



T.C.
KAHRAMANMARAŞ SÜTÇÜ İMAM ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

ORTAÖĞRETİM BAŞARISINI ETKİLEYEN
FAKTÖRLERİN KARAR AĞACI İLE
SINIFLANDIRILMASI

ÇİĞDEM GÜLER

YÜKSEK LİSANS TEZİ
ZOOTEKNİ ANABİLİM DALI

KAHRAMANMARAŞ 2017

T.C.
KAHRAMANMARAŞ SÜTÇÜ İMAM ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

ORTAÖĞRETİM BAŞARISINI ETKİLEYEN FAKTÖRLERİN
KARAR AĞACI İLE SINIFLANDIRILMASI

ÇİĞDEM GÜLER

Bu tez,
Zootekni Anabilim dalında
YÜKSEK LİSANS
Derecesi için hazırlanmıştır.

KAHRAMANMARAŞ 2017

Kahramanmaraş Sütçü İmam Üniversitesi Fen Bilimleri Enstitüsü öğrencisi Çiğdem GÜLER tarafından hazırlanan ‘‘ORTAÖĞRETİM BAŞARISINI ETKİLEYEN FAKTÖRLERİN KARAR AĞACI İLE SINIFLANDIRILMASI’’ adlı bu tez, jürimiz tarafından 20/07/2017 tarihinde oy birliği ile Zootekni Anabilim Dalında Yüksek Lisans tezi olarak kabul edilmiştir.

Doç. Dr. Mustafa ŞAHİN (DANIŞMAN)

Zootekni Anabilim Dalı

Kahramanmaraş Sütçü İmam Üniversitesi

Prof. Dr. Ercan EFE

Zootekni Anabilim Dalı

Kahramanmaraş Sütçü İmam Üniversitesi

Doç. Dr. Fatih ÜÇKARDEŞ (ÜYE)

Biyoistatistik ve Tıbbi Bilişim ABD

Adıyaman Üniversitesi

Yukarıdaki imzaların adı geçen öğretim üyelerine ait olduğunu onaylım.

Doç. Dr. Mustafa ŞEKKELİ

Fen Bilimleri Enstitüsü Müdürü

TEZ BİLDİRİMİ

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar dâhilinde elde edilerek ve düzenlenerek sunulduğunu, aynı zamanda tez yazım kurallarına uygun olarak hazırlanan bu çalışmada, alıntı yapılan her türlü kaynağa eksiksiz atıf yapıldığını bildiririm.

Çiğdem GÜLER

Bu çalışma Kahramanmaraş Sütçü İmam Üniversitesi, Bilimsel Araştırma Projeleri Koordinasyon Birimi tarafından desteklenmiştir.

Proje No:..... YLS

Not: Bu tezde kullanılan özgün ve başka kaynaktan yapılan bildirişlerin, çizelge, şekil ve fotoğrafların kaynak gösterilmeden kullanımı, 5846 sayılı Fikir ve Sanat Eserleri Kanunundaki hükümlere tabidir.

**ORTAÖĞRETİM BAŞARISINI ETKİLEYEN FAKTÖRLERİN KARAR AĞACI
İLE SINIFLANDIRILMASI
(YÜKSEK LİSANS TEZİ)**

Çiğdem GÜLER

ÖZET

Bu çalışmanın amacı ortaokul eğitim başarısını etkileyen faktörlerin karar ağacı yöntemi ile sınıflandırılmasıdır. Bu amaçla Kahramanmaraş Dulkadiroğlu ilçesine bağlı bulunan 44 ortaokuldan alınan veriler kullanılarak çalışmalar yapılmaktadır. Araştırmada kullanılacak olan değişkenler eğitim kurumlarında görev yapan öğretmenler ile görüşülerek belirlenmektedir. Araştırmada bağımlı değişken öğrenci başarısı (ÖB) ve bağımsız değişkenler ise; öğrencilerin kitap okuma alışkanlığı (KOA), sayısal yetenek becerisi (SYB), odaya sahip olması (OS), kardeş sayısı (KS), hastalık durumu (HD), baba iş durumu (BİD), anne iş durumu (AİD), yaşadığı konum (YK), bilgisayar bağımlılığı (BB), televizyon bağımlılığı (TB), baba eğitim durumu (BED), anne eğitim durumu (AED), okul öncesi eğitim durumu (OÖED), ailesi ile birliktelik durumu (AB) olarak belirlenmektedir. Karar ağaçlarında sınıflandırma yapmak için birçok algoritma geliştirilmektedir. Geliştirilen algoritmalar ile analizler yapılarak en iyi sınıflandırmayı sağlayan algoritma yapısıyla ağaç modellemesi oluşturulmaktadır. Karar ağaçlarında kullanılan CHAID, C5.0, C&R TREE ve QUEST algoritmaları uygulanarak ve elde edilen analiz sonuçlarına bakılarak en uygun sınıflandırıcı algoritma tespit edilmektedir. Söz konusu bu algoritmalara göre oluşturulan analiz sonuçları karşılaştırıldığında %90.23 ile en iyi performansı C5.0 karar ağacı algoritması sağlamıştır. Bu algoritma yapısının oluşturduğu ağaç modellemesine göre sınıflandırıcı kurallar belirlenmektedir. Elde edilen karar kurallarına göre öğrenci başarısını en fazla etkileyen bağımsız değişken anne iş (AİD) değişkenidir. Oluşturulan kurallarda, C5.0 algoritması ile elde edilen ağaç yapısında % 50 ve üzeri başarılı olan öğrenciler yer almaktadır. Çalışma sonunda 24 tane sınıflandırıcı kural elde edilmektedir.

Anahtar Kelime: Eğitim, Ortaokul, Veri Madenciliği, Karar Ağacı

Kahramanmaraş Sütçü İmam Üniversitesi

Fen Bilimleri Enstitüsü

Zootekni Anabilim Dalı, Ağustos/ 2017

Danışman: Doç. Dr. Mustafa ŞAHİN

Sayfa Numarası: 63



DETERMINATION OF THE FACTORS AFFECTING THE SUCCESS OF SECONDARY EDUCATION BY DECISION TREE

(POST GRADUATE THESIS)

Çiğdem GÜLER

ABSTRACT

The aim of this study is to classify the factors affecting the success of secondary education by decision tree analysis. In accordance with this purpose, studies were carried out through the data obtained from 44 secondary schools in Dulkadiroğlu, Kahramanmaraş. In this study, dependent variable is student success (SS) and independent variables affecting the dependent variable are reading habit (RH), numerical ability (NA), having a room (HR), sibling number (SN), health condition (HC), father job status (FJS), mother job status (MJS), the location he/she lives (LL), computer addiction (CA), television addiction (TA), father educational status (FES), mother educational status (MES), preschool education status (PES) and status of living with his/her own family (FL). A number of algorithms were developed in order to classify in decision trees. The best algorithms were determined to be used in these developed algorithms and by doing analyzes with these algorithms, the algorithm allowing the best classifying and tree modelling were formed. The most appropriate classifier algorithm was determined by practising CHAID, C5.0, C&R TREE and QUEST algorithms and looking at the analysis results. Among these classifiers, C5.0 decision tree algorithm provided the best performance. Classifier rules were determined according to tree modelling this algorithm structure formed. According to the decision rules obtained, the independent variable that affects the student's success is the mother work (MW) variable. In the generated rules, the tree structure obtained with the C5.0 algorithm is 50% and the students who are successful are found. At the end of the study, 24 classifier rules are obtained.

Key Words : Education, Data Mining, Decision Tree

Kahramanmaraş Sütçü İmam University
Graduate School in Natural and Applied Sciences
Department of Animal Science, July/ 2017
Supervisor: Doç. Dr. Mustafa ŞAHİN
Page Numbers: 63



TEŐEKKÜR

Çalıőmalarım sırasında beni yönlendiren, her türlü bilgi ve tecrübesi ile beni aydınlatan hocam, Doç. Dr. Mustafa ŐAHİN' e, tezin uygulamasında yardımlarını esirgemeyen Doç. Dr. Fatih ÜÇKARDEŐ' e ve maddi manevi desteklerini esirgemeyen babam, Mustafa GÜLER' e, annem, Güldane GÜLER' e, ve kardeşlerim' e, tez çalışmalarımda yanımda olan arkadaşım Esra YAVUZ' a sonsuz teşekkürlerimi sunarım.



İÇİNDEKİLER

ÖZET	I
ABSTRACT	III
TEŞEKKÜR	V
İÇİNDEKİLER	VIII
ŞEKİLLER DİZİNİ	VIII
ÇİZELGELER DİZİNİ	IX
EKLER DİZİNİ	X
SİMGELER VE KISALTMALAR DİZİNİ	XI
1. GİRİŞ	1
2. ÖNCEKİ ÇALIŞMALAR	3
3. MATERYAL VE YÖNTEM	8
3.1. Materyal	8
3.2. Metot	8
3.2.1. Veri Madenciliği	8
3.2.2. Veri Madenciliği Kullanım Alanları	11
3.2.3. Veri Madenciliği İle İstatistik Arasındaki İlişki	11
3.2.4. Veri Madenciliği Neden Kullanılır	12
3.2.5. Süreç Olarak Veri Madenciliği	13
3.2.6. Veri Madenciliği İle Karar Destek Sistemleri Arasındaki İlişki	14
3.2.7. Karar Ağacı	16
3.2.8. Karar Ağacı Kullanım Alanları	20
3.2.9. Karar Ağacı Kullanımının Avantajları Ve Dezavantajları	22
3.2.10. Karar Ağacının Uygulanma Koşulları	23
3.2.11. Karar Ağacı Oluşturma Süreci	24
3.2.12. Karar Ağacı Algoritmaları	25
3.2.12.1. Hunt Algoritması	26
3.2.12.2. ID3 Algoritması	26
3.2.12.2.1. Karar Ağacında Entropi	27
3.2.12.3. C4.5 Algoritması	28
3.2.12.4. C5.0 Algoritması	29
3.2.12.5. Sınıflandırma ve Regrasyon Ağaçları (C&RT)	30

3.2.12.6. CHAID Algoritması	31
3.2.12.7. Quest Algoritması.....	32
3.2.12.8. SLIQ Algoritması	32
3.2.12.9. Sprint Algoritması.....	33
4. BULGULAR.....	35
4.1. Modelin Oluřturulması Ve Sonu Üretimi	36
5. TARTIřMA	57
6. SONU	58
KAYNAKLAR	60
EKLER	
ÖZGEMIř	

ŞEKİLLER DİZİNİ

	<u>Sayfa No</u>
Şekil 1. Veri Madenciliği Yöntemleri.....	18
Şekil 2. PASW Akışı	37
Şekil 3. Verilerin Alınması	37
Şekil 4. Veri Tablosu	38
Şekil 5. Değişken Tiplerinin Belirlenmesi.....	39
Şekil 6. C5.0 ile Oluşturulan Karar Ağacı	40
Şekil 7. Düğüm 0, 1, 78, 2, 77, 79 ve 84.....	41
Şekil 8. Düğüm 3, 4, 45, 5, 10, 6, 7, 11, 44, 8, 9, 12 ve 43.....	42
Şekil 9. Düğüm 13, 14, 15, 40, 16, 17, 41 ve 42.....	43
Şekil 10. Düğüm 46, 80, 83, 47, 76, 81, 82, 48, 75, 49, 70, 50, 69, 71 ve 72.....	44
Şekil 11. Düğüm 18, 23, 19, 20, 24, 33, 21, 22, 25, 32, 34 ve 39.....	45
Şekil 12. Düğüm 51, 56, 73, 74, 52, 53, 57, 58, 54, 55, 59 ve 62.....	46
Şekil 13. Düğüm 26,31, 35, 38, 27, 30, 36, 37, 28 ve 29.....	47
Şekil 14. Düğüm 60, 61, 63, 68, 64, 65, 66 ve 67.....	48
Şekil 15. Başarı Değişkenine Etki Eden Faktörler.....	49

ÇİZELGELER DİZİNİ

	<u>Sayfa No</u>
Çizelge 1. Veri Madenciliği Kullanım Alanları	11
Çizelge 2. İstatistiksel Analiz ve Veri Madenciliği.....	12
Çizelge 3. ID3 Algoritması örneği başarı durum verisi	28
Çizelge 4. Karar Ağacı Algoritmaları	33
Çizelge 5. Başarı, kitap, sayısal, kendi odası, hastalık, baba iş durumu, anne iş bilgisayar, okul öncesi eğitim, aile ile durumu ve TV bağımlılığı değişkenlerine ait toplam ve yüzde değerler	35
Çizelge 6. Kardeş değişkenine ait toplam ve yüzde değerler	36
Çizelge 7. TV bağımlılığı değişkenine ait toplam ve yüzde değerler	36
Çizelge 8. Eğitim durumu değişkenine ait toplam ve yüzde değerler	36
Çizelge 9. QUEST, C&RT, CHAID ve C5.0 Analiz Sonuçları.....	40
Çizelge 10. Karar ağacı ile elde edilen kurallar	50

EKLER DİZİNİ

	<u>Sayfa No</u>
Ek 1. QUEST Algoritması Analiz Sonuçları	64
Ek 2. C&RT Algoritması Analiz Sonuçları	64
Ek 3. CHAID Algoritması Analiz Sonuçları	65
Ek 4. C5.0 Algoritması Analiz Sonuçları	65
Ek 5. Anket Çalışması.....	66



SİMGELER VE KISALTMALAR DİZİNİ

- ark.** : Arkadaşları
KA : Karar Ağaçları
ÖB :Öğrenci Başarısı
KOA : Kitap Okuma Alışkanlığı
SYB : Sayısal Yetenek Becerisi
OS : Odaya Sahip Olma
KS : Kardeş Sayısı
HD : Hastalık Durumu
BİD : Baba İş Durumu
AİD : Anne İş Durumu
YK : Yaşadığı Konum
BB : Bilgisayar Bağımlılığı
TB : Televizyon Bağımlılığı
BED : Baba Eğitim Durumu
AED : Anne Eğitim Durumu
OÖED: Okulöncesi Eğitim Durumu
AB : Aile ile Birliktelik Durumu
KDS : Karar Destek Sistemi
VM : Veri Madenciliği

1. GİRİŞ

İstatistik; arařtırmaların planlanmasında, verilerin elde edilmesinde, verilerin analiz edilmesinde, sonuçların yorumlanmasında ve sunulmasında kullanılmasının yanı sıra arařtırmalarda ortaya ıkabilecek birok problemi özmede kullanılan bir araçtır. Her geen gün veri miktarındaki artıştan dolayı bazı durumlarda kullanımı sınırlı olabilmektedir. Büyük veri yığınlarının analiz edilmesi ve elde edilen sonuçlara göre karar verilmesi oldukça zorlařtığından yapay sinir ağıları, karar ıkarımı, mantık programlama, karar ağaları, genetik algoritma gibi akıllı veri analiz yöntemlerini ieren veri madenciliğinin önemi ve uygulanması daha da önem kazanmaktadır. Veri sayısında meydana gelen bu aşırı artıştan dolayı alışıl gelmiş yöntemler kullanılarak bu veri yığınının anlamlı bir sonuç ıkarmak mümkün olmayacağı kanaatine varılarak, son yıllarda her alanda kurum ve kişilere destek olmak amacıyla veri madenciliği tekniği yaygın olarak kullanılmaya başlanmıştır.

Büyük miktarlarda tutulan verilerden, aslında var olan fakat farkındalık oluşturmayan saklı kalmış desenlerin açığa ıkarılması günümüzde odaklanılan konuların başında yer almaktadır. Veri madenciliği olarak bilinen bu kapsam, büyük çoğunluktaki veri arasında saklı olan kıymetli ve önemli bilgilerin ortaya ıkarılması ve bu bilgiler ışığında gelecekle ilgili varsayım oluşturulmasına imkân saėlayan fonksiyonların ve kuralların taranması süreci olarak ifade edilebilmektedir (Göken Alı, 2014).

Eğitim, günümüzde en fazla oluşumun ve gelişimin saėlandığı kapsamların başında yer almaktadır. Bu gelişimi saėlayabilmek için birok alıřma yapılmaktadır. Bu alıřmaları saėlıklı bir şekilde yürütebilmek için verilere ihtiyaç duyulmaktadır. Bu bilgiler doėrultusunda alıřmada, eğitim verilerinden yararlanılarak geleceğe ışık tutan deėerli bilgilere ulaşmak için sınıflandırıcı karar ağacı modeli oluşturularak kurallar elde edilmiştir. Veri madenciliği, eğitim alanındaki karmaşık sistemde bulunan saklı bilgileri ortaya ıkararak ve sistem başarısını destekler nitelikte olacak şekilde bilgilerin açıklığa kavuřturulması imkânını sunmuştur. Eğitim ile ilgili yapılan alıřmalarda Söz konusu olan birey ve toplum gelişimi olduėu için bu kapsamda yapılan arařtırmalar ve alıřmalar gelecek adına büyük bir öneme sahiptir. Birey ve topluma verilen önemin her geen gün artması sonucunda eğitimde gelişim göstermek durumundadır. Eğitim ile ilgili birok yenilikler yapılmıştır ve yapılmaya da devam edilmektedir. Bu gelişim

abalarında en doęru karar srecine ulařmak iin kullanılan karar aęaları sistemi olduka uygun olmaktadır. Eęitim bir sre sistemi olduęu iin bu srete bize ıřık tutabilecek en kk veri seti dahi kıymetlidir.

Bu tez alıřmasında, uygulama alanı olarak eęitimde ortaokul bařarısını etkileyen faktrlerin Karar aęacı (KA) yntemi ile sınıflandırılması saęlanmıřtır. Ortaokul eęitim srecinde olduka neme sahip bir dnemdir ve bu dnemin en saęlıklı biimde olabilmesi iin mevcut kuralların olması gerekir. KA modeli kullanılarak bu kurallar btn oluřturulmaktadır. alıřmada ortaokul bařarısını etkileyen faktrlerin sınıflandırılmasında KA yntemi kullanılmakta ve bir eęitim veri madencilięi alıřması saęlanmaktadır. Elde edilen KA modelinden oluřturulan kurallar eęitim alıřmalarına yn verecek niteliktedir.

2. ÖNCEKİ ÇALIŞMALAR

Veri madenciliği birçok kullanım alanına sahip olması, konunun popüler olmasına ve bu konuda birçok bilimsel çalışmaların yapılmasını da beraberinde getirmiştir. Veri madenciliği alanında birçok teorik/uygulama çalışmaları bulunmaktadır (Hilage ve Kulkarni, 2012). KA son yıllarda literatürde yaygın kullanımı olan bir sınıflandırma ve örüntü tanımlama algoritmasıdır. Ağaç yapılarının oluşturulmasında kullanılan kuralların sade ve anlaşılabilir olmasından dolayı bu yöntem yaygın bir kullanım alanına sahiptir. Karar ağaçlarının sınıflandırılmasında çok aşamalı veya ardışık bir yaklaşım kullanılmaktadır (Safavian, 1991).

Erdoğan ve ark. (2005), Çalışmalarında Maltepe Üniversitesi öğrencilerini kümeleme teknikleri ile kümelere ayırarak, üniversite giriş sınav puanı ile üniversite dersleri başarıları arasındaki ilişkiyi araştırmışlardır.

Im ve ark. (2005), Tasarladıkları bir anketi uygulayarak sınıflandırma ve kümeleme yöntemleri ile üstün yetenekli öğrencilerin yetenek türlerini ortaya çıkarmış, ayrıca benzer karakteristikteki öğrencileri gruplamışlardır.

Amershi ve Conati (2006), Bilgisayar tabanlı bir öğrenme ortamında öğrencileri davranışlarına göre kümelere ayırmışlardır. Böylelikle her öğrenci grubuna uyarlanabilir destek sağlanabilecektir.

Tissera ve ark. (2006), Çalışmalarında birliktelik analizlerini kullanarak güçlü bir şekilde ilişkili olan derslerin saptanabileceğini göstermişlerdir. Araştırma öğrencilerin ilişkili derslerden benzer notları alacağı varsayımı üzerine kurulmuştur. Çalışmanın sonuçları ders programlarının düzenlenmesinde kullanılabilir.

Bresfelean ve ark. (2008), çalışmada öğrenci başarısını desteklemek amacıyla veri madenciliği sınıflandırma ve veri kümeleme tekniklerini uygulayarak öğrenci profillerini tespit edip öğrenci başarısızlık nedenlerini ortaya koymuşlardır.

Gülpınar (2008), Avrupa Birliği ülkeleri ile Türkiye'nin ekonomik göstergelerini karar ağacı yöntemini kullanarak belirlemeye çalışmıştır. Sonuç olarak C4.5 karar ağacı algoritmasını kullanarak ağaç yapısını ve bu yapıdan karar ağacı kurallarını oluşturmuştur.

Halees (2008), çalışmada 151 öğrenciye yönelik olan bazı verileri kullanarak öğrenci tutumlarını veri madenciliği ile tespit ederek ve bulunan sonuçlara göre gereken iyileştirme çabaları yapılmıştır

Kayri (2008), çalışmada öğrencilerin performans takibinin sürekli olarak yapılabildiği ve bilgisayar sistemi aracılığıyla ürünler arası örüntülerin izlenebildiği, e-portfolio değerlendirmeleri için veri madenciliğinde kullanılan yöntemler uygulanabilir bir ölçme aracı olarak belirlenebilir.

Koyuncugil ve Özgülbaş (2008), İMKB’de işlem gören kobi’lerin güçlü ve zayıf yönler: CHAID karar ağacı uygulamasında kobilerin karlılıklarını ve finansal performans düzeylerinin artması sonucuna ulaşılmıştır.

Bozkır ve ark. (2009), Çalışmada 2008 yılı ÖSS dönemine ait ÖSYM’nin internet sitesinden sağlanan bilgi anketi ile elde edilen veriler veri madenciliği uygulanılarak analiz edilmiş ve ÖSS’de başarıyı temsil eden sözel, sayısal ve eşit ağırlık puanı ve ağırlıklı ortaöğretim başarı puanı gibi sonuçlarını etkileyen en önemli etmenlerin tespit edilmesi amaçlanmıştır.

Gaafar ve Khanmis (2009), Kahire Amerikan Üniversitesinde gerçekleştirilen çalışmada farklı veri ambarlarından elde edilen verileri kullanarak sağlanan veri yığınının değişik veri madenciliği teknikleri uygulanılarak mezun olabilecek, okulu bırakabilecek şekilde iki farklı öğrenci profilinin modellenmesi sağlanmıştır.

Ayesha ve ark. (2010), Çalışmasında öğrenci bilgilerinden final sınav notları tahmin edilmiş, böylece öğrencilerin erken uyarılmaları sağlanmıştır. Ayrıca öğrenciler kümelerine ayrılmıştır.

Kelley-Winstead (2010), Sınıflandırma ağaçları ile aynı öğretim yılını tekrar edecek öğrencileri tahmin eden bir çalışma gerçekleştirmiştir.

Zhang ve ark. (2010), çalışmada risk altındaki öğrencilere nasıl destek verilebileceğini, ders uygunluğunun nasıl belirlenebileceğini ve bulunan sonuçların öğrencilere nasıl uygulanılabileceğini araştırmışlardır.

Akçapınar ve ark. (2011), Çalışmalarında çevrim içi öğrenme ortamında Random Forest Regression (RFR) veri madenciliği yöntemi ile öğrencilerin algılanan yön duygusu yitirme durumunun tahmini için iki model önermişlerdir. Söz konusu çalışmanın sonuçları, öğrencilerin destek eğitimi ve bireysel yetenekleri fark ettirme programlarının son aşamalarında hangi alanlara yönlendirilmeleri gerektiğini belirlemek için sistematik ve bilimsel bir yöntem sunmaktadır.

Kumar ve Chadha (2011), Makalelerinde, veri madenciliği tekniklerinin kullanılması ile ilişkili derslerin tespit edilip bu sayede ders programlarının daha etkin düzenlenebileceğini, çeşitli eğitim programlarına kaç öğrencinin katılacağını tahmin edilebileceğini, öğrencilerinin performanslarının öngörülebileceğini, öğrencilerin sınav kâğıtlarındaki anormal değerlerin tespit edilebileceğini ve benzer özellikteki öğrencilerin gruplanarak sınıflar oluşturulabileceğini belirtmişlerdir.

Tsai ve ark. (2011), Bilgisayar yeterlilik sınavından kalacak öğrencileri tahmin etmişlerdir. Çalışmada hem kümeleme hem sınıflandırma teknikleri kullanılmıştır.

Baradwaj ve Pal (2012), Bilgisayar uygulamaları dersindeki akademik başarıyı tahmin etmişler, ayrıca başarı üzerinde etkili faktörleri saptamışlardır.

Bırtıl (2012), Çalışmada öğrencilerin başarısızlık sebeplerini tespit etmek için yapılan anket çalışması veri madenciliği yöntemlerinden kümeleme tekniği kullanılarak incelenmiş, öğrencilerin başarısız olmalarına neden olan etkenlerin aynı zamanda görüldüğünün tespit edilmesi ve bunlar arasındaki ilişkilerin saptanması amaçlanmıştır.

Dejaeger ve ark. (2012), Çalışmalarında sınıflandırma teknikleri ile öğrencilerin memnuniyetini etkileyen faktörleri saptamışlardır.

Erdem ve Kurt (2012), Çalışmada başarılı veya başarısız olarak tanımlanan öğrenciler tespit edilmiş ve bunun sonucunda uygun olacak çözüm önerileri sunulmuştur. Sıradan bakış açılarının aksine yeni ve farklı bakış açılarını ortaya çıkartmak amacıyla veri madenciliği uygulanmıştır.

Jormanainen ve Sutinen (2012), Robot derslerinde öğrencilerin robot programlaması için müdahaleye ihtiyaç duyup duymadıklarını tahmin eden bir model geliştirmişlerdir.

Leong ve ark. (2012), Öğrencilerin cep telefonu mesajları üzerinde metin madenciliği tekniklerini uygulayarak ders ve öğretmen değerlendirme yapmışlardır.

Sevindik ve ark. (2012), Çalışmalarında çağın gereksinimi haline gelen web tabanlı uzaktan eğitim sistemlerini veri madenciliği uygulamalarını kullanarak analiz edilmiştir.

Şen ve ark.(2012), Ortaöğretim yerleştirme testi skorunu tahmin eden bir model geliştirerek başarı üzerinde etkili olan faktörleri saptamışlardır.

Şen ve Uçar (2012), Veri madenciliği yöntemleri ile Karabük Üniversitesi Bilgisayar Mühendisliği Bölümü öğrencilerinin başarılarını çeşitli kriterleri kullanarak karşılaştırmışlardır. Çalışma uzaktan eğitimde örgün eğitime göre başarı oranının düştüğünü göstermektedir.

Taşdemir (2012), çalışmada ÖSYM tarafından gönderilen öğrenci verilerinden ve öğrencilerin öğrenim dönemindeki ders başarılarından faydalanılarak öğrencilerin akademik başarısına etki eden etkenler veri madenciliği tekniklerinden regresyon kullanılarak tespit edilmeye çalışılmıştır.

Yadav ve ark. (2012), Yükseköğretimde okuldan ayrılabilir öğrencileri tahmin eden bir çalışma gerçekleştirmişler, böylece öğrencileri okulda tutma programından en çok yararlanma ihtiyacı olan öğrencilerin kısa listesinin oluşturulabileceğini belirtmişlerdir.

Gülen ve Özdemir (2013), Çalışmada Ankara bilim ve sanat merkezinde 12 ve üstü yaşlarındaki üstün yetenekli öğrenim gören öğrencilere akademik benlik kavram ölçeği ve boş zamanları değerlendirme anketi yapılarak sağlanan verilerden üstün yetenekli öğrencilerin sıklıkla bir arada ilgi duydukları alanlar Apriori birliktelik algoritması uygulanarak bulmak hedeflenmiştir. Çalışma sonunda ulaşılan sonuçlardan Bilim ve Sanat Merkezlerindeki üstün yetenekli öğrencilerin eğitiminin bireysel gereksinimlerine göre farklılaştırılması ve ders programlarının daha etkili olacak şekilde yenilenmesi gibi mevzularda birçok yarar sağlayacağı düşünülmüştür.

Hark (2013), Çalışmada 905 öğrenciye anket yapılarak belirlenen verilerden veri madenciliği yöntemlerinden birliktelik kuralları kullanılarak etkileşimli tahta kullanımına yönelik öğrenci profilleri tespit edilmeye çalışılmıştır.

Şengür ve Tekin (2013), Çalışmada 2011 yılında mezun olan 127 bilgisayar ve öğretim teknolojileri öğretmenliği öğrencisinin 4 sene boyunca aldıkları 49 kültür ve mesleki dersin sene sonu notları kullanılarak mezuniyet notlarının tahmin edilmesi hedeflenmiştir.

Bilen ve ark. (2014), Çalışmada 2011 senesinde İstanbul şehrinde LYS sınavına katılan 42 farklı lise türünün sınavdaki başarı durumlarına göre kümelenebilirliği hedeflenmiştir. Kümeleme olayının gerçekleştirilmesinde ve hangi test türlerinin kümelere dağılımında daha etkili olduğunun saptanmasında veri madenciliği tekniklerinden kümeleme ve karar ağaçları kullanılmıştır.

Aydın ve Özkul (2015), Çalışmada öğrencilere ait değişik kaynaklardaki veriler toplanarak öğrenci başarısını tahmin etmek için veri madenciliği çalışması yapılmıştır. Uygulanan karar ağacı algoritmalarından en yüksek başarıyı C5.0 algoritması sağlamıştır. Çalışmanın sonucunda bilgi ve iletişim teknolojilerindeki gelişim öğrencilerin öğrenme başarısını etkilemektedir.

Cihan ve Kalıpsız (2015), Çalışmada veri madenciliği kullanılarak öğrencilerin proje anketlerinin sınıflandırılması yapılmış ve bu çalışma için uygun olan en başarılı algoritma tespit edilmiştir.

Emel ve Taşkın (2015), Çalışmada perakendeci işletme için işletmenin müşterilere göre kişileştirilmiş satış hareketlerini içeren veri tabanından yararlanarak ayrıntılı ve göreceli ölçüm sonuçlarını içeren bir satış analizi amaçlanmaktadır. C&RT karar ağacı algoritması kullanılmıştır. Bulunan karar kuralları ile işletmenin pazarlama stratejilerinin etkinliği de artırılabilir.

Özbay (2015), Çalışmada eğitim verilerini kullanarak veri madenciliği çalışması yapılmıştır. Çalışma sonunda öğrenci profilleri sınıflandırılmış, yaşanan problemlere çözüm üretilmeye çalışılmış, risk altındaki öğrenciler tespit edilmiş ve mezuniyet ortalamaları tahmin edilmeye çalışılmıştır.

3. MATERYAL VE METOT

3.1. Materyal

Kahramanmaraş ilindeki 44 farklı ortaokulda 4024 öğrenciye hazırlanan anket uygulanmakta ve anket sonuçları ile veri seti oluşturulmaktadır. Anket soruları öğretmenler ile görüşülerek belirlenmiş ve öğrenci başarısını etkileyen bağımsız değişkenler belirlenmiştir. Çalışmada 14 bağımsız değişken ve 1 bağımlı değişken ile ağaç modeli oluşturulmaktadır. Veri setinde bağımlı değişken; öğrenci başarısı (ÖB), bağımsız değişkenler; öğrencilerin kitap okuma alışkanlığı (KOA), sayısal yetenek becerisi (SYB), odaya sahip olması (OS), kardeş sayısı (KS), hastalık durumu (HD), baba iş durumu (BİD), anne iş durumu (AİD), yaşadığı konum (YK), bilgisayar bağımlılığı (BB), televizyon bağımlılığı (TB), baba eğitim durumu (BED), anne eğitim durumu (AED), okul öncesi eğitim durumu (OÖED), ailesi ile birliktelik durumu (AB) olarak belirlendi. Bağımlı değişken olan başarı öğrencilerin ortaöğretim boyunca aldığı belgeler baza alınarak belirlenmektedir. Veri madenciliği tekniklerinden olan KA tekniği kullanılarak öğrencilerden alınan veriler doğrultusunda, öğrenci başarısını etkileyen faktörler sınıflandırılmış ve sınıflandırma sonucunda sınıflandırma kuralları oluşturulmaktadır. Bu sınıflandırma işlemini sağlamak için karar ağaçlarında kullanılan SPSS tabanlı Clementine yazılımı kullanıldı.

3.2. Metot

KA veri madenciliğinde kullanılan bir teknik olması nedeniyle konuya veri madenciliği (VM) tanımlaması ile başlanmaktadır. Veri madenciliğinde karar ağaçları, ağaç yapılarının kuralları ifade edebilmesi, kurulumlarının ucuz ve yorumlanmasının kolay olması, veri tabanı sistemleri ile kolayca bütünleşebilmesinden, güvenilirliklerinin iyi olmasından dolayı sınıflama modelleri içerisinde yaygın olarak kullanılan, tahmin edici bir tekniktir.

3.2.1. Veri Madenciliği

1950'li yıllarda sayımlar yapmak için ilk bilgisayarlar kullanılmaya başlandı. 1960'lı yıllarda ise veri tabanı ve verilerin depolanması kavramı teknoloji dünyasında bir yere sahip oldu. Bilişim alanındaki gelişmelere paralel olarak, üretilen sayısal bilgilerin çoğalması veri tabanlarının daha fazla veriyi elde tutabilecek düzeye ulaşmasına neden olmaktadır (Gökçen Alıç, 2014). Söz konusu verilerdeki saklı

örüntülerin açığa çıkarılması, bunlardan kayda değer bilgilere ulaşarak karar işleyişine sunulması gerekliliği veri madenciliğinin özel bir araştırma alanı olmasına neden olmaktadır (Gökçen Alıç, 2014). Başka bir tanımlama ile veri madenciliği, veri grubunda saklı kalan ve farkındalık yaratması güç olan örüntüleri ortaya çıkaran geleceğe ışık tutabilecek düzeyde tahmin yapılmasını sağlayan kuralları üreten veri tabanı teknolojisi ve tekniklerinin uygulama alanıdır (Gökçen Alıç, 2014).

Veri madenciliği, bilgi teknolojisi ile paralel gelişim göstermektedir. Veri madenciliği son yıllarda toplumumuz için çok önemli bir yere sahiptir (Han ve Kamber, 2006).

Veri madenciliği, fazla miktarda veri bulunduran ve bu verilerle gelecekle ilgili tahmin yapabilmek için, bağıntı ve kuralların bilgisayar programlarını kullanarak oluşturulmasıdır (Babadağ, 2006).

Veri madenciliğinin bir başka tanımı ise, veri kümesinin içindeki verilerin anlamlandırılması ve yararlı olacak şekilde ilişkilendirilmesi, veriyi yeniden özetlemek için veri kümesinin incelenmesidir (Larose, 2005).

Veri madenciliği tahmin edici ve tanımlayıcı yöntemler olmak üzere iki sınıfta incelenmektedir. Tanımlayıcı yöntemler, büyük veri kümelerindeki ilişkileri tespit ederek, incelenen sistemin anlamını kavramaktır (Akpınar, 2000). Denetleme, Birliktelik Analizi Yöntemleri, tanımlayıcı yöntemlerden bir kaçıdır. Örneğin, 30 yaş üstü evli kişiler ile 30 yaş altı bekâr kişiler üzerinde yapılan ve ödeme performanslarını gösteren bir analiz tanımlayıcı modellere örnek verilebilir.

Tahmin edici yöntemler ise, sonuçları bilinen verilerden hareket edilerek bir model geliştirilmesi ve oluşturulan bu modelden yararlanılarak sonuçları bilinmeyen veri kümeleri için sonuç değerlerin tahmin edilmesidir (Akpınar, 2000). Tahmin edici yöntemler Regresyon (Eğri Uydurma) ve Sınıflandırma olarak ikiye ayrılmaktadır.

Veri madenciliğinde, tahmin edici ve tanımlayıcı modeller kullanılmaktadır. Tahmin edici modeller, sonuçları belli olan veriler kullanılarak bir model geliştirilmekte ve oluşturulan bu model kullanılarak sonuçları belli olmayan veri kümelerinin değerlerini tahmin edilmesinde kullanılmaktadır. Tanımlayıcı modeller ise, karar vermeye yardımcı olmak için elde bulunan verilerdeki örüntülerin tanımlanması sağlanarak model oluşturulmaktadır (Akpınar, 2000).

Veri madenciliğinde modeller kullanıldıkları yerlere göre; Sınıflama ve Regresyon Modelleri, Kümeleme Modelleri, Birliktelik Kuralı ve Ardışık Zamanlı Örüntü Modelleri olmak üzere üç grupta toplanmaktadır.

Sınıflama ve regresyon modelleri tahmin edici, Kümeleme Modelleri ve Birliktelik Kuralı -Ardışık Zamanlı Örüntüler ise tanımlayıcı modeller içerisinde yer almaktadır.

Tahmin etmede yararlanılan ve veri madenciliği teknikleri içerisinde en çok kullanım alanına sahip olan model, sınıflama ve regresyon modelleridir. Bu modeller önemli veri sınıflarını ortaya koyan veya gelecekte veri sonuçlarını tahmin etmek için modelleri kurabilen veri analiz yöntemidir.

Sınıflandırma, en çok bilinen ve kullanılan veri madenciliği modellerinden biridir. Resim, örüntü tanıma, hastalık tanıları, dolandırıcılık tespiti, kalite kontrol çalışmaları ve pazarlama konuları sınıflandırmanın çokça kullanıldığı alanlardır. Sınıflandırma tahminleyici bir modeldir. Örneğin, havanın ertesi gün nasıl olacağı ya da bir kutuda ne kadar mavi top olduğunun tahmin edilmesi aslında bir sınıflandırma işlemidir (Dunham, 2003).

Sınıflandırma öğrenme algoritmasına dayanan bir yöntem olup, öğrenmenin amacı da bir sınıflandırma modelinin oluşturulmasıdır.

Sınıflandırma işleminde mevcut sınıf veya bağımlı değişken hem sınıfsal hem de sürekli değer taşıyabilir, bu da regresyon ve çok terimli regresyona yaklaştığını göstermektedir (Akpınar 2000). Veri madenciliği çerçevesinde bu istatistiksel yöntemlerin dışında, sınıflama ve regresyon modellerinde, karar ağaçlarına dayalı algoritmalar, yapay sinir ağları, genetik algoritmalar, k-en yakın komşu algoritması ve nalve-bayes sınıflandırma algoritması gibi birçok teknik ve algoritma kullanılmaktadır (Ayık ve ark. 2007).

Bu tekniklerden karar ağaçları, hedef fonksiyonları hesaplayabilmek için kullanılan en güçlü ve en yaygın, sınıfları bilinen örnek veriden tümevarım yöntemiyle öğrenilen, ağaç şeklinde bir karar verme ve sınıflandırma araçlarından birisidir (Ma, 1998). Karar ağaçları, basit karar verme adımlarını uygulayarak, büyük yapılı verileri daha küçük veri gruplarına bölerek kullanılmaktadır. Her başarılı bölme işlemiyle sonuç gruplarının üyeleri bir diğeriyle benzer bir hale gelmektedir (Sun ve Hui, 2008). Yani

bu model, bir unsur hakkındaki gözlemlerin oluşturulmasından istenen değer ile ilgili sonuca ulaşabilmektedir.

3.2.2. Veri Madenciliği Kullanım Alanları

Veri madenciliği yığın bilgileri anlamlandırmaktır. Bilgisayar sisteminin kullanıldığı hemen hemen her alanda yaygın olarak kullanılır. Gelişen toplumla birlikte veri yığınları da çoğalmakta ve gün geçtikçe veri madenciliğine ihtiyaç artmaktadır.

Veri madenciliği; bankacılık, finans, pazarlama, e-ticaret, tıp gibi birçok alanda kullanılmaktadır (Eker, 2006). Veri madenciliğinin belirttiğimiz uygulama alanlarından ziyade daha birçok uygulama alanına sahip olması, konunun ne derece önemli olduğunun göstergesidir (Gökçen Alıç, 2014).

Çizelge 1. Veri Madenciliği Kullanım Alanları (Gökçen Alıç, 2014)

Satış/Pazarlama	-Pazar sepet analiziyle, hangi ürünlerin bir arada tüketildiği, tüketimde zamanlama olarak nasıl sıralandığına ilişkin bilgilerin üretilmesi. -Perakende firmalarında tüketicilerin satın alma tutumlarının belirlenmesi. -Tüketiciler arasındaki tutumların tespit edilmesi.
Bankacılık/Finans	Kredi kartı dolandırıcılığının belirlenmesi -Müşteri satın alma faaliyetlerinin tespiti ile müşterilerin sadık kalma derecelerinin belirlenmesi -Kredi kartı kullanıcılarının devamlılığının sağlanması -Kredi kartı kullanım alanlarının belirlenmesi -Değişik finansal göstergeler içerisindeki saklı bağın keşfedilmesi -Daha önce ki piyasa verilerinden yararlanılarak, hisse senedi alım satım niteliklerinin tespiti, hisse senedi fiyat tahmini
Taşımacılık	Depolar ve mağazalar arasındaki dağıtım programlarının tespit edilmesi ve yükleme örüntülerinin belirlenmesi
Tıp/İlaç	-Değişik hastalıklar için başarılı tıbbi teşhis ve tedavi örüntülerinin analiz edilmesi, tehlike boyutu irdelenmesi -İlaç dozajına verilen tepkinin tahmini, ilaç dozaj ve yan etkileri arasındaki ilişkinin tespiti

3.2.3. Veri Madenciliği İle İstatistik Arasındaki İlişki

Elde bulunan veri setinin dağılımını belirleme, aritmetik ortalama ve varyansını hesaplama gibi bazı istatistiksel durumlar veri madenciliği teknikleri olarak düşünülebilir. Bu istatistiksel kavramlar eldeki veri seti için tanımlayıcı ifadelerdir.

İstatistik tarihinde son dönemlerde bu hedef için farklı teknikler öne sürülmüştür. Bu teknikler istatistik biliminde Çok Boyutlu Analiz kapsamına girer. Genel olarak bilginin parametrik durumdan geldiğini düşünür (Yaralıoğlu, 2008).

Çizelge 2. İstatistiksel Analiz ve Veri Madenciliği (Ulusoy, 2013)

İstatistik	Veri Madenciliği (VM)
Genel olarak başlangıç noktaları bir hipotezdir.	Hipotezle başlama gibi bir zorunlulukta yoktur.
Kendi eşitliklerini geliştirerek hipotezleri eşleştirirler.	Veri madenciliği eşitlikleri sağlamak için algoritmaları otomatik olarak sağlamaktır.
Analizler sırasında gereksiz veriyi bulup temizler.	Veri madenciliğinde kirli veri zaten yoktur.
Yalnızca sayısal verilerin kullanıldığı analizlerdir.	Veri madenciliği sadece sayısal veriye dayanmaz. Daha geniş kapsamlı verilerin kullanımına açıktır.
Sadece kendi sonuçlarını açıklarlar ve gereken mercilere iletirler.	Veri madenciliği sonuçlarını açıklamak için mutlaka bir istatistikçiye ihtiyaç duyulmaktadır. Yani veri madenciliği sonuçlarını yorumlayabilmek basit bir eylem değildir.

3.2.4. Veri Madenciliği Neden Kullanılır

Veri madenciliği elde var olan verileri değerlendirerek geleceğe dair tahminler üretebilmek için kullanılır. Veri madenciliğinde asıl amaç göze çarpmayan saklı kalmış veya birey sebebi ile tespit edilemeyen durumların belirlenmesini sağlamak ve geçmiş deneyimlerden yola çıkarak geleceğe yön vermektir.

Veri madenciliği ile büyük veri depolarından oluşan veri tabanı sistemleri arasında saklı kalmış bilgilerin meydana çıkarılması olanaklıdır. Bunları yapabilmek için istatistik, cebir, modelleme teknikler, veri taban teknolojisi ve çeşitli bilgisayar uygulamaları kullanılır. Veri madenciliği çok büyük ölçütteki veri seti ile ilgilendiği için veri tabanları büyük önem arz etmektedir. Kullanılacak verinin saklama koşullarına uygun ve lazım olduğu an ulaşılabilir olması gereklidir. Çağımızda çok

tercih edilen veri depoları günlük uygulanan veri tabanlarının bütünleştirilmiş ve işlem yapmaya daha elverişli genel hatlarını saklı tutmayı amaçlar.

Veri madenciliği tek başına kullanılarak yol alabileceğimiz bir kapsam olmaktan ziyade yürüyeceğimiz yol için atacağımız adımları tespit etmek adına kullanılan yol gösterici bir tekniktir. Veri madenciliğinde uygulanan modelleri tahmin edici ve tanımlayıcı olarak açıklayabiliriz.

3.2.5. Süreç Olarak Veri Madenciliği

Veri işleme, hem veri tutma hem de veri madenciliği için oldukça önem arz etmektedir. Çünkü dünya verileri eksik, gürültülü ve tutarsız olma eğilimindedir. Veri işlemede; veri temizleme, veri birleştirme, veri dönüştürme ve veri indirgeme yer alır (Gülpınar, 2008).

Tanımlayıcı veri özetleme, veri ön işlemeye yönelik analitik temeli sağlar. Veri özetlemeye yönelik temel istatistiki ölçümler içerisinde verilerin ortalama değeri, ağırlıklı ortalama, medyan ve verilerin yayılımının ölçülmesine yönelik olarak aralık, kartiller ve kartiller arası aralık, varyans ve standart sapma yer almaktadır. Histogram, kutu grafiği, dağılım grafikleri, çifte dağılım grafikleri, dağılık grafikler ve dağılık grafik matrisleri gibi grafiksel gösterimler verilerin görsellik açısından ele alınmasını kolaylaştırır (Gülpınar, 2008).

- Veri Temizleme: veri temizleme modelinde, yarım kalan değerleri tamamlama, uç değerler açıklanırken gürültü oluşturan verileri onarma ve veride oluşan dengesizliklerin düzeltilmesi sağlanır. Genel olarak veri temizlemede uyumsuzluk denetimi ve veri dönüştürmeden oluşan iki aşamalı bir süreçtir (Han ve Kamber, 2006).
- Veri Bütünleştirme: Veri bütünleştirme, tutarlı bir veri deposu oluşturmak için değişik kaynaklardan alınan verileri bütün olarak analiz etmek üzere toplarlar (Han ve Kamber, 2006).
- Veri İndirgeme: Veri indirgemedede amaç veri sayısının ya da değişken sayısının indirgenmesidir. Bu indirgeme işleminde en az bilgi kaybı ile indirgenmiş bir veri gösterimi elde etmek için veri küp birleştirme, değişken alt küme seçimi, boyutsallık indirgemesi, sayısal indirgeme ve aralık belirleme gibi veri indirgeme

metotları kullanılabilir. Nümerik verilere yönelik veri aralığı tespit etme ve kavram hiyerarşilerinin otomatik oluşturulması; paketleme, histogram analizi, entropi tabanlı aralık belirleme, küme analizi ve sezgisel bölümlenmeli aralık belirleme gibi teknikler kullanılabilir (Han ve Kamber, 2006).

- **Veri Dönüştürme:** Bazen veri madenciliğini kullanabilmek için veriler üzerinde dönüşümler yapılabilir. Değişkeler arasında varyans ve ortalama birbirlerinden çok farklılık gösterdiği durumlarda büyük ortalama ya da varyansa sahip değişkenin diğer değişkenlere oranla baskınlığı artar ve diğer değişkenlerin önem derecesini düşürür. Çözümlemenin daha iyi olabilmesi içinde değişkenlerin durumu önemlidir. Bu sebepten ötürü değişkenlerin dönüştürülmesi gerekebilmektedir. Verilerin 0-1 arasında sayısal değere dönüştürme işlemi normalleştirme işlemi, verilerin yeni değerlere dönüştürülmesi standartlaştırma işlemidir (Han ve Kamber, 2006).
- **Veri Madenciliğinin Modelleme Aşamaları:** Belirlenen problem durumu için en uygun modelin tespit edilebilmesi için birçok deneme çalışması yapılmaktadır. Dolayısıyla veri kurma işlemi uzun bir süreç gerektirir (Han ve Kamber, 2006).

3.2.6. Veri Madenciliği İle Karar Destek Sistemleri Arasındaki İlişki

1960 tarihinden itibaren gelişen teknoloji ile birlikte karar alma sürecini etkileyen karar sistemleri önemli hale gelmektedir. Bunun sonucu olarak ilk karar sistemleri olan veri işlem sistemleri, yöneylem araştırmacıları ve bilgisayar bilim uzmanlarının önemli çalışmaları sonucu ortaya çıkmıştır. Daha sonra ise bunları yönetim bilgi sistemi, karar destek sistemleri (KDS) gibi sistemler takip etmektedir.

Karar verme aşamasında, kullanıcıya yardım edebilmek hedefiyle amaçlanan bilginin üretilmesi ve bu bilginin kullanıcıya yöneltilmesi için destek ve donanım araçlarının birlikte kullanılması ile oluşturulmuş sistemler Karar Destek Sistemleri (KDS)'dir. KDS' de amaç; verilerin daha iyi analiz edilmesi, daha iyi açıklanabilmesi ve böylece etki alanının artmasıdır. KDS kullanıcıya sonuca varma açısından destek olabilmekte ama kullanıcı yerine karar alamamaktadır. KDS grup içerisinde sıradan ve basit olmayan karara ulaşabilmeyi ve basitten karmaşığa doğru genel analizlerde destek sağlar. Veri, model ve yararlı yazılımları bir arada kullanarak özelleşmiş entegre

yönetim sistemleridir. KDS bir firmada kullanım örneği; yöneticilerin ve çalışanların herhangi bir konu üzerinde sağlıklı bir sonuca ulaşabilmelerini destekleyen ve sistem ile etkileşim halinde yapılan bilgisayar tabanlı bir bilişim sistemidir. KDS genellikle tam yapılandırılmamış sorunların çözümünde kullanılabilir fakat yapılanmış sorunlarda da kullanılmaktadır. KDS veri ve model tabanlıdır. KDS farklı kaynaklardan bilgileri derleyip düzenleyerek, karar vermeyi destekleyen ve bulguları analiz edip sonuçları yorumlayan, kullandığı modellerle kullanıcıya vermesi gereken en doğru karar için yol gösteren bilgisayar temelli bir sistemdir. Karar destek sürecinde dikkat edilmesi gereken hususlar vardır. Eğer bu hususlara dikkat edilmezse yanlış ve yetersiz karar almaya neden olur. Tabii ki bu hususları etkileyen etmenler bulunmaktadır ve bunlar; yetersiz zaman, veri toplayamama, analiz hataları gibi durumlar sıralanabilir. KDS bilgisayar temelli olduğu için zamandan tasarruf sağlayabilmektedir (Gülpınar, 2008). İnsan yeteneğinin yetersiz kaldığı sarmal problemlerin çözümlenebilmesi için oldukça uygun bir sistemdir. Genel olarak KDS için bilgiler veri ambarlarından temin edilir. Bu bilgiler için bilgi keşfi tekniği kullanılır. Bilgi keşfini kullanarak eldeki problem durumu için eski kullanılan modeller taranarak yeni ve özgün model oluşturulur. Yönetim birimlerindeki gelişmeler ile birlikte bilgi işlem modelleri de gelişmektedir. Sonuç olarak kurumlar kendilerine ait bilgi depoları oluşturmaktadır. Kurum kapsamındaki tüm birimlerin bilgisayar kullanımında yer alması, bilgi oluşturulmasında sorumlulukların belirlenmesi, bilgilerin depolanmasında ve sunulmasında sorumluluk bilincine sahip olunması, etkili bir veri tabanının bulunması ve veri ambarı alt yapısının sağlanması gibi düzenlemeler yapıldıktan sonra KDS uygulamaları kullanılabilir. Kuruluşların etkili ve sağlıklı kararlara ulaşabilmesi için ve bu kararları daha iyi muhafaza edebilmesi için son dönemlerde üst yönetim bilişim sistemleri ve üst yönetim destek sistemleri terimleri geliştirilmektedir. Veri madenciliği araçları KDS süreci için kullanılan araçlardır. KDS bazen kullanıcıya yetersiz kalabilmektedir. Bunun nedeni ise KDS sadece sahip olduğu bilgi ile sınırlıdır. Kısacası bu bilgisayar tabanlı KDS kullanıcı adına karar vermez sadece kullanıcıya kararında yol gösterir (Gülpınar, 2008). Çağımızda yaşanan hızlı gelişmeler karar verme sürecini zorlaştırmakta ve karar aşamasında belirsizlik oranı da yükselmektedir. Bunun sonucunda karar verme sürecinde yoğun emek ve yeterli süre gerekmektedir. Bazı durumlarda doğru karara ulaşabilmek için karar sürecinde sezgisel kararlar alınabilir ama bu kararlar her daim kullanışlı olmamaktadır. KDS uygulama olarak kullanışlı ve esnek olmalıdır. Sistemi oluşturan temel unsurlar ele alınan

probleme göre yeniden şekillenebilmelidir. Kullanıcı sistem gidişatını yönetebilir ve gereken durumlarda müdahale edebilir. KDS ile basit ve sıradan karar mekanizmasından uzaklaşılabilir. Bunun için veri, model ve etkileşimli zararsız yazılımlar birlikte kullanılır. Sonuç olarak KDS entegre bir sistemdir diyebiliriz. İlk üretilen karar destek yazılımları özel durumlar için düşünülmüş ve bu durumun sonucu olarak genel amaç için kullanılamama ve devamlı yenilik yapılma ihtiyacı gibi olumsuz durumlar meydana geldi. Çağımızda yetersiz kaynak ve sınırsız gereksinimler olduğu için üretilen projelerde alınacak kararların önemi artmakta ve dolayısıyla KDS nin önemi de paralel bir seyir izlemektedir. KDS de asıl amaç alınan kararın yeterliliğinden çok etkinlik alanını genişletmektir.

3.2.7. Karar Ağacı

Karar ağaçları, sınıflandırma amacıyla veri madenciliğinde en çok kullanılan tahmin edici bir tekniktir. Genellikle sınıflandırma, kümeleme, tahmin modellerinde ve sorunla ilgili araştırma alanını alt gruplara ayırmak için kullanılmaktadır (Quinlan, 1986). Örneğin, bir banka önceki dönemlerde vermiş olduğu kredilere ilişkin gerekli tüm verilere sahiptir. Bu verilerdeki bağımsız değişkenler kredi alan müşterilerin özellikleri, bağımsız değişken değeri ise alınan bu kredilerin geri ödenip ödenmediğidir. Eldeki bu verilere uygun olarak kurulan model, daha sonraki kredi taleplerinde müşterilerin özelliklerine göre verilecek olan kredilerin geri ödenip ödenemeyeceğinin tahmininde kullanılmaktadır. Bu teknikte sınıflandırma için bir ağaç oluşturulur, daha sonra, her bir kayıt bu ağaca uygulanır ve çıkan sonuca göre de bu kayıt sınıflandırılmaktadır. Birçok çalışmada eldeki problemle ilgili en iyi kararı verebilmek için bazı işlemleri yerine getirirken bir takım tekniklere ihtiyaç duyulmaktadır. Bu teknikler arasında kullanım açısından karar vericiye problemi anlamada kolaylık sağlayan, güvenilirliği yüksek olan ve sınıflama modelleri içinde son zamanlarda en yaygın olarak kullanılan karar ağacı tekniğidir (Özekeş ve Çamurcu, 2002).

Karar ağaçları Bierman ve Friedman tarafından 1973 yılında önerilmiş olup bağımlı değişken üzerindeki farklılıkların maksimize edilmesi amacıyla değişkenleri parçalayarak bir ağaç elde edilmesidir. Veri madenciliğinde ise kurulumunun ucuz olması, yorumlanmasının basit olması, veri tabanı sistemleri ile basitçe entegre edilmesi ve güvenlik probleminin az olması sebeplerinden dolayı karar ağaçları tekniği sınıflama modelleri içerisinde en fazla tercih edilen tekniktir (Safavian ve Landgrebe, 1991).

Karar ağaçları isminden de tahmin edildiği gibi eldeki verilerden sınıflandırıcılar üretmek için ve en fazla tercih edilen ağaç görünümüne sahip bir mantık tekniğidir. Karar ağacında tanımlanan problemin yanıtları gruplara dağılmaktadır. Birden fazla hareket durumunda en sağlıklı karara varabilmek amacıyla karar ağaçları oluşturulabilir. Karar ağacında; karar alma eyleminde bulunmak için kararı alan, hedef, karar ölçütü, stratejiler (S), olaylar (N) ve sonuç olacak şekilde ağaç oluşum elemanları ile işe başlanır.

Karar Alan: Ele alınan problem durumu için tercihte bulunan birey veya grubu belirtir.

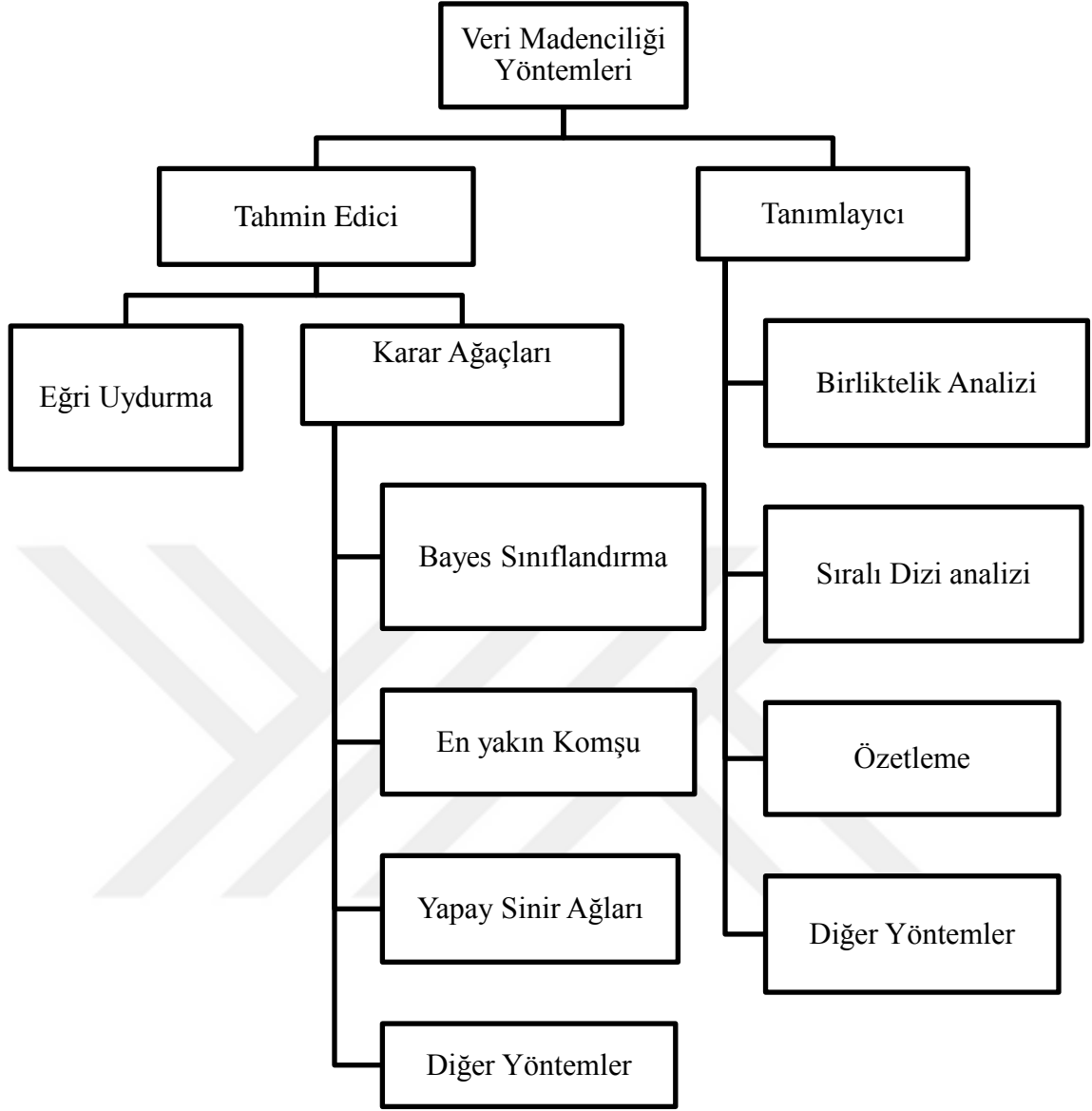
Hedef: Eldeki problem durumunun sonuçlandırma maksadıdır.

Karar Ölçütü: Karar verenin karar aşamasında kullanacağı değerler sistemidir.

Stratejiler (S): Karar vericinin tercihte bulunabileceği farklı alternatif eylemlerdir. Stratejiler kontrol altında tutulabilen ve karar vericinin yönetiminde bulunan değişkenler grubudur.

Olaylar (N): Karar vericinin elinde olmadan gerçekleşen durumlardır. Karar vericinin tercihte bulunurken istemsiz olarak etki alanında kaldığı durumlardır.

Sonuç: Her bir strateji veya olaydan meydana gelen değerleri belirtir. İçerisinde strateji, olay, sonuç değerlerinin bulunduğu karar matris tablosu oluşturulur. En mantıklı karar ölçütünü belirlemek için olasılık mantığının grafik gösterimi karar ağaçları kullanılarak belirlenir.



Şekil 1. Veri Madenciliği Yöntemleri

Karar ağaçları kullanılan değişken kesikli ise sınıflandırma ağacı, sürekli ise tahmin ağacı olarak isimlendirilmektedir. Literatürde karar ağacı tanımlamalarından bazıları:

- Bir karar ağacı modeli, veri keşfinde kullanılan öğrenme yöntemlerini en yaygın kullanan tekniklerden biridir. “Eğer-ise-aksi halde” yapısında temsil edilebilmektedir. Veri dağılımından önceki herhangi bir bilgiye gerek duymaz. Büyük veriler üzerinde iyi çalışmaktadır.
- Bir karar ağacı, basit karar kurallarını, büyük veri setlerine uygulayarak ard arda küçük veri setlerini oluşturmada kullanılabilen bir yapıdır.

- Bir karar ağacı algoritması, hedef alan veya değişken açısından her kaydı başarıyla sınıflandıran bir model geliştirmek amacıyla veri kümesini bölerek çalışır. Karar ağaçları, yineleyici bölme işlemi olarak adlandırılan işlemi gerçekleştirir. Verinin açıklanması, sınıflandırılması, regresyonu ve tahmini için güçlü bir araç sağlamaktadır.

- Karar ağaçları, sınıflandırma amacıyla veri madenciliğinde kullanılan ve heterojen veri setinde, veriyi açıklayan özelliklerinin bir kümesi verildiğinde, bu veri kümesini daha homojen küçük parçalara ayırmada kullanılan bir yöntemdir.

Karar ağacı modeli, araştırmacılara karar aşamasında hangi faktörleri göz önüne alması gerektiğini ve her bir faktörün kararın farklı çıktıları ile geçmişle nasıl bir ilişkisi olduğunun belirlenmesine yardımcı olmaktadır. Karar ağaçlarında oluşan model çok sade ve net olmaktadır.

Karar ağaçları, sınıflandırıcıları temsil etmek amacıyla kullanılan en yaygın yaklaşımlardan biri olarak kabul edilmektedir. Karar ağacı köklü bir ağaç oluşturan düğümlerden oluşmaktadır, yani kendisine doğru gelen ve kök olarak adlandırılan bir düğüme sahip yönlü bir araçtır. Diğer tüm düğümler tam olarak bir tane gelen kenara sahiptirler. Kenarlardaki düğümler iç ya da test düğümü, diğer tüm düğümler yaprak (terminal ya da karar düğümü olarak da bilinir) olarak adlandırılmaktadır. Bir karar ağacında her bir iç düğüm, girdi özniteliklerin değerlerinin belli bir ayrık fonksiyonuna göre, örnek uzayını iki ya da daha fazla alt uzaylara bölmektedir. Basit ve en sık görülen durumda, her bir test tek bir özneliği göz önünde bulundurmakta ve örnek uzayı özneliğin değerine göre bölümlendirilmektedir. Nümerik öznelik durumlarında, durum bir aralık ifade etmektedir.

Her yaprak en uygun hedef değeri temsil eden bir sınıfa atanmaktadır. Alternatif olarak, yaprak hedef özneliğin olasılığı olacak şekilde belli bir değere sahip olasılık vektörüne sahip olabilmektedir. Örnekler, ağacın kökünden yaprağa kadar bu örnekleri dolaşarak, yol boyunca testlerin sonucuna göre sınıflandırılabilir.

Sınıflama araçlarından en önemlilerinden olan karar ağaçlarında öğrenme algoritması oldukça basittir ve veriden sınıflandırıcıları oluşturmada kullanılan etkili yöntemlerden biridir. Uygulamalı istatistik literatüründe birçok karar ağacı tümevarım algoritması bulunmaktadır. Bu algoritmalar, bir seri girdi çıktı kümesinden karar ağacı oluşturan denetlenmiş öğrenme yöntemleridir. Karar ağaçları yalnızca kararları göstermekle kalmaz, aynı zamanda kararların açıklanmasını da sağlamaktadır. Bunu da

veri yığınlarını sıralı bir şekilde bölüp, daha düzenli ifade ederek yapmaktadır. Karar ağaçları değişkenlerin test edildiği yerlerde düğümler oluşturur ve bu düğümlerden açılan dallar, düğümdeki testin bütün olası sonuçlarını göstermektedir.

Karar ağaçları temelde iki aşamadan oluşmaktadır. Bunlardan ilki ağacın kurulması, ikincisi ise verilerin tek tek ağaca uygulanıp, sınıflandırılmasıdır. İstatistiksel yöntemlerin çoğunda verilerin oluşturduğu fonksiyonların araştırmacılar tarafından kolayca yorumlanabilir bir kural olması oldukça güçtür. Ancak karar ağaçlarında veriler oluşturulduktan sonra, ağaç kökten yaprağa doğru sınıflandırılarak kurallar oluşturulduğundan dolayı araştırmacılar tarafından kolayca yorumlanabilmektedir.

3.2.8. Karar Ağacı Kullanım Alanları

Karar ağaçlarının veri tabanı sistemleriyle kolayca uyum sağlayabilmeleri ve güvenilir olmasından dolayı sınıflama modelleri içerisinde yaygın olarak kullanılmaktadır. Genelde, belirli bir sınıfa üye olacak unsuların belirlenmesi, olayların tahmin edilmesi, bazı durumların yüksek, orta ve düşük risk grupları gibi çeşitli kategorilere ayrılması, gelecekte olabilecek olayların tahmin edilebilmesi için bazı kuralların oluşturulması, parametrik modellerin oluşturulması için gerekli olan değişken ve veri kümesinden faydalı olacakların seçilmesi, kategorilerin birleştirilerek sürekli değişkenlerin kesikli değişkenlere dönüştürülmesi, belirli alt gruplara ait olan ilişkilerin tanımlanması gibi birçok yerde kullanılmaktadır.

Karar ağacı kurulumunun ucuz olması, yorumlanmasının basit olması, veri taban sistemleri ile basit bir şekilde entegre edilebilmesi ve güvenlik problemi olmamasından dolayı veri madenciliğinde sınıflama amaçlı uygulamalarda oldukça sık kullanılan tekniklerden biridir. Karar ağaçları tahmin edici ve tanımlayıcı özelliklere sahip veri madenciliği projelerinde kullanılmaktadır. Bir sınıfın muhtemelen üyesi olabilecek elemanların tespit edilmesi, farklı durumların risk kategorilerine göre dağılması, yalnız belirli alt gruplara ait olan ilişkilerin tespit edilmesi, gelecek adına yaşanacak durumlar için tahmin edici kurallar sağlama ve kategorilerin bütünleştirilmesi gibi alanlarda karar ağaçları yöntemi kullanılmaktadır.

Karar ağacı analizleri ile yapılan çalışmalar;

- Belirli bir sınıfın muhtemelen elemanları olacak üyelerin tespit edilmesi
- Çeşitli olayların risk grup kategorilerine dağılması

- Çok miktardaki deęişkenlerden en önemli deęişkenin tespit edilip parametrik modellerin kurulmasında kullanılmaktadır.
- Yalnız belirli bir alt gruba ait ilişkilerin tespit edilmesi
- Kesiksiz deęişkenin süreksiz deęişkene dönüştürülmesi ve kategorilerin bütünleştirilmesinde kullanılır.
- Otomatik Bilgi İşlem alanı için karar ağacı kavramı çok yeni olmasına rağmen çok fazla öneme sahiptir. Kullanımda karar ağacı el ile veya bilgisayar ile değerlendirme yapar.

Karar ağacı verilerin çizelgeleridir, deęişkenler ve parametreler arası ilişki grafiğidir ve sabit ifadeler içerir. Karar ağaçları problem yapı grafiği olarak kullanılan ögeler içeren bir yapı çizelgesidir. Karar ağaçları karmaşık olayların seyirlerinin grafiğe dönüştürülmesinde kullanılabilir. Kısacası hem yardımcı hem de araç olarak kullanım alanları mevcuttur. Karar ağaçlarının yapısı verilerin organizasyonu için büyük bir öneme sahiptir. Yönlendirilmiş ağaç örnekleri genel olarak bağımsız bağlantının kurulması için kullanılabilir. Karar ağacındaki düğümler içlerinde veriler barındırır. Karar ağaçları görsel bir grafik örneğidir. İkili bir karar ağacında bir düğümden en fazla iki dal meydana gelmektedir. Devam eden bir örnek olarak soyağacı verilebilir. Karar ağacında karar düğümleri bir alandaki testi, dallar test içinde bulunan değerleri ve yapraklar ise sınıfı belirtir. Karar ağacı akış diyagramı şeklindeki bir ağaç modelidir. Ağaç yapısındaki kök kısmı en büyük düğümü belirtir. Karar ağaçları veri kümesindeki kuralların oluşturulmasıyla gelişen modelleme belirli deęişkenlerin aldığı değerlere dayanarak etkilenen deęişkene bağlı olarak gözlemlerin ne kadar iyi farklılaştırılabileceğine göre deęişkenlerin aldığı değerlere bağlı olan dallar ile kurallar oluşturulur. Bunun dışında karmaşık kuralların oluşturduğu karar ağaçları veri setini açıklayıcı nitelikte olmasına rağmen anlaşılması güçtür. Budama seviyesi tespit edilerek karmaşık kuralların üretilmesi mümkündür. Tek deęişkenli karar ağaçlarında ise tek deęişkenin durumunu her kara düğümünde inceleyerek eksenleri dik böler. Her dalda giriş uzayının tesadüfi bölmeler yapması ile doğrusal karar ağaçları oluşur. Çok katmanlı sinir ağlarının tesadüfen giriş uzayını bölmesi ile doğrusal olmayan karar ağaçları oluşur. Karar ağacında veri setinin sıralı bir şekilde bölünmesindeki amaç etkilenen deęişken üzerindeki farklılığın maksimize edilmesidir. Veri yapısına göre girdi veri deęişkenleri belirlenerek kullanılan karar ağacı algoritması ile çıktı veri

değişkenine yol açılır. Karar ağacında verileri belirli değişken değerlerine göre sınıflandırmaya yarayan algoritmalar kullanılır.

3.2.9. Karar Ağacı Kullanımının Avantajları Ve Dezavantajları

Karar ağaçlarının, kendini açıklayıcı özellikte olması, hem kategorik hem de sayısal verilerle işlem yapılabilir. Eksik değerleri içeren veri kümelerini analiz edebilir ve kolay yorumlanabilir olması gibi bir çok avantajının yanı sıra, birçok karar ağacının, bağımlı değişkenleri sadece kesikli değerler gerektirmesi, yapılan sınıflandırma sonucunda bölüm düğümlerinde daha az bilgi kalacağı için karar vermek, karar ağacı yaklaşımı için ciddi bir sorun olabilmesi gibi birçok dezavantajı da bulunmaktadır (Quinlan 1993).

Karar Ağacı Kullanım Avantajları

- Karar ağacının sonuç çıktıları hakkında yorum yapmak oldukça basittir.
- Karar ağaçları sonucunda mantıklı kurallar elde etmek basittir.
- Karar ağaçlarında değişkenlerin niteliği çok önemli değildir yani ayırık ve kesiksiz değerlerin kullanımı içinde oldukça elverişlidir.
- Karar ağaçlarının büyüklüğü fazla olsa da takibi kolaydır.
- Karar ağaçları gereken dönüşümleri sağlama açısından elverişlidir.
- Karar ağaçlarının oluşturduğu yaprak miktarı ideal sayıda ise üstün bir çaba göstermeksizin anlaşılabilir niteliktedir.
- Karar ağaçları eksik veri kümelerini tamamlama açısından başarılıdır.
- Sezgisel düşünce yöntemi ile karar ağacını anlamak mümkün olabilmektedir.
- Çok değişkenli tekniklerde istatistik varsayımların sağlanması şart olsa da karar ağaçlarında böyle bir durum şart değildir.
- Karar ağacı sonucunda elde edilen bütün kurallar uygulanmalıdır şeklinde bir zorunluluk arz etmemektedir.
- Karar ağaçları algoritmalarında bağımlı ve bağımsız değişkenler arasındaki bağların durumunun görselleştirilebilir olması.

Karar Ağacı Kullanımının Dezavantajları

- Karar ağaçları kesiksiz nitelik değerlerini tahmin etmede fazla başarı sağlayamamaktadır.

- Karar ağaçları öğrenme kümesi örneklerinin yetersiz ve sınıf miktarının aşırı olduğu durumlarda model oluşturmada yetersiz kalmaktadır.
- Karar ağacı uygulamalarında zaman ve mekân karmaşıklığı nedeniyle sorunlar yaşanabilmektedir.
- Karar ağacında budama uygulamalarında sorunlar yaşanabilmektedir.
- Karar ağacı uygulamalarında verinin dinamik olması durumunda sorunlar çıkabilir. Bundan dolayı değişkenin etki kümeleri kolaylık sağlayacak şekilde kategorilere ayrılmalıdır.
- Karar ağacında değişken miktarının artması karar almayı olumsuz etkileyebilmektedir.
- Karar ağacında doğru kurala ulaşabilmek için veri setini dengeli ayarlamak gerekmektedir. Gereğinden az olan veri miktarı durumunda hatalı sonuçlara varma ihtimali yükselir.
- Karar ağacında dalların sayısı arttıkça ağacı denetim altında tutmak zorlaşır.

3.2.10. Karar Ağacının Uygulanma Koşulları

Karar ağacında yöntem olarak tümevarım öğrenme yaklaşımını kullanmak için birkaç uyulması gereken husus vardır. Bunlar;

Değişken Değer Tanımı: Veri formatının analize uygun olacak biçimde düz olması gerekmektedir. Değişkenlerin ve oranların sabit toplanması için bir örnek hakkında bütün bilgilerin açıklanabilir nitelikte olması gerekir. Bu nitelikler farklı ve rakamsal değere sahip olabilir. Bir örneği ifade ederken kullanılacak olan özellikler önemlidir. Bu özellikler durumdan duruma farklılık arz etmemelidir. Gereken durumlarda kesikli değişkenler kesiksiz hale dönüştürülebilir olmalı veya kesiksiz değişkenler kesikli değişkene dönüştürülebilir olmalıdır. Bu işlemlerin algoritma kullanılarak da sağlanması gerekmektedir. Bu kısıtlama sonucu nitelik olarak farklılaşan bir yapıya sahip olan örneklerin tanım kümelerini kapsam dışı yapar.

Önceden Tanımlanmış Sınıflar: Eldeki örneklerin dâhil olması gereken kategoriler daha öncesinden tespit edilmelidir. Denetimli öğrenme olarak da makine öğrenme kaynaklarında yer almaktadır.

Ayrık Sınıflar: Örnek özel bir sınıfa dâhil olsa da olmasa da sınıflar kesinlikle belirlenmelidir. Beklenen durum sınıflardaki örnek miktarının çok fazla olmasıdır.

Yeterli Veri: Karar ağacı formatında sunulan tümevarımsal çıkarım, veride içerisinde örnekler tanımlama yoluyla sağlanır. Yeterli miktardaki örüntü tesadüflerden ayrışabiliyorsa yaklaşım kayda değerdir. Bu yaklaşımda yeterli örnek alma sebebi istatistik testler uygulandığı için bu testlerin etki alanını sağlamasına olanak tanımaktır. Sınıf özellikleri ve sınıflama model tipi gibi etmenler sonucu veri miktarı belirlenir. Bu etken miktarı ile veri miktarı arasında doğrusal bir ilişki bulunmaktadır.

Mantıksal Sınıflama Modelleri: Sadece karar ağacı veya karar kurallarının oluşturduğu sınıflandırıcılar tarafından kullanılan yöntemlerdir. Tanım olarak sınıf kavramını mantıksal bir ifade kısıtlayabilir. Bu mantıksal ifadenin temel unsurları belirli değişkenlerin değerleri hakkındaki beyan ifadelerdir. Güvenilir bir sınıf tanımlaması için bazı uygulamaların ağırlıklı özellikler veya bunların aritmetik kombinasyonları olmalıdır. Böyle durumlarda mantık modellerinin etki alanı sınırlıdır.

3.2.11. Karar Ağacı Oluşturma Süreci

1. Aşama: İlk aşamada bağımlı değişken ve bağımsız değişkenler tespit edilir.

2.Aşama: Bu aşamada kök kısmını etkisi altına alan bütün bağımsız değişkenler incelenir. Karar ağacında sürekli bölünmelerin olduğu süreç başlar. Bu bölünmeler arasında en mantıklı sınıflandırma bulunana dek devam etmektedir. Sınıflandırma kriteri bazen kullanıcı tarafından da belirlenebilir.

3. Aşama: Yukarıdaki işlemler bütün değişkenler için uygulandıktan sonra tahmin ihtimali en yüksek olan değişken tercih edilerek yaprak düğümleri oluşturulmaya başlanır.

Karar ağaçları kullanıcıya göre çok çeşitli uygulama tipleri mevcuttur. Basit, küçük, serbest, bağımlı, el ile hesaplanacak kadar küçük, bilgisayar ile hesaplanacak kadar büyük olabilmektedir. Karar ağacında belirlenen problemin cevabı gruplara dağılır. Soruya verilecek bir ölçüt tespit edildikten sonra sınırlar arasındaki risk maksimize edilecek şekilde yanıtlar dağılmaktadır. En iyi parçalanmayı bulmak için her soruda bu aşamalar yenilenir. Bir soru için grup oluşturulup ve risk ihtimali düzenlendikten sonra oluşan yeni gruplar için uygulama tekrarlanır. Bu uygulamalar istatistik anlamlılık sağlanana dek devam ettirilir ve bulunamaz ise son verilir. Dağıtma işlemi tamamlandıktan sonra ise gruplar içerisinde yer alan denetimlerin oranına göre değerlendirmesi yapılır.

3.2.12. Karar Ağacı Algoritmaları

Karar ağacı çok büyük veri setlerinden bir takım istatistik anlamlılık sağlayacak şekilde kurallar oluşturmak için karar kuralları uygulanarak daha küçük veri setleri haline dönüştürmek için kullanılan yapıdır (Albayrak ve Yılmaz, 2009). Bunun için veri madenciliği alanında karar ağacı uygulayabilmek için bir takım algoritmalar sağlanmıştır. Öğrenme algoritması karar ağacı uygulaması sağlamak için oldukça elverişlidir. Meydana gelen asıl bilginin gösterimi zahmetsizce anlaşılabilir. Uygulamalı istatistikte yetmişli yıllarda bilim insanları karar ağacını uygulayabilmek için farklı teknikler sağlamışlardır. Karar ağacında anlamlılık sağlamak için kullanılan algoritmalar değersiz değişkenleri otomatik bir şekilde saf dışı bırakarak yeni öğrenme döneminde değişken seçimini sağlamaktadır. Karar ağaçları algoritmaları bir bağımlı değişken ve bu bağımlı değişkeni yorumlamaya yardımcı olacak bağımsız değişkenler ile uygulama alanına başlar. Karar ağaçlarında sınıflandırma işlemini gerçekleştirmek için başlıca algoritmalar; CHAID, C&RT, ID3, C4.5, MARS, QUEST, C5.0, SLIQ, SPRINT algoritmalarıdır.

Algoritma, anlam ifade etmeyen değişkenleri sistemsal olarak çıkararak yeni öğrenme döneminde değişken tercihini kendisi belirler. (Gülpınar, 2008). Karar ağacı modellerinin başlangıcı AID (Automatic Interaction Detector) yöntemidir. Bu süreç içerisinde çeşitli algoritmalar geliştirilmiştir. AID algoritması, karar ağacı tabanlı ilk algoritma ve yazılım olarak 1970'li yılların başında Morgan ve Sonquist isimli bilimciler tarafından tercih edilmiştir.

Karar ağaçları algoritmaları, bir bağımlı değişken ve bu değişkeni açıklamaya yönelik olan bağımsız değişkenler olarak iki grup değişken ile uygulamaya başlanır (Albayrak ve Yılmaz, 2009). Karar ağacı algoritmaları bir sınıflandırma modeline gereksinim duyan tahminleyici olarak kullanılır. Problemlerin en sağlıklı bir şekilde çözümlenebilmesi için değişkenler farklı alanlara dağılacak şekilde tasarlanmıştır. Karar ağaçlarında bazen modellen çalışma prensibinden ziyade bulunan sonucun ya da sınıflandırmanın tutarlılığı önem arz etmektedir. Bazen ise yapılan uygulama sonucunda bulunan kararın asıl sebebini anlamlandırmak önemlidir. Kısacası sağlıklı uygulama için kullanılacak olan verilerin iyi tanımlanması çok önemlidir. Bu uygulama alanlarında birçok karar ağaçları algoritmaları kullanılabilir.

3.2.12.1. Hunt Algoritması

Hunt algoritmasında karar ağacı, eğitim kayıtlarının sıra ile daha homojen alt gruplara ayırarak tekrarlı bir biçimde büyütürken yapılandırılmaktadır. Hunt algoritması, eğitim verilerindeki her değişken değerinin kombinasyonu var olması ve her bir kombinasyonun yalnız bir sınıf etiketine sahip olması durumunda çalışmaktadır. Fakat bu durum birçok uygulamada başarılı olamamaktadır.

Hunt algoritmasının en büyük eksikliği ağaca başlanacak olan kök düğümünün tesadüfi tercih yapılarak başlanmasıdır. Bu tesadüfi seçim olası bütün ağaçları ortaya koymak için büyük bir çaba ve zaman kaybına neden olmaktadır. Bu olumsuz durumlardan ötürü Hunt algoritması geliştirilerek ID3 algoritması oluşturulmuştur.

3.2.12.2. ID3 Algoritması

ID3 algoritması karar ağaçlarının ilk geliştirilen algoritmasıdır. (Gülpınar, 2008: 70). Karar ağaçlarının ilk algoritması olması dolayısıyla çok basit bir algoritmadır. Bu algoritma bilgi kazancını bölme kriteri olarak kullanır. J.Ross Quinlan, 1970'li yılların sonunda Hunt'ın 'böl ve elde et' (Divide and Conquer) algoritmasını geliştirerek bu algoritmayı oluşturmuştur. Hunt algoritmasında yapılan en önemli değişiklik rastgele yapılan değişken seçimini değiştirmek olmuştur. Hunt algoritmasının en büyük yetersizliği değişken seçiminin gelişigüzel yapılmasıdır. Quinlan değişken tercihi aşamasında entropi yöntemini kullanmıştır.

ID3, önceleri satranç oyununda kullanılan stratejileri öğrenmek için kullanılmış ve daha sonra ise hem akademik, hem de sanayi alanında birçok probleme çözüm üretmek için kullanılmıştır. Zamanla bu algoritma önemli gelişimler göstermiştir (Gülpınar, 2008).

Karar ağacı oluşturulurken her bir niteliğin bilgi kazancı hesaplanır ve en fazla bilgi kazancı sağlayan nitelik kök düğüm olarak belirlenir.

ID3 algoritmasıyla oluşturulan karar ağacında işleme öncelikle ağacın kök düğümündeki düzenleme ile başlanır. Bu düzenleme için bir değişken seçilir. Daha sonra her bir değişken değeri için bir dal oluşturulur. Oluşturulan bu dallanmalar sayesinde seçilen örneğe yeni özellikler kazandırılmış olur ve bu örnek alt kümeler de yeni oluşturulan alt düğüme yerleştirilir. Bir düğüme bulunan bütün örnekler tek bir sınıfa ait olana dek her bir alt düğüme algoritma tekrar ederek yazılır. Karar ağacı

yaprağındaki her yol, bir sınıflandırma kuralı oluşmasını sağlar. Bu şekilde yukarıdan aşağıya dek oluşturulan ağaç yapısında önem bir husus, düğümdeki niteliğin seçim aşamasıdır. ID3 ve C4.5 algoritmalarında değişken tercihi bir düğümdeki örneklere uygulanan entropi ölçüt bilgisini minimuma indirgeme temeline dayanmaktadır.

Karar ağacında Entropi, veri kümesindeki belirsiz durum, beklenmedik ya da tesadüfi durum miktarını ölçmede kullanılır. Tabi ki bir kümedeki verilerin hepsi belirli bir sınıfa ait iken belirsiz bir durum söz konusu dâhilinde olamaz. Böyle bir durumda entropi sıfır olacaktır. Karar ağacında sınıflandırmanın amacı, eldeki veri setini, nihai alt kümenin bütün elemanları aynı sınıfa ait olacak şekilde tekrarlı bir şekilde bölümlenektir (Gülpınar, 2008).

Veri madenciliği karar ağacı oluşturma sisteminde en sık kullanılan algoritmalarından olan ID3 algoritmasında temel prensip, nesnelere niteliklerinin değerlerini test ederek sınıflandırma yapmasıdır.

ID3 algoritması eksik olmayan veri ambarlarını dikkate alarak analiz yapar ve veri tabanında bazı bilgilerin yer almaması sıkça karşılaşılan durumlardır. Çünkü nitelik belirli bir duruma bağlı değildir (Gülpınar, 2008).

3.2.12.2.1. Karar Ağacında Entropi

Entropi veri kümesindeki belirsizliği ölçmek için kullanılmaktadır. Entropi, belirsizliğin, rastgeleliğin ve beklenmeyen bir durumun meydana gelme olasılığıdır. Entropi 0-1 arasında değere sahiptir. Tüm olasılıklar eşitlendiğinde entropi en yüksek değerine ulaşacaktır. Entropi sınıflandırmada, örneklerin hepsi aynı sınıfa aitse, Entropi= 0 (homojen), örnekler sınıflar arasında eşit bir şekilde dağılmış ise Entropi= 1 (heterojen), şayet örnekler sınıflar arasında rastgele dağılmış ise, $0 < \text{entropi} < 1$ değerlerine sahip olmaktadır. Yani bir kümedeki verilerin tamamı bir sınıfa ait ise belirsizlik yoktur ve bu durumda entropi sıfır değerini alır.

Karar ağaçlarında sınıflandırmada maksat, eldeki mevcut veri kümesini, alt kümenin bütün elemanları aynı sınıfa ait olacak biçimde tekrar tekrar bölümlenmesini sağlamaktır.

Karar ağacı yöntemi kullanılarak, öğrenme ve sınıflandırma işlemi yapılarak verilerin kategorize edilmesi sağlanır. Öğrenme basamağında önceden bilinen bir eğitim verisi, model oluşturmak için sınıflama algoritması tarafından analiz edilmektedir.

Öğrenilen model karar ağacı ya da sınıflama kuralları olarak ifade edilmektedir. Sınıflamada test verisi, karar ağacının veya sınıflandırma kurallarının doğruluğunu belirlemede kullanılmaktadır. Şayet doğruluk kabul edilebilir ise kurallar, yeni verilerin sınıflandırılmasında kullanılmaktadır. Eğitim verisinde hangi alanların hangi sırada kullanılarak ağacın oluşturulacağına yani ağaca başlanacak kök düğümün belirlenmesinde entropi yöntemi kullanılmaktadır. Bir alanın entropi ölçüsü ne kadar fazla ise, o alanın kullanılmasıyla ortaya çıkan sonuç o kadar belirsizdir. Bu nedenle karar ağacı kökünde entropi ölçüsü en az olan alanlar kullanılmaktadır.

Çizelge 3. ID3 Algoritması Örneği Başarı Durum Verisi

OKUL	AİLE EĞİTİM DURUMU	KARDEŞ SAYISI	ÖZEL ODA DURUMU	BAŞARI
Özel okul	Çok iyi	Üç	Var	Orta
Devlet okulu	Orta	Beş	Yok	Yüksek
Devlet okulu	Kötü	Beş	Yok	Kötü
Özel okul	İyi	Sıfır	Var	Çok yüksek
Köy okulu	Kötü	Sekiz	Yok	Orta
Devlet okulu	Çok iyi	İki	Var	Yüksek
Köy okulu	Orta	Sekiz	Var	Çok yüksek
Özel okul	İyi	Dört	Yok	Kötü
Güneşli	Soğuk	Normal	Yok	Evet

Bu ID3 örneğinde hava durumu, sıcaklık, nem oranı, rüzgâr bağımsız değişkenleri ile oyun oyna bağımlı değişkeni kullanılarak karar ağacı oluşturulmaya çalışılmıştır.

3.2.12.3. C4.5 Algoritması

1993 yılında Quinlan isimli araştırmacı tarafından ortaya çıkarılmıştır. Araştırmacı ‘ Programs For Machine Learning’ isimli kitaplarda C4.5 algoritmasını ele almıştır (Emel ve Çağatan, 2005).

ID3 algoritmasının bir üst versiyonu olarak C4.5 algoritması üretilmiştir. C4.5 algoritması üretiminden sonra çok fazla tercih edilen bir karar algoritması olmuştur. ID3 algoritması ‘ bilgi kazancı’ ölçütünü kullanırken, C4.5 ise ‘ kazanç oranı’ ölçütünü kullanır.

Kazanç oranı: bir kategorik özelliğin olası durum çeşitliliği ne kadar fazla olursa o özelliğin bilgi kazancı gereksiz bir şekilde fazla olur. Ve bu durum ağacın

güvenilirliğine zarar verir. Bu gibi özellikler işe yaramadıkları gibi bilgi kazancı fazla olan özelliklerinde önüne geçip veride saklı kalmış kuralların açığa çıkarılmasını da engeller.

Quinlan, değer çeşitliliği yüksek olan durumların bilgi kazancını indirgeyerek algoritmanın gereksiz bazı kurallar oluşturmasını engellemek için C4.5 algoritmasını geliştirmiştir.

C4.5 algoritmasının ID3 algoritmasından farkları;

- Eksik veri: karar ağacı yapıldığında eksik olan veri kolaylıkla ihmal edilebilmektedir. Bu şekilde kazanç oranı sadece geçerli parametre için bir değere sahip diğer durumlara bakılarak hesaplanabilmektedir. Eksik bir parametre değeri olan kaydı sınıflandırabilmek için söz konusu durumun değeri, diğer kayıtların sahip olduğu parametre değerine bakılarak tahminde bulunulabilmektedir.

- Sürekli veri: sayısal değere sahip değişken içerisinde uygun eşik değeri tespit edildikten sonra ikili veya daha fazla bölünme ile veri seti bölünebilir.

- Budama: C4.5 algoritmasında iki türlü budama mevcuttur. Bunlar alt ağaç değiştirmeli ve alt ağaç yükseltmedir. Alt ağaç değiştirmeli; uygulanan bu teknikte değiştirme sonucunda, başlangıçtaki ağacınkine yakın hata oranı elde edilebiliyorsa alt ağaç, yaprak düğüm ile değiştirilebilir. Bu teknik ağacın altındaki kökünden yukarıdaki köküne dek uygulanabilir. Alt ağaç yükseltme; bu teknik ile budamada alt ağaç, en fazla kullanılan kendi alt ağacıyla değiştirilir. Bu şekilde alt ağaç, mevcut konumdan ağacın daha üst noktasındaki düğüme çıkarılır. Yine de yapılan bu değişiklik için hata oranındaki artışı belirlemek gereklidir.

- Ayırma: oluşan hata çok fazla uyumdan kaynaklanıyorsa bu durum C4.5 algoritması tarafından geliştirilen teknik ile halledilmeye çalışılır (Gülpınar, 2008).

3.2.12.4. C5.0 Algoritması

ID3 algoritmasının bir üst versiyonu olarak C4.5 ve C5.0 algoritması geliştirilmiştir. C4.5 algoritmasının bazı eksik yönlerini gidermek amacıyla C5.0 algoritması geliştirilmiştir. C5.0 algoritması büyük veri setleri için oldukça uygun kullanım alanına sahiptir. Kullanım sırasında bütün aşamalarda tüm özelliklerin kontrolü sağlanır. Bütün özelliklerin normalize edilmiş bilgi kazancını hesaplar ve en iyi bilgi kazancını sağlayan özellik ağaç yapısına dâhil edilir.

C5.0 algoritması her ne kadar C4.5 algoritmasına benzerlik gösterse de üstünlüğü daha da belirgindir.

3.2.12.5. Sınıflandırma ve Regrasyon Ağaçları (C&RT)

1984 yılında Breiman, Freidman, Olshen ve Stone tarafından yazılan kitapta C&RT algoritmasına yer verilmiştir. Bu şekilde bu algoritma istatistik bilimine kazandırılmıştır (Emel ve Çağatan, 2005).

C&RT (Classification and Regression Trees) algoritmasında her aşamada ilgili kümenin, kendisinden daha homojen olacak şekilde iki alt kümeye ayrılması sağlanır. Bu ayırım uygulaması kategorik bağımlı değişkenler için gini, twoing, sürekli değişkenler için en küçük kareler sapması (Least-Squared Deviation) indeks hesaplamalarına göre yapılmaktadır.

Gini: bir sıklık dağılımının eşitsizlik miktarı değerlerini tespit eder. Gini de amaç her defasında en büyük veri kümesini temin etmektir. Bu şekilde en iyi bölme durumu sağlanacaktır. Ayrıca bölme işlemi gerçekleştirildikten sonra işe yaramayan kısım izole edilmektedir.

Twoing: gini ye göre daha düzenli bir yapıdadır. Bunun nedeni ise her defasında ana ve yavru düğümlerin %50 sini içermeye çalışmasıdır. Bu sebeple gini ye kıyasla bir yavaşlama söz konusudur.

C&RT algoritması, etkilenen değişkenin kategorik yada nümerik olmasına bağlı olarak ya sınıflandırma veya regresyon ağaçları üreten parametrik olmayan ikili karar ağacı yapısını oluşturmaktadır. Bu algorithmada en sağlıklı bölme niteliğinin belirlenmesinde entropi indeksi kullanılmaktadır. Bu algoritmanın C4.5 algoritmasından farkı yapısı, ayırma kriterleri, budama metotları ve eksik değerlerle alakalı uygulanan tekniklerdir. Bu algorithmada hedef değişken nominal, ordinal, sürekli olabilmektedir.

C&RT algoritması kullanılarak oluşturulan ağaç yapısında bütün veri kümesi kullanılarak başlanır ve tekrar-tekrar iki alt dal oluşturmak amacı ile bütün tahmin değişkenlerinin kullanılması şartı ile veri kümelerinin alt kümelerinin dağılımları yapılarak ağaç yapısı elde edilir. Bu algoritmanın en büyük avantajlarından birisi eksik değerleri halletmek için tamamen otomatik bir mekanizma içermesidir. Bu algoritma ile oluşturulan ağaç yapısında sadece ikili sistemde ayrılan ağaçlar elde edilmektedir. Çoklu ayırma işlemi sağlayamadığı için ayırma kriterlerini basite indirger.

Bu algoritma hem regrasyon hem de sınıflandırma için çok fazla bölünme kurallarını barındırması, sürekli ve kategorik ayırıcıların ayrı ele alınması, çok düzeyli kategorik ayırıcıların özel olarak işleme alınması ve eksik verilerin temin edilmesi gibi sebeplerden dolayı karmaşık bir algoritmadır. Ağaç büyüme süreci ve budanma süreci de karmaşıktır.

3.2.12.6. CHAID Algoritması

Chaid (chi-Square Automatic Interaction Detector) algoritması, C&RT algoritmasına benzer özellikler gösterir. C&RT algoritmasından farkı ise veriyi bölümlere ayırırken optimum bölümleri seçmek için kullanılan entropy veya gini metriği yerine chi-square testi uygulamasıdır.

CHAID algoritması 1980 yılında Kass tarafından bulunmuştur. Bu algoritmada amaç en başarılı bölmeyi tespit etmek için istatistik olarak anlamlı bir farklılığın olmadığı, amaç değişkenine uyum sağlayan çiftlerde tahmin değişkeninin olası kategori çiftini bütünleştirilmesiyle elde edilmiştir.

En uygun olan bölmeyi tespit etmek için tahmin değişkenleri amaç değişkenine uyum sağlayan bir çiftin içerisinde istatistik olarak anlamlı bir fark kalmayınca dek bütünleştirilmesidir (Albayrak ve Yılmaz, 2009).

Chi-kare Otomatik Etkileşim Tespiti, Kass tarafından 1980 yılında bulunan en köklü ve en kullanışlı ağaç sınıflandırma tekniklerinden biridir. Değişkenlerdeki kombinasyonları tespit eden bu teknikte sürekli etkileyen değişkenler, otomatik olarak sıralı alt sınıflara bölünür.

CHAID algoritması, bir popülasyonun, etkilenen değişkendeki varyasyonu gruplar içi en aza ve gruplar arası en fazla olacak biçimde farklı alt gruplara veya bölümlere tekrarlı olacak şekilde ayırım yapmayı sağlayan bir tekniktir.

Bu algoritmada ki-kare tekniği sayesinde, ilgi seviyesine göre farklılık içeren grupları ayrı olacak şekilde sınıflandırma sağlanır. Sonuç olarak ağacın yaprakları, veride bulunan farklı yapı miktarı kadar dallanma sağlayacaktır.

CHAID algoritması ile oluşturulan ağaç yapısında budama işlemi gerçekleştirilmez. Bu algoritmada eksik değerlerin hepsi tek bir kategori olarak düşünülür ve o şekilde işlem görür.

CHAID, kategorik bir şekilde bulunan verilerin sınıflandırmasında kullanılan açıklayıcı bir algoritma yapısıdır. En uygun dağılımı tespit etmek için ki-kare istatistiği

kullanılmaktadır. Bu algoritma daha büyük veri kümeleri için uygunluk sağlayan, kısmen basit bir algoritmayı esas alan ve ikili olmayan ağaçların oluşumunu sağlar.

Değişken değerleri kategorik şekilde olan ve etkileyen değişkenler ile kategorik olarak ölçülebilen sonuçlar arasında bağ arandığında oldukça faydalı sonuçlar bulan CHAID, kullanıcılara birçok kolaylık sunmaktadır.

CHAID algoritmasını, ID3, C4.5, C&RT algoritmalarından farklı yapan unsur ikili değil çoklu ağaç üretebilmesidir (Albayrak ve Yılmaz, 2009).

CHAID algoritmasının yararları;

- Bağımlı ve bağımsız değişkenin ölçü tipi nominal, ordinal veya aralıklı seçilebilir.
- Bağımsız değişkenlerin hepsini aynı düzeyde ölçmesi gerekmemektedir.
- Bağımsız değişkenlerdeki kayıp değerler sabit olmayan kategori olarak elde tutulabilir.
- En doğru istatistiksel kriter uygulanırsa, şans başarısı çok düşük olacak biçimde ağaç modellemesi elde edilebilir, şeklinde sıralanabilir.

3.2.12.7. Quest Algoritması

Bu algoritmada etkilenen değişkenin nominal seçilerek, ikili ağaç yapısı oluşturan bir yapıdır.

QUEST (Quik, Unbiased, Efficient, Statistical Tree) algoritması bir değişkenli ve doğrusal bileşim bölmelerini destekler niteliktedir.

QUEST algoritması değişken tercihi önyargısı ve hesaplama bütçesi gereği detaylı araştırma tekniklerinden çok daha uygun olduğu tespit edilmiştir (Gülpınar, 2008).

3.2.12.8. SLIQ Algoritması

SLIQ (Supervised Learning In Quest), IBM Quest tarafından geliştirilen, büyük veri gruplarını bölümlere ayıran ve karar ağacı oluştururken ön sıralama yöntemini kullanan bir algoritmadır. Bu yöntem ile her düğümdeki sıralama maliyetini büyük oranda önlemiş olacaktır. Sayısal ve kategorik özellikte olan verilerin sınıflandırılmasında kullanılır. Aynı zamanda bu algoritma ile bellekte tutulması güç olan çok büyük veri setlerinde kullanmak için oldukça uygundur.

3.2.12.9. Sprint Algoritması

SPRINT (Scalable Parallelizable Induction of Decision Trees) algoritması, çok büyük veri grupları için oldukça uygun bir kullanıma sahiptir. Bu algoritma ana bellek ile eğitim veri tabanının boyutu arasındaki bütün bağı yok eder ve sınıf ve kayıt numarasını tutan farklı bir nitelik liste yapısı uygular. Bu algoritma farklı veri yapıları kullanması nedeniyle SLIQ algoritmasından farklılık arz eder.

Çizelge 4. Karar Ağacı Algoritmaları ve Özellikleri (Ulusoy, 2013)

Karar ağacı algoritmaları	Özellikler	Referans
C&RT	<p>Gini' ye dayalı ikili bölme işlemi vardır.</p> <p>Son ya da uç olmayan her bir düğümde iki tane dal bulunur.</p> <p>Budama eylemi ağaç yapısının karmaşıklığına bağlıdır.</p> <p>Sınıflandırma ve regresyona uygun ve destekler nitelikte bir yapıya sahiptir.</p> <p>Sürekli olan amaç değişkenleri ile çalışma yapar. Verinin hazırlık aşaması vardır.</p>	BREİMA N ET AL (1984)
C4.5 ve C5.0 (ID3 algoritmalarının geliştirilmiş versiyonları)	<p>Ağaç yapısı düğümlerden çıkan dallanmalar sayesinde oluşur. Oluşan dallanma sayısı tahmincinin kategori sayısına denktir.</p> <p>Yalnız bir sınıflayıcı da birden çok karar ağacını birleştirir. Bilgi kazancı eylemini ayırma işlemi için tercih eder.</p> <p>Her yapraktaki yanılma payına göre budama gerçekleştirilir.</p>	QUİNL N (1993)
CHAID	<p>Bölme eyleminde ki-kare testini uygular.</p> <p>Dallanma sayısı iki veya tahmincinin belirlediği katagori sayısı arasında değişkenlik gösterir.</p>	KASS (1980)
SLIQ	<p>Çabuk ölçeklenebilir bir sınıflayıcıdır.</p> <p>Ağaç budama eylemini hızla yapacak kapasiteye sahiptir.</p>	IBM Alme dan Araştırma Merkezi (1996)
SPRINT	<p>Veri setinin fazla olması durumunda tercih edilen bir algoritmadır.</p> <p>Bir nitelik değerine göre bölme eylemi gerçekleşir.</p> <p>Bütün bellek kısıtlamaları üzerinde nitelik listesi ve veri yapısı kullanarak uygulama yapar.</p>	SHAFER ATAL (1996)

Daha önce SPSS Clementine olarak bilinen PASW Modeller (SPSS tarafından 2009 yılında bu ismi almıştır), IBM tarafından geliştirilen SPSS' in kurumsal veri madenciliği yazılımıdır. Kullanıcıya kolaylıkla veri madenciliği algoritmalarını kullanabilmelerine imkân sunar. Görsel bir ara yüze sahip olan yazılım, Windows işletim sisteminde çalıştırılabilmektedir. Yazılım ticari lisansa sahiptir ve en popüler veri madenciliği paketlerinden birisidir (Gökçen Alıç, 2014).



4. BULGULAR

4024 tane ortaöğretim öğrencilerinden alınan veriler doğrultusunda karar ağaçları modellemesi yapılmaktadır. Çalışmada kullanılan değişkenler; öğrenci başarısı (ÖB), aile ile birliktelik durumu (AB), TV bağımlılığı (TB), kendine ait oda durumu (OD), kardeş sayısı (KS), hastalık durumu (HD), baba iş durumu (BİD), anne iş durumu (AİD), yaşadığı konum (YK), bilgisayar bağımlılığı (BB), baba eğitim durumu (BED), anne eğitim durumu (AED), okulöncesi eğitim durumu (OÖED) olarak belirlenmektedir. Bu veriler kullanılarak karar ağacı oluşturulmaktadır. Karar ağacında kullanılan algoritmalarda C&RT, CHAID, QUEST ve C5.0 algoritmaları uygulanmış ve elde edilen analiz sonuçlarına göre en yüksek başarı yüzdesine sahip olan C5.0 algoritması olmaktadır. C5.0 karar ağacı modellemesine göre sınıflandırma kuralları oluşturulmaktadır. Veri setine ait toplam ve yüzde değerleri Çizelge 5, 6, 7 ve 8'de görüldüğü gibidir.

Çizelge 5. Başarı, Kitap Okuma Alışkanlığı, Sayısal Yetenek, Oda Durumu, Hastalık, Baba İş Durumu, Anne İş Durumu, Bilgisayar Bağımlılığı, Okul Öncesi Eğitim, Aile ile Birlikte Yaşama Durumu ve TV Bağımlılığı değişkenlerine ait toplam ve yüzde değerler.

Değişken	Kodlama	Sınıflandırma	Toplam	Yüzde
Başarı	0	Başarısız	1687	41,9
	1	Başarılı	2337	58,1
Kitap	0	Okumayan	214	5,3
	1	Okuyan	3810	94,7
Sayısal	0	Yeteneksiz	694	17,2
	1	Yetenekli	3330	82,8
Oda Durumu	0	Yok	1799	44,7
	1	Var	2225	55,3
Hastalık	0	Yok	3735	92,8
	1	Var	289	7,2
Baba İş Durumu	0	Çalışmıyor	174	4,3
	1	Çalışıyor	3850	95,7
Anne İş Durumu	1	Çalışmıyor	2808	69,8
	0	Çalışıyor	1216	30,2
Bilgisayar Bağımlılığı	1	Yok	652	16,2
	0	Var	3372	83,8
Okul Öncesi Eğitim	0	Almamış	1703	42,3
	1	Almış	2321	57,7
Aile ile Birliktelik Durumu	0	Aile ile yaşamıyor	214	5,3
	1	Aile ile yaşıyor	3810	94,7
TV Bağımlılığı	0	TV bağımlılığı yok	694	17,2
	1	TV bağımlılığı var	3330	82,8

Çizelge 6.Kardeş değişkenine ait toplam ve yüzde değerler.

Kardeş Durumu	Kardeş Sayısı	Toplam	Yüzde
	0	16	,4
	1	269	6,7
	2	1746	43,4
	3	1334	33,2
	4	316	7,9
	5	119	3,0
	6	108	2,7
	7	63	1,6
	8	40	1,0
	9	4	,1
	11	9	,2

Çizelge 7.TV bağımlılığı değişkenine ait toplam ve yüzde değerler.

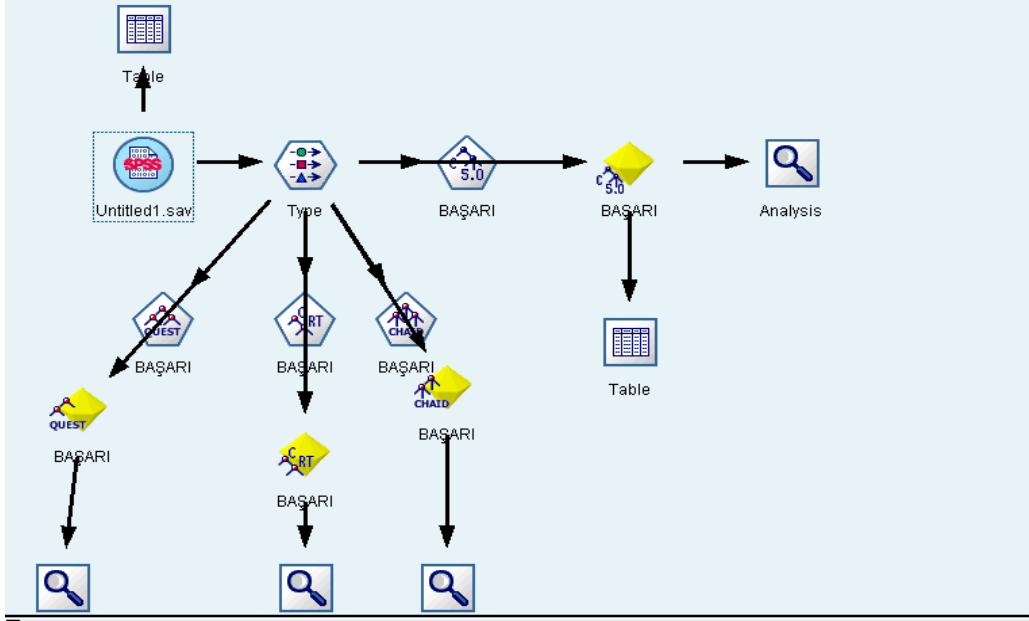
Değişken	Kodlama	Sınıflandırma	Toplam	Yüzde
Yaşadığı konum	1	İl merkez	1054	26,2
	2	İlçe	1029	25,6
	3	Köy	1941	48,2

Çizelge 8.Eğitim durumu değişkenine ait toplam ve yüzde değerler.

Değişken	Kodlama	Sınıflandırma	Baba		Anne	
			Toplam	Yüzde	Toplam	Yüzde
Eğitim Durumu	0	Okur-yazar değil	51	1,3	48	1,2
	1	İlkokul	10	,2	39	1,0
	2	Ortaokul	564	14,0	1170	29,1
	3	Lise	1226	30,5	1174	29,2
	4	Üniversite	1273	31,6	902	22,4
	5	Üniversite Üzeri	900	22,4	691	17,2

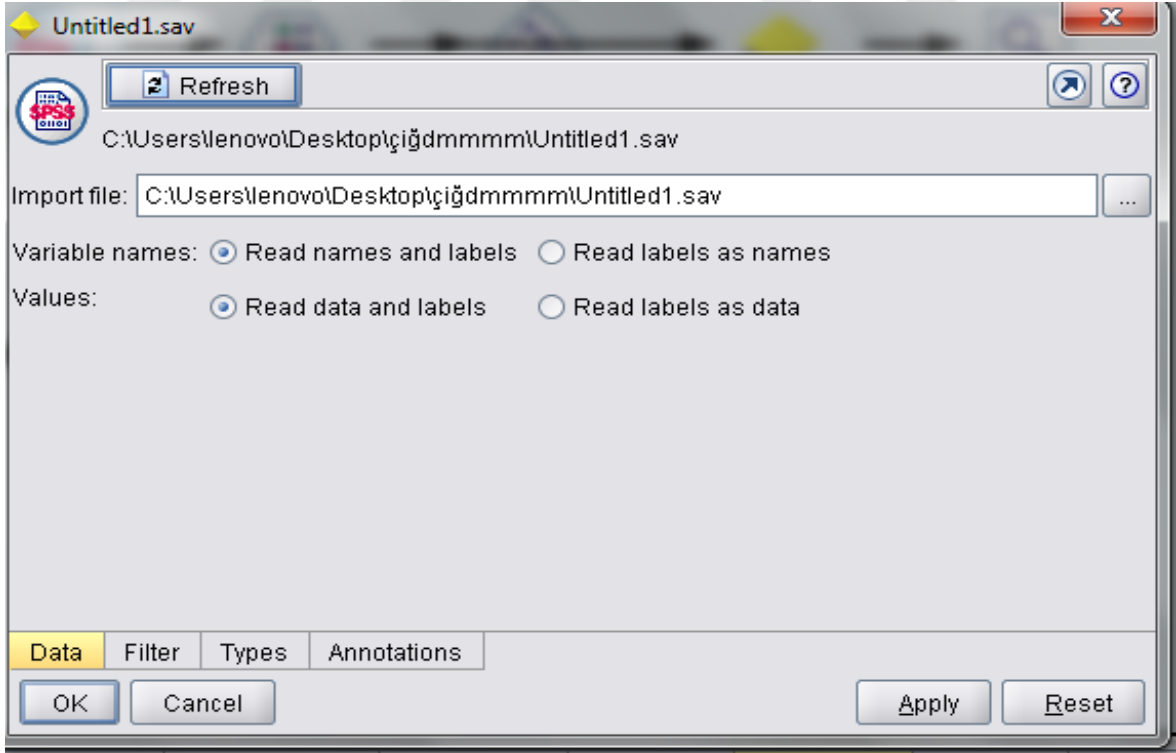
4.1. Modelin Oluşturulması Ve Sonuç Üretimi

Öğrenci başarısını etkileyen faktörlerin sınıflandırılmasına yönelik çalışmada PASW akışı (stream) Şekil 2.'de verilmiştir. Her bir düğüm (node) işlevi sırasıyla aşağıda açıklanmıştır.



Şekil 2. PASW Akışı

Verilerin Okunması: PASW, SQL ve SAS gibi farklı veri tabanları, Excel, XML, SPSS istatistik dosyası, sabit ve değişken dosya gibi farklı kaynaklardan verileri tedarik etme kapasitesine sahiptir. Öğrenci başarısını etkileyen faktörlere ait veriler SPSS istatistik dosyasından (*.sav) alınmıştır.



Şekil 3. Verilerin Alınması

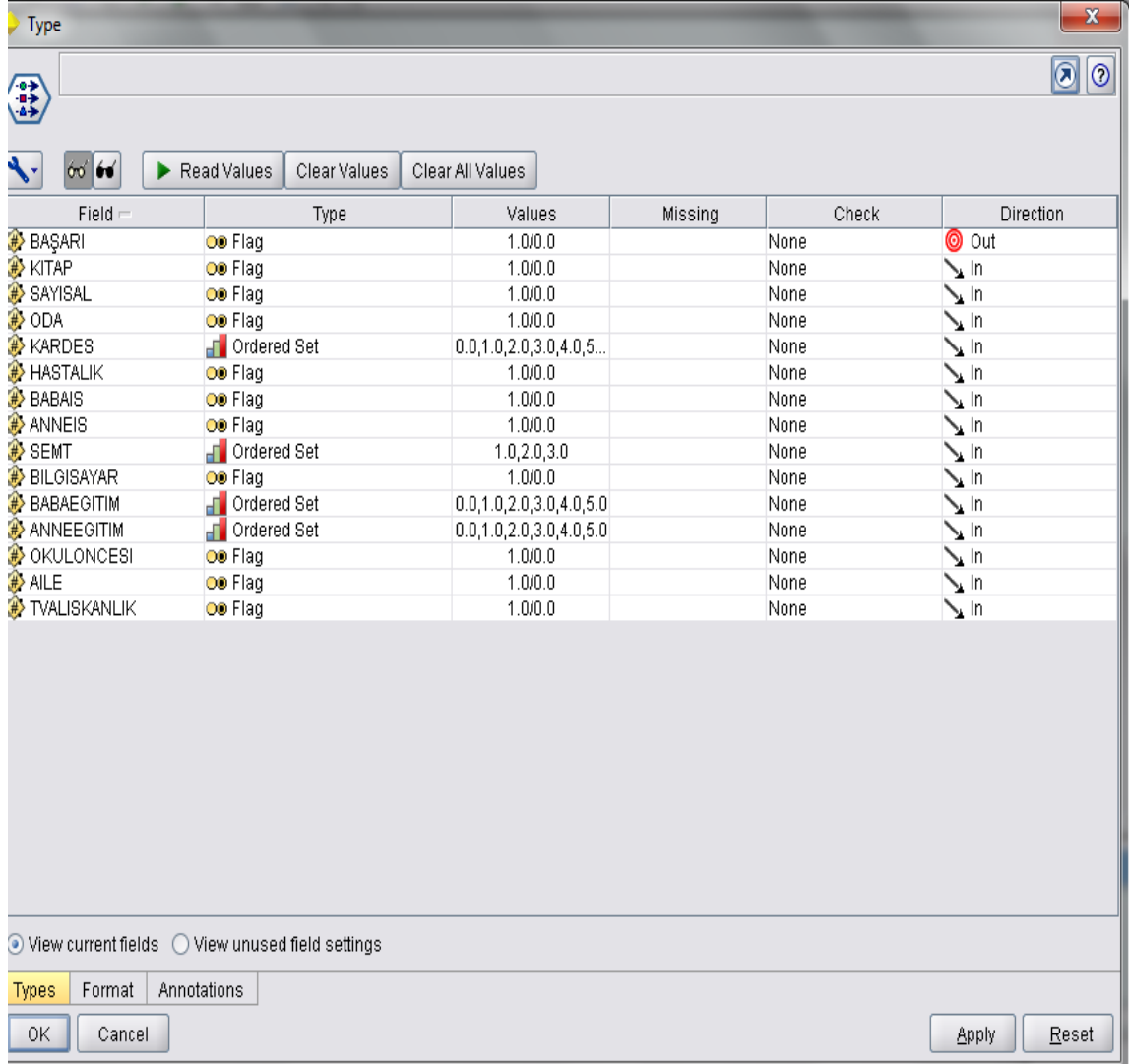
Veri Tablosu: Tablo (table) düğümü ile eldeki veri görüntülenebilir bir şekilde sunulmuştur. Şekil 4.' te verilere ait tablo verilmiştir. Veri tabanında 15 değişken (field) ve 4024 kayıt (record) bulunmaktadır.

	BAŞARI	KİTAP	SAYISAL	ODA	KARDES	HASTAL...	BABAIS	ANNEIS	SEMT	BILGISA...	BABAE...	ANNEE...	OKULO...	AILE	TVALIS...
1	0.00	1.00	1.00	1.00	2.00	0.00	1.00	1.00	2.00	1.00	4.00	4.00	1.00	1.00	1.00
2	0.00	1.00	0.00	1.00	3.00	0.00	1.00	0.00	2.00	1.00	5.00	4.00	0.00	1.00	0.00
3	0.00	1.00	1.00	1.00	2.00	0.00	1.00	0.00	2.00	1.00	4.00	4.00	1.00	1.00	1.00
4	1.00	1.00	0.00	1.00	2.00	0.00	1.00	0.00	2.00	1.00	5.00	4.00	1.00	1.00	0.00
5	0.00	1.00	1.00	1.00	2.00	1.00	1.00	0.00	2.00	0.00	4.00	2.00	1.00	1.00	1.00
6	0.00	1.00	1.00	1.00	2.00	0.00	1.00	1.00	2.00	1.00	3.00	2.00	0.00	1.00	1.00
7	0.00	1.00	0.00	0.00	4.00	1.00	1.00	0.00	2.00	0.00	3.00	2.00	0.00	1.00	0.00
8	1.00	1.00	1.00	1.00	2.00	0.00	1.00	1.00	2.00	1.00	4.00	2.00	0.00	1.00	1.00
9	0.00	1.00	0.00	0.00	2.00	0.00	1.00	0.00	2.00	1.00	4.00	4.00	0.00	1.00	0.00
10	1.00	1.00	0.00	0.00	6.00	0.00	1.00	0.00	2.00	1.00	2.00	2.00	1.00	1.00	0.00
11	0.00	1.00	0.00	0.00	3.00	1.00	0.00	0.00	2.00	0.00	4.00	4.00	1.00	1.00	0.00
12	1.00	1.00	0.00	0.00	3.00	0.00	1.00	0.00	2.00	1.00	5.00	4.00	1.00	1.00	0.00
13	0.00	1.00	1.00	1.00	2.00	0.00	1.00	0.00	2.00	1.00	4.00	2.00	1.00	1.00	1.00
14	1.00	1.00	0.00	1.00	2.00	0.00	1.00	0.00	2.00	1.00	5.00	4.00	1.00	1.00	0.00
15	1.00	1.00	0.00	0.00	3.00	0.00	1.00	0.00	2.00	1.00	5.00	1.00	1.00	1.00	0.00
16	1.00	1.00	0.00	1.00	2.00	0.00	1.00	0.00	2.00	1.00	2.00	1.00	1.00	1.00	0.00
17	0.00	1.00	0.00	0.00	2.00	0.00	1.00	0.00	2.00	1.00	2.00	2.00	1.00	1.00	0.00
18	1.00	1.00	0.00	0.00	2.00	0.00	1.00	1.00	2.00	1.00	5.00	5.00	1.00	1.00	0.00
19	1.00	1.00	1.00	0.00	3.00	0.00	1.00	0.00	2.00	1.00	2.00	2.00	1.00	1.00	1.00
20	1.00	1.00	0.00	0.00	2.00	0.00	1.00	0.00	2.00	1.00	5.00	5.00	0.00	1.00	0.00
21	1.00	1.00	1.00	1.00	2.00	0.00	1.00	0.00	2.00	1.00	4.00	2.00	0.00	1.00	1.00
22	1.00	1.00	0.00	1.00	3.00	0.00	1.00	0.00	2.00	1.00	2.00	2.00	0.00	1.00	0.00
23	0.00	1.00	1.00	0.00	2.00	0.00	1.00	0.00	2.00	1.00	3.00	2.00	0.00	1.00	1.00
24	1.00	1.00	1.00	1.00	3.00	0.00	1.00	0.00	2.00	1.00	4.00	4.00	0.00	1.00	1.00
25	1.00	1.00	1.00	0.00	4.00	0.00	1.00	0.00	2.00	1.00	4.00	2.00	0.00	1.00	1.00
26	0.00	1.00	0.00	1.00	1.00	0.00	1.00	0.00	2.00	0.00	4.00	4.00	0.00	1.00	0.00
27	1.00	1.00	0.00	0.00	2.00	0.00	1.00	0.00	2.00	1.00	2.00	5.00	0.00	1.00	0.00
28	0.00	0.00	0.00	1.00	3.00	0.00	1.00	1.00	2.00	1.00	2.00	4.00	1.00	0.00	0.00
29	0.00	1.00	0.00	1.00	2.00	0.00	1.00	0.00	2.00	1.00	4.00	4.00	1.00	1.00	0.00
30	0.00	1.00	1.00	1.00	2.00	1.00	1.00	0.00	2.00	0.00	4.00	4.00	0.00	1.00	1.00
31	1.00	1.00	1.00	1.00	2.00	0.00	1.00	0.00	2.00	1.00	2.00	4.00	0.00	1.00	1.00
32	1.00	1.00	1.00	1.00	2.00	0.00	1.00	1.00	2.00	1.00	5.00	5.00	1.00	1.00	1.00
33	0.00	0.00	1.00	0.00	2.00	0.00	1.00	0.00	2.00	0.00	3.00	3.00	0.00	0.00	1.00
34	0.00	0.00	0.00	0.00	2.00	1.00	0.00	0.00	2.00	0.00	2.00	2.00	1.00	0.00	0.00
35	1.00	1.00	1.00	1.00	2.00	0.00	1.00	0.00	2.00	1.00	2.00	4.00	0.00	1.00	1.00
36	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00	2.00	1.00	4.00	4.00	1.00	1.00	0.00
37	1.00	1.00	1.00	0.00	2.00	0.00	1.00	0.00	2.00	1.00	2.00	3.00	0.00	1.00	1.00

Şekil 4. Veri Tablosu

Değişken Tiplerinin Belirlenmesi: Sınıflandırma probleminin çözümü için öncelikle hangi değişkenlerin girdi (input), hangi değişkenin ise çıktı (out) olduğunun tespit edilmesi gereklidir. Aynı zamanda değişkenlerin ölçüt düzeylerinin de belirtilmesi gerekir. Tip (type) düğümü ile bu işlemler gerçekleştirilmiştir. (Şekil 5) Amaç değişkeni out, diğer değişkenler ise input olarak tayin edilmiştir. Başarı (flag), kitap (flag), sayısal

(flag), oda (flag), kardeş (ordered set), hastalık (flag), babaiş (flag), anneiş (flag), semt (ordered set), bilgisayar (flag), babaegitim (ordered set), anneegitim (ordered set), okulöncesi (flag), aile (flag), tvalışkanlık (flag) olarak belirlenmiştir.



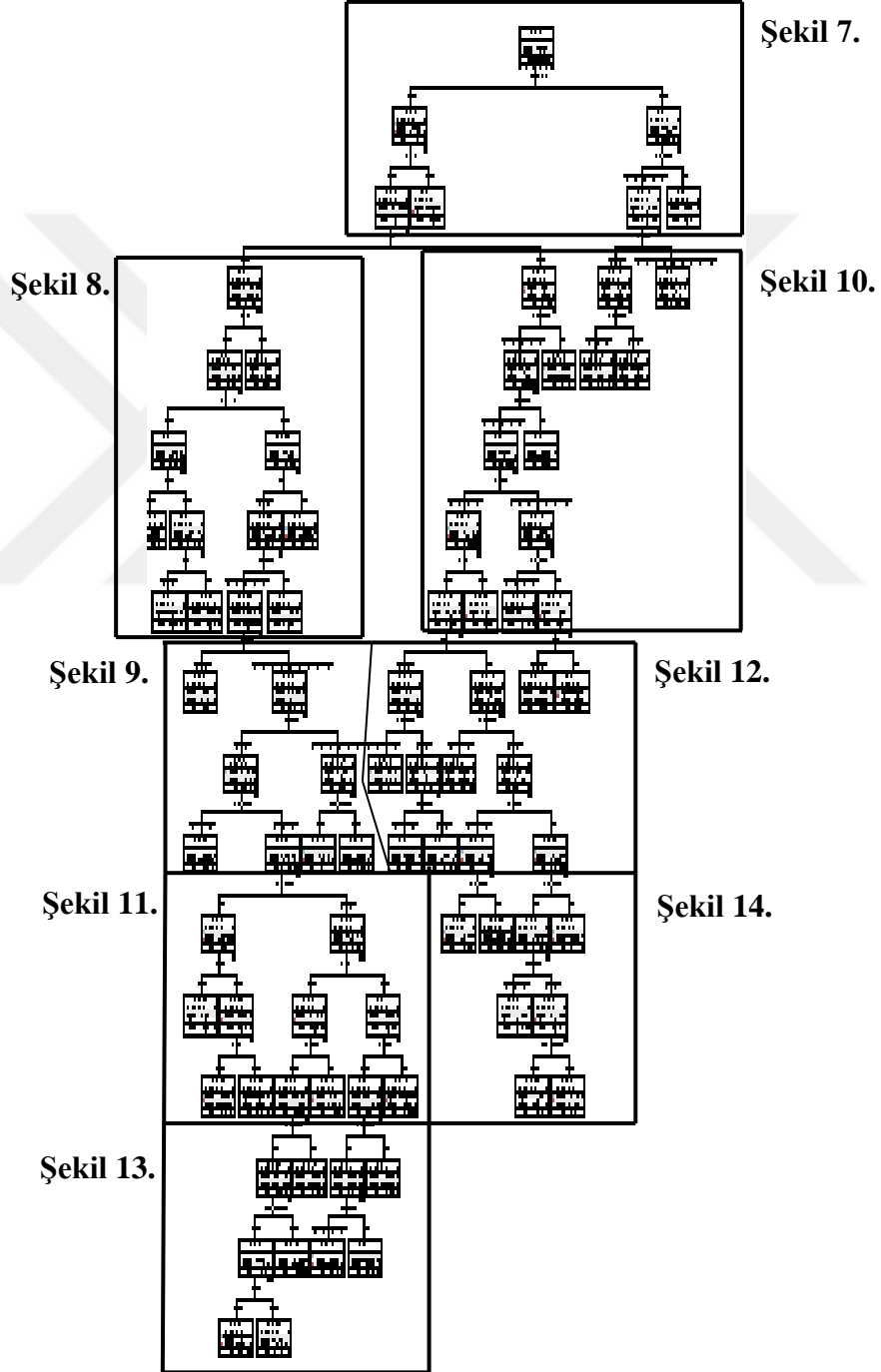
Şekil 5. Değişken Tiplerinin Belirlenmesi

Uygun Yöntemin Belirlenmesi: öğrenci başarısını etkileyen faktörlerin sınıflandırılmasında en iyi performansı sağlayan sınıflandırıcı algoritması, “ yapılan analiz sonuçlarına dayanarak C5.0 olduğu görülmüştür. Karar ağaçlarında yaygın olarak kullanılan algoritmalar QUEST, C&RT, CHAID ve C5.0 ile analizler yapılarak ve elde edilen analiz sonuçları karşılaştırılarak en başarılı sınıflandırmayı C5.0 algoritması sağladığı görülmektedir. Bu algoritmaların analiz sonuçları sırasıyla Ek 1, 2, 3, 4’de ve Çizelge 9’da verilmiştir.

Çizelge 9. QUEST, C&RT, CHAID ve C5.0 Analiz Sonuçları

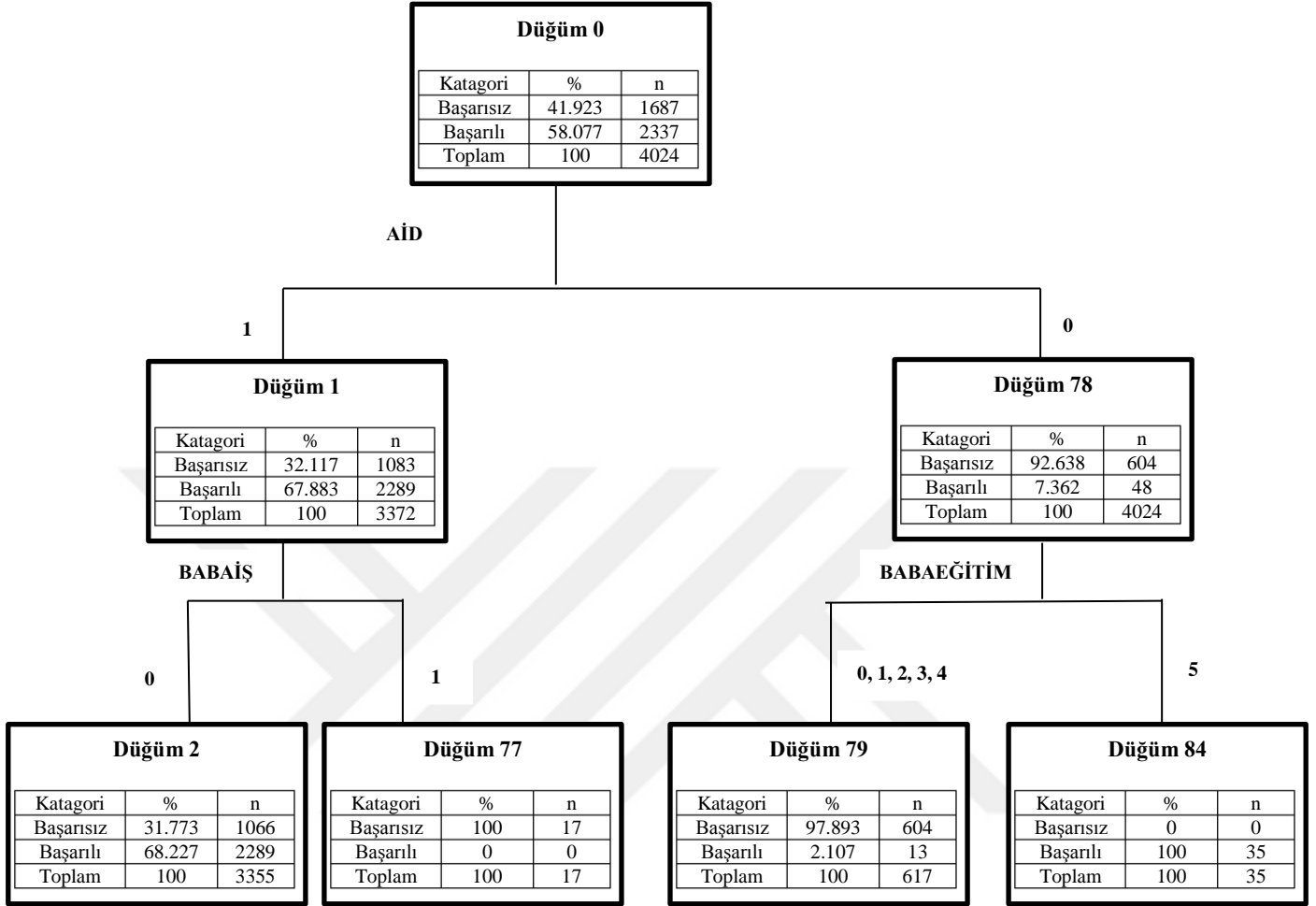
Algoritma	Doğru	Yüzde	Yanlış	Yüzde
QUEST	3238	90.47	786	19.53
C&RT	3292	81.81	732	18.19
CHAID	3328	92.7	696	17.3
C5.0	3131	90.23	393	9.77

Karar Ağacı ve Kuralların Üretilmesi

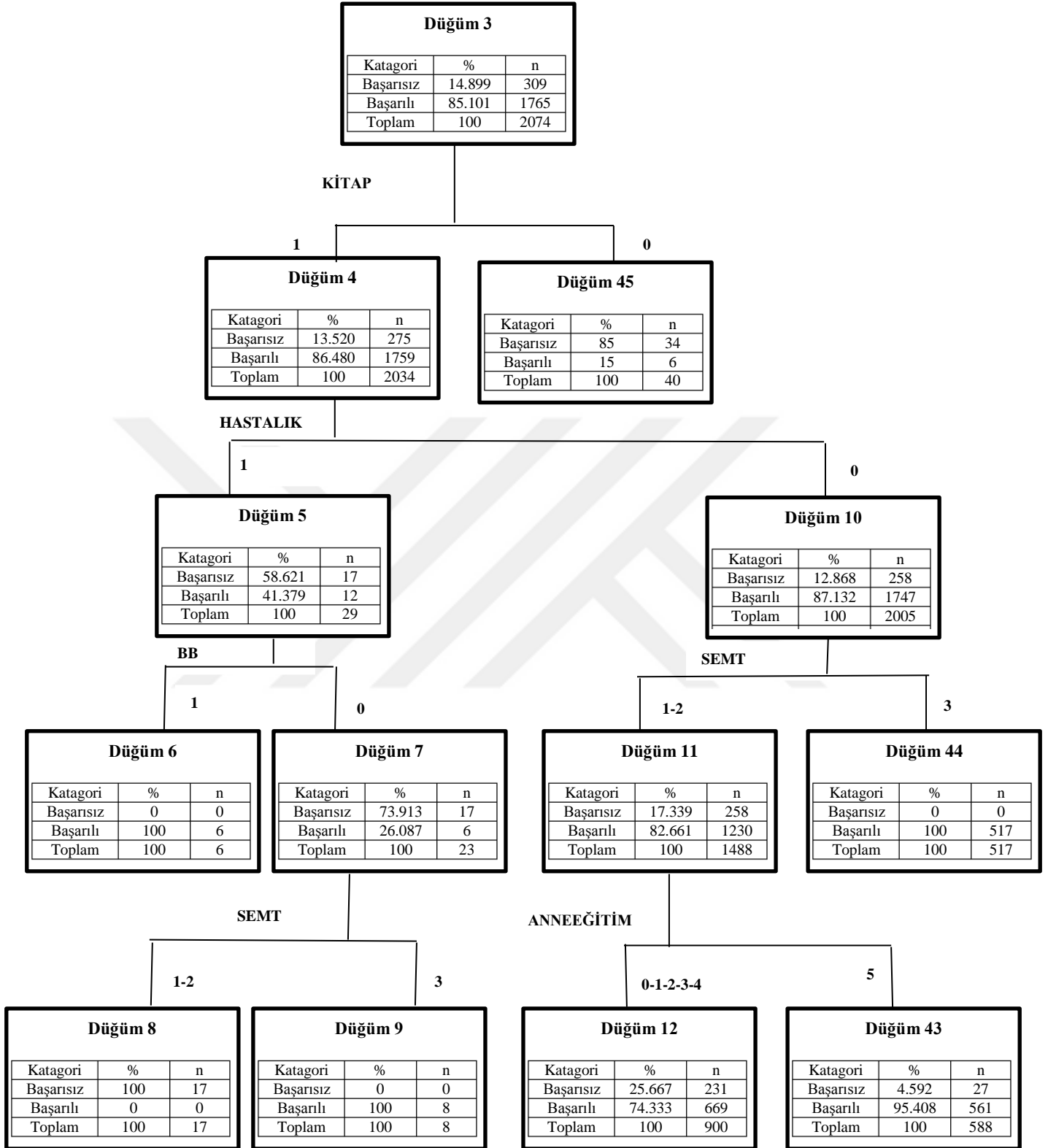


Şekil 6. C5.0 ile Oluşturulan Karar Ağacı

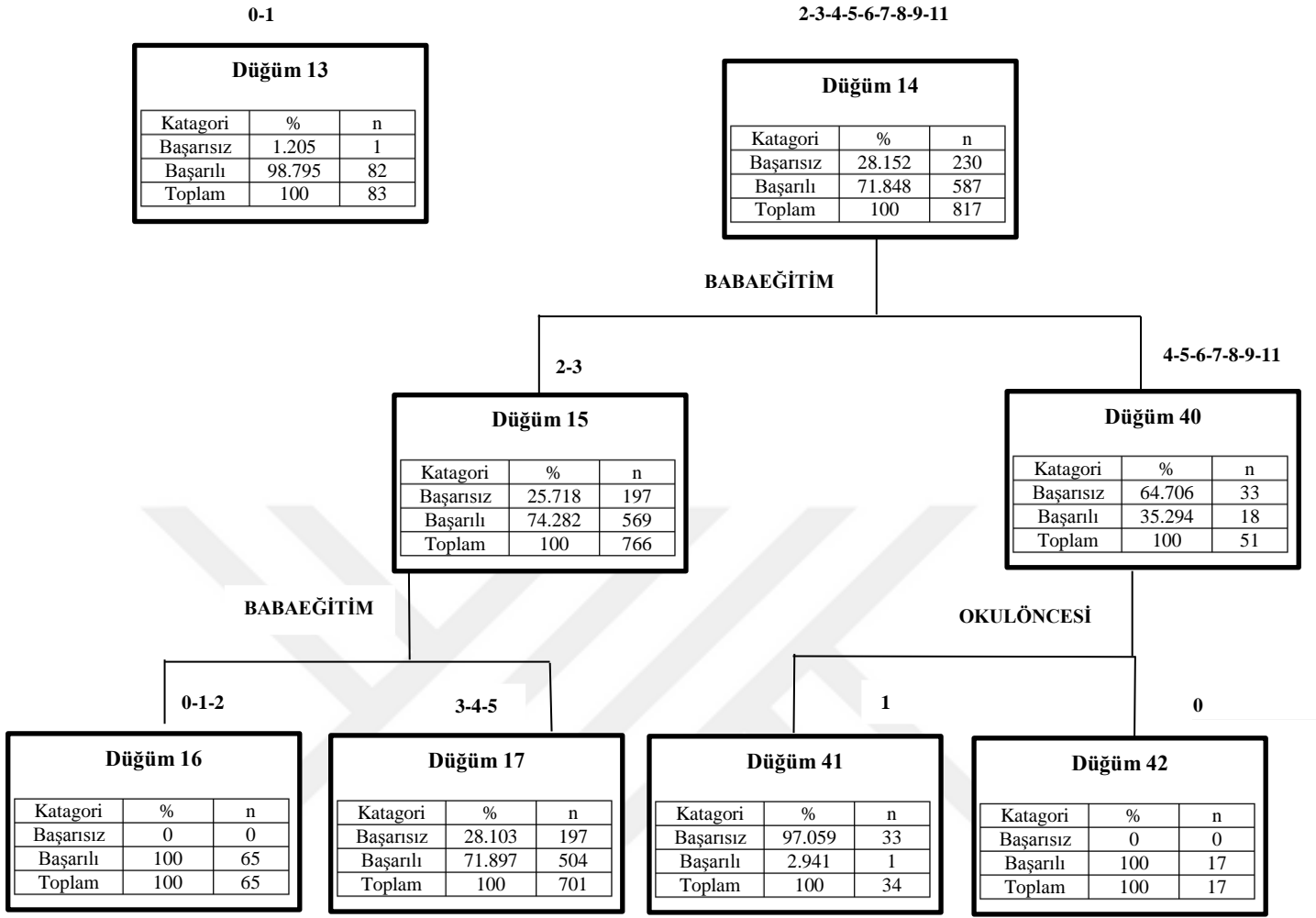
Karar ağacına ait ayrıntılar Şekil 7, 8, 9, 10, 11, 12, 13 ve 14’de görülmektedir.



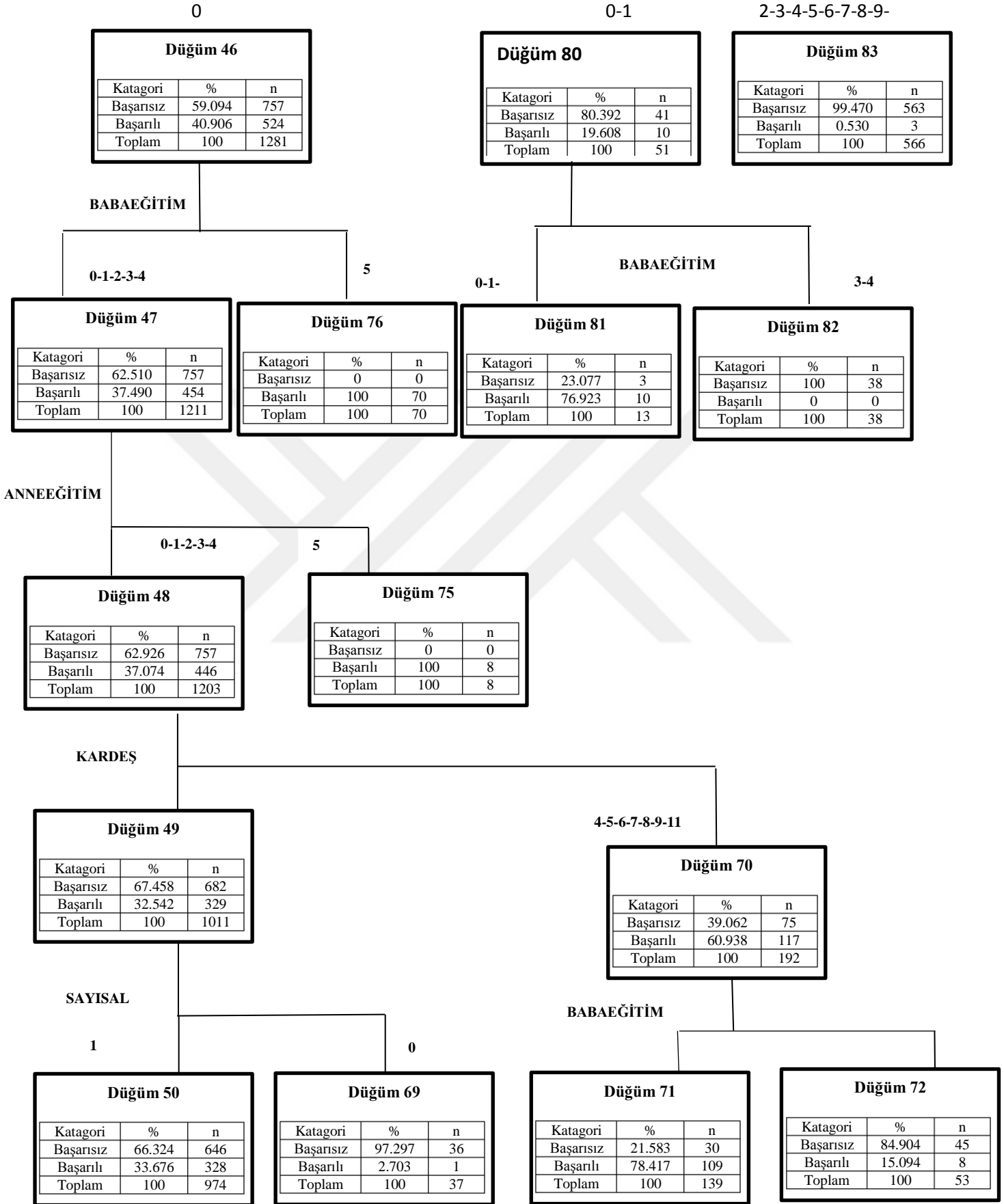
Şekil 7. düğüm 0, 1, 78, 2, 77, 79 ve 84



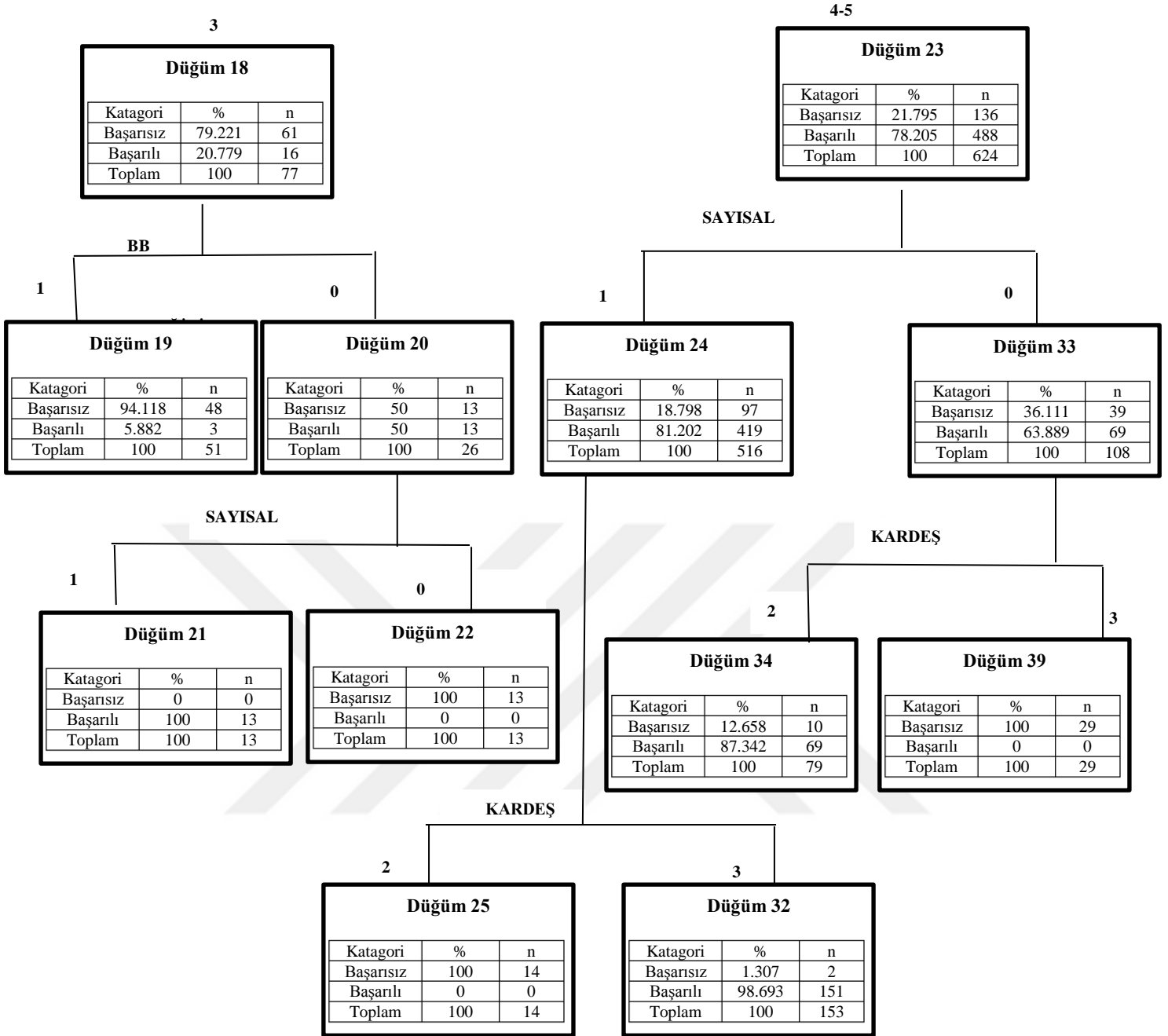
Sekil 8. Düğüm 3, 4, 45, 5, 10, 6, 7, 11, 44, 8, 9, 12 ve 43



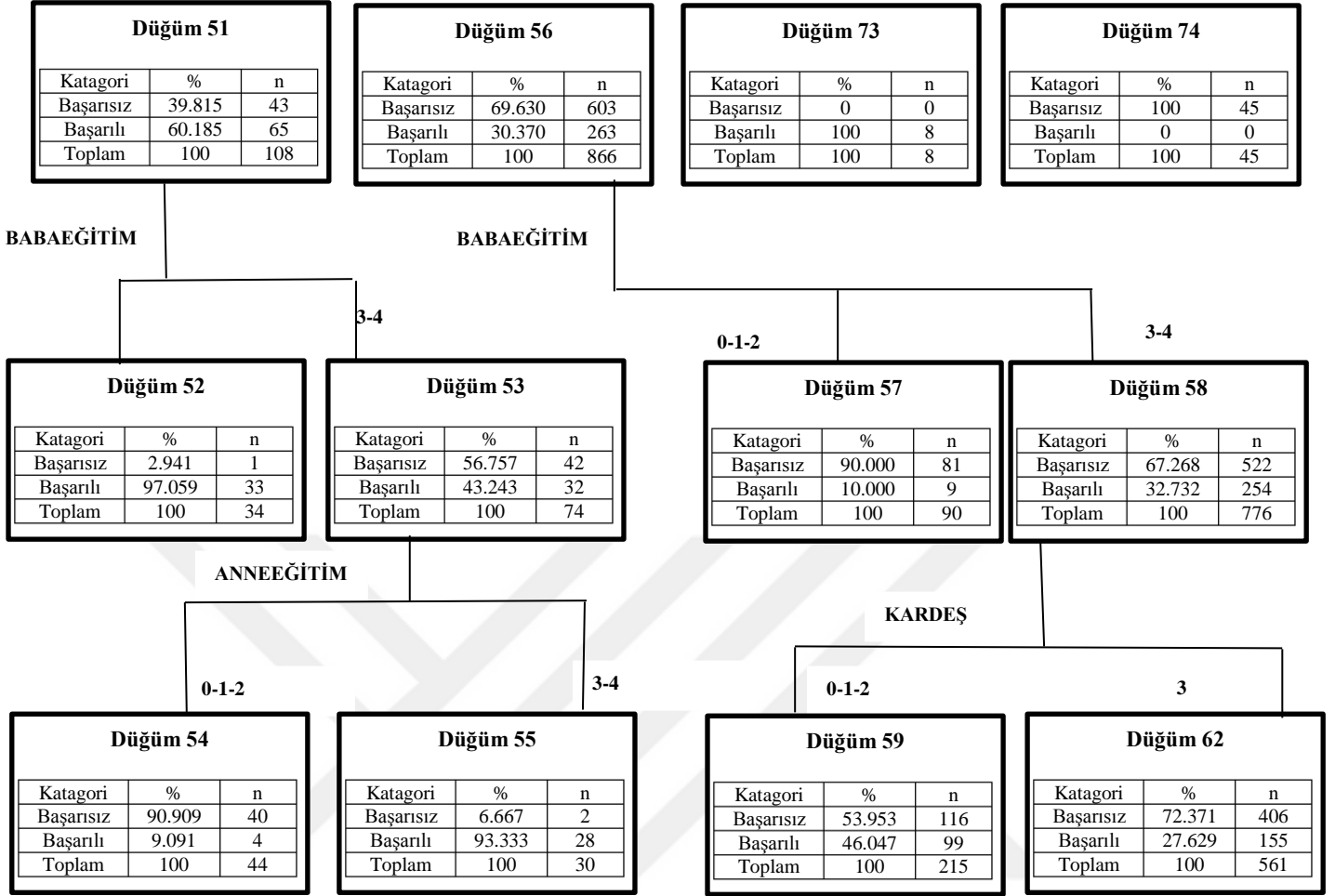
Sekil 9. Düğüm 13, 14, 15, 40, 16, 17, 41 ve 42



Şekil 10. Düğüm 46, 80, 83, 47, 76, 81, 82, 48, 75, 49, 70, 50, 69, 71 ve 72

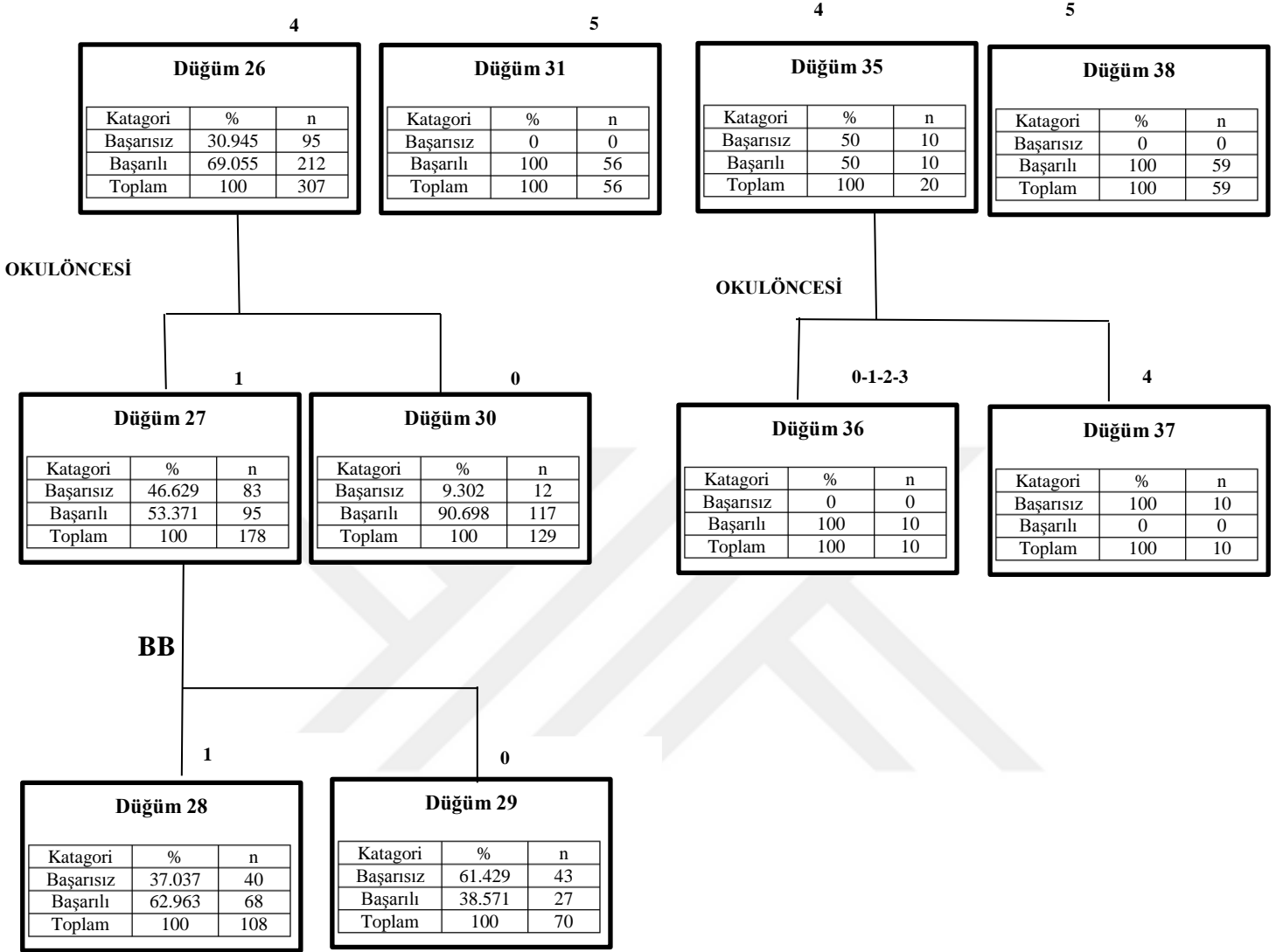


Sekil 11. Düğüm 18, 23, 19, 20, 24, 33, 21, 22, 25, 32, 3 ve 39

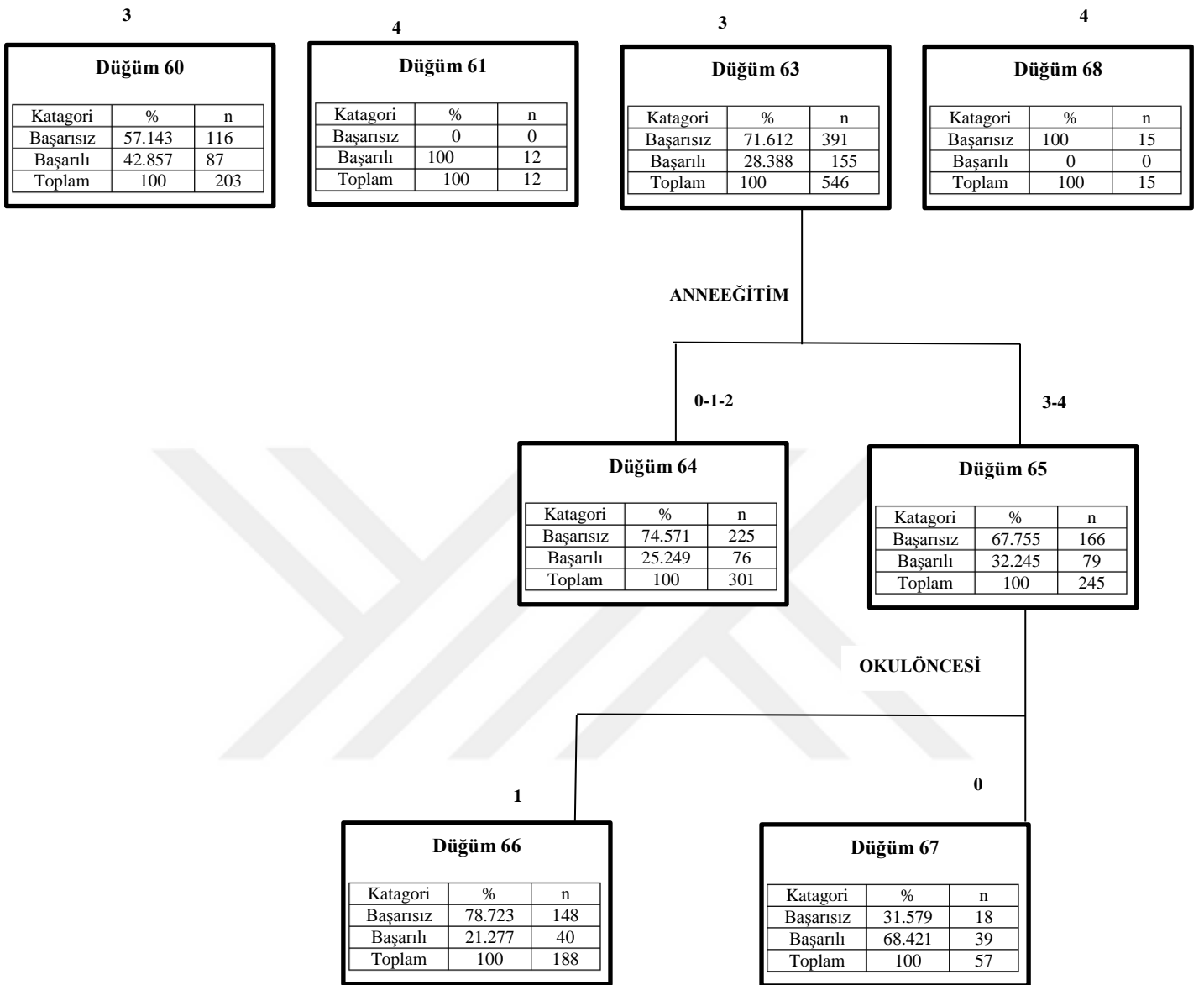


Sekil 12. Düğüm 51, 56, 73, 74, 52, 53, 57, 58, 54, 55, 59 ve 62

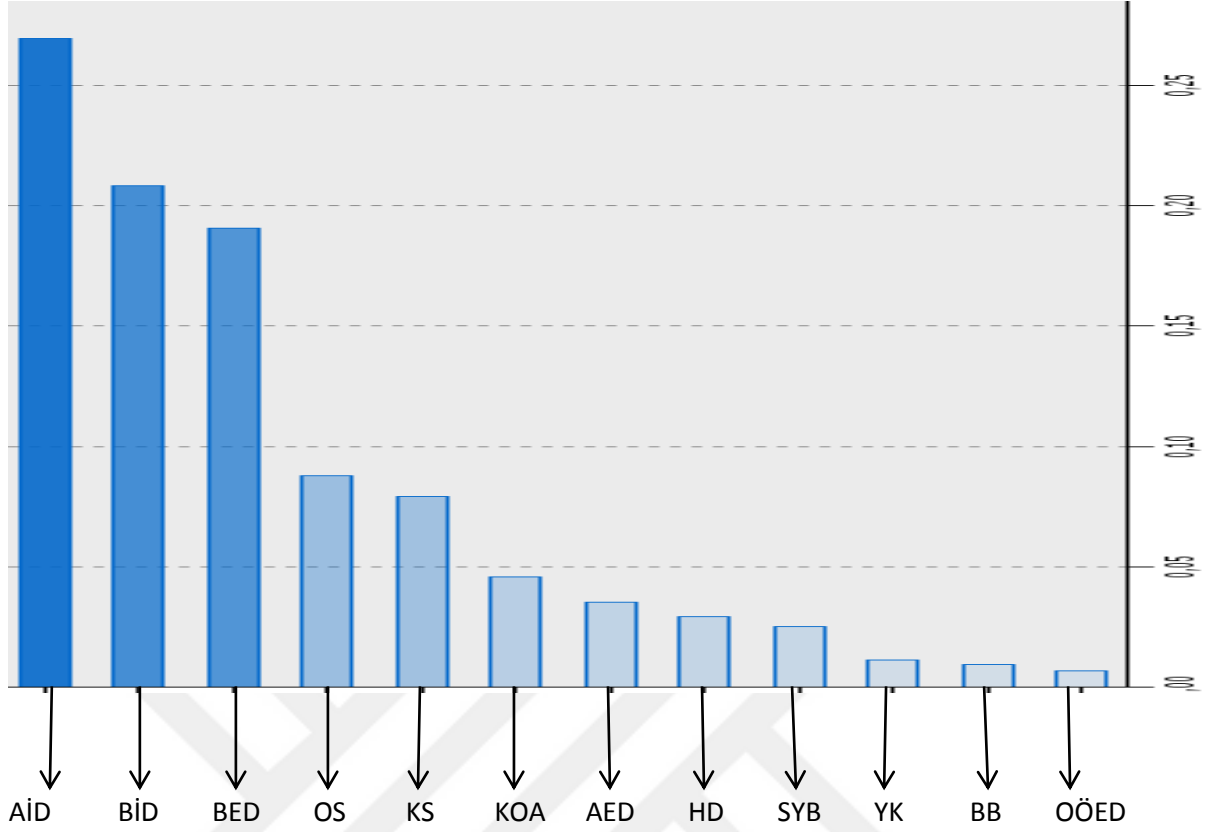
BABAEĞİTİM



Sekil 13. Düğüm 26,31, 35, 38, 27, 30, 36, 37, 28 ve 29



Sekil 14. Düğüm 60, 61, 63, 68, 64, 65, 66 ve 67



Şekil 15. Başarı Değişkenine Etki Eden Faktörler

Analize toplam 14 girdi (input) değişkeniyle başlanmıştır. Bu değişkenlerden 12 tanesi ağaç yapısında yer almıştır. Karar ağacında yer alan bu değişkenlerin karar kuralları üretiminde yeterli olduğu görülmektedir. Başlangıç dallanma AİD üzerinde olmuştur. C5.0 algoritması ile oluşturulan ağaç modelinin % 90.23 başarılı bir performans sağladığı görülmektedir. C5.0 karar ağacı sınıflandırıcısından elde edilen kurallar aşağıda verilmiştir. Şekil 15’de görüldüğü üzere başarı durumuna etki eden en önemli değişken anne iş durumudur.

Çizelge 10. Karar Ağacı ile Elde Edilen Kurallar

<p><u>KURAL 1</u> AİD=1 BİD=1 OS=1 KOA=1 HD=1 BB=0 YK=3</p>	<p>Annesi çalışmayan, babasının işi olan, kendine ait odası olan, kitap okuma alışkanlığı olan, geçirmiş olduğu hastalığı olan, bilgisayar bağımlılığı olan, yaşadığı konum köy olan öğrencilerden 6 öğrenci belirlenmiş ve bu 6 öğrencinin de başarılı sınıf içerisinde olduğu tespit edilmiştir. % 100 başarı performansı ile sınıflandırma işlemi yapılmıştır. (düğüm=9)</p>
<p><u>KURAL 2</u> AİD=1 BİD=1 OS=1 KOA=1 HD=1 BB=1</p>	<p>Annesi çalışmayan, babasının işi olan, kendine ait odası olan, kitap okuma alışkanlığı olan, geçirmiş olduğu hastalığı olan, bilgisayar bağımlılığı olmayan 6 öğrenci belirlenmiş ve bu 6 öğrencinin de başarılı sınıf içerisinde olduğu tespit edilmiştir. % 100 başarı performansı ile sınıflandırma işlemi yapılmıştır. (düğüm=6)</p>
<p><u>KURAL 3</u> AİD=1 BİD=1 OS=1 KOA=1 HD=0 YK= 1 ve 2 AED= 0, 1, 2, 3, 4 KS= 2, 3, 4, 5, 6, 7, 8, 9, 11 KS= 2, 3 BED= 3, 4, 5 BED= 3 BB= 0 SYB= 1</p>	<p>Annesi çalışmayan, babasının işi olan, kendine ait odası olan, kitap okuma alışkanlığı olan, geçirmiş olduğu hastalığı olmayan, yaşadığı konum il ve ilçe olan, annesinin eğitim durumu (olmayan-ilkokul-ortaokul-lise-üniversite) olan, kardeş sayısı 13 öğrenci belirlenmiş ve bu 13 öğrencinin de başarılı sınıf içerisinde olduğu tespit edilmiştir. % 100 başarı performansı ile sınıflandırma işlemi yapılmıştır. (düğüm=21)</p>
<p><u>KURAL 4</u> AİD=1 BİD=1 OS=1 KOA=1 HD=0 YK= 1 ve 2 AED= 0, 1, 2, 3, 4 KS= 2, 3, 4, 5, 6, 7, 8, 9, 11 KS= 2, 3 BED= 3, 4, 5 BED=4,5 SYB=1 KS=2 BED=4 OÖED=1 BB=1</p>	<p>Bu grupta toplam 108 öğrenci belirlenmiştir. Bu öğrencilerin 40 tanesi başarısız ve 68 tanesi başarılı sınıfta yer almıştır. (düğüm=28)</p>

<p><u>KURAL 5</u> AİD=1 BİD=1 OS=1 KOA=1 HD=0 YK= 1 ve 2 AED= 0, 1, 2, 3, 4 KS= 2, 3, 4, 5, 6, 7, 8, 9, 11 KS= 2, 3 BED= 3, 4, 5 BED=4,5 SYB=0 KS=2 BED=4 AED=0, 1, 2, 3</p>	<p>Bu grupta toplam 10 öğrenci belirlenmiştir. Bu öğrencilerin 0 tanesi başarısız ve 10 tanesi başarılı sınıfında yer almıştır. (düğüm=36)</p>
<p><u>KURAL 6</u> AİD=1 BİD=1 OS=1 KOA=1 HD=0 YK=3</p>	<p>Bu grupta toplam 517 öğrenci belirlenmiştir. Bu öğrencilerin 0 tanesi başarısız ve 517 tanesi başarılı sınıfında yer almıştır. (düğüm=44)</p>
<p><u>KURAL 7</u> AİD=1 BİD=1 OS=0 BED=5</p>	<p>Bu grupta toplam 70 öğrenci belirlenmiştir. Bu öğrencilerin 0 tanesi başarısız ve 70 tanesi başarılı sınıfında yer almıştır. (düğüm=76)</p>
<p><u>KURAL 8</u> AİD=0 BED=5</p>	<p>Bu grupta toplam 35 öğrenci belirlenmiştir. Bu öğrencilerin 0 tanesi başarısız ve 35 tanesi başarılı sınıfında yer almıştır. (düğüm=84)</p>
<p><u>KURAL 9</u> AİD= 0 BED= 0, 1, 2, 3, 4 KS= 0, 1 BED= 0, 1, 2</p>	<p>Bu grupta toplam 13 öğrenci belirlenmiştir. Bu öğrencilerin 3 tanesi başarısız ve 10 tanesi başarılı sınıfında yer almıştır. (düğüm=81)</p>
<p><u>KURAL 10</u> AİD=1 BİD=1 OS=0 BED=0, 1, 2, 3, 4 AED=5</p>	<p>Bu grupta toplam 8 öğrenci belirlenmiştir. Bu öğrencilerin 0 tanesi başarısız ve 8 tanesi başarılı sınıfında yer almıştır. (düğüm=75)</p>

<p><u>KURAL 11</u> AİD=1 BİD=1 OS=0 BED=0, 1, 2, 3, 4 AED=0, 1, 2, 3, 4 KS=4, 5, 6, 7, 8, 9, 11 BED=0, 1, 2, 3</p>	<p>Bu grupta toplam 139 öğrenci belirlenmiştir. Bu öğrencilerin 30 tanesi başarısız ve 109 tanesi başarılı sınıfında yer almıştır. (düğüm=71)</p>
<p><u>KURAL 12</u> AİD=1 BİD=1 OS=0 BED=0, 1, 2, 3, 4 AED=0, 1, 2, 3, 4 KS=4, 5, 6, 7, 8, 9, 11 BED=4 YK=1, 2</p>	<p>Bu grupta toplam 8 öğrenci belirlenmiştir. Bu öğrencilerin 0 tanesi başarısız ve 8 tanesi başarılı sınıfında yer almıştır. (düğüm=73)</p>
<p><u>KURAL 13</u> AİD=1 BİD= 1 OS=0 BED=0, 1, 2, 3, 4 AED=0, 1, 2, 3, 4 KS=0, 1, 2, 3 SYB=1 YK=3 BED=3, 4 KS=3 BED=3 AED=3, 4 OÖED=0</p>	<p>Bu grupta toplam 57 öğrenci belirlenmiştir. Bu öğrencilerin 18 tanesi başarısız ve 39 tanesi başarılı sınıfında yer almıştır. (düğüm=67)</p>
<p><u>KURAL 14</u> AİD=1 BİD=1 OS=1 KOA=1 HD=0 YK=1, 2 AE=0, 1, 2, 3, 4 KS=2, 3, 4, 5, ,7, 8, 9, 11 KS=2,3 BED=3, 4, 5 BED=4,5 SYB=1 KS=2 BED=4 OÖED=0</p>	<p>Bu grupta toplam 129 öğrenci belirlenmiştir. Bu öğrencilerin 12 tanesi başarısız ve 117 tanesi başarılı sınıfında yer almıştır. (düğüm=30)</p>

<p><u>KURAL 15</u> AİD=1 BİD=1 OS=1 KOA=1 HD=0 YK=1, 2 AED=0, 1, 2, 3, 4 KS=2, 3, 4, 5, 6, 7, 8, 9, 11 KS=2, 3 BED=3, 4, 5 BED=4, 5 SYB=0 KS=2 BED=5</p>	<p>Bu grupta toplam 59 öğrenci belirlenmiştir. Bu öğrencilerin 0 tanesi başarısız ve 59 tanesi başarılı sınıfında yer almıştır. (düğüm=38)</p>
<p><u>KURAL 16</u> AİD=1 BİD=1 OS=0 BED=0, 1, 2, 3, 4 AED=0, 1, 2, 3, 4 KS=0, 1, 2, 3 SYB=1 YK=3 BED=3, 4 KS=0, 1, 2 BED=4</p>	<p>Bu grupta toplam 12 öğrenci belirlenmiştir. Bu öğrencilerin 0 tanesi başarısız ve 12 tanesi başarılı sınıfında yer almıştır. (düğüm=61)</p>
<p><u>KURAL 17</u> AİD=1 BİD=1 OS=0 BED=0, 1, 2, 3, 4 AED=0, 1, 2, 3, 4 KS=0, 1, 2, 3 SYB=1 YK=1,2 BED=0, 1, 2</p>	<p>Bu grupta toplam 34 öğrenci belirlenmiştir. Bu öğrencilerin 1 tanesi başarısız ve 33 tanesi başarılı sınıfında yer almıştır. (düğüm=52)</p>

<p><u>KURAL 18</u> AİD=1 BİD=1 OS=1 KOA=1 HD=0 YK=1, 2 AED=0, 1, 2, 3, 4 KS=2, 3, 4, 5, 6, 7, 8, 9, 11 KS=2, 3 BED=3, 4, 5 BED=4, 5 SYB=1 KS=3</p>	<p>Bu grupta toplam 153 öğrenci belirlenmiştir. Bu öğrencilerin 2 tanesi başarısız ve 151 tanesi başarılı sınıfında yer almıştır. (düğüm=32)</p>
<p><u>KURAL 19</u> AİD=1 BİD=1 OS=1 KOA=1 HD=0 YK=1,2 AED=0, 1, 2, 3, 4 KS=2, 3, 4, 5, 6, 7, 8, 9, 11 KS=2, 3 BED=3, 4, 5 BED=4, 5 SYB=1 KS=2 BED=5</p>	<p>Bu grupta toplam 56 öğrenci belirlenmiştir. Bu öğrencilerin 0 tanesi başarısız ve 56 tanesi başarılı sınıfında yer almıştır. (düğüm=31)</p>
<p><u>KURAL 20</u> AİD=1 BİD=1 OS=1 KOA=1 HD=0 YK=1, 2 AED=0, 1, 2, 3, 4 KS=2, 3, 4, 5, 6, 7, 8, 9, 11 KS=2, 3 BED=0, 1, 2</p>	<p>Bu grupta toplam 65 öğrenci belirlenmiştir. Bu öğrencilerin 0 tanesi başarısız ve 65 tanesi başarılı sınıfında yer almıştır. (düğüm=16)</p>

<p><u>KURAL 21</u> AİD=1 BİD=1 OS=1 KOA=1 HD=0 YK=1, 2 AED=0, 1, 2, 3, 4 KS=2, 3, 4, 5, 6, 7, 8, 9, 11 KS=4, 5, 6, 7, 8, 9, 11 OÖED=0</p>	<p>Bu grupta toplam 17 öğrenci belirlenmiştir. Bu öğrencilerin 0 tanesi başarısız ve 17 tanesi başarılı sınıfta yer almıştır. (düğüm=42)</p>
<p><u>KURAL 22</u> AİD=1 BİD=1 OS=0 BED=0, 1, 2, 3, 4 AED=0, 1, 2, 3, 4 KS=0, 1, 2, 3 SYB=1 YK=1,2 BED=3,4 AED=3, 4</p>	<p>Bu grupta toplam 30 öğrenci belirlenmiştir. Bu öğrencilerin 2 tanesi başarısız ve 28 tanesi başarılı sınıfta yer almıştır. (düğüm=55)</p>
<p><u>KURAL 23</u> AİD=1 BİD=1 OS=1 KOA=1 HD=0 YK=1, 2 AED=5</p>	<p>Bu grupta toplam 588 öğrenci belirlenmiştir. Bu öğrencilerin 27 tanesi başarısız ve 561 tanesi başarılı sınıfta yer almıştır. (düğüm=43)</p>

<p><u>KURAL 24</u></p> <p>AİD=1</p> <p>BİD=1</p> <p>OS=1</p> <p>KOA=1</p> <p>HD=0</p> <p>YK=1, 2</p> <p>AED=0, 1, 2, 3, 4</p> <p>KS=0, 1</p>	<p>Bu grupta toplam 83 öğrenci belirlenmiştir. Bu öğrencilerin 1 tanesi başarısız ve 82 tanesi başarılı sınıfta yer almıştır. (düğüm=13)</p>
---	--

5. TARTIŞMA

Bu çalışmada veri madenciliği tekniklerinden olan karar ağaçları ile eğitim veri madenciliği çalışması yapılmıştır. Çalışmada kullanılan değişkenler; öğrenci başarısı (ÖB), aile ile birliktelik durumu (AB), TV bağımlılığı (TB), kendine ait oda durumu (OD), kardeş sayısı (KS), hastalık durumu (HD), baba iş durumu (BİD), anne iş durumu (AİD), yaşadığı konum (YK), bilgisayar bağımlılığı (BB), baba eğitim durumu (BED), anne eğitim durumu (AED), okulöncesi eğitim durumu (OÖED) olarak belirlenmiştir. Uygulamada SPSS Clementine veri madenciliği yazılımı kullanılmıştır. Analize toplam 15 girdi değişkeni ile başlanmış ve 12 niteliğin ağaç yapısında yer aldığı tespit edilmiştir. Yani C5.0 algoritmasının değişken sayısını indirgeyerek bu sonuçları elde etmesi C5.0 karar ağacı tekniğinin performans üstünlüğünü gösterir ve bu sonuç Gökçen Alıç (2014)' in elde ettiği sonuçlar ile uyum sağlamaktadır. Uygulama kapsamında kullanılacak olan verilerin ait olduğu nitelikler, öğrenci başarısını etkileyen faktörler olarak tespit edilmiştir. Karar ağaçlarında kullanılan birçok algoritma mevcuttur. Bu çalışmada C&RT, C5.O, QUEST VE CHAID algoritmaları denenmiş ve bu algoritmalarından elde edilen analiz sonuçları karşılaştırılarak en yüksek başarıyı sağlayan algoritmanın C5.0 algoritması olduğu tespit edilmiş ve bu sonuç Aydın (2015)' nın elde ettiği algoritma sonuçları ile uyumlu olduğu tespit edilmiştir.

C5.0 algoritması analiz sonuçları çizelge 9'da verilmiştir. Bu sonuçlara göre C5.0 algoritması % 90.23 başarı ve % 9.77 başarısızlık sağlamaktadır. Yani 4024 öğrenciden 3631 tanesini doğru ve 393 öğrenciyi yanlış sınıflandırma yapmaktadır. C5.0 karar ağacı sınıflandırıcısından, % 50 ve üzeri başarılı olan öğrenciler için toplam 24 adet karar kuralları oluşturulmaktadır. Elde edilen sınıflandırma kuralları eğitim çalışmalarına ışık tutabilecek niteliktedir.

Cihan (2015), öğrenci proje anketlerini sınıflandırmada en iyi algoritmanın belirlenmesi adlı makalesinde, sınıflandırma algoritmalarından öğrenci projelerinden elden edilen veri setini sınıflandırmada en başarılı algoritma olarak C4.5 algoritması olduğunu tespit etmiştir ve bu sonuç burada bulunan sonuç ile uyumsuz bulunmuştur.

6. SONUÇ

Bilgi teknolojilerindeki gelişmeler, bu kapsamda kuruluşların bilgi sistemlerini kurma ihtiyaçları, çok yüksek miktarda verinin sağlanmasına ve depolanmasına sebep olmaktadır.

Söz konusu olan verilerdeki saklı kalmış örüntülerin (pattern) tespit edilmesi, bunlardan anlamlı ve işe yarayacak bilgilerin elde edilerek karar mekanizmalarına sunulması ihtiyacı veri madenciliğinin önemli bir araştırma alanı olmasına sebep olmaktadır.

Eğitim veri madenciliğinin hedefi, veri madenciliği tekniklerini kullanarak eğitim alanındaki saklı bilgilerin keşfedilmesi ve önemli örüntülerin açığa çıkarılmasıdır. Bu örüntüler arkasındaki mekanizmalar net olarak anlaşılmamış olsa bile, örüntülerin tanımlanması mümkün olabilmektedir.

Kahramanmaraş ilinde bulunan Dulkadiroğlu ilçesine ait olan 44 ortaokuldan alınan veriler doğrultusunda çalışmalar yapılmıştır. Araştırmada kullanılacak olan değişkenler eğitim kurumlarında görev yapan öğretmenler ile görüşülerek belirlenmiştir. 4024 öğrenciye daha önce hazırlanan anket uygulanmış ve anket sonuçlarına göre ağaç modellemesi elde edilmiştir. Çalışma sonucunda 24 karar kuralı oluşturulmuştur. Elde edilen karar kurallarına ve Şekil 15' e dayanarak öğrenci başarısını en fazla etkileyen faktörün anne iş durumu (AİD) olduğu görülmektedir. 24 karar kuralından 22 tanesinde annenin çalışmıyor olması öğrenci başarısını etkileyen en büyük faktör olduğu tespit edilmiştir. Yani annesi çalışmayan öğrencilerin daha başarılı olduğu sonucuna varılmıştır. Şekil 15' e göre başarıyı etkileyen ikinci faktör ise BİD olmuştur. 22 karar kuralında BİD yer almaktadır ve babası çalışan öğrencilerin daha başarılı olduğu tespit edilmiştir. Şekil 15' e göre başarıyı etkileyen üçüncü faktör ise BED olmuştur. Yani babası üniversite ve üzeri eğitime sahip olan öğrencilerin daha başarılı olduğu görülmektedir. Problemin çözümünde IBM PASW (predictive analyticss oftware) Modeller 14.0 kullanılmaktadır. Analize toplam 15 girdi değişkeni ile başlanmaktadır. Uygulama kapsamında kullanılacak olan verilerin ait olduğu nitelikler, öğrenci başarısını etkileyen faktörler olarak tespit edilmiştir. Çalışmada kullanılan değişkenler; öğrenci başarısı (ÖB), aile ile birliktelik durumu (AB), TV bağımlılığı (TB), kendine ait oda durumu (OD), kardeş sayısı (KS), hastalık durumu (HD), baba iş durumu (BİD), anne iş durumu (AİD), yaşadığı konum (YK), bilgisayar bağımlılığı (BB), baba eğitim

durumu (BED), anne eğitim durumu (AED), okulöncesi eğitim durumu (OÖED) olarak belirlenmiştir. Bu çalışmada veri madenciliği tekniklerinden olan karar ağaçları ile çalışma yapılmaktadır. Karar ağaçlarında kullanılan birçok algoritma mevcuttur. Bu çalışmada C&RT, C5.0, QUEST VE CHAID algoritmaları denenmiş ve bu algoritmalarından elde edilen analiz sonuçları karşılaştırılarak en yüksek başarıyı sağlayan algoritmanın C5.0 algoritması olduğu tespit edilmekte ve oluşturulan ağaç yapısı C5.0 ile sağlanmaktadır. Bu algoritma ile oluşturulan ağaç yapısına göre 15 nitelik ile başlanmıştır ve 12 niteliğin ağaç yapısında yer aldığı tespit edilmektedir. Başlangıç dallanma AİD niteliğinde olmaktadır.

C5.0 karar ağacı sınıflandırıcısından, %50 ve üzeri başarılı olan öğrenciler için toplam 24 adet kural oluşturulmaktadır. Elde edilen sınıflandırma kuralları eğitim çalışmalarına ışık tutabilecek niteliktedir.

KAYNAKLAR

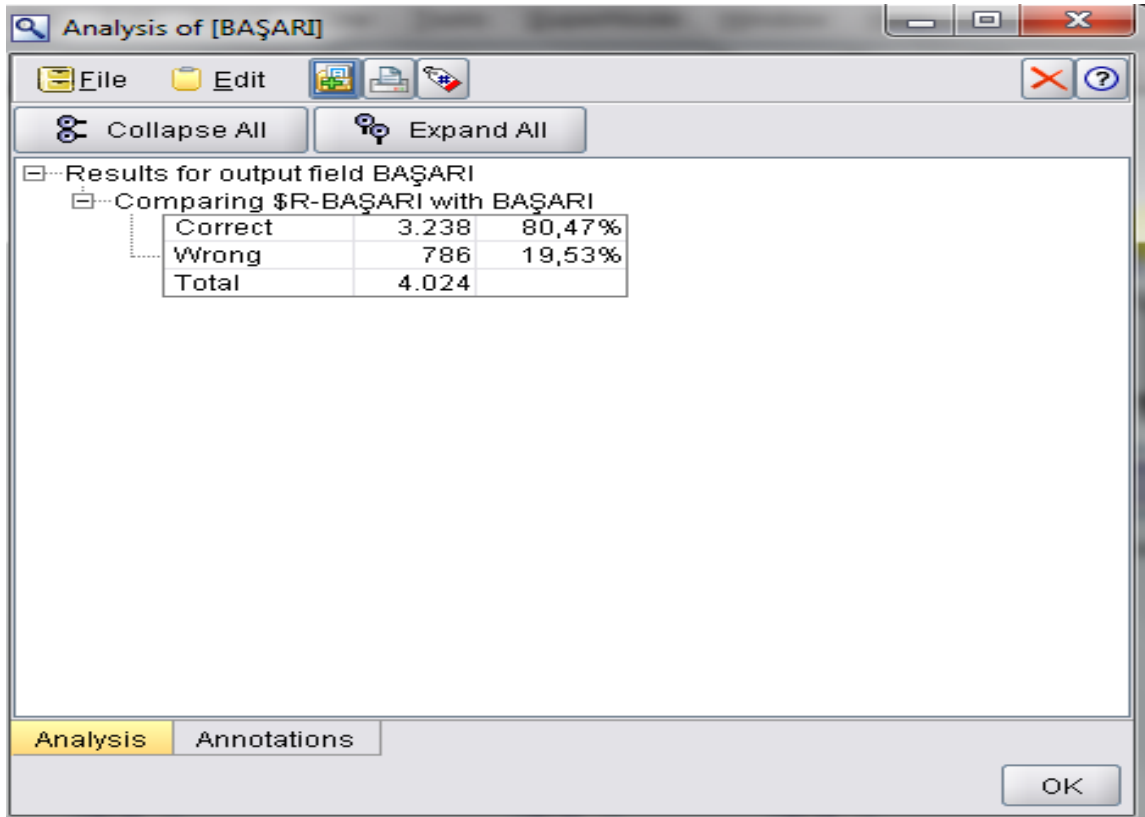
- Akçapınar, G., Coşgun, E., Altun, A., 2011. Prediction of Perceived Disorientation in Online Learning Environment with Random Forest Regression. In *EDM* (pp. 259-264).
- Akpınar, H., 2000. Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği. İstanbul Üniversitesi, İşletme Fakültesi Dergisi. 29(1), s 1-22.
- Albayrak A.S., Yılmaz, Ş.K., 2009. Veri Madenciliği: Karar Ağacı Algoritmaları ve İmkb Verileri Üzerine bir Uygulama. Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 4(1), s. 31-52.
- Amershi, S., Conati, C., 2006. Automatic Recognition of Learner Groups in Exploratory Learning Environments. In *Intelligent Tutoring Systems* (pp. 463-472). Springer Berlin Heidelberg.
- Aydın, S., Özkul, A. E., 2015. Öğrenci Bilgi Sisteminde Değerlendirmenin Veri Madenciliği ile Yapılması. Eğitim ve Öğretim Araştırmaları Dergisi (Journal of Research in Education and Teaching). 4(3).
- Ayesha, S., Mustafa, T., Sattar, A. R., Khan, M. I., 2010. Data Mining model for Higher Education System. *European Journal of Scientific Research*, 43(1), s 24-29.
- Ayık, Y. Z., Özdemir, A. ve Yavuz, U., 2007. Lise Türü ve Lise Mezuniyet Başarısının, Kazanılan Fakülte İle İlişkisinin Veri Madenciliği Tekniği İle Analizi. Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, Cilt 10, Sayı 2.
- Babadağ K., 2006, Zeki Veri Madenciliği: Ham Veriden Altın Bilgiye Ulaşma Yöntemleri, *Industrial Application Software*, 85-87
- Baradwaj, B. K., & Pal, S., 2012. Mining educational data to analyze students' performance. ArXiv preprint arXiv:1201-3417.
- Bırtıl, F.S., 2012. kız meslek lisesi öğrencilerinin akademik başarısızlık nedenlerinin veri madenciliği tekniği ile analizi., Afyon Kocatepe Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Anabilim Dalı, Yüksek Lisans Tezi.
- Bilen, O., Hotaman, D., Aşkın, O.E. ve Büyüklü, A.H., 2014. LYS Başarılarına Göre Okul Performanslarının Eğitsel Veri Madenciliği Teknikleriyle İncelenmesi: 2011 İstanbul Örneği. Eğitim ve Bilim, cilt 39, sayı 172, s 78-94.
- Bozkır, A.S., Sezer, E. ve Gök, B., 2009. Öğrenci Seçme Sınavında (ÖSS) Öğrenci Başarımını Etkileyen Faktörlerin Veri Madenciliği Yöntemleriyle Tespiti. 5. Uluslararası İleri Teknolojiler Sempozyumu (IATS'09), 13-15 Mayıs, Karabük Üniversitesi, Karabük. s 37-43.
- Bresfelean, P., Bresfelean, M. Ghisoiu, N., 2008. 'Determining Students' Academic Failure Profile Founded on Data Mining Methods', Proceedings of the ITI 2008 30th International Conference on Informational Conference on Information Technology Interfaces, s 23-26.

- Cihan, P., Kalıpsız, O., 2015. Öğrenci Proje Anketlerini Sınıflandırmada En Başarılı Algoritmanın Belirlenmesi. TBV- Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi, 8(1), s 41-49.
- Dejaeger, K., Goethals, F., Giangreco, A., Mola, L., Baesens, B., 2012. Gaining Insight Into Student Satisfaction Using Comprehensible Data Mining Techniques. European Journal of Operational Research. 218(2), 548-562.
- Dunham, M. H., 2003. Data Mining Introductory and Advanced Topics. Southern Methodist University. Pearson Education Inc. 92-102.
- Eker, H., 2006. Veri Madenciliği veya Bilgi Keşfi. http://www.bilgiyonetimi.org/pages/mkl_gos.php?nt=538.
- Emel, G. G., Taşkın, Ç., 2015. Veri Madenciliğinde Karar Ağaçları ve Bir Satış Analizi Uygulaması. Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi, 6(2), s. 221-239.
- Erdoğan, Ş. Z., Timor, M., 2005. Öğrenci Veritabanında bir Veri Madenciliği Uygulaması. Havacılık ve Uzay Teknolojileri Dergisi, 2 (2), s 53-57.
- Gaaafar, L., Khanmis, M., 2009. Applications Of Data Mining for Educational Decision support. Proceedings of the 2009 Industrial Engineering Research Conference, s.228-233.
- Gökçen Alıç, Z. H., 2014. Gazi Üniversitesi Bilişim Enstitüsü, Yönetim Bilişim Sistemleri Anabilim Dalı, Yüksek Lisans Tezi.
- Gülen, Ö. ve Özdemir, S., 2013. Veri Madenciliği Teknikleri ile Üstün Yetenekli Öğrencilerin İlgi Alanlarının Analizi. Journal of Gifted Education Research, 1(3), s 215-226.
- Gülpınar, V., 2008. Avrupa Birliği Ülkeleri ile Türkiye'nin Ekonomik Göstergelerinin Karar Ağacı Yöntemi ile Kararlaştırılması. Marmara Üniversitesi Sosyal Bilimler Enstitüsü Ekonometri Anabilim Dalı İstatistik Bilim Dalı. Yüksek Lisans Tezi.
- Halees, A., 2008. Mining students data to analyze learning behavior: a case study, departman of computer science, islamic university, paletsine
- Han, J., ve Kamber, M., 2006. Data Mining Concepts and Techniques (2nd ed). San Fransisco: Elsevier.
- Hark, C., 2013. Öğrencilerin Akıllı Tahtaya İlişkin Tutumlarının İncelenmesine Yönelik bir Veri Madenciliği Uygulaması. Fırat Üniversitesi. Eğitim Bilimleri Enstitüsü, Elazığ. Yüksek Lisans Tezi.
- Hilage, T.A., Kulkarni, R.V., 2012. Review of literature on data mining. International Journal of Research and Reviews in Applied Sciences 10(1), 107-114
- Im, K. H., Kim, T. H., Bae, S., Park, S. C., 2005. Conceptual modeling with neural network for giftedness identification and education. In Advances in Natural Computation (pp. 530-538). Springer Berlin Heidelberg.

- Jormanainen, I., Sutinen, E., 2012. Using Data Mining to Support Teacher's Intervention in a Robotics Class. In Digital Game and Intelligent Toy Enhanced Learning (DIGITEL), Fourth International Conference on (pp. 39-46). IEEE.
- Kayri, M., 2008. Elektronik Portfolyo Değerlendirmeleri için Veri Madenciliği Yaklaşımı. Yüzüncüyıl Üniversitesi, Eğitim Fakültesi Dergisi, 5(1), s 98-110.
- Kelley-Winstead, D., 2010. New Directions in Education Research: Using Data Mining Techniques to Explore Predictors of Grade Retention. ProQuest LLC.
- Koyuncugil, A. S., Özgülbaş, N., 2008. İMKB’de işlem gören Kobi’lerin Güçlü ve Zayıf Yönleri: CHAID Karar Ağacı Uygulaması. Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi. 23(1), s 1-21.
- Kumar, V., Chadha, A., 2011. Yükseköğretimde Veri Madenciliği Teknolojilerinin Uygulamalarına Yönelik Ampirik Bir Çalışma. Uluslararası İleri Bilgisayar Bilimi ve Uygulamaları Dergisi, 2(3): 80-84.
- Kurt, Ç., Erdem, O. A., 2012 Öğrenci Başarısını Etkileyen Faktörlerin Veri Madenciliği Yöntemiyle İncelenmesi. Gazi Üniversitesi, Teknik Eğitim Fakültesi, Bilgisayar Sistemleri Anabilim Dalı, Politeknik Dergisi.5(2), s.111-116.
- Larose, D. T., 2005. Discovering Knowledge in Data: An Introduction to Data Mining, John Wiley and Sons, Inc., pp. 42-70.
- Leong, C. K., Lee, Y. H., Mak, W. K., 2012. Mining Sentiments in SMS Texts for Teaching Evaluation. Expert Systems with Applications, 39(3), 2584-2589.
- Ma, Y., 1998. Data Warehousing, OLAP, And Data Mining: An Integrated Strategy For Use At FAA, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.
- Quinlan, J.R., 1986. Induction of Decision Trees. Journal of Machine Learning, Cilt 1, s 81-106.
- Quinlan J.R., 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann, San Mateo, CA, s 302.
- Özbay, Ö., 2015. Veri Madenciliği Kavramı ve Eğitimde Veri Madenciliği Uygulamaları. Inesjournal Uluslararası Eğitim Bilimleri Dergisi, sayı 5, s 262-272.
- Özekeş, S., Çamurcu, A. Y., 2002. Veri Madenciliğinde Sınıflama ve Kestirim Uygulaması, Marmara Üniversitesi Fen Bilimleri Dergisi, sayı 18, s 1-17.

- Safavian S.R. Landgrebe D., 1991. A Survey of Decision Tree Classifier Methodology,.IEEE Transactions on Systems Man and Cybernetics, 21, 660-674.
- Sevindik, T., Kayışlı, K., Ünlükahraman, O., 2012. Web Tabanlı Eğitimde Veri Madenciliği. Turkish Journal of Computer and Mathematics Education. 3(3), s 183-193.
- Sun, J., Hui L., 2008. Data Mining Method for Listed Companies, Financial Distress Prediction. Knowledge-Based Systems, 21, No. 1.
- Şen, B., Uçar, E., Delen, D., 2012. Orta Öğretim Yerleştirme-Test Puanlarını Öngörme ve Analiz Etme: Bir Veri Madenciliği Yaklaşımı. Uygulamalı Uzman Sistemler, 39(10), 9468-9476.
- Şen, B., Uçar, E., 2012. Uzaktan Öğretim Öğrencilerinin Bilgisayar Mühendisliği Bölümünün Başarılarının Veri Madenciliği Yöntemleri ile Değerlendirilmesi. Prosedür Teknolojisi, 1, 262-267.
- Şengür, D., Tekin, A., 2013. Öğrencilerin Mezuniyet Notlarının Veri Madenciliği Metotları İle Tahmini. Bilişim Teknolojileri Dergisi, 6(3): 7-16.
- Ulusoy, G., 2013. Karar Ağacı Analizi ile AB Genişleme Kriterlerinin Değerlendirilmesi. Marmara Üniversitesi, Sosyal Bilimler Enstitüsü, Ekonometri Anabilim Dalı, İstatistik Bilim Dalı. Yüksek Lisans Tezi.
- Taşdemir, M., 2012. Öğrenci Başarısına Etki Eden Faktörlerin Regresyon Analizi İle Tespiti, Dicle Üniversitesi Sosyal Bilimler Enstitüsü İşletme Anabilim Dalı, Yüksek Lisans Tezi.
- Tissera, W. M. R., Athauda, R. I., Fernando, H. C., 2006. Discovery of Strongly Related Subjects in the Undergraduate Syllabi Using Data Mining. In Information and Automation, s 57-62.
- Tsai, C. F., Tsai, C. T., Hung, C. S., Hwang, P. S., 2011. Mezuniyet için Gerekli bir Bilgisayar Yeterlilik Sınavından Başarısız Olma Riskini Taşıyan Öğrencileri Belirlemek için Veri Madenciliği Teknikleri. Australasian journal of educational technology, 27(3), 481-498.
- Yadav, S. K., Bharadwaj, B., Pal, S., 2012. Mining Education Data to Predict Student's Retention: A comparative Study. arXiv preprint arXiv: 1203-2987.
- Yaralioğlu, K., 2008. www.deu.edu.tr/userweb/k.yaralioğlu/dosyalar/ver_mad.doc, s 5.
- Zhang, Y., Oussena, S., Clark T., Kim, H., 2010. ‘ Use Data Mining to Improve Student Retention in Higher Education: a case study’ , in iceis 2010: Proceedings of the 12th International Conference on Enterprise Information Systems Integration, Insticc, Funchal, Portugal, s 190-197.

Ek 1. QUEST Algoritması Analiz Sonuçları



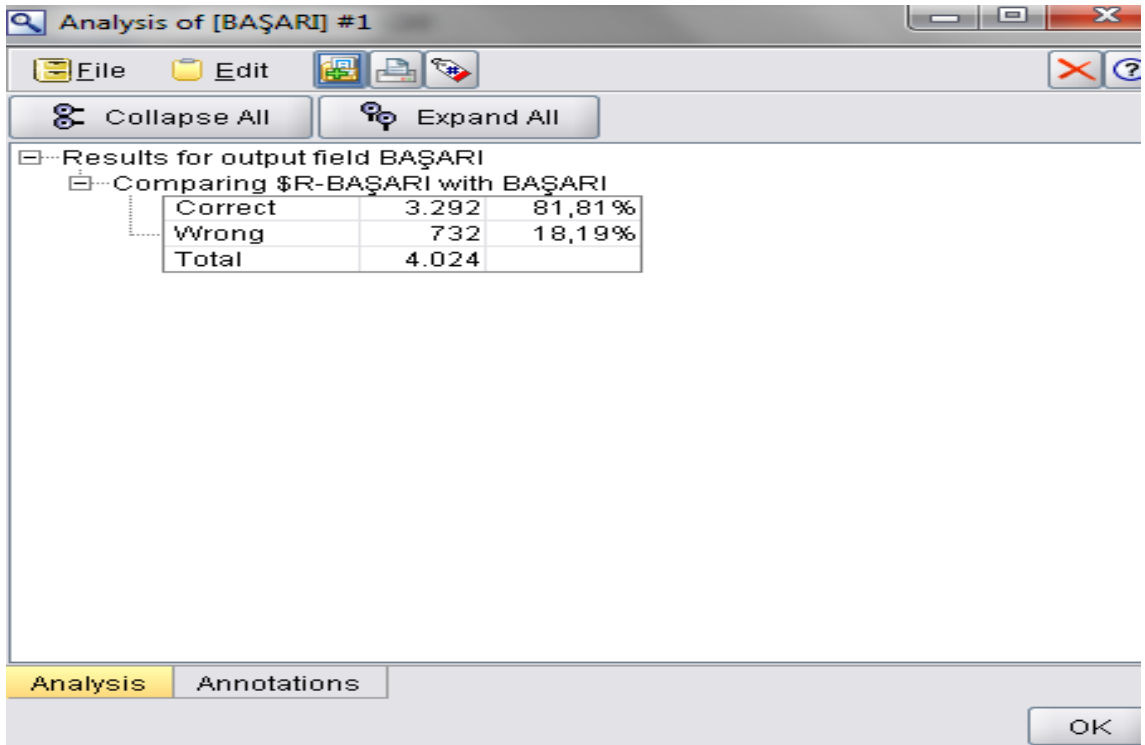
The screenshot shows a window titled "Analysis of [BAŞARI]". It features a menu bar with "File" and "Edit" options, and icons for file operations. Below the menu bar are "Collapse All" and "Expand All" buttons. The main content area displays a tree view with the following structure:

- Results for output field BAŞARI
 - Comparing \$R-BAŞARI with BAŞARI
 - Correct: 3.238 (80,47%)
 - Wrong: 786 (19,53%)
 - Total: 4.024

At the bottom of the window, there are tabs for "Analysis" (selected) and "Annotations", and an "OK" button.

Category	Count	Percentage
Correct	3.238	80,47%
Wrong	786	19,53%
Total	4.024	

Ek 2. C&RT Algoritması Analiz Sonuçları



The screenshot shows a window titled "Analysis of [BAŞARI] #1". It features a menu bar with "File" and "Edit" options, and icons for file operations. Below the menu bar are "Collapse All" and "Expand All" buttons. The main content area displays a tree view with the following structure:

- Results for output field BAŞARI
 - Comparing \$R-BAŞARI with BAŞARI
 - Correct: 3.292 (81,81%)
 - Wrong: 732 (18,19%)
 - Total: 4.024

At the bottom of the window, there are tabs for "Analysis" (selected) and "Annotations", and an "OK" button.

Category	Count	Percentage
Correct	3.292	81,81%
Wrong	732	18,19%
Total	4.024	

Ek 3. CHAID Algoritması Analiz Sonuçları

Analysis of [BAŞARI] #2

File Edit

Collapse All Expand All

Results for output field BAŞARI

Comparing \$R-BAŞARI with BAŞARI

Correct	3.328	82,7%
Wrong	696	17,3%
Total	4.024	

Analysis Annotations

OK

Ek 4. C5.0 Algoritması Analiz Sonuçları

Analysis of [BAŞARI] #5

File Edit

Collapse All Expand All

Results for output field BAŞARI

Comparing \$C-BAŞARI with BAŞARI

Correct	3.631	90,23%
Wrong	393	9,77%
Total	4.024	

Coincidence Matrix for \$C-BAŞARI (rows show actuals)

	0.000000	1.000000
0.000000	1.551	136
1.000000	257	2.080

Confidence Values Report for \$CC-BAŞARI

Range	0,571 - 0,998
Mean Correct	0,913
Mean Incorrect	0,722
Always Correct Above	0,993 (12,85% of cases)
Always Incorrect Below	0,571 (0% of cases)
90,23% Accuracy Above	0,0
2,0 Fold Correct Above	0,957 (74,58% of cases)

EK 5. Anket Çalışması

Her gün en az 1 saat kitap okurum.	
<input type="checkbox"/> EVET	<input type="checkbox"/> HAYIR
Matematik ve sayılarla uğraşmayı severim.	
<input type="checkbox"/> EVET	<input type="checkbox"/> HAYIR
Kendime ait odan var mı?	
<input type="checkbox"/> EVET	<input type="checkbox"/> HAYIR
Geçirmiş olduğun herhangi bir hastalığın var mı?	
<input type="checkbox"/> VAR	<input type="checkbox"/> YOK
Baban çalışıyor mu?	
<input type="checkbox"/> ÇALIŞIYOR	<input type="checkbox"/> ÇALIŞMIYOR
Annen çalışıyor mu?	
<input type="checkbox"/> ÇALIŞIYOR	<input type="checkbox"/> ÇALIŞMIYOR
Her gün en az 2 saat bilgisayarla uğraşırım.	
<input type="checkbox"/> EVET	<input type="checkbox"/> HAYIR
Okul öncesi eğitim aldın mı?	
<input type="checkbox"/> ALDIM	<input type="checkbox"/> ALMADIM
Ailem ile birlikte yaşıyorum.	
<input type="checkbox"/> EVET	<input type="checkbox"/> HAYIR
Her gün en az 2 saat TV izlerim.	
<input type="checkbox"/> EVET	<input type="checkbox"/> HAYIR

Kaç kardeşiniz?

0 1 2 3 4 5 6 7 8 9 10 11 12

Yaşadığınız yer?

İL İLÇE KÖY

Babanızın eğitim durumu nedir?

OKURYAZAR DEĞİL İLKOKUL ORTAOKUL
LİSE ÜNİVERSİTE ÜNİVERSİTE ÜZERİ

Babanızın eğitim durumu nedir?

OKURYAZAR DEĞİL İLKOKUL ORTAOKUL
LİSE ÜNİVERSİTE ÜNİVERSİTE ÜZERİ

ÖZGEÇMİŞ

Kişisel Bilgiler

Adı, soyadı : Çiğdem GÜLER
Uyruğu : T.C.
Doğum tarihi ve yeri : 02.07.1992, KAHRAMANMARAŞ
e-posta : gulercgdm@gmail.com

Eğitim

Derece	Eğitim Birimi	Mezuniyet tarihi
Yüksek lisans	KSÜ /Zootekni Bölümü	2017
Lisans	Giresun Üniversitesi/ Matematik Bölümü	2014
Lise	19 Mayıs Lisesi	2010

İş Deneyimi

Yıl	Yer	Görev
2015-2017	Bağlarbaşı İmam Hatip Ortaokulu	Matematik Öğretmeni

Yabancı Dil

İngilizce

Makale

A. Tatlıyer, E. Yavuz, M. Şahin, A. Ünalın, Ç. Güler, M. Sözeyatarlar, M. Farsak, XII. Ulusal Zootekni Öğrenci Kongresi konferansı dahilinde " XII. Ulusal Öğrenci Kongresi " bildiri kitapçığındaki "Alman Alaca X Kıl Keçisi Melezlerinde Bireysel Aktasyon Eğrilerinin Modellenmesi", 106 ss, Isparta, Türkiye, Mayıs 2016.