bu bölümlerin seçilmesi ağaç modelinin yapısını değiştirmez ancak çıktıda farklılıklar yaratır.



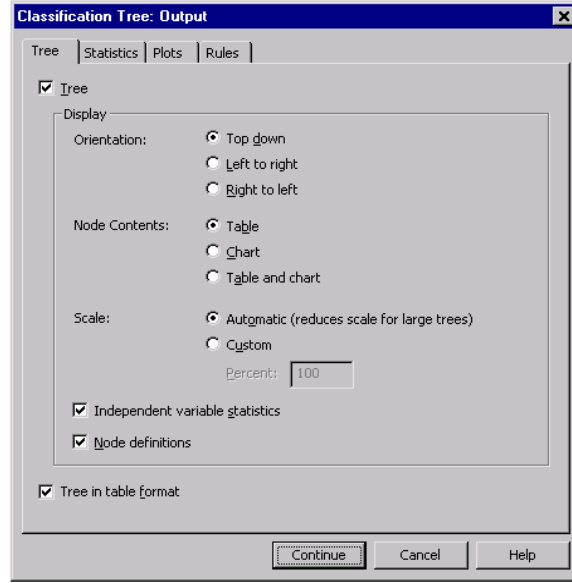
Şekil 4.4 : Hedef bölümlerin seçilmesi.

Açıklanan örnek için ağacın yapısının basit olması için ebeveyn ve çocuk sayısının belli sayıda olması istenmiştir. Bunun için 'Criteria' butonuna basıldığında minimum sayıda durum ebeveyn düğümler için 400, çocuk düğümler için 200 olarak belirlenmiştir.



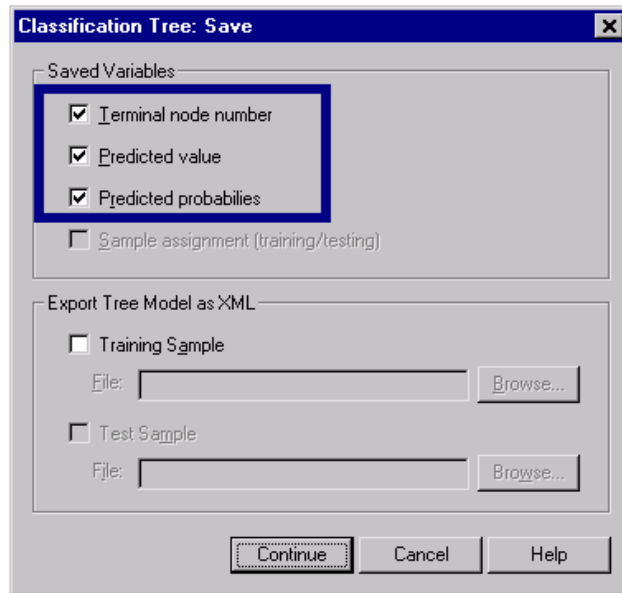
Şekil 4.5 : Düğüm sayısının belirlenmesi.

Modelin çıktısını özelleştirmek için alt kısımdaki 'Output' butonuna basılır. 'Tree' sekmesinde 'Tree in table format' seçilir. 'Plot' sekmesi tıklanır. Burada 'Gain' ve 'Index' seçilir. 'Plot' sekmesi bir veya birden fazla hedef bölümler seçilmeden çalışmaz.



Şekil 4.6 : Model çıktısının özelleştirilmesi.

Modelin tahminiyle ilgili bilgi içeren değişkenler kaydedilebilir. Bu örnek için her bir durum için tahmin edilen kredi puanı kaydedilip daha sonra bu tahminler gerçek değerleriyle karşılaştırılabilir. Açılan pencerede ilk üç durum seçilir. (Şekil-6)



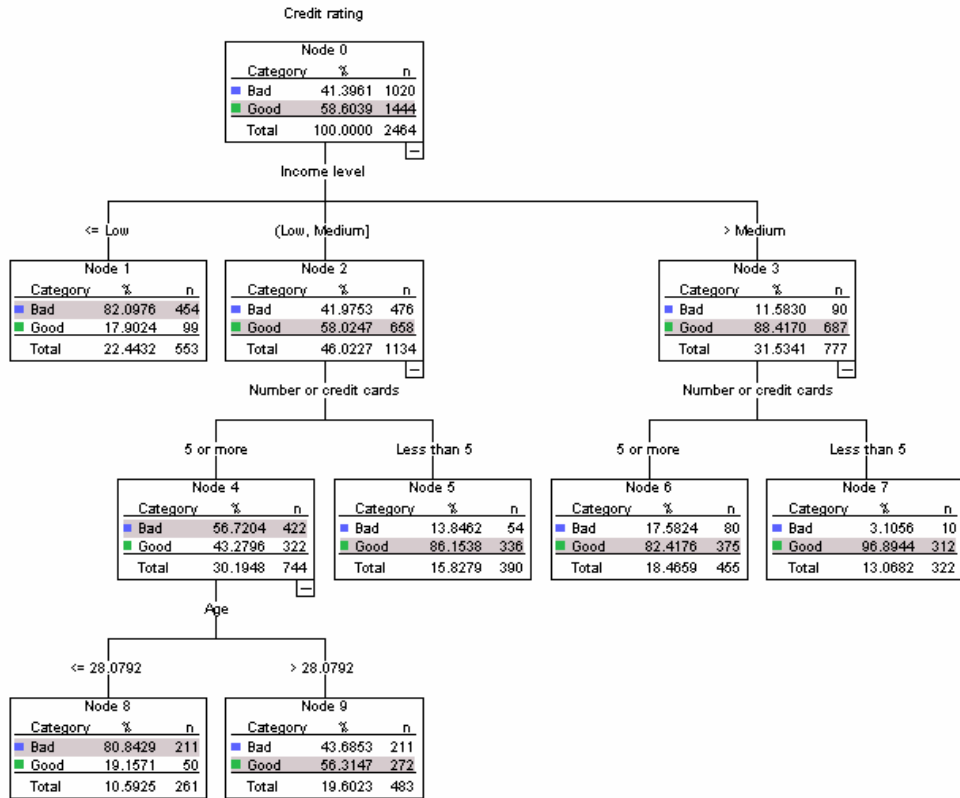
Şekil 4.7 : Tahmin değerlerinin kaydedilmesi.

‘OK’ butonuna basılarak model çalıştırılır. Programın çıktısına bakıldığında özet tabloda istenen özellikler ve model sonucuyla ilgili genel bir bilgi verilir.

Specifications	Growing Method	CHAID
	Dependent Variable	Credit rating
	Independent Variables	Age, Income, Credit cards, Education, Car loans
	Validation	NONE
	Maximum Tree Depth	3
	Minimum Cases in Parent Node	400
	Minimum Cases in Child Node	200
	Independent Variables Included	Age, Income, Credit cards
Results	Number of Nodes	10
	Number of Terminal Nodes	6
	Depth	3

Şekil 4.8 : Model sonucu ile ilgili genel bilgi.

Şekil 4.9’da karar ağacının grafik hali gösterilmektedir. CHAID algoritmasına göre kredi puanlamasını (credit rating) en iyi açıklayan değişken gelir seviyesi (income level)dir. Düşük gelir seviyesi tek önemli değişken gelir seviyesidir. Banka müşterilerinin %82’si kredilerini geri ödememiştir.



Şekil 4.9 : Örnek karar ağacı.

Orta ve yüksek gelir seviyelerinde ise ikinci en önemli değişken kredi kartı sayısı (number of credit cards)dır. Orta gelir seviyesindeki 5 veya daha fazla karta sahip olanlar için tahmin edici bir değişken daha sözkonusudur: yaş (Age). Yaşları 28 ve

28'den küçük olan %80'den fazla müşteri kötü kredi skoruna sahiptir. 28 yaşından büyük olanların neredeyse yarısı kötü kredi skoruna sahiptir.

Şekil 4.10'da gösterilen ağaç çizelgesinde ise grafik modeldeki bilgiler çizelge halinde gösterilmektedir. Her bir düğüm için

- Bağımlı değişkenin her bir bölümü için durumların sayıları ve yüzdeleri gösterilmektedir.
- Bağımlı değişken için tahmin kısımları gösterilmektedir.
- Her bir düğüm için ebeveyn düğümünün numarası gösterilmektedir. Burada düşük gelir seviyesini gösteren 1. düğüm hiçbir düğümün ebeveyni değildir. Çocuk düğümü yoktur.

Node	Bad		Good		Total		Predicted Category	Parent Node
	N	Percent	N	Percent	N	Percent		
0	1020	41.4%	1444	58.6%	2464	100.0%	Good	
1	454	82.1%	99	17.9%	553	22.4%	Bad	0
2	476	42.0%	658	58.0%	1134	46.0%	Good	0
3	90	11.6%	687	88.4%	777	31.5%	Good	0
4	422	56.7%	322	43.3%	744	30.2%	Bad	2
5	54	13.8%	336	86.2%	390	15.8%	Good	2
6	80	17.6%	375	82.4%	455	18.5%	Good	3
7	10	3.1%	312	96.9%	322	13.1%	Good	3
8	211	80.8%	50	19.2%	261	10.6%	Bad	4
9	211	43.7%	272	56.3%	483	19.6%	Good	4

Şekil 4.10 : Örnek karar ağacına ait çizelge gösterimi.

Şekil 4.11'de ise açıklayıcı değişkenler için ki-kare değeri, serbestlik derecesi ve anlamlılık seviyesi gösterilmektedir. Son sütunda ise bağımsız değişkenlerin ele alınan düğüm için değerleri gösterilmiştir.

Primary Independent Variable				
Variable	Sig.	Chi-Square	df	Split Values
Income level	.000	662.457	2	<= Low
Income level	.000	662.457	2	(Low, Medium]
Income level	.000	662.457	2	> Medium
Number or credit cards	.000	193.113	1	5 or more
Number or credit cards	.000	193.113	1	Less than 5
Number or credit cards	.000	38.587	1	5 or more
Number or credit cards	.000	38.587	1	Less than 5
Age	.000	95.299	1	<= 28.0792
Age	.000	95.299	1	> 28.0792

Şekil 4.11 : Açıklayıcı değişkenlerin değerleri.

Şekil 4.12’de modeldeki çocuğu olmayan düğümler için bir özet çizelge sunulmaktadır.

- ‘Node N’ her bir terminal düğümdaki (çocuğu olmayan düğüm) durum sayısını ve ‘Node Percent’ ise her bir düğümdaki durum sayısının toplam durum sayısı içindeki yüzdesini gösterir.
- ‘Gain N’ ise her bir terminal düğümdaki hedef kategorinin durum sayısını ve ‘Gain Percent’ ise bunun tüm durumlar içindeki yüzdesini göstermektedir. bu örnek için hedef kategori kötü kredi skoru olarak seçilmişti.
- ‘Response’ ise düğümdaki hedef kategorideki durum sayısının yüzdesidir.
- Kategorik bağımlı değişkenler için ‘Index’ hedef kategorinin ‘Response’ değerinin tüm kategoriye oranıdır. ‘Index’ %100’den büyük ise hedef kategoride daha çok değişkenin olduğunu gösterir.

Node	Node		Gain		Response	Index
	N	Percent	N	Percent		
1	553	22.4%	454	44.5%	82.1%	198.3%
8	261	10.6%	211	20.7%	80.8%	195.3%
9	483	19.6%	211	20.7%	43.7%	105.5%
6	455	18.5%	80	7.8%	17.6%	42.5%
5	390	15.8%	54	5.3%	13.8%	33.4%
7	322	13.1%	10	1.0%	3.1%	7.5%

Growing Method: CHAID
Dependent Variable: Credit rating

Şekil 4.12 : Açıklayıcı değişkenlerin değerleri.

Şekil 4.13’de ‘Risk’ ve ‘Classification’ çizelgeleri ile modelin ne derece iyi olduğu gösterilmektedir. Risk tahmini 0.205 ile verinin %20.5’inin yanlış tahmin edildiği anlaşılmaktadır. ‘Classification’ tablosu ‘Risk’ tablosu ile tutarlıdır. %79.5 ile doğru tahmin edilen müşteri sayısı gösterilmektedir. Bunun yanı sıra bu tabloda kötü skora sahip olanların %35’inin yanlış olarak tahmin edildiği görülmektedir.

Risk	
Estimate	Std. Error
.205	.008

Growing Method: CHAID
Dependent Variable: Credit rating

Observed	Predicted		
	Bad	Good	Percent Correct
Bad	665	355	65.2%
Good	149	1295	89.7%
Overall Percentage	33.0%	67.0%	79.5%

Growing Method: CHAID
Dependent Variable: Credit rating

Şekil 4.13 : Risk ve sınıflandırma çizelgeleri.

Veri kümesine dört yeni değişken eklenmiştir. ‘NodeID’ ile terminal düğümün numarası gösterilmektedir. ‘PredictedValue’ ile bağımlı değişkenin durumunu göstermektedir. Burada 0 ile kötü, 1 ile iyi kredi skoru verilmiştir. ‘PredictedProbability’ müşterinin hangi kategoriye ait olma olasılığını verir. Örneğin birinci düğümde müşterilerin %82’si kötü kategoride, %18’i ise iyi kategoride tahmin edilmiştir.

Category	%	n
Bad	82.0976	454
Good	17.9024	99
Total	22.4432	553

	PredictedProbability_1	PredictedProbability_2
1	.44	.56
2	.81	.19
3	1.00	.00
4	1.00	.00
5	.90	.10
6	.90	.10
7	.90	.10

Şekil 4.14 : Değişkenlerin tahmini olasılıkları.

4.2 Korelasyon Analizi

İki değişken arasındaki karşılıklı ilişkiyi bulmak için korelasyon analizi uygulanır. Bir değişkendeki değişimin diğer bir değişkendeki değişimin ne kadarını açıkladığını korelasyon katsayısı göstermektedir. Korelasyon ile karşılıklı ilişkinin varlığı gösterilirken sebep sonuç ilişkisi gösterilmez.

N birimden oluşan bir kümede her birim için x ve y değerleri mevcut olsun. Bu iki değişkenin birlikte değişme derecesi Pearson katsayısı (ρ) ile gösterilir.

$$\rho = \frac{Cov(x,y)}{\sigma_x \sigma_y} \quad (4.1)$$

formülü ile hesaplanır. Burada σ_x ve σ_y terimleri X ve Y'nin standart sapmaları olup X ile Y arasındaki varyans

$$Cov(x, y) = \frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y}) \quad (4.2)$$

Örnek kütle n 'den korelasyon katsayısı hesaplanırsa

Toplam kareler toplamı:

$$SS_T(x, x) = \sum_{i=1}^n (x_i - \bar{x})^2 \quad (4.3)$$

$$SS_T(y, y) = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (4.4)$$

$$SC_T = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (4.5)$$

Bu ara formüllerden korelasyon katsayısı r şu şekilde hesaplanır:

$$r = \frac{SC_T}{\sqrt{SS_T(x, x)} \sqrt{SS_T(y, y)}} \quad (4.6)$$

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{[\sum_{i=1}^n (x_i - \bar{x})^2][\sum_{i=1}^n (y_i - \bar{y})^2]}} \quad (4.7)$$

Korelasyon katsayısı -1 ile 1 arasında bir değer olup pozitif değerler değişkenlerden biri artarken (azalırken) diğerinin de artması (azalırken), negatif değerler değişkenlerden biri artarken (azalırken) diğerinin azalması (artması) ve sıfır değeri ise iki değişken arasında ilişkinin olmadığını gösterir.

İki değişken arasındaki ilişkinin derecesi bilinse bile bir testle bu ilişkinin geçerli olup olmadığına bakılmalıdır.

$$H_0: r = 0, H_1: r \neq 0$$

Serbestlik derecesi $n - 2$ olan bir t-testi ile korelasyon katsayısının geçerliliği test edilebilir.²

$$t = \frac{r\sqrt{(n-2)}}{\sqrt{(1-r^2)}} \quad (4.8)$$

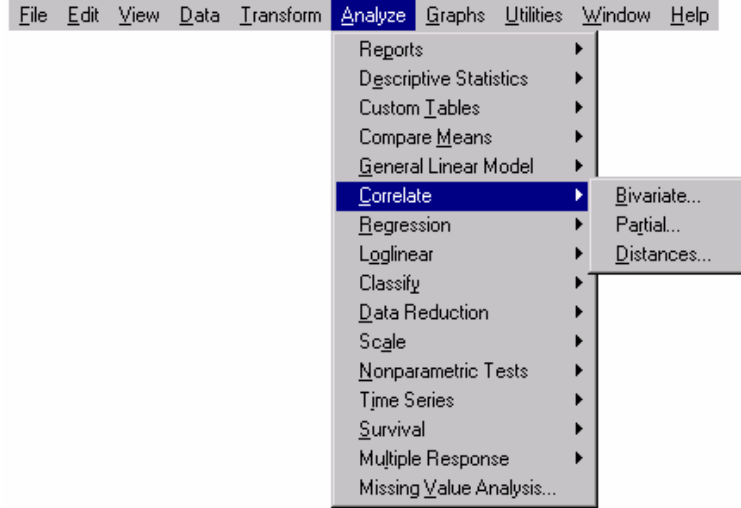
4.2.1 SPSS ile korelasyon analizi uygulaması

Satışları arttırmak için motorlu araç tasarımcıları müşterilere göre aracın hangi özelliğinin daha önemli olduğu araştırılmaktadır. Örneğin yakıt verimliliği satışa

² Bir korelasyon katsayısının zayıf çıkması anlamsız olması anlamına gelmez. Anlamlılığı belirleyen sadece r değeri değildir.

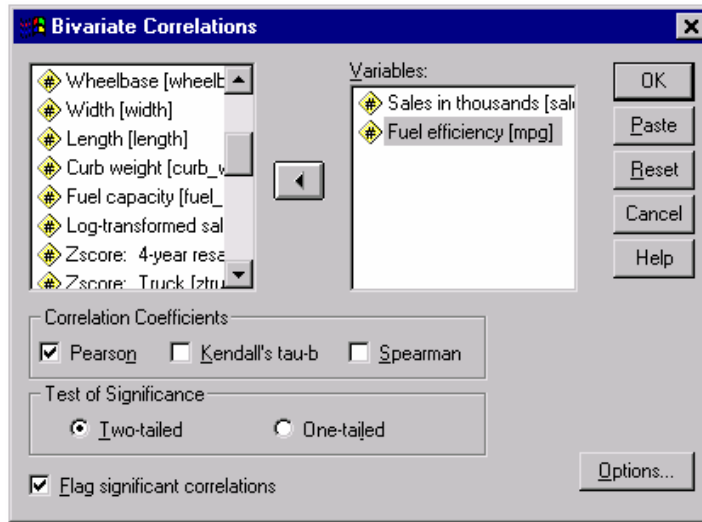
göre ne kadar önemlidir? Geçmiş satış verilerine bakarak satış ile yakıt verimliliği arasındaki korelasyona bakmak mümkündür.

SPSS programı açılıp veri kümesi girildikten sonra sırasıyla ‘Analyze’, ‘Correlate’ ve ‘Bivariate’ seçilir.



Şekil 4.15 : Korelasyon analizi için programın açılması.

Açılan ekranda ‘Sales in thousands’ ve ‘Fuel efficiency’ değişkenler olarak seçilir.



Şekil 4.16 : Korelasyon analizi için değişkenlerin seçilmesi.

‘OK’ butonuna basılarak yeni pencerede ‘Sales in thousands’ ve ‘Fuel Efficiency’ değişkenlerine ait korelasyon matrisi görülür.

		Sales in thousands	Fuel efficiency
Sales in thousands	Pearson Correlation	1	-.017
	Sig. (2-tailed)	.	.837
	N	157	154
Fuel efficiency	Pearson Correlation	-.017	1
	Sig. (2-tailed)	.837	.
	N	154	154

Şekil 4.17 : Satış ve yakıt verimliliğine ait korelasyon matrisi.

Çizelgedeki değerlere göre iki değişken arasındaki korelasyon -0.017 çıkmıştır. Fakat p –değeri 0.837, 0.10 dan büyük çıkmıştır. Korelasyon anlamlı değildir. Demekki tasarımcıların çalışmalarında yakıt verimliliğine dikkat etmelerine gerek yoktur çünkü satışlar üzerinde belli bir etkisi olmadığı görülmüştür.

Tasarımcıların araştırmak istediği bir diğer konu ise araçların beygir gücünün satışların üzerinde etkisinin olup olmadığı olsun. Daha önce bahsedildiği üzere SPSS programı kullanılırsa aşağıdaki şekildeki çıktı elde edilir. Burada da daha önceki örnekte olduğu gibi korelasyon negatif değerli çıkmıştır ancak yapılan testle bu iki değişken arasındaki ilişki anlamlı çıkmıştır.

		Sales in thousands	Horsepower
Sales in thousands	Pearson Correlation	1	-.198*
	Sig. (2-tailed)		.013
	N	157	156
Horsepower	Pearson Correlation	-.198*	1
	Sig. (2-tailed)	.013	
	N	156	156

*. Correlation is significant at the 0.05 level (2-tailed).

Şekil 4.18 : Satış ve beygir gücüne ait korelasyon matrisi.

4.3 Eşdoğrusallık analizi

Bağımlı değişkeni daha iyi açıklamak için modele yeni bir bağımsız değişken eklediğimizde bu değişkenin sadece bağımlı değişken üzerinde değil diğer bağımsız değişkenler üzerinde de etkisi vardır. Eşdoğrusallık (collinearity) korelasyonda olduğu gibi bu defa iki bağımsız değişken arasında ilişkiyi temsil eder. Çoklu eşdoğrusallık ise üç veya üçten fazla değişken arasındaki korelasyonu ifade eder. Eşdoğrusallığı ifade eden iki ölçüt vardır: ‘Tolerance’ ve ‘Variance Inflation Factor’.

‘Tolerance’ bir bağımsız değişkenin diğer bağımsız değişkenler tarafından açıklanamamasının ölçütüdür. ‘Tolerance’ değeri ne kadar küçükse ele alınan bağımsız değişken diğer bağımsız değişkenler tarafından o kadar büyük ölçüde açıklanabilir. Eşdoğrusallığın bir diğer öltü olan ‘Variance Inflation Factor’ ise basitçe ‘Tolerance’ değerinin tersidir. Örneğin; ‘Tolerance’ değeri 0.75 olan bir değişken için ‘Variance Inflation Factor’ değeri $1/0.75=1.33$ olur. Demekki yüksek derecede eşdoğrusallık düşük ‘Tolerance’ seviyesi ve yüksek ‘Variance Inflation Factor’ değeri ile mümkündür.

4.3.1 SPSS ile eşdoğrusallık analizi uygulaması

Daha önce sözü edilen araç tasarımcıları aracın beygir gücünün satış miktarına etkisi olup olmadığını araştırmak istemiş ve bunun için korelasyon analizi uygulamışlardır. Aynı zamanda aracın fiyatının da satış miktarını etkileyip etkilemediği merak edilmiştir. Araç fiyatının modele dahil edilmesiyle satış üzerindeki etkisini araştırmak için öncelikle diğer bağımsız değişken olan beygir gücüyle aralarında ilişki olup olmadığına bakılmalıdır.

SPSS programı açılıp veriler girildikten sonra sırasıyla ‘Analyze’, ‘Regression’ ve ‘Linear’ seçilir. Açılan pencerede bağımlı değişken kısmına ‘Sales in thousands’, bağımsız değişkenler kısmına ‘Horsepower’ ile ‘Price in thousands’ yerleştirilir. Daha sonra ‘Statistics’ butonuna basılarak ‘Collinearity Diagnostics’ kutucukları işaretlenir. Elde edilen program çıktısında aşağıdaki şekilde gösterilen modele dahil edilen değişkenler gösterilmektedir.

Variables Entered/Removed^b

Model	Variables Entered	Variables Removed	Method
1	Price in thousands, Horsepower	.	Enter

a. All requested variables entered.

b. Dependent Variable: Sales in thousands

Şekil 4.19 : Eşdoğrusallık analizinde kullanılan değişkenler.

Aşağıdaki şekilde ise eşdoğrusallık ölçütleri olan ‘Tolerance’ ve ‘VIF’ değerleri görülmektedir. Görüldüğü gibi ‘Tolerans’ seviyeleri düşük, ‘VIF’ değerleri ise

yüksektir. Ayrıca t –testinin sonucu da göz önüne beygir gücü değişkeninin satış miktarını açıklamakta anlamsız olduğu görülmüştür. Bunun nedeni modele sonradan eklenen fiyat değişkeni ile beygir gücü değişkeni arasında korelasyonun olmasıdır.

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B		Collinearity Statistics	
	B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1									
(Constant)	69.754	19.678		3.545	.001	30.876	108.631		
Horsepower	.246	.170	.204	1.445	.151	-.090	.582	.295	3.392
Price in thousands	-2.269	.673	-.476	-3.370	.001	-3.600	-.939	.295	3.392

a. Dependent Variable: Sales in thousands

Şekil 4.20 : Eşdoğrusallık analizinin sonuç çizelgesi.

5. UYGULAMA

Bu tez çalışmasında telekomünikasyon sektöründe karşılaşılan iptal sorununa yönelik bir yaklaşım ele alınmıştır. Gelişen teknolojik koşulların ve beraberinde getirdiği rekabetin bir sonucu olarak telekomünikasyon endüstrisini müşteri odaklı bir hizmet anlayışı içine sokmuştur. Dolayısıyla abone iptali odaklanılan en önemli konulardan biri haline gelmiştir. Bu çalışmada telekom sektöründe abone iptaline sebep olan en önemli faktörler belirlenmeye çalışılmıştır. Bunun için Bayes ağlarından yararlanılmıştır. Bayes ağlarıyla abone iptalini etkileyen değişkenler belirlendikten sonra çeşitli senaryolar oluşturularak iptal eden abonelerin özellikleri incelenmiştir.

Çalışma boyunca izlenen adımlar şunlardır:

Verilerin toplanması

Verilerin hazırlanması

Sürekli değişkenlerin kesikli hale getirilmesi

Değişkenler arasındaki ilişkinin gösterilmesi

Problemin Bayes ağlarıyla modellenmesi

Senaryo analizi.

5.1 Verilerin Toplanması

Tez çalışmasında kullanılan veriler Türkiye'nin önde gelen telekomünikasyon şirketlerinden birinden alınmıştır. Uzun süredir telekomünikasyon piyasasının içinde olan bu şirket saygın ve köklü bir konuma sahiptir. Abonelerine ait oldukça fazla büyüklükteki verileri 'Oracle' veri tabanında tutmaktadır. Bu çalışmada kullanılan veriler şirketin Veri Ambarı bölümünde çalışan uzmanların izniyle veritabanından elde edilmiştir. Ocak 2008-Haziran 2008 tarihleri arasında kapsayan toplam 2000 aboneye ait veri alınmıştır. Alınan veriler 'Microsoft Excel 2007' programına aktarılmıştır. Veritabanındaki ilk 15 aboneye ait veriler ekte verilmiştir.

5.2 Verilerin Hazırlanması

Verilerin yapılacak olan analize uygun hale getirilmesi için belli bir hazırlık aşaması gerekmektedir. Değişkenlere ait ayrıntılı açıklama ve yapılan değişiklikler şöyledir.

Yerleşim: çalışmada ele alınan abonelerin sürekli ikamet ettikleri yere ait verilerdir. Kesikli bir değişken olup 0 ile abonenin kırsal bölgede, 1 ile abonenin şehirde yaşadığı belirtilmektedir. Kırsal kesimde sabit hatlara duyulan gereksinimden dolayı bu kesimde iptal oranının düşük, şehirde yaşayan aboneler için seçeneklerin fazlalığından dolayı iptal etmeye daha eğilimli oldukları görülmektedir. Bu yüzden iptal oranıyla abonelerin yerleşim yerleri arasında bir ilişki olduğu anlaşılmaktadır.

Yaş: daha önce yapılan araştırmalarda yaş ile iptal oranı arasında bir ilişki olduğu gösterilmiştir. Hung ve diğerlerinin yaptıkları çalışmada Tayvan'da yaşayan kablosuz iletişim kullanan abonelerden 45 ile 48 yaşları arasında olanların iptale daha eğilimli oldukları gösterilmiştir [7]. Türkiye'de ise 35 yaş ve altının iptale daha eğilimli olduğu görülmektedir. 35 yaşın üstündekiler ise sahip oldukları alışkanlıklarını devam ettirip aynı telekom operatörüne bağlı kalarak daha sadık müşteri olmuşlardır.

Abonelik süresi: bu değişken abone iptalini etkileyen en önemli değişkenlerden biridir. Alınan verilere bakıldığında kısa süredir abone olanların iptal etmeye daha yatkın oldukları ve uzun süredir abone olanların ise daha sadık müşteriler olduğu görülmüştür.

Tarife türü: telekom şirketi aboneleri için üç farklı tarife sunmaktadır. İşyerleri gibi kurumsal yapıdaki yerler için kurumsal tarifi, ev telefonu olan aboneler için birinci ve ikinci tip olmak üzere iki farklı tarife bulunmaktadır. Tarifelerin sabit ücretleri ve dakika başına ödenen ücretleri birbirinden farklıdır. Bu değişkenin iptale doğrudan bir etkisi olmamakla birlikte konuşma süresine ve fatura tutarına doğrudan etkisi vardır.

Ortalama fatura tutarı: abonelerin fatura tutarlarının iptale etkisi önemlidir. Abonelere ait 6 aylık fatura tutarlarının ortalaması alınarak tek bir değişkene indirgenmiştir. 6 farklı değişken kullanmak yerine bunların ortalaması alınarak fatura bedelinin açıklayıcı özelliği arttırılmıştır.

Fatura eğilimi: iptal eden abonelerin faturalarına bakarak hangi davranış tipinde oldukları anlaşılmak istenmiştir. Bunun için aylık fatura tutarlarının ardışık olarak

farkları alınmıştır. Birinci ay ile ikinci ay, ikinci ay ile üçüncü ay vb arasındaki farklara bakılmıştır. Farkların pozitif çıktığı durumların sayısı toplanmıştır. En fazla 5 tane pozitif fark olacaktır. Aylar arasında 4 ve 5 kez pozitif farkın olduğu abonelerin için artan eğilimli, 3 kez pozitif farkın olduğu aboneler için sabit eğilimli, 2 ve 1 tane pozitif fark ile hiç pozitif farkın olmadığı aboneler için azalan eğilimli abone olarak tanımlanmıştır. Böyle bir tanımlama içine girilmesinin sebebi ise abonelerin davranışlarını daha iyi açıklayabilmektir. Verilere bakıldığında görülmüştür ki ortalama faturası düşük olup azalan fatura eğilimli olan aboneler ortalama faturası düşük olup fatura eğilimi yüksek olanlara göre daha fazla iptal etmişlerdir. Dolayısıyla ortalama fatura değişkeninin iptali daha iyi açıklayabilmesi için böyle bir değişken oluşturulmuştur.

Ortalama konuşma süresi: abonelerin ele alınan 6 ay boyunca her ay kaç konuştuğu dakikaların ortalamasını gösteren değişkendir. İptali açıklamak için önemli değişkenlerden biridir. Bir abonenin ortalama konuşma süresinin düşük olması onu iptale yaklaştıran bir durumdur. Diğer türlü, abonenin ortalama konuşma süresinin yüksek olması iptalden uzaklaşmasına neden olmaktadır.

Ortalama arama sıklığı: bir abonenin 6 ay boyunca aradığı farklı operatörlerin sıklığının ortalama sayısını gösteren değişkendir. Ortalama arama sıklığı doğrudan iptali etkilememekle birlikte ortalama konuşma süresini ve ortalama fatura tutarını etkilemektedir. Bir abone çok fazla farklı numarayı ararsa konuşma süresi ve fatura bedeli artmaktadır.

İptal: gözlemlenen 6 ayın sonunda abonenin hattını iptal edip etmediğini gösteren değişkendir. Çalışma boyunca üstünde durulan bağımlı değişken olup 0 olduğu durumlar abonenin iptal etmediğini, 1 olduğu durumlar ise abonenin iptal ettiğini ifade etmektedir.

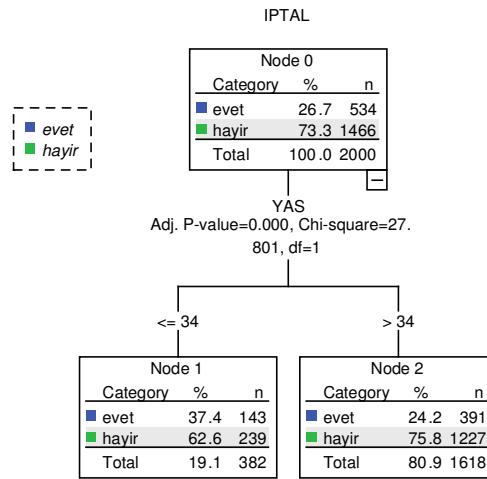
Verilerin hazırlık aşamasından sonra elde edilen yeni veri tabanındaki ilk 15 aboneye ait veriler ekte verilmiştir.

5.3 Sürekli Değişkenlerin Kesikli Hale Getirilmesi

Bayes ağlarında kullanılacak değişkenlerin tümünün kesikli durumda olması gerekmektedir. Böylece her bir değişkenin ayrı ayrı durumlarına göre koşullu olasılıklar

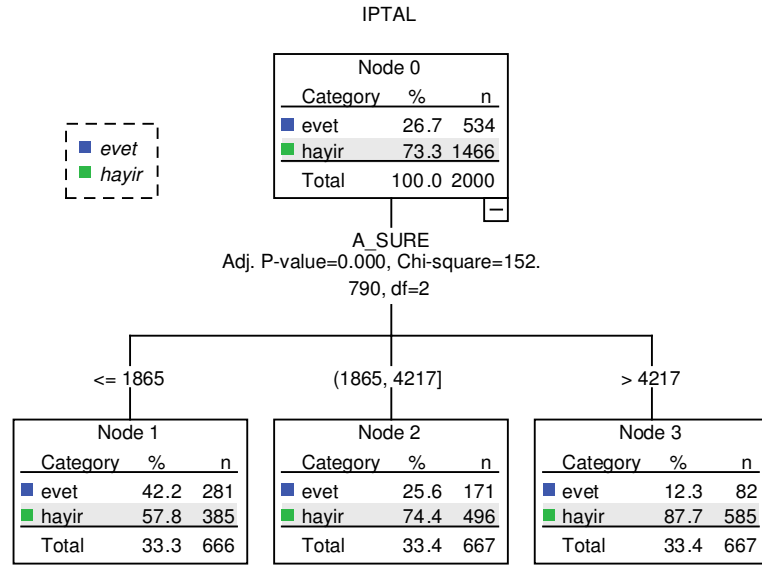
hesaplanabilir. Bu çalışmada yer alan bağımsız değişkenlerden yaş (YAS), abonelik süresi (A_SURE), ortalama fatura tutarı (ORT_FATURA), ortalama konuşma süresi (ORT_K_SURE) ve ortalama arama sıklığı (ORT_SIKLIK) sürekli değişkenlerdir. Bu değişkenleri kesikli hale getirebilmek için SPSS programında yer alan CHAID algoritmasından yararlanılmıştır. Bağımlı değişken olarak iptal seçildikten sonra kesikli hale getirilmek istenen değişken seçilir. Bütün bağımsız değişkenlerin en fazla üç sınıfa ayrılması istenmiştir.

Aşağıda yaş değişkenine ait karar ağacı görülmektedir. Burada Node 0 olarak gösterilen ilk düğümde tüm veri kümesi için iptal oranları yüzdesel ve sayısal olarak ifade edilmiştir. Burada yaş değişkeni 34 den küçük ve eşit ile 34 den büyük olarak iki sınıfa ayrılmıştır. Çocuk düğümlerde de baba düğümde olduğu gibi iptal oranları gösterilmektedir.



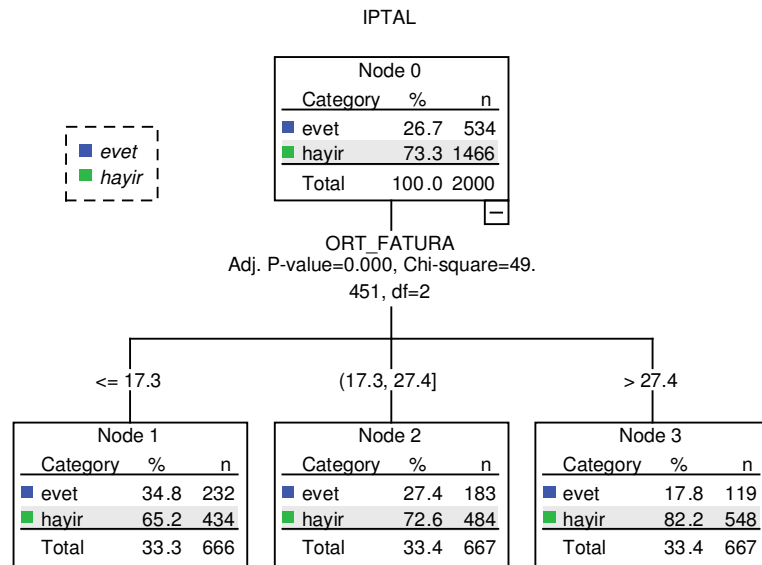
Şekil 5.1 : Yaş değişkeninin karar ağacında sınıflandırılması.

Abonelik süresine ait karar ağacı ise aşağıdaki gibidir. Buna göre abonelik süresini belirten değişken 1865 günden az, 1865 gün ile 4217 gün arası ve 4217 günden fazla olmak üzere üç sınıfa ayrılmıştır. Düğümler sırasıyla kısa, orta ve uzun süreli abonelik sürelerini işaret etmektedir.



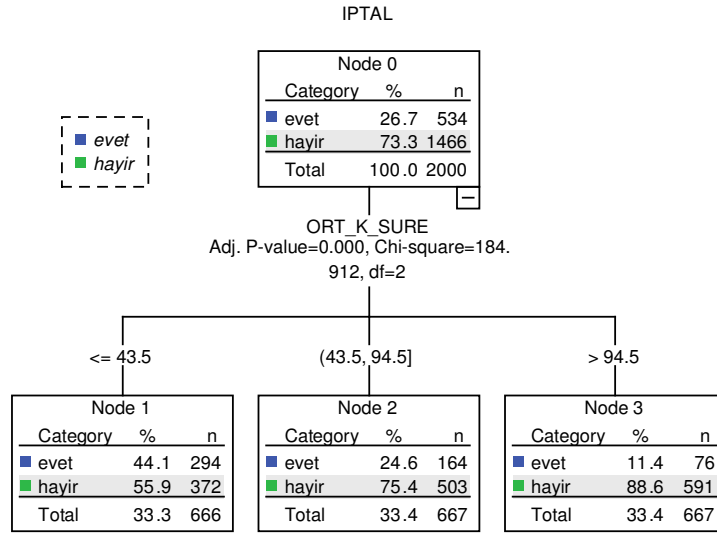
Şekil 5.2 : Abonelik süresi değişkeninin karar ağacında sınıflandırılması.

Ortalama fatura tutarını gösteren değişken aşağıdaki karar ağacında fatura tutarı 17.3 liradan küçük, 17.3 ile 27.4 lira arası ve 27.4 liradan büyük olmak üzere sınıflara ayrılmıştır. Sınıflar sırasıyla düşük, orta ve yüksek fatura tutarlarını göstermektedir.



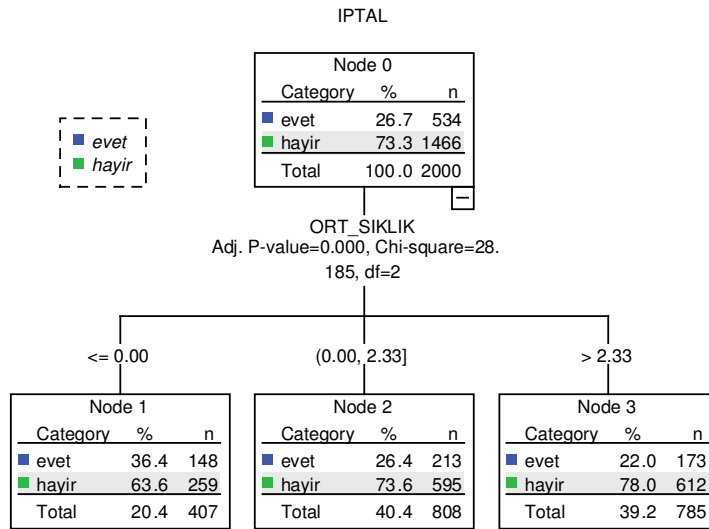
Şekil 5.3 : Ortalama fatura tutarı değişkeninin karar ağacında sınıflandırılması.

Ortalama konuşma süresini gösteren değişkene ait karar ağacında bu değişken 43.5 dakikadan az konuşan, 43.5 ile 94.5 dakika arasında konuşan ve 94.5 dakikadan fazla konuşan aboneler olarak sınıflandırılmıştır. Sınıflar sırasıyla az, orta, çok konuşan aboneleri göstermektedir.



Şekil 5.4 : Ortalama konuşma süresi değişkeninin karar ağacında sınıflandırılması.

Aşağıdaki şekilde ortalama arama sıklığına ait karar ağacı görülmektedir. Bu değişken düşük, orta ve yüksek olmak üzere üç sınıfa ayrılmıştır. 0 a eşit olanlar, 0 ile 2.33 arasında olanlar ve 2.33 den büyük olanlar olarak ayrılmıştır.



Şekil 5.5 : Ortalama arama sıklığı değişkeninin karar ağacında sınıflandırılması.

Sürekli olan değişkenler kesikli hale getirildikten sonra ‘Microsoft Excel 2007’ programında yeni bir veri tabanı oluşturulmuştur. Yeni veritabanından ilk 15 aboneye ait veriler ekte verilmiştir. Burada kesikli hale getirilen değişkenlerin sınıflarına 0, 1 ve 2 değerleri atanmıştır.

5.4. Değişkenler Arasındaki İlişkinin Gösterilmesi

Değişkenler arasındaki nedensel ilişkilerin ortaya çıkarılması Bayes ağlarının oluşturulmasının başlangıç aşamasıdır. Bu ilişkilerin oluşturulması için önce korelasyon ve çoklu eşdoğrusallık analizleri yapılmıştır. Daha sonra telekom şirketinde veri ambarı departmanında çalışan uzmanlardan birine sorular yöneltilmiştir. Bu sorular genellikle ‘abonelik süresinin iptal üzerine etkisi var mıdır?’ tarzında sorulardır. Böylece korelasyon ve eşdoğrusallık analizleri ile uzman görüşlerinden yararlanarak bir nedensel harita oluşturulmuştur.

Korelasyon analizi uygulandığında aşağıdaki sonuçlar elde edilmiştir:

Correlations										
		YERLESİM	ORT_SIKLIK	ORT_K_SURE	YAS	TARIFE	IPTAL	A_SURE	ORT_FATURA	EGILİM
YERLESİM	Pearson Correlation	1	.154**	.340**	-.015	-.184**	-.085**	-.019	.333**	.054*
	Sig. (2-tailed)		.000	.000	.515	.000	.000	.397	.000	.016
	N	2000	2000	2000	2000	2000	2000	2000	2000	2000
ORT_SIKLIK	Pearson Correlation	.154**	1	.305**	.017	-.188**	-.110**	.033	.391**	.074**
	Sig. (2-tailed)	.000		.000	.434	.000	.000	.142	.000	.001
	N	2000	2000	2000	2000	2000	2000	2000	2000	2000
ORT_K_SURE	Pearson Correlation	.340**	.305**	1	-.021	-.398**	-.304**	.011	.765**	.128**
	Sig. (2-tailed)	.000	.000		.359	.000	.000	.616	.000	.000
	N	2000	2000	2000	2000	2000	2000	2000	2000	2000
YAS	Pearson Correlation	-.015	.017	-.021	1	.028	-.118**	.376**	-.045*	.047*
	Sig. (2-tailed)	.515	.434	.359		.203	.000	.000	.043	.037
	N	2000	2000	2000	2000	2000	2000	2000	2000	2000
TARIFE	Pearson Correlation	-.184**	-.188**	-.398**	.028	1	.062**	.045*	-.508**	.021
	Sig. (2-tailed)	.000	.000	.000	.203		.005	.043	.000	.350
	N	2000	2000	2000	2000	2000	2000	2000	2000	2000
IPTAL	Pearson Correlation	-.085**	-.110**	-.304**	-.118**	.062**	1	-.276**	-.156**	-.312**
	Sig. (2-tailed)	.000	.000	.000	.000	.005		.000	.000	.000
	N	2000	2000	2000	2000	2000	2000	2000	2000	2000
A_SURE	Pearson Correlation	-.019	.033	.011	.376**	.045*	-.276**	1	-.033	.042
	Sig. (2-tailed)	.397	.142	.616	.000	.043	.000		.140	.059
	N	2000	2000	2000	2000	2000	2000	2000	2000	2000
ORT_FATURA	Pearson Correlation	.333**	.391**	.765**	-.045*	-.508**	-.156**	-.033	1	.061**
	Sig. (2-tailed)	.000	.000	.000	.043	.000	.000	.140		.007
	N	2000	2000	2000	2000	2000	2000	2000	2000	2000
EGİLİM	Pearson Correlation	.054*	.074**	.128**	.047*	.021	-.312**	.042	.061**	1
	Sig. (2-tailed)	.016	.001	.000	.037	.350	.000	.059	.007	
	N	2000	2000	2000	2000	2000	2000	2000	2000	2000

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Şekil 5.6 : Değişkenler arasındaki korelasyon matrisi.

Elde edilen sonuca göre iptal değişkenini en çok etkileyen değişkenler ortalama konuşma süresi, abonelik süresi ve eğilimdir. Çoklu eşdoğrusallık analizinde ise bağımsız değişkenlerin birbirlerini ne kadar etkiledikleri gösterilmektedir. Aşağıdaki tabloda ‘variance inflation factor’ sütununa göre abonelik süresi, yaş, ortalama fatura tutarı, ortalama konuşma süresi ve ortalama arama sıklığı değişkenleri diğer değişkenlerle ilişkilidir.

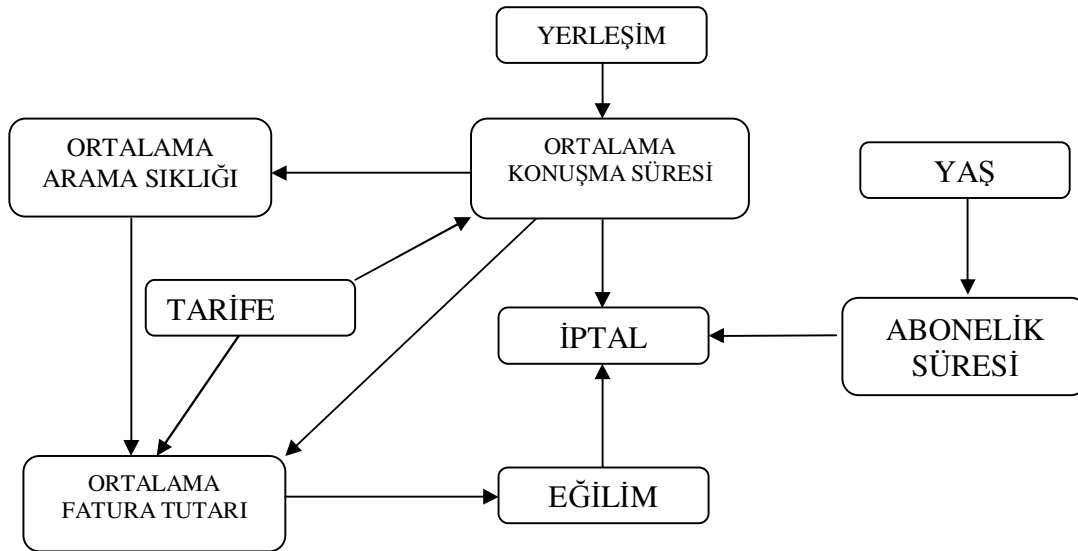
Coefficients ^a									
Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B		Collinearity Statistics	
	B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	.486	.052	9.398	.000	.384	.587		
	YERLESİM	-.038	.022	-1.690	.091	-.081	.006	.933	1.071
	YAS	.002	.001	2.579	.010	.000	.003	.702	1.424
	A_SURE	-5.6E-005	.000	-.297	.127	.000	.000	.689	1.452
	TARİFE	.030	.019	.034	.127	-.008	.068	.763	1.310
	EGİLİM	-.150	.011	-.276	.000	-.171	-.129	.978	1.022
	ORT_FATURA	.003	.001	.192	.000	.002	.004	.250	3.996
	ORT_K_SURE	-.001	.000	-.318	.000	-.001	-.001	.336	2.978
	ORT_SIKLIK	.001	.001	.025	.921	-.002	.004	.526	1.902

a. Dependent Variable: İPTAL

Şekil 5.7 : Eş doğrusallık analizi sonucu.

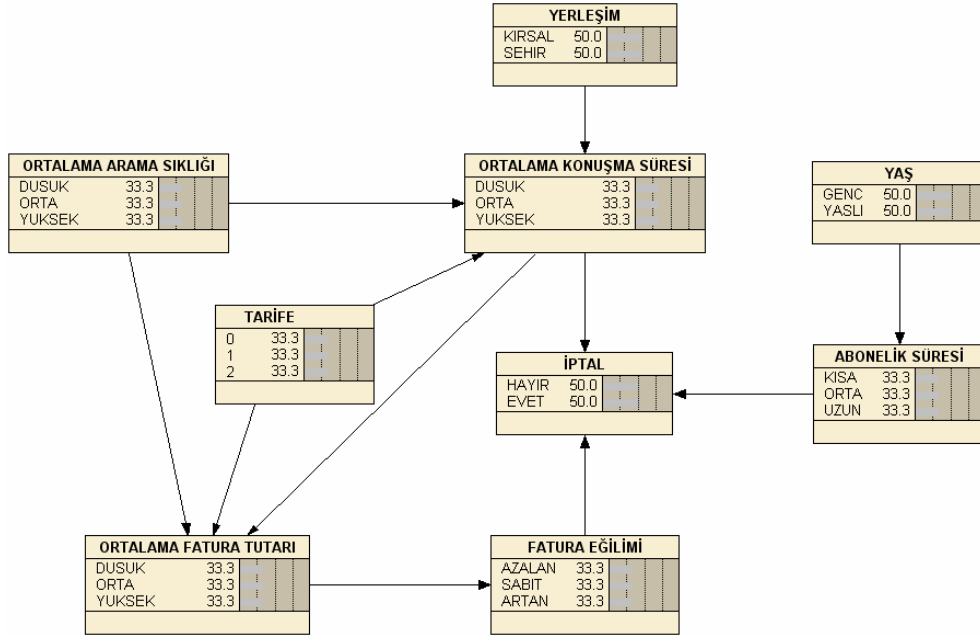
5.5 Problemin Bayes Ağları İle Modellenmesi

Uzman görüşleri alındıktan sonra aşağıdaki nedensel harita elde edilmiştir.



Şekil 5.8 : Nedensel Harita.

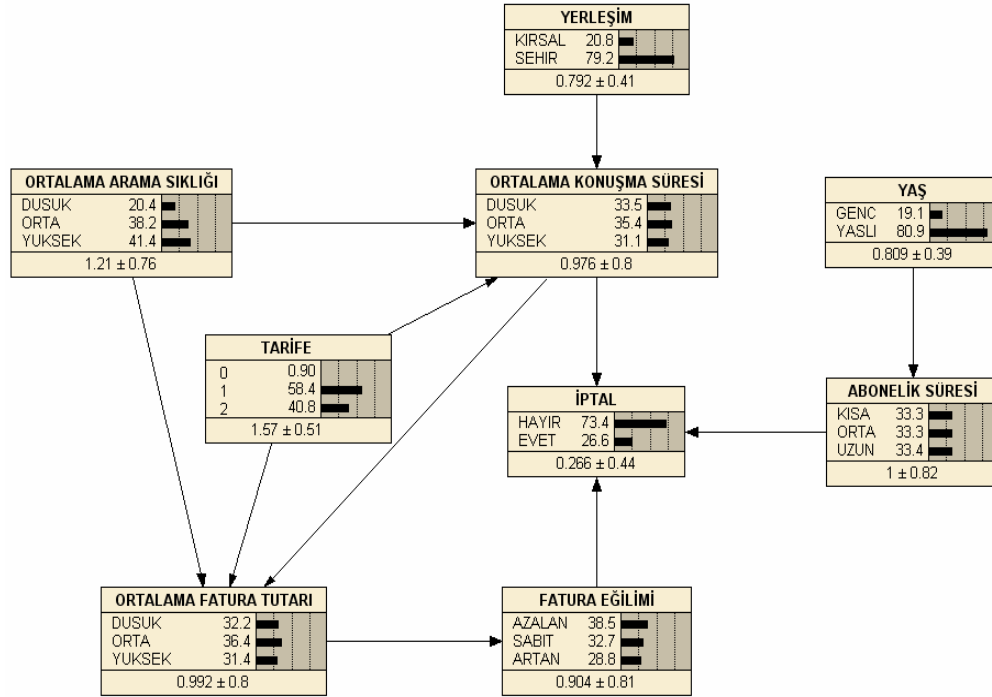
Bayes ağlarında modeli kurmak için yeni bir ağ sayfası açılıp yukarıda görülen nedensel haritadaki her bir düğüm ve ilişkili oldukları diğer düğümleri göstermek için kullanılan oklar ile çizilmiştir. Her bir düğümün üzerine çift tıklanarak açılan pencerede düğümlerin durumları ve durumların aldıkları değerler girilmiştir. Böylece Netica programında elde edilen ilk durumdaki Bayes ağı aşağıdaki gibidir:



Şekil 5.9 : İptal analizi için Bayes ağı.

Veri tabanındaki verileri Bayes ağında kullanabilmek için veritabanındaki değişkenler nedensel haritadaki sırayla düzenlenmelidir. Bunun öncesinde Excel dosyasının ilk satırına (A1 hücresine) // ~->[CASE-1]->~ yazılır. Böylece bu dosya 'Netica' programında kullanılabilir hale gelir. Daha sonra ikinci satırdan itibaren birinci sütuna YERLESIM değişkeni, ikinci satırda devam eden sütunlara nedensel haritanın ikinci seviyesindeki ORT_SIKLIK, ORT_K_SURE ve YAS değişkenleri, üçüncü seviyedeki TARIFE, İPTAL ve A_SURE değişkenleri, dördüncü seviyedeki ORT_FATURA ve EGILIM değişkenleri sütunlara sırasıyla eklenir. 'Netica' programının 'Excel' dosyasından veri okuyabilmesi için dosyanın 'Text (Tab delimited, *.txt)' uzantısıyla kaydedilmesi gerekir. Bu işlemler yapıldıktan sonra oluşan dosyadaki ilk 15 aboneye ait veriler aşağıdaki görünümündedir:

Verileri 'Netica'ya aktarmak için 'Cases' menüsü altında 'Incorp Case File' seçeneği seçilir. Veriler girildikten sonra program derlendiğinde aşağıdaki ağ ortaya çıkar:



Şekil 5.10 : Derlenmiş Bayes ağı.

5.6 Senaryo Analizi

Bayes ağlarından elde edilen sonuçlara göre farklı senaryolar analiz edilmiştir. Senaryolar elde edilirken aralarında korelasyon katsayısı yüksek olan değişkenlere öncelik verilerek üç farklı senaryo elde edilmiştir. 1. senaryoda ortalama konuşma süreleri düşük, abonelik süreleri kısa ve fatura eğilimi azalan abone sınıfı ele alınmıştır. 2. senaryoda ortalama arama sıklığının, ortalama konuşma süresi ve ortalama fatura tutarı ile birlikte iptal üzerindeki etkisi incelenmiştir. 3. senaryoda abonelerin tarifelerin iptal etme oranına etkisi incelenmiştir. Her üç senaryo analizinin sonucunda kampanya önerileri sunulmuştur ve kampanya sonucunda beklenen abone ve gelir kazanımı hesaplanmıştır. Aşağıda her üç senaryo da ayrıntılı incelenmiştir.

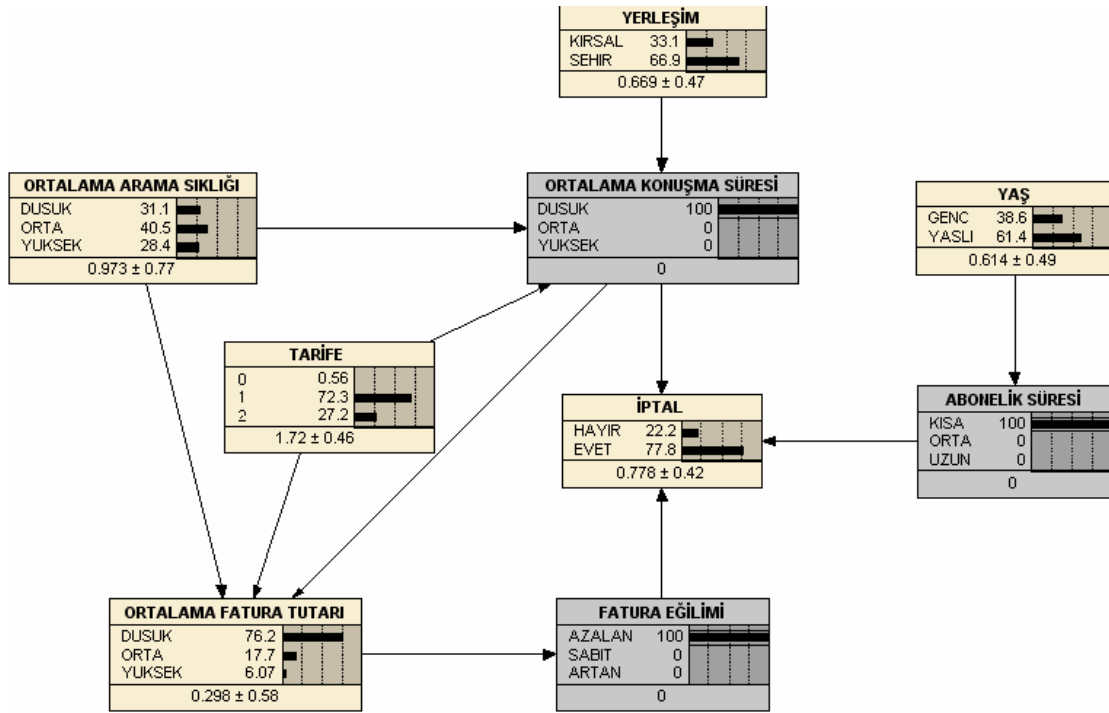
5.6.1 Senaryo 1

Bayes ağlarından iptal için elde edilen koşullu olasılık tablosuna bakıldığında en yüksek iptal olasılığının olduğu abone grubu ortalama konuşma süresi düşük, fatura eğilimi azalan ve abonelik süresi kısa özelliklerine sahiptir ve bu abone grubu bundan sonraki kısımlarda birinci tip abone grubu olarak isimlendirilecektir. Bayes ağlarında oluşan yapıda aynı tip abonelerin ortalama faturaları düşük, genellikle şehirde yaşayan ve çoğunluğu birinci tip hat kullanıcılarından oluşan bir kesim olduğu görülmektedir. (Şekil 5.10)

ORT_K_SURE	A_SURE	EGILIM	0	1
0	0	0	22.222	77.778
0	0	1	38.776	61.224
0	0	2	79.592	20.408
0	1	0	44.828	55.172
0	1	1	54.930	45.070
0	1	2	84.783	15.217
0	2	0	59.341	40.659
0	2	1	80.303	19.697
0	2	2	90.909	9.091
1	0	0	36.111	63.889
1	0	1	50.685	49.315
1	0	2	80.519	19.481
1	1	0	61.194	38.806
1	1	1	84.706	15.294
1	1	2	91.071	8.929
1	2	0	87.143	12.857
1	2	1	90.588	9.412
1	2	2	94.872	5.128
2	0	0	61.728	38.272
2	0	1	79.104	20.896
2	0	2	98.611	1.389
2	1	0	87.013	12.987
2	1	1	83.333	16.667
2	1	2	98.571	1.429
2	2	0	94.286	5.714
2	2	1	97.500	2.500
2	2	2	100.00	1.4e-025

Şekil 5.11 : Senaryo 1 için olasılık çizelgesi.

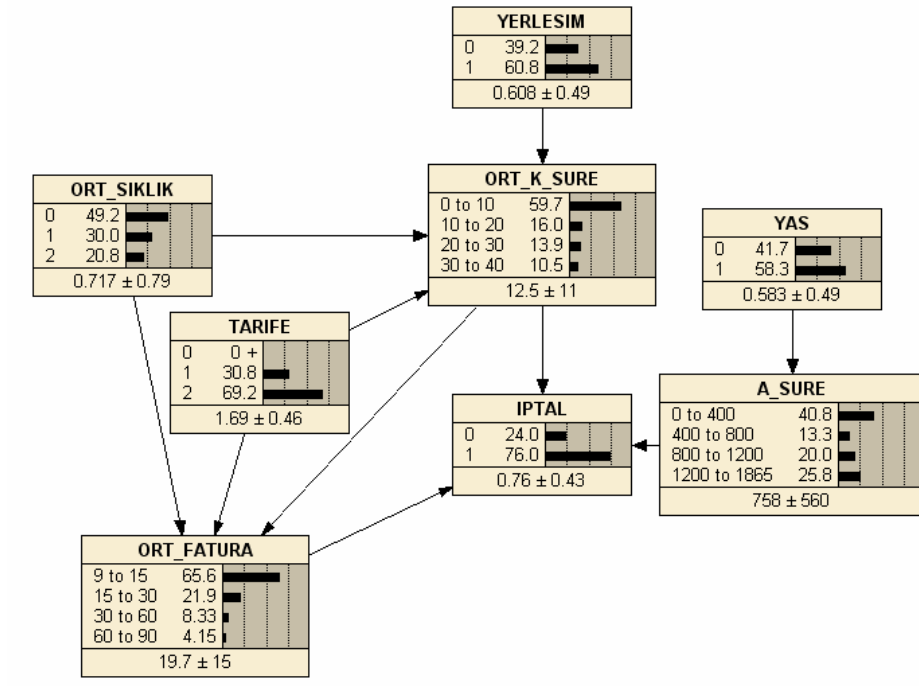
Şekil 5.11'e göre ortalama konuşma süreleri düşük ve kısa süreli abone olanlardan fatura eğilimi sabit olanların iptal oranları %61.2 ye, faturaları artan eğilimde olanların ise %20.4 e düşmektedir. Aynı şekilde abonelik süresi uzadıkça ve ortalama konuşma süreleri uzadıkça iptal oranlarının düştüğü görülmektedir.



Şekil 5.12 : Senaryo 1 için Bayes ağı.

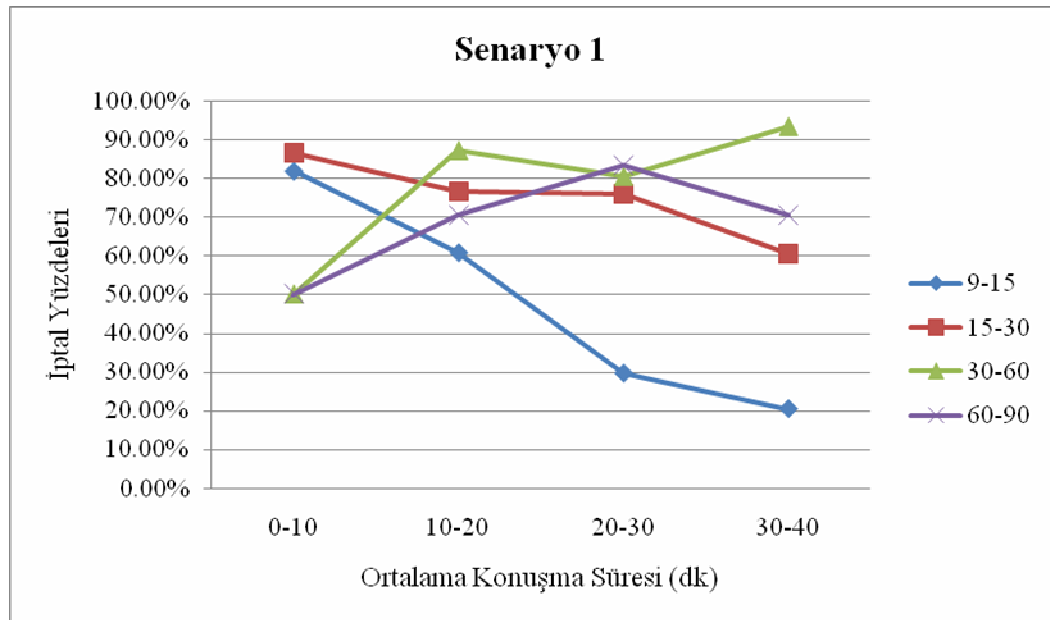
Birinci tip abone grubunun iptal yüzdeleri azaltacak kampanyayı belirlemek için bu müşteri segmentinin özelliklerine göre aşağıdaki Bayes ağı Netica’da elde edilmiştir.

Şekil 5.13’deki Bayes ağında ortalama konuşma süresi düşük (ayda ortalama 40 dakikadan az konuşan) olan abonelerin konuşma süreleri kendi içinde dört alt duruma ayrılmıştır. Bunun sebebi birinci tip abonelerin düşük ortalama konuşma süresini daha ayrıntılı bir şekilde ortaya koyabilmektir. Aynı şekilde kısa süredir abone olanların durumlarını daha ayrıntılı incelemek için abonelik süresi değişkeni dört alt gruba ayrılmıştır. Birinci tip aboneler sadece azalan eğilimde olduklarından eğilim düğümü Bayes ağından çıkartılmıştır. Bunun yerine ortalama fatura tutarlarının iptale etkisi incelenmiştir. Birinci tip abonelerin ortalama fatura tutarları 9 ile 90 Lira arasında değişmektedir. Bu aralık 9-15, 15-30, 30-60 ve 60-90 olacak şekilde bölünmüştür.



Şekil 5.13 : Senaryo 1 için ayrıntılı Bayes ağı.

Aşağıdaki grafikte yatay ekseninde ortalama konuşma süreleri ve dikey ekseninde ise ortalama fatura tutarlarına göre iptal yüzdeleri görünmektedir.



Şekil 5.14 : Senaryo 1 için iptal yüzdeleri gösteren grafik.

Şekil 5.14’de verilen grafiğe göre 9-15 arasında ortalama faturaya sahip abonelerin iptal etme yüzdesi ortalama konuşma süreleri arttıkça azalmaktadır. Aynı durum 15-30 aralığında ortalama faturaya sahip aboneler için de geçerlidir. Nitekim 30-60 ve 60-90 aralığındaki ortalama faturaya sahip aboneler farklı iptal eğilimleri içindedir. Buna göre 9-15 ve 15-30 aralığında ortalama faturaya sahip aboneler için benzer kampanyalar yürütülebilir. Onar dakikalık aralıklarla ortalama konuşma süresi arttıkça abonelerin iptal oranı düştüğüne göre 9-15 ve 15-30 aralığındaki abonelere faturaları değişmeyecek şekilde bir ay boyunca 10 dakika bedava konuşma süresi verilebilir. Bu durumda abonelerin iptal yüzdeleri azalacaktır. Her ne kadar bedava konuşma kampanyasına katılan abonelerin fatura tutarları artmasa da 6 ayın sonunda iptal eden abonelerin iptal etmedikleri durumda şirkete sağlayacağı bir kazanç olacaktır.

Çizelge 5.1 : 9-15 TL faturası olan aboneler için beklenen kazanç analizi.

Ort. konuşma süresi	Abone sayısı	Ort. fatura tutarı	Kampanya öncesi iptal yüzdesi	Kampanya sonrası iptal yüzdesi	Kampanya sonrası elde tutulacak abone sayısı	Beklenen kazanç
0-10	63	11.58	%82	%61	13	150.94
10-20	9	12.69	%61	%30	3	38.07
20-30	2	10.96	%30	%20	1	10.96
30-40	6	13.50	%20			
Toplam Kazanç						199.97

Çizelge 5.1’ e göre ortalama konuşma süreleri 0-10 dk arasında toplam 63 abone olup ortalama fatura tutarları 11.58 TL dir. Bu abonelerin kampanya öncesi iptal oranı %82 dir.10 dakika bedava süre vermeyi planladığımız kampanya sonrasında bu gruptaki aboneler 10-20 dk arasında ortalama konuşma süresi olan abonelerin grubuna gireceği için iptal oranları %61’e düşeceği düşünülmektedir. Dolayısıyla iptal oranı %82’den %61’e düşeceği için kampanya sonrasında 13 abonenin iptal etmeyip elde tutulacağı hesaplanmıştır. Elde tutulan abonelerin sayısı bu gruptaki ortalama fatura tutarı ile çarpılınca (11.58 TL) beklenen kazanç 150.94 TL olarak bulunmaktadır. Benzer

hesaplamalar 10-20 dk ve 20-30 dk ortalama konuşma süresi olan aboneler için yapıldığında kampanya sonrası elde tutulacak toplam kazançlar sırasıyla 38.07 TL ve 10.96 TL'dir. Böylece elde tutulacak abonelerle toplam 199.97 TL kazanç sağlanacaktır. Bunun yanında uzun vadede abonenin kazanımı sağlanmış olacaktır.

Aynı durumu 15-30 TL arasında ortalama faturaya sahip aboneler için düşünürsek aşağıdaki tablo elde edilmektedir.

Çizelge 5.2 : 15-30 TL faturası olan aboneler için beklenen kazanç analizi.

Ort. konuşma süresi	Kampanya öncesi abone sayısı	Ort. fatura tutarı	Kampanya öncesi iptal yüzdesi	Kampanya sonrası iptal yüzdesi	Kampanya sonrası elde tutulacak abone sayısı	Beklenen kazanç
0-10	9	17.83	%87	%77	1	17.83
10-20	5	20.61	%77	%76	1	20.61
20-30	8	19.94	%76	%60	1	19.94
30-40	9	18.68	%60			
Toplam kazanç						58.38

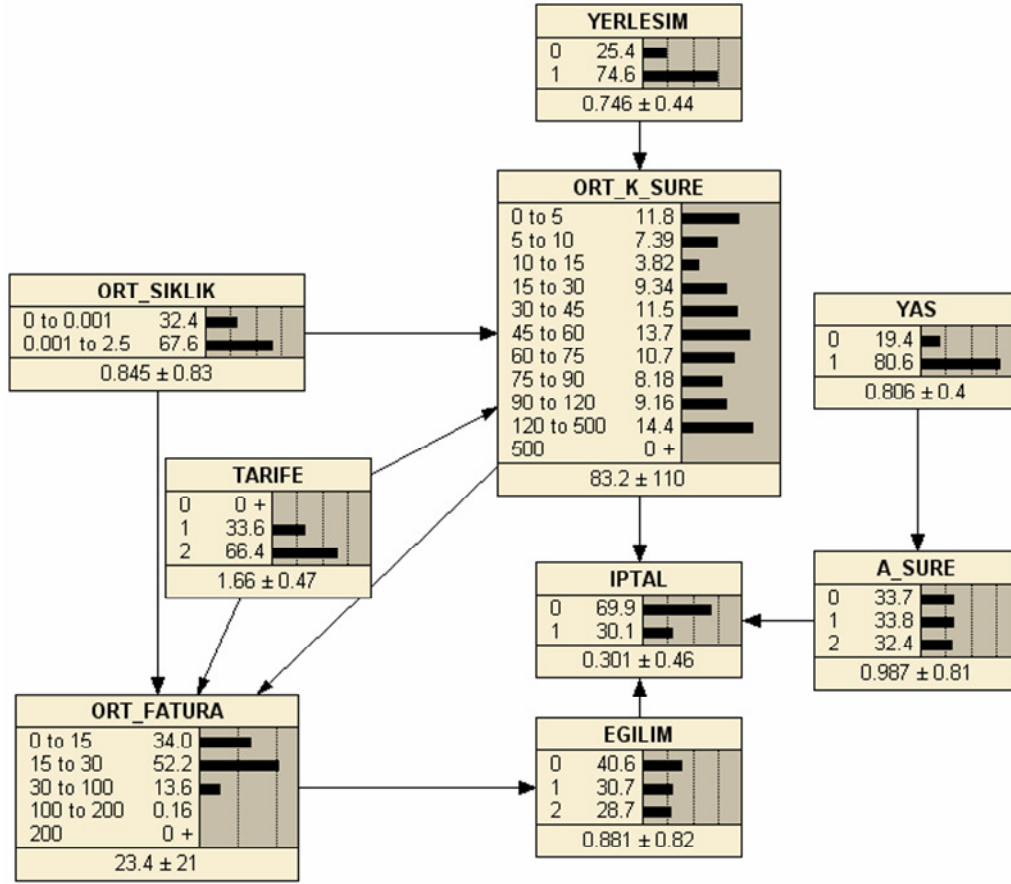
15-30 TL arasında ortalama faturaya sahip birinci tip abonelerden kampanya sonrasında elde edilecek kazanç 58.38 TL'dir.

Ortalama fatura tutarı 30-60 ve 60-90 TL arasında olan birinci tip abonelerin genel özelliği abonelik sürelerinin 400 günden az olan yeni aboneler olmalarıdır. Ortalama konuşma süreleri düşük olduğu halde faturaları yüksek gelen bu abonelere yönelik farklı bir kampanya anlayışı içine girilmelidir. Bu abonelere abonelik başlangıcından itibaren altı ay boyunca fatura indirimi yapan bir kampanya sunulabilir.

5.6.2 Senaryo 2

Bayes ağlarında ele alınan değişkenlerden ortalama arama sıklığı bir abonenin bir ay boyunca kaç defa farklı operatörü aradığını göstermektedir. Kurulan ilk modelde ortalama arama sıklığının iptal üzerine doğrudan bir etkisi olmamakla birlikte ortalama

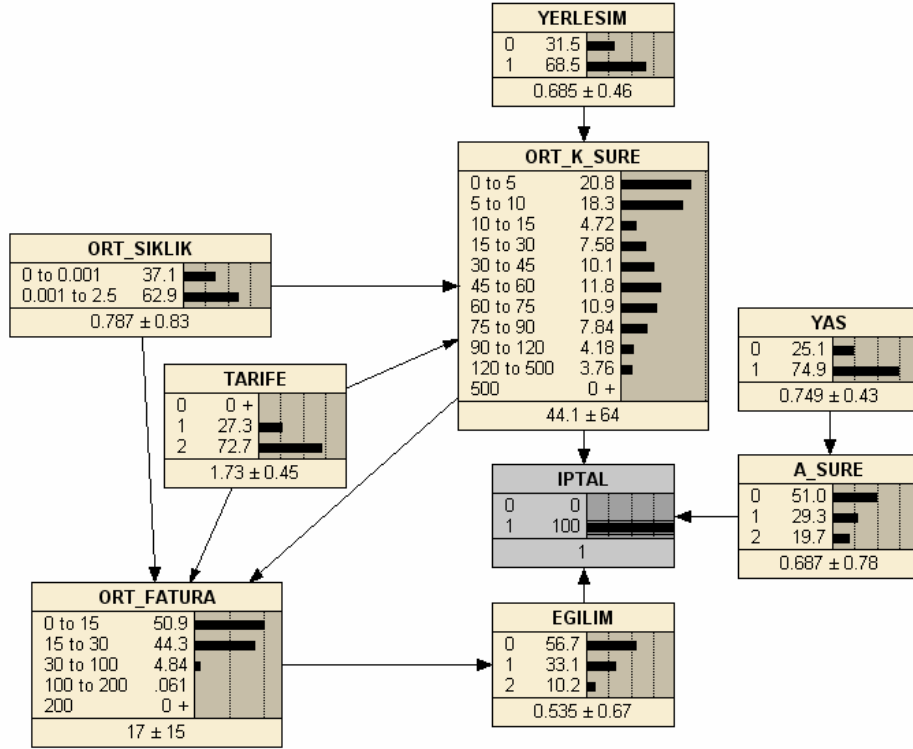
konuşma süresini ve ortalama fatura tutarını etkilemektedir. Diğer iki değişkenle birlikte ortalama arama sıklığının iptal üzerindeki etkisi incelenebilir ve bu duruma yönelik kampanyalar geliştirilebilir.



Şekil 5.15 : Senaryo 2 için derlenmiş Bayes ağı.

Bu aşamada incelenecek abonelerin özelliklerine göre ikinci tip abone olarak adlandırılacaktır. Farklı operatörü ortalama arama sıklığı kendi içinde farklı operatörü arayan ve aramayan aboneler olmak üzere iki alt gruba ayrılmıştır. Mevcut operatörün haricinde farklı operatörleri aramanın karşılığında abone daha fazla ücretlendirmeye tabidir. Yani farklı operatördeki bir aboneyle yapılan telefon görüşmesinin dakika başına birim ücreti aynı operatördeki bir aboneyi aramaktan her zaman daha pahalıdır. Bununla birlikte Şekil 5.16’de görülen Bayes ağında iptal edenlerin %62.9’unun farklı operatörü arayan aboneler olduğu görülmektedir. Ayrıca bu abonelerin %95.2’sinin ortalama 30

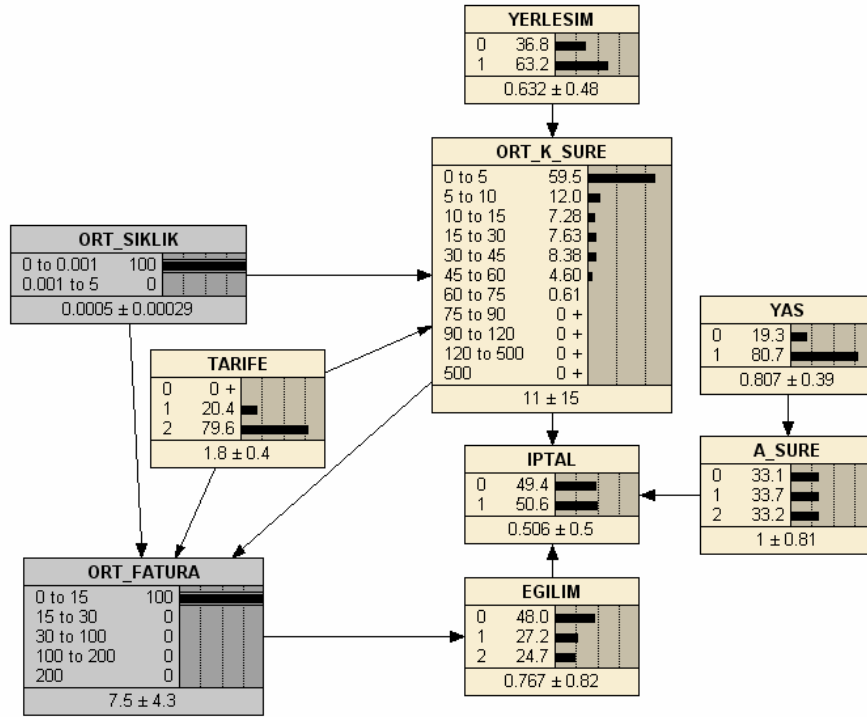
TL fatura bedeli ödediği ve % 39.1'nin ortalama konuşma süresinin en fazla 10 dakika olduğu da dikkat çekicidir.



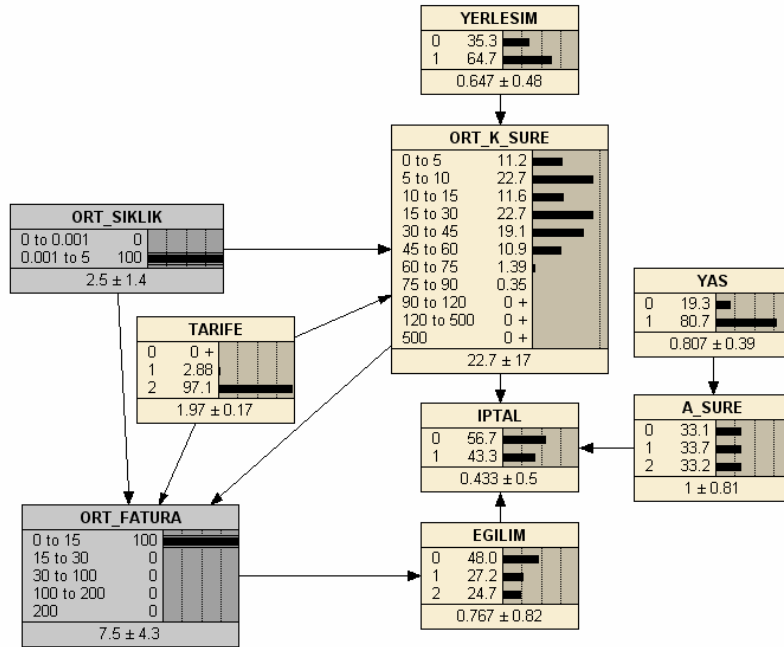
Şekil 5.16 : Senaryo 2 için ayrıntılı Bayes ağı.

Mevcut operatöründen farklı operatörü aramayan ve ortalama faturaları 0-15 aralığında değişen abonelere ait ortalama konuşma sürelerini ve iptal oranlarını inceleyen Bayes ağı Şekil 5.17’de görülmektedir. Buna göre ortalama faturası 0-15 aralığında olan ve farklı operatörleri aramayan insanların %59.5’inin ortalama konuşma süresinin 0-5 dakika arasında olduğu ve bu abonelerin iptal etme oranının %50.6 olduğu anlaşılmaktadır.

Diğer bir yandan farklı operatörü arayan ve ortalama faturaları 0-15 aralığında değişen abonelere ait ortalama konuşma sürelerini ve iptal oranlarını inceleyen Bayes ağı Şekil 5.18’de görülmektedir. Buna göre ortalama faturası 0-15 aralığında olan ve farklı operatörleri arayan insanların iptal etme oranı %43.3’dür.

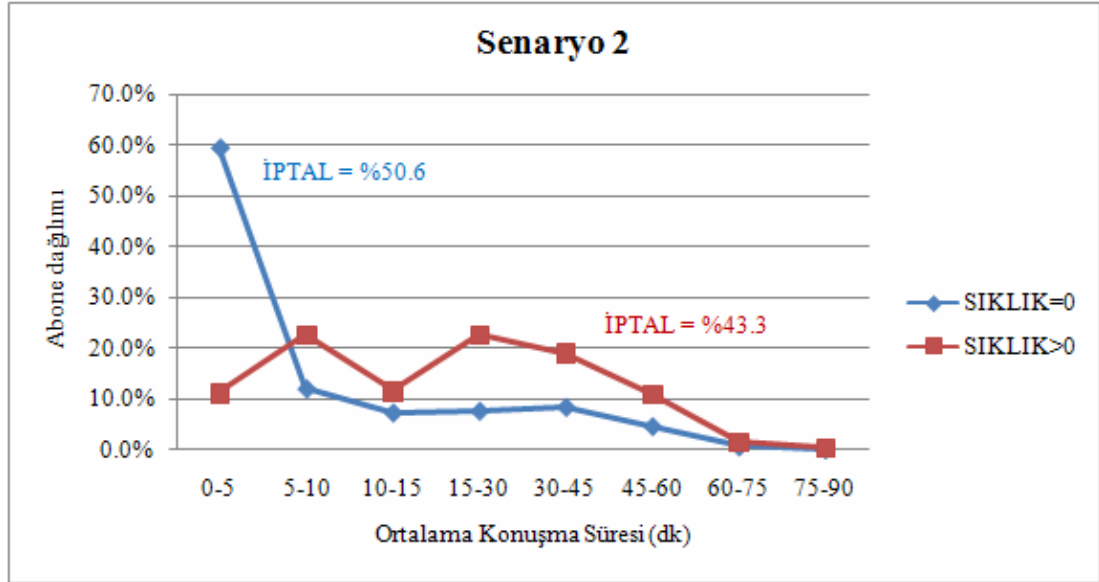


Şekil 5.17 : Farklı operatörleri aramayan aboneler için Bayes ağı.



Şekil 5.18 : Farklı operatörleri arayan aboneler için Bayes ağı.

Bayes ağıının sonuçlarından yararlanarak ortalama fatura tutarları 0-15 arasında değişen ve farklı operatörleri arayan ve aramayan abonelerin ortalama konuşma sürelerine göre iptal yüzdeleri aşağıdaki grafikte gösterilmiştir.



Şekil 5.19 : Senaryo 2 için iptal yüzdelerini gösteren grafik.

Yukarıdaki grafikte yatay eksenle abonelerin ortalama konuşma süreleri, dikey eksenle ise abonelerin dağılımı verilmektedir. Örneğin ortalama fatura bedeli 0-15 arasında olan ve farklı operatörü aramayan abonelerin %60'a yakını 0-5 dakika ortalama konuşma süresine sahiptir. Bu abonelerin altıncı ayın sonunda iptal etme olasılığı %50.6' dır. Ortalama konuşma süresi arttıkça farklı operatörü aramayan abonelerin sayısı azalmaktadır. Aynı şekilde ortalama fatura tutarı 0-15 aralığında değişen abonelerden farklı operatörleri arayanlardan 0-5 dakika arasında konuşanların oranı %12' dir. Bu abonelerin altıncı ayın sonunda iptal etme olasılıkları farklı operatörü aramayan abonelere göre daha düşüktür ve %43.3'tür. Bu durumda yapılacak kampanya iptal oranı daha fazla olduğu için farklı operatörü aramayan abonelere yönelik olmalıdır. Abonelerin fatura tutarlarında değişiklik olmaksızın farklı operatörleri aramanın ilk beş dakikası bedava yapılırsa bu aboneler mevcut operatörlerinden farklı operatörleri aramaya ikna olacaklardır. Grafikten de anlaşıldığı üzere farklı operatörü arayan sınıfında olduklarında iptal olasılıkları düşmektedir. Altı ay boyunca eğilimleri izlenen bu abonelere altıncı ayda bu şekilde bir kampanya yapılırsa iptal etmek yerine aboneliğe

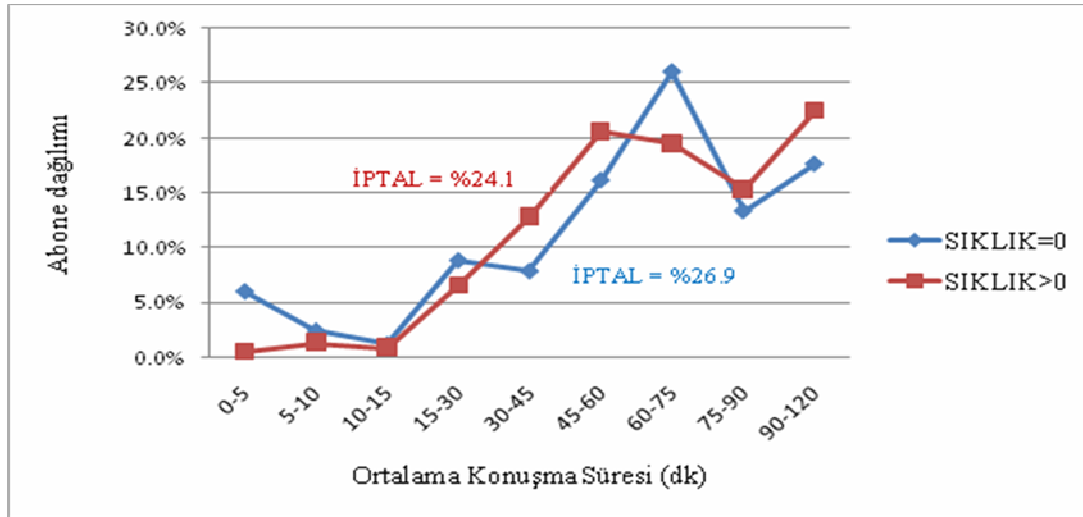
devam edenlerden elde edilecek bir gelir oluşacaktır. Aşağıdaki tabloda bu aboneler için kazanç analizi yapılmıştır.

Çizelge 5.3 : 0-15 TL faturası olan aboneler için beklenen kazanç analizi.

	Kampanya öncesi abone sayısı	Ort. fatura tutarı	Kampanya öncesi iptal yüzdesi	Kampanya sonrası iptal yüzdesi	Elde tutulacak abone sayısı	Beklenen kazanç
Farklı operatörü aramayan	187	11.66	%51	%43	15	174.9
Farklı operatörü arayan	294	12.45	%43			

Farklı operatörü aramayan abonelere uygulanan kampanyayla iptal etmesi engellenirse 174.9 TL ek kazanç sağlanacaktır. Ayrıca bu abonelerin sadakati sağlanmış olacaktır.

Aynı durum ortalama fatura tutarları 15-30 arasında değişen aboneler için de mevcuttur. Bu aboneler için Bayes ağlarından yararlanılarak oluşturulan grafik Şekil 5.20'deki gibidir. Grafikte farklı operatörleri arayan ve aramayan abonelerin ortalama konuşma sürelerine göre iptal yüzdeleri görülmektedir.



Şekil 5.20 : 15-30 TL arası faturası olan aboneler için iptal yüzdelerini gösteren grafik.

Ortalama faturaları 15-30 TL arasında değişen, farklı operatörleri aramayan abonelerin iptal etme oranı %26.9' dur. Aynı ortalama faturaya sahip abonelerden farklı operatörü arayanların iptal etme oranı daha düşük olup %24.1'dir. Ortalama faturaları 0-15 arasında değişen abonelere uygulanan kampanya bu abone segmentine de uygulanırsa elde tutulması beklenen abone sayısı ve bnlardan elde edilecek kazanç aaşğıdaki tabloda gösterilmiştir.

Çizelge 5.4 : 15-30 TL faturası olan aboneler için beklenen kazanç analizi.

	Kampanya öncesi abone sayısı	Ort. fatura tutarı	Kampanya öncesi iptal yüzdesi	Kampanya sonrası iptal yüzdesi	Elde tutulacak abone sayısı	Beklenen kazanç
Farklı operatörü aramayan	132	19.56	%27	%24	4	78.24
Farklı operatörü arayan	529	20.47	%24			

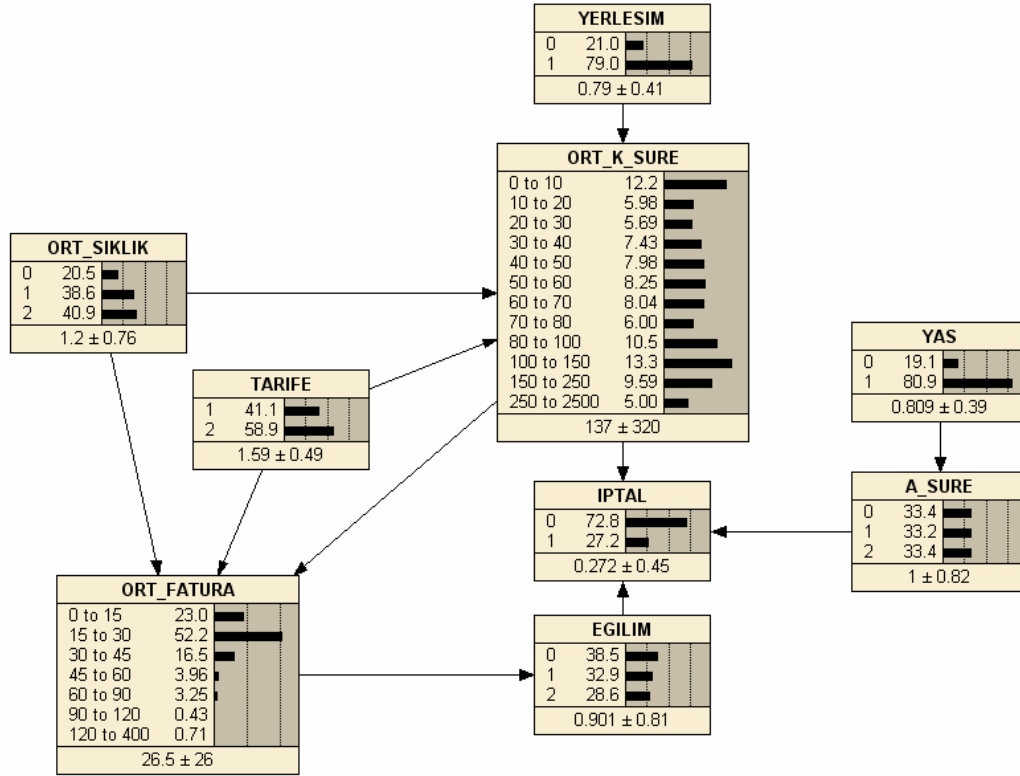
Bu durumda kampanya sonrasında elde edilecek kazanç 78.24 TL olmaktadır.

5.6.3 Senaryo 3

Mevcut tarife iptal etme oranını büyük ölçüde etkilemektedir. Müşteriler ilk kez abonelik için başvurduklarında tarife türü olarak 1.tip hat verilmektedir. Tarife değişikliği abone istediği takdirde yapılabilmektedir. Bireysel aboneler kurumsal tip abonelik haricinde 1. ve 2. tip tarifelerden birini seçebilmektedir. Her bir tarifenin sabit ücreti ve birim ücreti farklıdır. 2. tip hat telefonla fazla görüşme yapmayan aboneler için uygun bir tarifiedir. 1. tip hatta göre sabit ücreti daha ucuzdur. Ancak yapılan görüşmelerde dakika başına alınan birim ücret 1. tip hatta göre daha fazladır. Veri kümesindeki abonelerin %0.9'u kurumsal, %40.75'i 1.tip ve %58.35'i 2. tip hatta sahiptir. Kurumsal hattaki abonelerin iptal etme oranı %16.6, 1.tip hattaki aboneler için %23 ve 2. tip hattakiler için %28.9'dur.

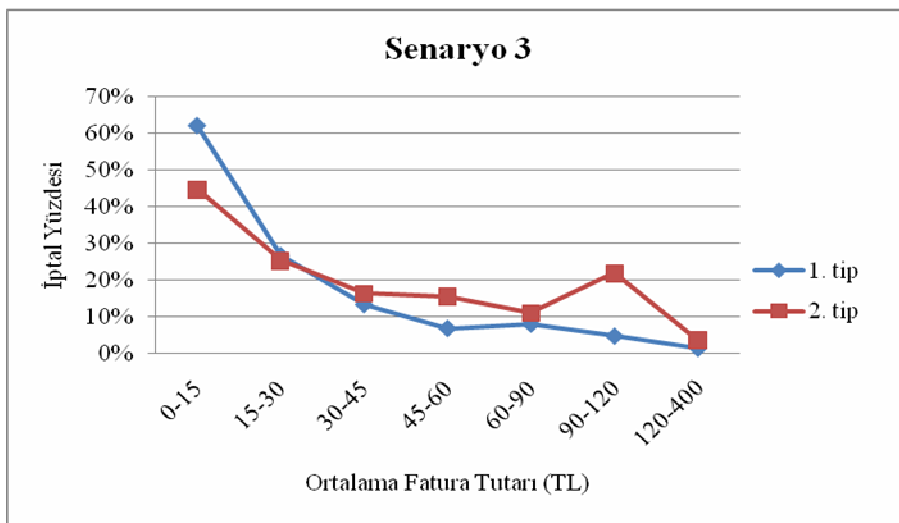
Bu senaryo analizinde toplam abone sayısı içindeki yüzdesi düşük olduğundan kurumsal hattı analiz dışında bırakılmıştır. Bayes ağı oluşturulurken tarife değişkeninin etkilediği

ortalama konuşma süresi ve ortalama fatura tutarları daha ayrıntılı incelenmek için alt gruplara ayrılmıştır.



Şekil 5.21 : Senaryo 3 için derlenmiş Bayes ağı.

Bayes ağından elde edilen sonuçlara göre aşağıdaki grafik elde edilmiştir.



Şekil 5.22 : Tarife bazında ortalama fatura tutarı ve iptal yüzdeleri.

Şekil 5.22’de yatay ekseninde abonelerin ortalama fatura tutarları, dikey ekseninde ise ortalama fatura aralığındaki 1. tip ve 2. tip hat kullanıcılarının iptal yüzdeleri görülmektedir. Örneğin; ortalama faturası 0-15 aralığında olan 1. tip hat kullanıcılarının iptal yüzdesi % 62, 2. tip hat kullanıcılarının iptal yüzdesi %45’dir. Ortalama fatura tutarı 15-30 aralığına yaklaştıkça 1. tip ve 2. tip hat kullanıcılarının iptal yüzdeleri birbirine yaklaşmaktadır. Bu aralıktan sonra ortalama fatura aralıkları arttıkça 2. tip hat kullanıcılarının iptal yüzdesi 1. tip hat kullanıcılarından daha fazla olmaktadır. Bu durumda ortalama fatura tutarları 0-15 ve 15-30 arasında değişen aboneler için farklı kampanya, ortalama faturaları 15-30’dan büyük olanlar için farklı kampanya yürütülmelidir. 2. tip hat telefonla fazla konuşma yapmayanlar için uygun bir tarife olduğundan 1. tip hatta olup fazla arama yapmayan abonelerin 2. tip hatta geçmesine yönelik bir kampanya başlatılabilir. Bu kampanyayla 2. tip hat kullanıp fazla görüşme yapmayan aboneler yüksek sabit ücret yükünden kurtarılabilir. Bu durum ilk başta kazancın azalmasına sebep olacak gibi dursa da elde tutulacak abonelerle abone sayısı artacağından kazanç sağlanacaktır. Aşağıdaki çizelgede kazanç analizi yapılmıştır.

Çizelge 5.5 : 0-30 TL arası faturalar için tarife değişiminden beklenen kazanç analizi.

Fatura aralığı	Tarife türü	Ort. fatura tutarı	Kam. öncesi abone sayısı	Kam. öncesi iptal yüzdesi	Kam. sonrası iptal yüzdesi	Elde tutulacak abone sayısı	Beklenen kazanç
0-15	1. tip	14.40	38	%62	%45	7	101
	2. tip	11.97	450	%45			
15-30	1. tip	23.23	376	%27	%25	5	116
	2. tip	20.94	587	%25			
Toplam kazanç							217

Çizelgeye göre kampanya sonrasında elde tutulacak abonelerle toplam 217 TL ek kazanç sağlanacaktır.

Ortalama fatura tutarı 30 TL’den fazla olan 2. tip hat kullanıcılarının iptal yüzdesi 1. tip hat kullanıcılarına göre artış içinde olduğu grafikten görülmektedir. Faturası yüksek aralıklarda olan 2. tip hat kullanıcılarına standart hatta geçmelerine yönelik bir kampanya yapılabilir. Bu kampanyayla 2.tip hatta çok konuşan aboneler 1. tip hatta geçerek dakika başına daha az birim ücret ödeyecektir. Bu durum ilk başta şirket için kazanç azalması gibi dursa da elde tutulan abonelerle kazanç sağlanacaktır. Çizelge 5.6’da kazanç analizi yapılmıştır.

Çizelge 5.6 : 30-400 TL arası faturalar için tarife değişiminden beklenen kazanç analizi.

Fatura aralığı	Tarife türü	Ort. fatura tutarı	Kam. öncesi abone sayısı	Kam. öncesi iptal yüzdesi	Kam. sonrası iptal yüzdesi	Elde tutulacak abone sayısı	Beklenen kazanç
30-45	1. tip	36.73	239	%13			
	2.tip	35.05	100	%16	%13	3	105
45-60	1. tip	51.61	73	%7			
	2.tip	49.22	15	%16	%7	2	98
60-90	1. tip	71.95	62	%8			
	2.tip	69.43	12	%11	%8	1	69
90-120	1. tip	104.72	10	%5			
	2.tip	103.75	2	%22	%5	1	104
120-400	1. tip	171.24	15	%1			
	2.tip	149.96	3	%4	%1	1	150
Toplam kazanç							527

Bu çizelgeye göre elde tutulacak abonelerle toplam 527 TL ek kazanç sağlanacaktır.

6. SONUÇ VE ÖNERİLER

Telekomünikasyon sektörü için abone kaynaklı iptal oranının giderek artması bir çok araştırma için dikkat çekici bir konu olmuştur.

Bu çalışmada abone iptaline sebep olan faktörler belirlenmeye çalışılmıştır. Bu amaçla Bayes ağları yöntemi kullanılmıştır. Değişkenler arasındaki nedensellik ilişkisine dayandığı için bu yöntemin kullanılmasına karar verilmiştir. Daha önce yapılan çalışmalarda sıklıkla kullanılan karar ağaçları yöntemi, aralarında korelasyon olan değişkenlerin sınıflandırılmasında etkin bir yöntem olamamaktadır. Aynı şekilde çok kullanılan yöntemlerden biri olan yapay sinir ağlarının ise veriye karşı hassasiyeti fazladır. Ayrıca yapısında bulunan saklı katmanlardan dolayı değişkenler arasındaki nedensel ilişkiler görülmemektedir. Dolayısıyla bu yöntemler tercih dışı bırakılmıştır.

Bayes ağlarında kullanılmak üzere 2000 aboneden oluşan veri kümesi Türkiye’de faaliyet gösteren telekomünikasyon şirketlerinden birinden alınmıştır. Veri kümesinde abonelere ait yerleşim yeri, yaş, abonelik süresi, tarife türü, son altı aya ait fatura tutarları, son altı aya ait konuşma süreleri, son altı aya ait farklı operatörü arama sıklığı ve bu abonelerin altıncı ayın sonunda iptal edip etmediğini gösteren değişkenler mevcuttur. Verileri Bayes ağlarında kullanmaya hazırlamak için son altı aya ait fatura tutarları, konuşma süreleri ve arama sıklıklarının ortalamaları alınarak her biri tek bir değişken haline dönüştürülmüştür. Bu değişkenlere ek olarak abonelerin fatura eğilimleri yeni bir değişken olarak eklenmiştir. Bu değişken ile abonelerin altı aylık fatura verilerine bakarak artan, azalan ve sabit eğilimde oldukları belirlenmiştir.

Bayes ağlarında değişkenler arasındaki nedensel ilişkiler yönlü, döngüsüz bağlantılarla gösterilmektedir. Değişkenler arasındaki bu bağlantıları saptayabilmek için hazırlanan veri kümesindeki veriler öncelikle korelasyon analiziyle incelenmiştir. Korelasyon analizi sonucunda abonelerin iptali ile en yüksek korelasyona sahip olan değişkenler ortalama konuşma süresi, abonelik süresi ve fatura eğilimleridir. Değişkenler arasındaki yönlü bağlantılar, korelasyon analizi sonucunda çıkan ilişkiler ve telekomünikasyon şirketinde çalışan uzmanların görüşleri ile birleştirilerek oluşturulmuştur. Böylece Bayes

ağının temelini oluşturan bir nedensel harita ortaya çıkmıştır. Nedensel haritanın Netica programında derlenebilmesi için sürekli değişkenler SPSS programında CHAID karar ağacı algoritmasıyla kesikli hale getirilmiştir. Karar ağacı algoritmasının kullanılması verilerin daha etkin bir şekilde kullanılmasını sağlamıştır. Çalışmanın devamında Netica’da derlenen Bayes ağının sonuçlarına bakılarak farklı senaryolar oluşturulmuştur.

Bayes ağının sonuçlarına göre üç farklı senaryo oluşturulmuştur. Birinci senaryoda iptalin en yüksek olasılıkla görüldüğü durum ele alınmıştır. Bu gruba dahil olan abonelerin fatura eğilimleri azalan olup ortalama konuşma süreleri ve abonelik süreleri kısadır. Ortalama konuşma süreleri ve abonelik süreleri kısa olan abonelerin durumları ile fatura eğilimi azalan abonelerin ortalama fatura tutarları daha ayrıntılı olarak incelenmiştir. Yapılan analize göre abonelere ayda 10 dakika bedava konuşma hakkı verilmesine dair bir kampanya yapılabileceği gösterilmiştir. Bu kampanya ile iptal oranının azalacağı vurgulanarak sadakati sağlanan aboneler sayesinde beklenen kazanç hesaplanmıştır.

İkinci senaryoda iptal değişkeni ile korelasyonu yüksek olan değişkenlerden farklı operatörleri ortalama arama sıklığı ve ortalama konuşma sürelerine göre analiz yapılmıştır. Aboneler farklı operatörleri arayanlar ve aramayanlar olarak ikiye ayrılmıştır. Ortalama arama sıklığının etkilediği ortalama konuşma süresi ve ortalama fatura tutarları daha ayrıntılı olarak incelenmiştir. Farklı operatörleri arayan abonelerin iptal oranları aramayanlara göre daha düşüktür. Dolayısıyla yapılacak kampanyanın farklı operatörleri aramayan abonelere yönelik bir uygulama olmasına karar verilmiştir. Farklı operatörleri aramaların ilk beş dakikası bedava yapıldığında abonelerin iptal oranlarının düşeceği öngörülmüştür. Bununla beraber elde tutulan abonelerden elde edilecek beklenen kazanç hesaplanmıştır.

Üçüncü senaryo için abonelerin tarifeleri ele alınmıştır. Tarife değişkeniyle bağlantılı olan ortalama konuşma süresi ve ortalama fatura tutarlarının ayrıntılı olarak ele alınması için yeni bir Bayes ağı oluşturulmuştur. Birinci tip tarifeye sahip çok az abone olduğu için analizin dışında bırakılmıştır. Ortalama fatura tutarları 30 TL’nin altında olan ikinci tip tarifeye bağlı olan abonelerin iptal oranları, üçüncü tip tarifeye sahip abonelerden daha fazladır. Bu nedenle bu gruptaki abonelere üçüncü tip tarifeye geçmeleri için

öneriler sunulmuştur. Böylece daha az fatura ödeyecek olan bu abonelerin iptal oranları düşecek ve şirkete uzun vadede katkıları sağlanmış olacaktır. Diğer bir yandan ortalama fatura tutarları 30 TL'nin üstünde olan üçüncü tip tarifeyle sahip abonelerin iptal oranı, ikinci tip tarifeyle bağlı abonelerin iptal oranından daha fazla olduğu görülmüştür. Bu yüzden üçüncü tip tarifeyle bağlı abonelere ikinci tip tarifeyle geçmeleri önerilmesine karar verilmiş ve sadakati sağlanan aboneler sayesinde beklenen kazanç hesaplanmıştır.

Bu tez çalışmasının sonucunda telekomünikasyon sektöründe abonelerin iptal etme sebepleri Bayes ağları yöntemi ile ortaya konmaya çalışılmıştır. Değişkenler arasındaki nedensel ilişkilerle bir model oluşturulmasını sağladığı için Bayes ağları tercih edilmiştir. Abonelerin ortalama konuşma süreleri, ortalama fatura tutarları, farklı operatörleri ortalama arama sıklığı ve bağlı oldukları tarife türünün iptali açıklayan önemli değişkenler olduğu görülmüştür. Bu tarz çalışmalarda verinin kalitesi ve miktarı çok önemlidir. Bu yüzden daha kaliteli, güvenilir ve daha çok veriyle daha nitelikli bir çalışma yapılabilir. Bunların yanında Bayes ağlarına sonradan eklenen değişkenlerle model geliştirilebilir. Abonelerin telekomünikasyon şirketinin sunduğu ek hizmetlerden (internet vb.) yararlanıp yararlanmadığına dair veriler, günün farklı periyotlarında yapılan görüşmelerin sayısı, aboneye gelen aramaların sayısı, haftanın günlerine göre yapılan aramaların gün içindeki (akşam 8.00-sabah 8.00 veya sabah 8.00-akşam 8.00 gibi) dağılımı gibi değişkenler eklenerek daha kapsamlı bir model oluşturulabilir. Bunların yanında farklı yöntemler kullanılarak veri kümesi oluşturulabilir. Müşteri memnuniyetini sorgulayan anket sonuçları ile abonelik iptalini etkileyen önemli etmenler analize eklenebilir. Ayrıca kampanya öncesi ve kampanya sonrası yapılacak anketlerle abone davranışları incelenebilir.

Bu çalışmada rekabetin giderek arttığı telekomünikasyon sektöründe abone iptalini etkileyen önemli etmenler bulunmaya çalışılmıştır. İleride yapılacak çalışmalarla telekomünikasyon sektöründe iptal sorunu için yeni açılımlar getirilebilir.

KAYNAKLAR

- [1] http://www.iec.org/online/tutorials/acrobat/bus_int.pdf, alındığı tarih 12.02.2009
- [2] <http://www.sequent.com>, alındığı tarih 09.01.2009
- [3] **Richeldi, M., and Perrucci, A.**, 2002. Churn analysis case study, *Telecom Italia Lab* Report, Torino, Italy.
- [4] **Xia, G., and Jin, W.**, 2008. Model of customer churn prediction on support vector machine. *Systems Engineering – Theory & Practice*, **28** (1), 71-77.
- [5] **SAS International**, 2001. Predicting Churn, SAS Institute, Cary, NC.
- [6] **Kim, H., and Yoon, C.**, 2004. Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market. *Telecommunications Policy*, **28**, 751-765.
- [7] **Hung, S., Yen, D., and Wang, H.**, 2006. Applying data mining to telecom churn management. *Expert Systems with Applications*, **31**, 515-524.
- [8] http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6V03-4, alındığı tarih 15.05.2008
- [9] **Wei, C., and Chiu, I.**, 2002. Turning telecommunications call details to churn prediction: A data mining approach, *Expert Systems with Applications*, **23**, 103-112.
- [10] **Ferreira, J., B., Vellasco, M., Pacheco, M., A., and Barbosa, C.H.**, 2004. Data mining techniques on the evaluation of wireless churn. In ESANN'2004. Proceedings of the European Symposium on Artificial Neural Networks, Bruges, Belgium, 28-30 April.
- [11] **Mozer, M., C., Wolniewicz, R., Grimes, D., B., Johnson, E., and Kaushansky, H.**, 2000. Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry, *IEEE Transactions on Neural Networks*, **11** (3), 690-696.
- [12] **Duke Teradata**, 2002. Teradata Center for Customer Relationship Management. , http://www.teradataduke.org/news_t_2.html, alındığı tarih 07.04.2008.
- [13] **Nath, S., V., and Behara, R., S.**, 2002. Customer churn analysis in the wireless industry: A data mining approach, alındığı tarih 28.08.2008.
- [14] **Karahoca, A., and Kara, A.**, 2006. Comparing clustering techniques for telecom churn management, *The 5th WSEAS International Conference on Telecommunications and Informatics*, İstanbul, Türkiye, Mayıs 27-29.
- [15] **Jahanzeb, S., and Jabeen, S.**, 2007. Churn management in the telecom industry of Pakistan: A comparative study of Ufone and Telenor. *Database Marketing & Customer Strategy Management*, **14** (2), 120-129.

- [16] **Sohn, S., Y., and Kim, Y.,** 2008. Searching customer patterns of mobile service using clustering and quantitative association rule. *Expert Systems with Applications*, **34**, 1070-1077.
- [17] **Heckerman, D.,** 1995. A tutorial on learning Bayesian networks, *Technical Report, MSR-TR-95-06*, Microsoft Research, Redmond, WA.
- [18] **Lee, E., Park, Y., Shin, J., G.,** 2009. Large engineering project risk management using Bayesian belief network, *Expert Systems with Applications*, **36**, 5880-5887.
- [19] **Chang, W., Lee, K., J.,** 2009. Bayesian belief network for box-office performance: A case study on Korean movies. *Expert Systems with Applications*, **36**, 280-291.
- [20] **Kim, H., and Park, Y.,** 2008. The impact of R&D collaboration on innovative performance in Korea: A Bayesian network approach. *Scientometrics*, **75** (3), 535-554.
- [21] **Ülengin, F., Önsel, Ş., Topçu, Y., İ., Aktaş, E., Kabak Ö.,** 2007. An integrated transportation decision support system for transportation policy decisions: The case of Turkey. *Transportation Research Part A*, **41**, 80-97.
- [22] **Ouali, A., Cherif, A., R., Krebs, M.,** 2006. Data mining based Bayesian networks for best classification. *Computational Statistics & Data Analysis*, **51**, 1278-1292.
- [23] **Şahin, Ş., Ö., Ülengin, F., Ülengin, B.,** 2006. A Bayesian causal map for inflation analysis: The case of Turkey. *European Journal of Operational Research*, **175**, 1268-1284.
- [24] **Jensen, F., V.,** 2001. Bayesian Networks and Decision Graphs, Springer-Verlag, New York.
- [25] **Akdeniz, F.,** 2004. Olasılık ve İstatistik, Nobel Kitabevi, Adana, Türkiye.
- [26] **Nakip, M.,** 2006. Pazarlama Araştırmaları, Teknikler ve (SPSS Destekli) Uygulamalar, Seçkin Yayıncılık, Ankara, Türkiye.
- [27] **Hair, J., F., Balck, W., C., Babin, B., J., Anderson, R., E., and Tatham, R., L.,** 2006. Multivariate Data Analysis, Pearson Prentice Hall, Upper Saddle River, New Jersey.

EKLER

Ek A.1: Veri tabanındaki değişkenlere ait açıklamalar

Ek A.2: Veritabanındaki ilk 15 aboneye ait veriler

Ek B : Verilerin hazırlık aşamasından sonra elde edilen yeni veri tabanındaki ilk 15 aboneye ait veriler

Ek C : İlk 15 aboneye ait kesikli hale getirilen veriler

Ek A.1: Veri tabanındaki değişkenlere ait açıklamalar

DEĞİŞKEN	TANIMI
Abone No	
YERLESİM	Abonenin Yaşadığı Bölge Kırsal(0), Şehir (1)
YAS	Abonenin Yaşı
A_SURE	Abonenin Gün Sayısı
TARIFE	Abonenin Bağlı olduğu Tarife Türü (0-kurumsal tarife), (1-1.tip bireysel tarife), (2-2. Tip bireysel tarife)
FATURA_1	1. Aydaki Fatura Tutarı
FATURA_2	2. Aydaki Fatura Tutarı
FATURA_3	3. Aydaki Fatura Tutarı
FATURA_4	4. Aydaki Fatura Tutarı
FATURA_5	5. Aydaki Fatura Tutarı
FATURA_6	6. Aydaki Fatura Tutarı
K_SURE_1	1. Aydaki Toplam Konuşma Süresi
K_SURE_2	2. Aydaki Toplam Konuşma Süresi
K_SURE_3	3. Aydaki Toplam Konuşma Süresi
K_SURE_4	4. Aydaki Toplam Konuşma Süresi
K_SURE_5	5. Aydaki Toplam Konuşma Süresi
K_SURE_6	6. Aydaki Toplam Konuşma Süresi
SIKLIK_1	1. Ayda Farklı Operatörü Arama Sıklığı
SIKLIK_2	2. Ayda Farklı Operatörü Arama Sıklığı
SIKLIK_3	3. Ayda Farklı Operatörü Arama Sıklığı
SIKLIK_4	4. Ayda Farklı Operatörü Arama Sıklığı
SIKLIK_5	5. Ayda Farklı Operatörü Arama Sıklığı
SIKLIK_6	6. Ayda Farklı Operatörü Arama Sıklığı
IPTAL	7. Ayda İptal Etme/Etmeme Durumu (İptal Etmeme 0, İptal Etme 1)

Ek A.2 : Veritabanındaki ilk 15 aboneye ait veriler

ABONE_NO	YERLESİM	YAS	A_SURE	TARIFE
1	1	50	927	1
2	0	41	4257	2
3	1	56	21061	1
4	1	29	808	1
5	0	41	2921	1
6	1	25	1765	1
7	1	41	6407	2
8	0	45	3222	1
9	0	80	4937	1
10	1	38	79	1
11	1	29	1577	1
12	0	41	1228	1
13	1	28	2976	2
14	1	49	1786	1
15	1	50	1379	2

FATURA_1	FATURA_2	FATURA_3	FATURA_4	FATURA_5	FATURA_6
92.5	53	40.25	14.5	24.25	19
11.75	17.75	28.5	25.75	22.79	20.52
67.5	54.75	19.5	79	85.75	78.25
34.25	38	29.25	51.85	43.55	30.5
15.75	28.25	14.25	50.84	24.74	23.25
33	18.75	18	28.25	41.07	20
9	9	9	9.5	12.63	9.04
14.25	15.5	14.75	19.32	14.5	16.75
14.25	20.25	24	43.09	33.25	34
10	10	12	16	14	14.25
24.75	27.75	26	24.55	25.23	24.5
14.25	18	17.25	25.84	19.3	25.25
12.5	28	26.25	15	26.53	31.01
40.5	46.5	36	38.39	35.72	32.5
40.25	38	25.75	24.2	28.76	29.69

K_SURE_1	K_SURE_2	K_SURE_3	K_SURE_4	K_SURE_5	K_SURE_6
676.85	320.35	233.65	4	87.95	35.69
40.13	40.98	52.78	30	61.05	33.2
264.79	184.68	399.21	363.62	398.94	434.22
132.82	184.06	217.01	332.33	267.46	173.52
29.88	12.62	6.23	152.6	46.25	48.57
127.77	24.43	13.7	61.51	91.03	15.63
7.03	3.18	3.1	3.75	14.33	1.4
0	7.01	0.7	26.65	2.49	9.46
47.65	80.28	83.03	141.72	89.62	86.5
0	0	0	0	0	0
62.59	78.04	77.37	63.91	57.31	49.79
7.72	49.92	28.88	44.92	31.88	49.68
58.79	97.03	52.88	8.86	43.25	47.86
303.21	233.85	171.9	230.08	165.1	173.3
197.33	165.02	96.07	77.92	130.7	121.32

SIKLIK_1	SIKLIK_2	SIKLIK_3	SIKLIK_4	SIKLIK_5	SIKLIK_6	IPTAL
4	1	0	0	2	2	1
10	6	6	6	4	7	0
28	21	28	36	36	29	0
4	6	1	0	1	0	0
4	3	0	10	9	5	1
7	2	4	7	9	4	0
1	0	0	0	2	1	0
0	1	0	2	0	1	1
2	2	3	7	3	7	1
0	0	0	0	0	0	1
3	4	1	1	6	2	0
2	8	12	9	2	12	1
2	3	1	0	2	1	0
7	10	8	4	2	6	0
3	2	1	2	5	5	0

Ek B: Verilerin hazırlık aşamasından sonra elde edilen yeni veri tabanındaki ilk 15 aboneye ait veriler

YERLESİM	YAS	A_SURE	TARIFE	ORT_FATURA	EGILIM	ORT_K_SURE	ORT_SIKLIK	IPTAL
1	50	927	1	40.58	0	226.42	1.50	1
0	41	4257	2	21.18	0	43.02	6.50	0
1	56	381	1	64.13	0	340.91	29.67	0
1	29	808	1	37.90	0	217.87	2.00	0
0	41	2921	1	26.18	0	49.36	5.17	1
1	25	1765	1	26.51	0	55.68	5.50	0
1	41	6407	2	9.70	0	5.47	0.67	0
0	45	3222	1	15.85	1	7.72	0.67	1
0	80	4937	1	28.14	2	88.13	4.00	1
1	38	79	1	14.54	0	0.00	0.00	1
1	29	1577	1	25.46	0	64.84	2.83	0
0	41	1228	1	19.98	1	35.50	7.50	1
1	28	2976	2	23.22	1	51.45	1.50	0
1	49	1786	1	38.27	0	212.91	6.17	0
1	50	1379	2	31.11	0	131.39	3.00	0

Ek C: İlk 15 aboneye ait kesikli hale getirilen veriler

YERLESİM	YAS	A_SURE	TARIFE	ORT_FATURA	EGILIM	ORT_K_SURE	ORT_SIKLIK	IPTAL
1	1	0	1	2	0	2	1	1
0	1	1	2	1	0	0	2	0
1	1	0	1	2	0	2	2	0
1	0	0	1	2	0	2	1	0
0	1	1	1	1	0	1	2	1
1	0	0	1	1	0	1	2	0
1	1	2	2	0	0	0	1	0
0	1	1	1	0	1	0	1	1
0	1	2	1	2	2	1	2	1
1	1	0	1	0	0	0	0	1
1	0	0	1	1	0	1	2	0
0	1	0	1	1	1	0	2	1
1	0	1	2	1	1	1	1	0
1	1	0	1	2	0	2	2	0
1	1	0	2	2	0	2	2	0

ÖZGEÇMİŞ

Ad Soyad: Pınar Kişioğlu

Doğum Yeri ve Tarihi: Bolu, 29.01.1985

Adres: İTÜ, İşletme Fakültesi, Endüstri Mühendisliği Bölümü

Lisans Üniversitesi: İTÜ Matematik Mühendisliği

