

**T.C.
ONDOKUZ MAYIS ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**



PSİKOMETRİK BİR ARAÇ OLARAK VERİ MADENCİLİĞİ

ESMA AKDEMİRCİ

YÜKSEK LİSANS TEZİ

**T.C.
ONDOKUZ MAYIS ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

YÜKSEK LİSANS TEZİ

PSİKOMETRİK BİR ARAÇ OLARAK VERİ MADENCİLİĞİ

ESMA AKDEMİRCİ

İSTATİSTİK ANABİLİM DALI

**SAMSUN
2018**

Her hakkı saklıdır.

TEZ ONAYI

Esma Akdemirci tarafından hazırlanan “Psikometrik Bir Araç Olarak Veri Madenciliği” adlı tez çalışması 11/06/2018 tarihinde aşağıdaki jüri tarafından Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı’nda Yüksek Lisans tezi olarak kabul edilmiştir.

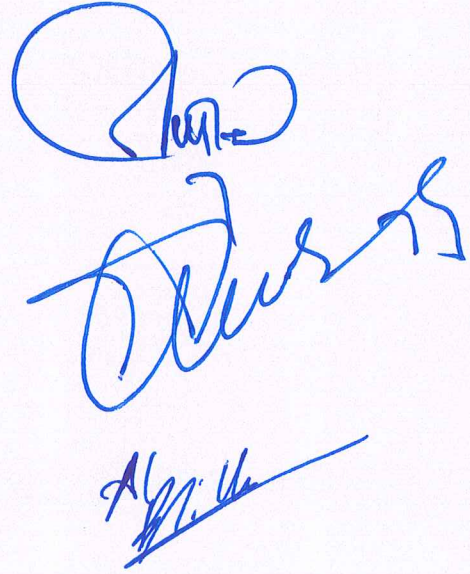
Danışman Doç. Dr. Taner Tunç
İstatistik Anabilim Dalı

Jüri Üyeleri

Başkan Prof. Dr. Bozkurt KOÇ
Ondokuz Mayıs Üniversitesi
Psikoloji Anabilim Dalı

Üye Doç. Dr. Taner TUNÇ
Ondokuz Mayıs Üniversitesi
İstatistik Anabilim Dalı

Üye Doç. Dr. N. Alp ERİLLİ
Sivas Cumhuriyet Üniversitesi
Ekonometri Anabilim Dalı



Yukarıdaki sonucu onaylarım. ... / ... / 20..

Prof. Dr. Bahtiyar ÖZTÜRK
Enstitü Müdürü

TEZ ONAYI

Esma Akdemirci tarafından hazırlanan “Psikometrik Bir Araç Olarak Veri Madenciliği” adlı tez çalışması 11/06/2018 tarihinde aşağıdaki jüri tarafından Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı’nda Yüksek Lisans tezi olarak kabul edilmiştir.

Danışman Doç. Dr. Taner Tunç
İstatistik Anabilim Dalı

Jüri Üyeleri

Başkan Prof. Dr. Bozkurt KOÇ
Ondokuz Mayıs Üniversitesi
Psikoloji Anabilim Dalı

Üye Doç. Dr. Taner TUNÇ
Ondokuz Mayıs Üniversitesi
İstatistik Anabilim Dalı

Üye Doç. Dr. N. Alp ERİLLİ
Sivas Cumhuriyet Üniversitesi
Ekonometri Anabilim Dalı

Yukarıdaki sonucu onaylarım. ... / ... / 20..

Prof. Dr. Bahtiyar ÖZTÜRK
Enstitü Müdürü

ETİK BEYAN

Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırladığım bu tez içindeki bütün bilgilerin doğru ve tam olduğunu, bilgilerin üretilmesi aşamasında bilimsel etiğe uygun davrandığımı, yararlandığım bütün kaynakları atıf yaparak belirttiğimi beyan ederim.

11/06/2018

Esma AKDEMİRCİ

Anneme, Babama ve Nişanlım Musa Atar'a



ÖZET

Yüksek Lisans Tezi

PSİKOMETRİK BİR ARAÇ OLARAK VERİ MADENCİLİĞİ

Esmâ Akdemirci

Ondokuz Mayıs Üniversitesi

Fen Bilimleri Enstitüsü

İstatistik Anabilim Dalı

Danışman: Doç. Dr. Taner TUNÇ

Veri madenciliği, verilerde daha önce bilinmeyen gizli örüntüleri açığa çıkaran ve bu bilgilerden yararlı sonuçlar elde ederek geleceğe yönelik tahmin yapmaya imkân tanıyan dinamik bir süreçtir. Karar Ağacı yöntemi, kolay anlaşılan kurallar üretmesi, kuralların görselleştirilebilmesi ve farklı değişken yapılarına uygun güçlü algoritmalara sahip olması sebebiyle en sık kullanılan Veri Madenciliği tekniklerindendir.

Veri madenciliği, büyük ve karmaşık yapıları veri gruplarından yeni, anlamlı, geçerli, anlaşılır ve potansiyel olarak yararlı verileri çıkarma ve bunları karar destek sürecinde bütünleştirme işlemidir.

Bu çalışmada Ondokuz Mayıs Üniversitesi'nde öğrenim gören öğrencilerin demografik bilgilerini içeren öğrenci bilgi formu, bireyin belirli nitelik ve yetenekleri bakımından kendini nasıl algıladığını belirlemek için Rosenberg Benlik Saygısı Envanteri, bireyin ani değişiklik gösteren heyecansal reaksiyonları değerlendirmede oldukça duyarlı bir araç olan Durumluk Kaygı Ölçeği, kişinin genelde, yaşama eğilimi gösterdiği kaygının sürekliliğini ölçmek için kullanılan Sürekli Kaygı Ölçeği ve problem çözme becerilerini belirlemek amaçlı Problem Çözme Envanteri'nden oluşan bir anket formu uygulanmış, verinin çözümlenmesi için bir veri madenciliği sistemi oluşturmak amaçlanmıştır. Analizde “Rosenberg Benlik Saygısı Envanteri”, “Problem Çözme Envanteri” hedef değişken olarak ele alınmış ve bu değişkenler üzerinde hangi demografik değişkenlerin etkili olduğu belirlenmek istenmiştir.

Çalışmada veri madenciliğinde kullanılan karar ağacı algoritma tekniklerinden CART algoritması ile bütünsel görsel modelleme gereci olan SPSS CLEMENTINE 12.0 paket programı kullanılarak sonuçlar elde edilmiştir.

Haziran 2018, 107 sayfa

Anahtar Kelimeler: Veri madenciliği, Karar ağacı, CART, Rosenberg benlik saygısı envanteri, Durumluluk kaygı ölçeği, Sürekli kaygı ölçeği, Problem çözme envanteri

ABSTRACT

Master's Thesis

DATA MINING AS A PSYCHIOMETRIC VEHICLE

Esma Akdemirci

Ondokuz Mayıs University
Graduate School of Sciences
Department of Statistics

Supervisor: Assoc. Prof. Dr. Taner TUNÇ

Data mining is a dynamic process that opens up previously unknown hidden patterns in data and allows us to make predictions about the future by obtaining useful results from this information. The Decision Tree method is the most commonly used Data Mining technique because it has easy-to-understand rules, visualization of rules, and powerful algorithms for different variable structures.

Data mining is a decision support system that searches for information patterns between data. Another change; valid, clear and potentially useful design from the data, and integrating them in the decision support process. Here the designation specifies the relationship between the data.

In this study, the student information form containing the demographic information of students studying at Ondokuz Mayıs University, the Rosenberg Self-Esteem Inventory to determine how the individual perceives himself in terms of his specific qualities and abilities, the State Anxiety Scale, a tool that is highly sensitive to assessing the individual's suddenly changing emotional reactions, In general, the questionnaire consisted of the Trait Anxiety Scale, which was used to measure the continuity of the lethargic tendency to live, and the Problem Solving Inventory, which was used to determine problem solving skills, and a data mining system was designed to analyze the data. In the analysis, "Rosenberg Self-Esteem Inventory" and "Problem Solving Inventory" were considered as target variables and it is aimed to determine which demographic variables are effective on these variables.

The results were obtained using the SPSS CLEMENTINE 12.0 package program, which is a holistic visual modeling registrant with the CART algorithm from decision tree algorithm techniques used in data mining in the study.

Haziran 2018, 107 pages

Key Words: Data mining, Decision tree, CART, Rosenberg self esteem inventory, Situation anxiety scale, Trait anxiety scale, Problem solving inventory

ÖNSÖZ VE TEŞEKKÜRLER

Tez çalışmam boyunca engin bilgilerini esirgemeyen, verdiği cesaret ve gösterdiği sabırla sadece seminer çalışmam için değil, akademik hayata atılan ve atılacak olan her adımda pay sahibi olan değerli danışman hocam Sayın Doç. Dr. Taner TUNÇ'a sabrı ve hoşgörüsü için teşekkür ederim.

Tez çalışmamın tamamında bana destek veren, benimle birlikte heyecan yaşayıp strese giren Nişanlım Musa ATAR'a ve her zaman yanımda olan ailem'e; eğitim hayatım boyunca her zaman destek verdikleri için ve maddi manevi yardımlarını esirgemedikleri için teşekkür ederim.

Haziran 2018, Samsun

Esma Akdemirci

İÇİNDEKİLER DİZİNİ

ÖZET.....	i
ABSTRACT.....	ii
ÖNSÖZ VE TEŞEKKÜRLER.....	iii
İÇİNDEKİLER.....	iv
SİMGELER VE KISALTMALAR.....	v
ŞEKİLLER DİZİNİ.....	vi
ÇİZELGELER DİZİNİ.....	vii
1.GİRİŞ.....	1
1.1. Literatür Taraması.....	4
1.2. Veri Madenciliğinin Tarihçesi.....	7
2. VERİ MADENCİLİĞİNİN TANIMI.....	9
2.1. Veri Madenciliği Süreci.....	14
2.2. Veri Hazırlama.....	15
2.2.1. Veri temizleme.....	16
2.2.2. Veri bütünleştirme.....	17
2.2.3. Veri indirgeme.....	18
2.2.4. Veri dönüştürme.....	18
2.3. Veri Madenciliği ve İstatistik.....	19
2.4. Veri Madenciliği Yöntemleri.....	22
2.4.1. Sınıflandırma.....	22
2.4.2. Kümeleme.....	23
2.4.3. Tahmin-Öntahmin.....	23
2.4.4. Benzer gruplama.....	23
2.5. Veri Madenciliği Metotları ve Algoritmalar.....	24
2.5.1. Karar ağacı ile sınıflandırma.....	24
2.5.1.1. Karar ağaçlarının elde edilme süreci.....	25
2.5.1.2. Örneklerin kullanımı.....	27
2.5.1.3. Öğrenme süreci.....	28
2.5.1.4. Karar ağaçlarının elde edilmesi.....	29
2.5.1.5. Karar ağacının kullanım alanları.....	32
2.5.1.6. Karar ağacının avantajları ve dezavantajları.....	32
2.5.2. Karar ağacı yapısı.....	33
2.5.2.1. Karar ağacı oluşturma aşamaları.....	35
2.5.2.2. Karar ağaçlarının karşılaştırılması.....	35
2.5.3. Karar ağacı algoritmaları.....	36
2.5.3.1. Hunt algoritması.....	37
2.5.3.2. ID3 algoritması.....	38
2.5.3.2.1. Karar ağacında entropi.....	39
2.5.3.2.2. Bölünme bilgi ve kazanç ölçütü.....	40
2.5.3.2.3. Kazanç oranı.....	41
2.5.3.3. C4.5 algoritması.....	41
2.5.3.3.1. Bilgi kazancı.....	44
2.5.3.3.2. C4.5 algoritmasında budama.....	45
2.5.3.4. Sınıflandırma ve regresyon ağaçları (CART).....	49
2.5.3.4.1. Ağaç yapısı.....	51

2.5.3.4.2. Twoing algoritması	51
2.5.3.4.3. Gini algoritması	52
2.5.3.4.4. Regresyon ağaçları.....	53
2.5.3.6. CHAID algoritması	54
2.5.3.7. Ayrıntılı CHAID	57
2.5.3.8. QUEST algoritması.....	58
2.5.4. CART, CHAID ve QUEST karar ağacı algoritmalarının karşılaştırılması..	58
3. BENLİK SAYGISI, PROBLEM ÇÖZME VE KAYGI İLE İLGİLİ KAVRAMLAR	61
3.1. Benlik	61
3.1.1. Benlik kavramı	61
3.1.2. Benlik saygısı	62
3.1.3. Benlik saygısını etkileyen faktörler.....	63
3.2. Problem ve Problem Çözme.....	63
3.2.1. Problem çözmei etkileyen faktörler	64
3.3. Kaygı	66
3.3.1. Durumluk ve sürekli kaygı kavramları.....	67
3.3.2. Kaygının nedenleri	67
4. UYGULAMA.....	69
4.1. Araştırmanın Amacı	69
4.2. Yöntem	69
4.2.1. Araştırma modeli.....	69
4.2.2. Araştırmanın evreni ve örnekleme.....	70
4.2.3. Veri toplama araçları	75
4.2.2.1. Sosyodemografik anket formu	75
4.2.2.2. Rosenberg benlik saygısı ölçeği	75
4.2.2.3. Problem çözme envanteri	77
4.2.2.4. Surekli-durum kaygı envanteri.....	77
5. BULGULAR VE TARTIŞMA	81
6. SONUÇ VE ÖNERİLER	99
KAYNAKLAR.....	103
EKLER	109
ÖZGEÇMİŞ	

SİMGELER VE KISALTMALAR

CART	Sınıflandırma ve Regresyon Ağaçları
CHAID	Ki-kare otomatik etkileşim tespit
KA	Karar Ağaçları
QUEST	Çabuk Objektif Etkili İstatistik Ağacı



ŞEKİLLER DİZİNİ

Şekil 2.1. Veri madenciliğinin kronolojik görünümü	10
Şekil 2.2. Veri madenciliği süreci.....	15
Şekil 2.3. Veri bütünleştirme	17
Şekil 2.4. Disiplinler arası ilişki.....	20
Şekil 2.5. Sınıflandırma ağaçlarında öğrenme süreci	29
Şekil 2.6. Karar ağacının elde edilmesi.....	30
Şekil 2.7. Karar ağacında yer alan bileşenler.....	30
Şekil 2.8. İkili dallanma için karar ağacı örneği	34
Şekil 2.9. Karar ağacında budama örneği	47
Şekil 5.1. Bağımlı değişken olarak Benlik Saygısı alındığında CART algoritması Clementine ekran çıktısı	91
Şekil 5.2. Bağımlı değişken olarak Benlik Saygısı alındığında CART algoritması Clementine ekran çıktısı.....	94

ÇİZELGELER DİZİNİ

Çizelge 1.1. Veri madenciliğinin tarihçesi	8
Çizelge 2.1. Veri madenciliğinin gelişim zaman çizelgesi	9
Çizelge 2.2. İstatistik ile veri madenciliği arasındaki farklar.....	21
Çizelge 2.3. Cinsiyet, meslek ve yaş niteliklerine ait eğitim kümesi.....	28
Çizelge 2.4. Karar ağacı algoritmalarının karşılaştırılması.....	60
Çizelge 4.1. Çalışmaya katılan öğrencilerin fakültelere göre dağılımı.....	70
Çizelge 4.2. Çalışmaya katılan öğrencilerin demografik özellikleri.....	72
Çizelge 5.1. Benlik saygısı, problem çözme, sürekli ve durum kaygı değişkenlerinin cinsiyet açısından değerlendirilmesi	81
Çizelge 5.2. Benlik saygısı, problem çözme, sürekli ve durum kaygı değişkenlerinin bilgisayara sahip olup-olmama açısından değerlendirilmesi	82
Çizelge 5.3. Benlik saygısı, problem çözme, sürekli ve durum kaygı değişkenlerinin okunan fakülteye göre değerlendirilmesi	83
Çizelge 5.4. Benlik saygısı, problem çözme, sürekli ve durum kaygı değişkenlerinin sınıflara göre değerlendirilmesi.....	85
Çizelge 5.5. Benlik saygısı, problem çözme, sürekli ve durum kaygı değişkenlerinin anne eğitim durumu düzeylerine göre değerlendirilmesi.....	86
Çizelge 5.6. Benlik saygısı, problem çözme, sürekli ve durum kaygı değişkenlerinin öğrencilerin günde kullandıkları internet miktarına göre değerlendirilmesi.....	88
Çizelge 5.7. Benlik saygısı, problem çözme, sürekli ve durum kaygı değişkenlerinin aile ilişki düzeylerine göre değerlendirilmesi	89
Çizelge 5.8. Benlik saygısı, problem çözme, sürekli ve durum kaygı değişkenlerinin günde izlenen TV miktarına göre değerlendirilmesi	90

1.GİRİŞ

Yoğun veri akışı hayatın her alanında kendisini hissettirmektedir. Örneğin; hayatımızın bir parçası olan cep telefonları ile bir günde alınan ve gönderilen metin mesajlarının sayısı dünya nüfusundan fazladır. Amerika Birleşik Devletleri'nde günde 3000'den fazla kitap yayınlanmaktadır. Ayrıca New York Times gazetesinin bir haftalık yayınları içerisinde yer alan bilgi ortalama bir insanın 18. yüzyılda hayatı boyunca karşılaştacağı bilgi miktarından fazla olduğu ileri sürülmektedir. Üretilen teknik bilgi miktarı günümüzde; her iki yılda bir, ikiye katlanmaktadır. Bilgisayar dünyasında 2013 senesine kadar, insan beyninin hesaplama kapasitesinin üzerine çıkabilecek bir süper bilgisayarın oluşturulabileceğinden söz edilmiştir. Bu örneklerden de görülebileceği gibi bilgi ve veri kavramları hem çok güncel, hem de üzerinde titizlikle çalışılması gereken konular olarak ortaya çıkmaktadır.

Tek başına veri; herhangi bir anlamı olmayan ancak işlendiği zaman yeni açılımlara olanak veren bir olgu olarak karşımıza çıkmaktadır (Terlemez, 2008). Akıp giden verilerin kontrollü ve düzenlenerek anlamlı bir hale getirilmesi günümüz teknoloji, eğitim ve sanayi sektörleri için oldukça önemli hale gelmektedir. Verinin anlaşılması ve yığın olarak gelen verinin kontrol altında tutulması önceliklidir.

Veri akışını kontrol etmek günümüz teknolojisinde çok kolaylaşmış olsa da, veri yapılarının etkin bir şekilde analiz edilemeyecek kadar karmaşık olması ve veri akışının süreklilik göstermesi gibi nedenlerle yönetilmesi ve anlamlı kullanılabilir gizli kalmış bilgiler ve ilişkileri ortaya çıkartarak, istenilen yönde kullanılması oldukça zorlaşmaktadır. Bu ve benzeri nedenlerle, günümüzde veri akışının küçük bir kısmı kullanılabilir hale gelebilmektedir.

Tüm organizasyonlar için saniyelerin bile önem taşıdığı günümüz rekabetçi dünyasında bu veri akışından en yüksek düzeyde yararlanmak ve bunun için ortaya çıkan zorlukların üstesinden gelmek ne şekilde olursa olsun; başarı açısından önem kazanmaktadır.

Sürekli akan veri içinde farkı örüntüler olabilir. Bazı örüntüler, örneğin; bir otomobilin yürüyen aksamının gösterdiği arızalar, güneşin her gün doğup batmasındaki örüntüler gibi sayılabilir. Bazı örüntüler ise; hava olayının her zaman aynı şekilde ortaya çıkmaması ve insanların yaygın olan davranışındaki örüntüler

gibi tekrarlı olmayan ancak aynı örüntünün önceki davranışlarından hareketle tanımlanabilen, örüntülerdir. Bu örüntüleri kelimelerle ve sayılarla sembolize etmek, davranışları sembolik nesneler olarak tanımlanmasını mümkün kılar (Terleme, 2008)

Büyük verinin analiz edilmesinde en temel yöntemlerden biri olan veri madenciliği büyük verinin içindeki bilinmeyen, geçerli ve uygulanabilir bilgiyi kaynağından çıkartarak verilerin önceden güvenilirliği kanıtlanmış istatistiksel tekniklerle karar destek sürecine entegre eder. (Doğan ve Arslante, 2016).

Veri madenciliği, veriler arasında ilişki arayan karar destek sistemidir. Bir başka deyişle; veriler arasından yeni, geçerli, anlaşılır ve potansiyel olarak yararlı ilişki çıkarma ve bunları karar destek sürecinde bütünleştirme işlemidir (Gölçe, 2010).

Bilgisayar biliminin yeni bir dalı olan veri madenciliği yapısı ve işleyişi bakımından çok fazla ilgiye sahiptir (Ertuğrul vd, 2013). Bu nedenden dolayı veri madenciliği son yıllarda oldukça dikkat uyandırmıştır. Ticari ve bilimsel keşifler için büyük potansiyeli ile yeni sorunlarla uğraşan, yeni bir teknoloji haline gelmiştir (Koldere, 2008).

Veri madenciliği kendi başına bir çözüm değil çözüme ulaşmak için verilecek karar sürecini destekleyen, problemi çözmek için gerekli bilgileri sağlamaya yarayan bir araçtır (Gölpinar, 2008).

Veri madenciliği, pek çok analiz aracı kullanımıyla veri içerisinde örüntü ve ilişkileri keşfederek, bunları geçerli tahminler yapmak için kullanan bir süreç izler (Kurt ve Erdem, 2012).

Veri madenciliği yöntemleri; sınıflandırma, kümeleme, tahmin-öngörü ve benzer gruplar olarak sıralanmaktadır. Bu yöntemlerden biri olan sınıflandırma istatistiksel veri analizi, örüntü tanıma gibi birçok alanda oldukça sık kullanılmaktadır (Dondurmacı, 2011).

Yaşamış olduğumuz bu zaman diliminde her alışverişte, her bankacılık işleminde, her türlü kamusal alandaki işlemlerde kayıt edilen veriler bulunmaktadır. Ayrıca işletmelerin ve devletin kendi bünyelerinde yapılan işlemler sonucunda da sakladığı veriler, görüntü ve ses cihazlarından elde edilen çoklu ortam verileri gibi birçok veri sürekli olarak depolanmakta ve depolanan bu veriler çok hızlı boyutlarda artmaktadır. Fakat bu veriler istenildiği şekilde değerlendirilememekte, hızla büyüyen bilgi yığınları şekline dönüşmektedir (Özçakır ve Çamurcu, 2007).

Veri madenciliği günümüzde işletmeler için çok önemli hale gelmiştir. Çok

büyük ölçekli veriler, farklı alanlardaki büyük ölçekli veri tabanları içerisinde değerli verileri bulunduran bir veri madeni gibi düşünülebilir. Bu büyüklükteki verilerin analizi, bu analiz sonucunda daha anlamlı bilgi elde etme ve elde edilen bilgiyi yorumlama işi, insan yeteneği ve ilişkisel veri tabanlarının yapabileceklerini aşmaktadır.

Pazarlama stratejilerinin belirlenmesi günümüzde, müşteri profilinin çıkarılması, risk analizi ve yöntemi, kredi taleplerinin belirlenmesi, hisse senedi fiyatı tahmini ve genel piyasa analizlerinde uygulama alanı bulan veri madenciliği, pazarlama, bankacılık, perakendecilik, sigortacılık, telekomünikasyon sektörü gibi pek çok alanda bilgi üretimine katkıda bulunmaktadır. Bu alanlar dışında hava tahminleri, toprak kullanımı ve şehir planlaması gibi uygulamaların yanında, endüstri, sağlık, güvenlik, eğitim ve daha birçok sektörde de uygulama alanı bulabilmektedir.

Veri madenciliği, bankacılık alanında kredi taleplerinin değerlendirilmesi, risk analizi, usulsüzlüklerin tespiti; finansal alanda ise hisse senedi fiyat tahmini, finansal piyasaların analizi, alım satım stratejilerinin iyileştirilmesi gibi alanlarda giderek önem kazanmaktadır. Veri madenciliği, hisse senedi piyasalarında da kullanılmaktadır. Çeşitli yöntemlerle denenmiş bu uygulamalarda, yenileri önerilmiş veya performans sınamaları gerçekleştirilmiştir.

1956 yılından günümüze gelinceye kadar gelişim gösteren ve çevrim içi veri akışı olan alanlarda etkin olan veri madenciliği finans piyasalarının iç yapısını daha iyi anlamak, hisse senedi fiyatlarını temel alarak finansal araçları sınıflamak, yakın gelecekteki fiyat hareketliliklerini önceden tahmin etmek gibi amaçlarla özellikle hisse senetleri piyasalarında giderek artan bir önem kazanmaktadır. Özellikle çevrimiçi veri akışı olan alanlarda daha çok uygulama imkânı bulan veri madenciliğinin, hisse senedi piyasalarında da çeşitli uygulama örnekleri bulunmaktadır (Terlemez, 2008).

Bu çalışmada veri madenciliği konusu ele alınmış ve psikometrik alanda uygulama yapılmıştır. Çalışmanın ilk bölümünde veri madenciliği hakkında kısa bir giriş yapılmış ve veri madenciliğinin tarihsel gelişimi hakkında bilgi verilmeye çalışılmıştır.

Çalışmanın ikinci bölümünde veri madenciliğinin tanımı, veri madenciliği süreci, veri hazırlama, veri madenciliği ve istatistik, veri madenciliğinin temel amacı, veri madenciliği metot ve algoritmaları hakkında bilgi verilmeye çalışılmıştır.

Çalışmanın üçüncü bölümünde benlik, benlik saygısı, problem çözme ve kaygı kavramları hakkında bilgi verilmiştir.

Çalışmanın dördüncü bölümünde ise veri madenciliğinin psikolojide kullanımı üzerine bir çalışma yapılmıştır. Ondokuz Mayıs Üniversitesi'nde öğrenim gören öğrencilere demografik bilgilerini içeren öğrenci bilgi formu, bireyin belirli nitelik ve yetenekleri bakımından kendini nasıl algıladığı ile nasıl olabileceği ya da nasıl olması gerektiği gibi davranışlarını belirlemek için Rosenberg Benlik Saygısı Envanteri, bireyin ani değişiklik gösteren heyecansal reaksiyonları değerlendirmede oldukça duyarlı bir araç olan Durumluk Kaygı Ölçeği, kişinin genelde, yaşama eğilimi gösterdiği kaygının sürekliliğini ölçmek için kullanılan Sürekli Kaygı Ölçeği ve problem çözme becerilerini belirlemek amaçlı Problem Çözme Envanteri'nden oluşan bir anket uygulanmıştır. Sonuçlar CART algoritmasına göre analiz edilmiş ve yorumlanmıştır.

1.1. Literatür Taraması

Veri madenciliği alanında yapılan çalışmalar aşağıdaki gibi sıralanabilir.

Doğan ve Özdamar (2003) ailelerin çocuk isteğine etki eden faktörlere ulaşmada bağımsız değişkenlerin birleşmiş kategorilerini ve alt gruplarını CHAID analizi ile incelemiştir.

Akpınar (2000) veri madenciliği modelleri, veri tabanlarında bilgi keşfi ve karar ağaçları algoritması hakkında bilgi sunmaktadır.

Koyuncugil ve Özgülbaş (2009) yaptıkları araştırmada sağlıkta veri madenciliğinin kullanımı konusunda alt yapı oluşturma ve sağlık profesyonellerine sağlık sektöründe veri madenciliğinin kullanımı ile ilgili karar verme süreçlerine bakış kazandırmaya çalışmıştır.

Koyuncugil ve Özgülbaş (2006) küreselleşme sürecinde önemli yeri olan KOBİ'lerin finansal olarak güçlü ve zayıf yönlerinin belirlemesini amaçlamıştır.

Koyuncugil (2007) borsada sektörel risk tespitine yönelik erken uyarı özelliği sağlayan ve öncü göstergeleri belirleyebilen bir model elde edilmiş ve modelin kullanım ve yorumuna yönelik bir yöntem sunmayı hedeflenmiştir.

Pehlivan (2006) veri madenciliği hakkında kısa bilgi vererek daha sonra karar ağaçlarının genel özelliklerinden, kullanım alanlarından, güçlü yönlerinden bahsetmiştir.

Dolgun vd (2009) yapısal olmayan verinin metin ve web madenciliği yöntemleri ile yapısal hale dönüştürülmesi sonucu modele dahil edilmesinin, model başarısına yapacağı katkıyı analiz edilmiştir.

Kavzoğlu ve Çölkesen (2010) karar ağaçlarının sınıflandırma performansı güncel bir LandsatETM+ uydu görüntüsü kullanılarak detaylı şekilde analiz etmiştir ve elde edilen sonuçların karar ağaçlarının uydu görüntülerinin sınıflandırılmasında etkin bir yöntem olduğunu gösterilmişlerdir..

Baykasoglu (2005) yaptığı çalışmada veri madenciliği tekniği, yöntemleri ve uygulama alanları tartışılarak çimento sektöründe yapılan bir uygulama çalışması tanıtılmıştır.

Tosun (2006) da kredi kartı müşterilerinin kaybedilme sebeplerinin bulunması için veri madenciliğinden faydalanarak sonuçlara ulaşılmış ve böylece müşterinin neden kaybedildiği bilgisinin yanı sıra, hangi tür müşterilerin daha sık kaybedildikleri tahmin edilmeye çalışılmıştır.

Köktürk vd (2009), veri madenciliği metotları, işlevi, tıpta veri madenciliği hakkında bilgi sunmaktadır.

Vahaplar ve İnceoğlu (2001) tarafından yapılan çalışmada, veri madenciliği tarihçesi, günümüzde kullanım alanları ve eksikleri tespit edilmeye çalışılmıştır. Veri madenciliğinin elektronik ticarete kazandıracağı faydalar belirtilerek, veri madenciliği teknikleri ve yöntemleri incelenmiştir.

Bindak'ın (2005) tarafından ilköğretim okulu öğrencilerinin matematik kaygısını belirlemek üzere bir matematik kaygı ölçeği geliştirilmiş ve ölçeğin geçerlik ve güvenirliğine ait bulgular sunulmuştur.

Bölükbaş vd (1998) yapmış oldukları çalışmada durumluluk ve süreklilik kaygısı envanteri kullanılarak ameliyathane ortamının çalışan ekip üzerinde yarattığı anksiyete düzeyinin sosyodemografik özelliklerle ilişkisi araştırılmıştır.

Sarıcaoğlu vd (2005) 12 saatlik gece ve gündüz vardiyaları halinde çalışan anestezi asistanlarında vardiya sonunda dikkat, öğrenme ve bellek gibi bilişsel fonksiyonları ve anksiyete değişimlerini değerlendirmektedir

Bacanlı vd (2009), Beş Faktör kuramından yola çıkılarak, bu faktörlere uygun sıfat çiftleri kullanılarak madde sayısı az, iki uçlu bir ölçeğin geliştirilmesidir. Sıfatlara Dayalı Kişilik Testi (SDKT)'nin faktör yapısını belirlemek için Direct Oblimin döndürme ile temel bileşenler analizi yönteminden yararlanılmıştır

Başaran vd (2009) yaptıkları araştırmada 14 yaş grubu sporcularında sürekli ve durumluluk kaygı düzeylerinin bazı değişkenlere göre incelenmesi hakkında bilgi verilmiştir.

Chakrabarti'nin (2003), web verisinin geniş dünyasından bilgi üretme tekniklerine ayrılan ilk kitaptır. Chakrabarti, özellikle Web madenciliği ile ilgili zorluklarla ilişkili olarak düşük seviyeli makine öğrenme tekniklerini incelemiştir. Daha sonra, kitabın son bölümünü makine öğrenimini sistematik bir şekilde sürdürmek için altyapı ve analizleri bir araya getiren uygulamalara ayırmıştır.

Kusiak vd (2000) tarafından yapılan çalışmada, tıbbi ve mühendislik uygulamalarında sonuçları tahmin etme problemi tartışılmıştır. Sorun bir veri madenciliği yaklaşımı ile çözülmüştür.

Han ve Kamber (2001) yaptıkları araştırmada, veri madenciliği kavramlar ve teknikler, çeşitli uygulamalarda kullanılacak olan toplanmış verileri veya bilgileri işleme konusundaki kavramlar ve teknikler anlatılmaktadır. Özellikle, veri madenciliğini ve toplanan veriden bilgi keşfetmek için kullanılan araçlar açıklanmaktadır. Büyük veri kümelerinin tekniklerinin fizibilitesi, kullanışlılığı, etkililiği ve ölçeklenebilirliği üzerine odaklanmaktadır.

Bidgoli (2004) Web tabanlı eğitim teknolojileri, öğretmenlerin öğrencilerin nasıl öğreneceğini (tanımlayıcı çalışmalar) ve hangi öğrenme stratejilerinin en etkili olduğu (nedensel / tahmini eğitim çalışmaları) hakkında bilgi vermektedir. Web tabanlı eğitim sistemleri, öğrencilerin nitelikleri arasında çok ilginç ilişkiler bulmak ve hem öğretmenler hem de öğrenciler için eğitim sistemlerinin daha doğru olmasını sağlamak için veri madenciliği ve bilgi keşfi tekniklerinden yararlanmıştır.

Quinlan (1986) çeşitli sistemlerde kullanılan karar ağaçlarının sentezine yönelik bir yaklaşımı özetler ve böyle bir sistemi ID3 olarak detaylı olarak açıklar.

Bounsaythip ve Rinta-Runsala (2001) raporunda müşteri ilişkileri yönetimi (CRM), özellikle de müşteri segmentasyonu ve müşteri profili oluşturma ile karar vermeyi desteklemek için veri madenciliği araçlarının nasıl kullanıldığı sorunlarını inceler. İlk önce, müşterilerin davranışlarını tahmin etme yolundaki adımları, veri toplama ve hazırlama, bölümlendirme ve profil modelleme gibi işlemleri hakkında bilgi verir ve daha sonra, küme bulma, karar ağaçları, sinir ağları, ilişki kuralı ve ardışıl (senkron) model dahil en çok kullanılan veri madenciliği yöntemleri hakkında bilgi sunar.

Giudici (2003), verilerin organizasyonu, özet istatistikleri, model şartnamesi, model değerlendirmesi, web sitesi ziyaretçilerini tanımlama, pazar sepeti analizi, müşteri memnuniyetinin tanımlanması, küçük kredi riskinin öngörülmesi, e-öğrenme öğrenci performansının öngörülmesi, müşteri yaşam boyu değeri öngörme, operasyonel risk yönetimi hakkında bilgi vermektedir.

Mehta vd (1996) yeni bir karar ağacı tabanlı sınıflandırma algoritması hakkında bilgi sunmaktadır. Bu algoritma, birçok işlemcinin çalışmasına olanak tanıyarak kolayca paralel hale getirilecek şekilde tasarlanmıştır.

1.1. Veri Madenciliğinin Tarihçesi

İnsanoğlu geçmişten günümüze gelinceye kadar sürekli verileri yorumlayıp bilgi edinmeye çalışmış ve bunun için çeşitli donanımlar oluşturmuştur. Bu donanımlar bilginin taşınmasına olanak sağlamaktadır. Zamanla bütün alanlarda bilgi toplamaya başlamış ve kronolojik olarak gelişimi Çizelge 1.1’ de özetlenmiştir.

Çizelge 1. 1. Veri madenciliğinin tarihçesi (Aldana, 2000)

Gelişim adımları	Cevaplanan karar problemi	Kullanabilen teknolojiler	Ürün sayıcıları	Karakteristikler
Veri toplama (1960'lar)	"Benim toplam karım geçen 5 yılda ne kadardı?"	Bilgisayarlar, teypler, diskler	IBM, CDC	Geriye dönük, statik veri dağılımı
Veri erişimi (1980'ler)	"İngiltere'de geçen Mart ayında birim satışları ne kadardı?"	İlişkisel Veritabanları, SQL, ODBC	Oracle, Sybase, Informix, IBM, Microsoft	Kayıt düzeyinde, geriye dönük dinamik veri dağılımı
Veri ambarlama ve karar destek sistemleri (1990'lar)	"İngiltere'de geçen mart birim satışları ne kadardı?"	OLAP, Çok Boyutlu Veritabanı Sistemleri, Veri ambarları	Pilot, Comshare, Arbor, Cognos, Microstrategy	Çoklu düzeylerde, geriye dönük dinamik veri dağılımı
Veri madenciliği (günümüz)	"Gelecek ay Boston'daki birim satışlar muhtemelen ne olabilir, niçin?"	İleri düzeyde algoritmalar, çok işlemcili bilgisayarlar, büyük veritabanları	Pilot, Lockheed, IBM, SGI, SPSS, SAS, Microsoft vs.	Geleceğe dönük, proaktif enformasyon dağıtımı

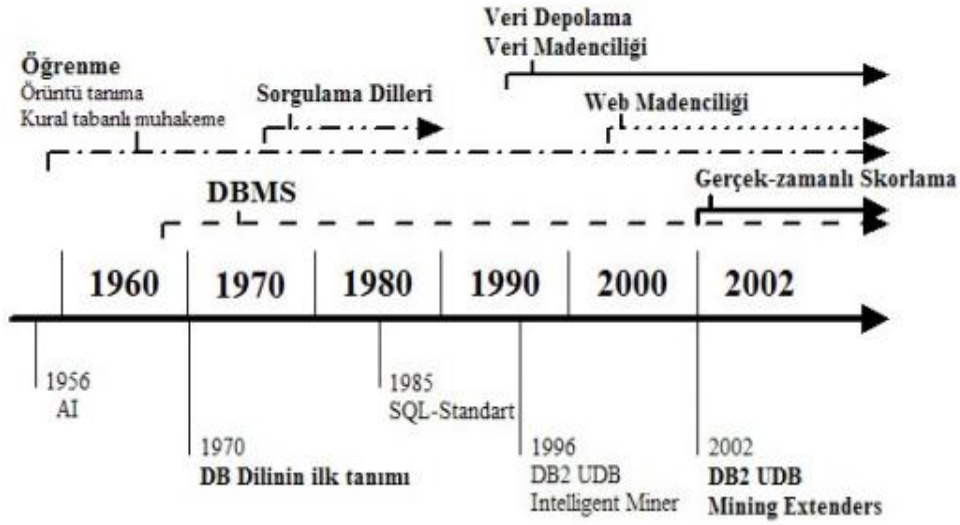
Aldana (2000), veri madenciliği evrelerini dört evrede incelemiş ve veri madenciliğini; işletmelerin, kurumsal verilerini optimize etmesini sağlayan bir yol olarak tanımlamıştır.

2. VERİ MADENCİLİĞİNİN TANIMI

Veri Madenciliği daha önceden bilinmeyen, geçerli ve uygulamaya konulabilir bilgiyi elde etmek ve bu hedefe ulaşmaya yardımcı olacak veriyi veri kaynağından derleyerek ortaya çıkartır, analiz eder ve karar destek sürecine uyarlar. Hedefin ortaya koyulabilmesi için Çizelge 2.1.’deki gelişimi gösteren istatistiksel ve matematiksel tekniklerle, örüntü tanımlama, yapay zekâ, veri tabanı gibi teknolojileri bir araya getirerek Şekil 2.1.’de verilen kronolojik görünümü kazanan veri madenciliği büyük miktardaki statik veya dinamik veri içinde yeni ilişkiler, eğilimler ve örüntüler bulma sürecinde ilerler (Terlemez, 2008).

Çizelge 2. 1. Veri madenciliğinin gelişim zaman çizelgesi (Terlemez, 2008)

Zaman	Alan	Katkı
1920’li yılların başı	İstatistik	Maksimum benzerlik tahmini
1940’li yılların başı	Yapay Zekâ	Sinir ağları
1950’li yılların başı		En yakın komşuluk
		Tek bağlantı
1950’li yılların sonu	Yapay Zekâ	Perseptron
	İstatistik	Yeniden örnekleme, yanlılık indirgeme, jackknife tahminci
1960’li yılların başı	Veri Tabanı	Yığın raporlama
1960’li yılların ortası		Karar ağaçları
	İstatistik	Sınıflandırma içim doğrusal modeller
	Bilgi Erişimi	Benzerlik ölçüleri
	Bilgi Erişimi	Kümeleme
	İstatistik	Açıklayıcı veri analizi
1960’li yılların sonu		İlişkisel veri modelleri
1970’li yılların başı	Bilgi Erişimi	Akıllı bilgi erişimi sistemleri
1970’li yılların ortası	Yapay Zekâ	Genetik algoritmalar
1970’li yılların sonu	İstatistik	Eksik veri ile tahmin(EM algoritmaları)
	İstatistik	K- ortalama kümeleme
1980’li yılların başı	Yapay Zekâ	Kohonen kendine örgütleyen eşleme
1980’li yılların ortası	Yapay Zekâ	Karar ağacı algoritmaları
1990’lı yılların başı	Veri Tabanı	Birliktelik kuralı algoritmaları
		Web ve arama metotları
1990’lı yıllar	Veri Tabanı	Veri ambarlama
	Veri Tabanı	Çevrimiçi analitik işleme (OLAP)



Şekil 2.1. Veri madenciliğinin kronolojik görünümü

Veri madenciliği kavramı birçok teorisyen tarafından ele alınmıştır. Bu teorisyenler geldikleri bilim kollarının etkisi ile değişik veri madenciliği tanımları ortaya koymuşlardır. Temel olarak bir tek veri madenciliği kavramı olmasına rağmen farklı bilim dallarının, problemlerinin çözümüne farklı öneriler getirmesinin sonucu burada da kendisini göstermiştir. Farklı veri madenciliği tanımları aşağıda sırasıyla verilmiştir (Koldere, 2008).

Fieldman ve Sanger (2006) veri madenciliğini; işletme performansını artırmak amacıyla büyük veri tabanlarından keşfedilmemiş bilgiyi elde etmek için kullanılan bir işletme zekâsı olarak tanımlamışlardır.

Silahtaroglu (2013) veri madenciliğini istatistik ve matematik yöntemler ile veri yığınlarını eleayarak elde edilen yeni ve anlamlı örüntüler, ilişkiler ve eğilimler belirlenmesi işlemi olarak tanımlamaktadır.

Jacobs'un (1999) veri Adenciliğini, ham verinin tek başına sunamadığı bilgiyi çıkaran veri analizi süreci olarak tanımlamıştır.

Hand (1998) veri madenciliğini; veri tabanı teknolojisi, makine öğrenme, örüntü tanıma, istatistik ile etkileşimli yeni bir disiplin ve geniş veri tabanlarından daha önceden tahmin edilemeyen ilişkilerin ikincil analizi olarak tanımlanmıştır.

Doğan ve Türkoğlu (2007) veri madenciliğini; bilgisayar programları kullanılarak büyük veri yığınları arasından gelecekle ilgili tahminde bulunmamızı sağlayacak bağlantıların aranması işlemi olarak tanımlamıştır.

Veri madenciliğinin iki adet birincil hedefi olduğu söylenebilir. Bunlar sırası ile kestirim ve tanımdır. Kestirimi amaç edinen, yönlendirilmiş veri madenciliği (directed), kestirimci veri madenciliği (predictive) veya kontrollü veri madenciliği (supervised); hedef değişkenler ile bağımsız değişkenler arasındaki ilişkileri tanımlayarak, hedef değişkenlerin bilinmeyen veya ilerideki değişkenlerini kestirerek, belli bir soruyu cevaplayacak şekilde düzenlenmesidir. Yönlendirilmiş veri madenciliği sonucunda, elde edilmiş olan veri seti ile betimlenen sistemin modeli elde edilmiş olur. Başka bir şekilde açıklamak gerekirse, araştırmacı neyi aradığını bildiğinde uygulanan teknik, yukarıdan aşağı doğru ilerleyen bir veri madenciliği türüdür. Tanımlamayı amaç edinen yönlendirilmiş veri madenciliği (undirected), tanımlayıcı veri madenciliği veya kontrolsüz veri madenciliği (unsupervised) ise, elde edilebilir verilere dayanarak, günümüze kadar bilinmeyen, araştırmacı tarafından yorumlanabilecek bilginin tanımlanabilmesi umuduyla veri üzerinde uygulaması test edilen veri madenciliği türüdür. Farklı bir söylem ile bir bilgi ortaya çıkarsa bu bilginin önem arz edip etmediği araştırmacıya bırakılarak, bilgi çıkarması yapan aşağıdan yukarı doğru ilerleyen veri madenciliği türüdür. Yönlendirilmemiş veri madenciliği ve yönlendirilmiş veri madenciliği birbirlerinden bağımsız iki ayrı başlık değildir. Temel veri madenciliği bir bileşimini içermeye gayret eder. Çünkü kestirimci bir model kurarken bile yönlendirilmemiş teknikleri kullanarak yardımcı olabilmektedir (Terlemez, 2008).

Veri madenciliği, istatistik veya başka alanlarda farklı isimler ile birlikte birçok bilim dalında kullanılmış ve bilgisayar teknolojisindeki hızlı gelişmelerden sonra, veri arkeolojisi, veri taranması, veri balıkçılığı, öz bilgi çıkarımı veya veri tabanlarında öz bilgi keşfi gibi farklı isimler ile de anılmıştır. Bu isimler içerisinde en garip olanı veri balıkçılığıdır. Bu isimle anılmasının sebebi ise, veri madenciliği verilerin belirli bir örüntü bulunması beklentisiyle taranması deniz dibinin taranarak kum ve çamurdan arındırıldıktan sonra yararlı deniz ürünlerinin elde edilmesine benzetilmesidir.

Veri madenciliği; büyük sayıda verinin (gigabyte ve hatta terabytelarca veri) incelendiği veritabanlarındaki ilişkileri, eğilimleri ve ilginç örüntüleri keşfetmede ve veri ambarlarında oluşan açıklanmamış bilginin çıkarımında kullanılan bir süreçtir. Veri madenciliği uygulamalarında aşağıdaki soruların cevaplanması gerekmektedir (Koldere, 2008);

- Analiz edilecek veri seti veri madenciliği için uygun mu? Veri setinin işlem yapılabilir formatta olması gerekmektedir. Genellikle farklı kaynaklarda ve farklı formatlarda bulunan veri dosyaları, bir araya getirilerek analiz edilebilmektedir.
- Analiz edilecek olan veri seti istenen bütün faktörleri kapsıyor mu? Veri setinin istenen bütün faktörleri kapsıyor olması tahminleri doğruluk derecesinin oranını arttırmaktadır.
- Analiz edilecek veri setinin büyüklüğü yeterli mi? Çoğu zaman problem, veri setinin küçüklüğünden değil çok büyük olan veri setinin temsil gücünün zayıf olmasından kaynaklanmaktadır. Eğer veri setinin temsil gücü yüksek ise, birkaç bin gözlem veri madenciliği için yeterli olmaktadır.

Veri Madenciliğine Neden İhtiyaç Duyulmuştur?

Bilgisayar sistemleri sayesinde sigorta, bankacılık ve borsa gibi birçok alandaki bilgiler günümüzde çeşitli veri tabanlarında kaydedilmektedir. Kaydedilen bütün bu verilere bir de uydulardan elde edilen veriler eklendiğinde, ulaşılan veri hacminin inanılmaz boyutlara vardığı bir gerçektir. Veri madenciliği, eldeki yapısız veriden, anlamlı ve kullanışlı bilgiyi ortaya çıkarmaya yönelik çalışmalarının bütünü olmuştur. Sonuç olarak yıllar ilerledikçe ortaya çıkan veri yığınlarına bir düzen verme, farklı bir anlamda potansiyel olarak kullanışlı bilgi haline getirme amacıyla veri madenciliği ortaya çıkmıştır. Veri madenciliğini büyük hacimde veri bulunan yerde kullanmak mümkündür. Günümüzde veri madenciliği uygulamaları karar verme sürecine ihtiyaç duyulan bir çok yerde yaygın olarak kullanılmaktadır (Koldere, 2008).

Günümüzde veri madenciliğinin başlıca uygulama alanları aşağıdaki gibi sayılabilir: Pazarlama alanında

- Müşteri segmentasyonu
- Müşterilerin demografik özellikleri arasındaki bağlantıların kurulması
- Çeşitli pazarlama kampanyaları
- Mevcut müşterilerin elde tutulması
- Yeni müşterilerin kazanılması
- Pazar sepeti analizi
- Çapraz satış analizleri ve satış tahminleri

- Müşteri değerlendirme ve müşteri ilişkileri yönetimi

Bankacılık alanında

- Farklı finansal göstergeler arasındaki gizli ilişkilerin bulunması
- Kredi kartı dolandırıcılıklarının tespiti
- Kredi taleplerinin değerlendirilmesi
- Usulsüzlük tespiti
- Risk analizleri

Sigortacılık alanında

- Yeni poliçe talep edecek müşterilerin tahmin edilmesi
- Sigorta dolandırıcılıklarının tespiti
- Riskli müşteri tipinin belirlenmesi

Perakendecilik alanında

- Satış noktası veri analizleri
- Alış-veriş sepeti analizleri
- Tedarik ve mağaza yerleşim optimizasyonu

Borsa alanında

- Hisse senedi fiyat tahmini
- Genel piyasa analizleri
- Alım-satım stratejilerinin optimizasyonu

Telekomünikasyon alanında

- Kalite ve iyileştirme analizleri
- Abonelik tespitleri
- Hataların yoğunluk tahminleri

Sağlık ve ilaç alanında

- Test sonuçlarının tahmini
- Ürün geliştirme
- Tıbbi teşhis
- Tedavi sürecinin belirlenmesi

Endüstri alanında

- Kalite kontrol analizleri
- Lojistik
- Üretim süreçlerinin optimizasyonu

2.1. Veri Madenciliği Süreci

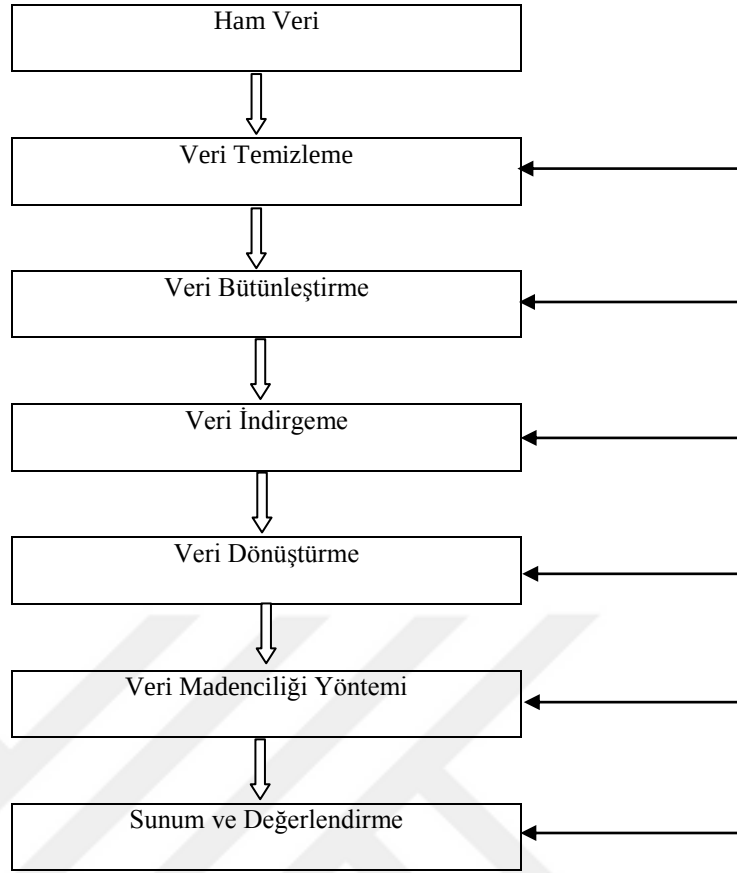
Veri hazırlama ile ilgili işlemleri veri madenciliği kavramı içinde düşünürsek, veri madenciliğinin bir süreç olarak değerlendirilmesi gerekir. Soyut kazılar yaparak veri yığını arasından veriyi çıkarmak, bilgi keşif sürecinde örüntüleri ayrıştırarak süzmek ve bir sonraki adıma hazır hale getirmek bu sürecin bir parçasıdır (Savaş vd, 2012).

Veri madenciliği büyük boyutlu verilerden yararlı, işe yarar ve anlamlı bilginin elde edilmesinde gerekli olan işlemlerin modellenmesi için yöntemler ve algoritmalar sunar (Gülçe, 2010).

Veri madenciliği süreci aşağıda belirtilen adımları içermektedir:

- Veri temizleme
- Veri bütünleştirme
- Veri indirgeme
- Veri dönüştürme
- Veri madenciliği algoritmasını uygulama
- Sonuçları sunum ve değerlendirme

Veri madenciliği sürecinin aşamaları Şekil 2.2.'de ifade edilmiştir. Burada bulunan ham veri, henüz herhangi bir işleme tabi tutulmamış (ön işlem öncesi) veridir. Ham veri sırasıyla temizleme, bütünleştirme, dönüştürme ve indirgeme gibi işlemlerden geçirildikten sonra, veri madenciliği algoritması uygulanarak sonuçlara ulaşılmaktadır.



Şekil 2.2. Veri madenciliği süreci

Veri madenciliği verisi bir veri ambarından geliyorsa, bu işlemlere gerek kalmayabilir. Bu tarz işlemler veri ambarı hazırlanırken yerine getirilir. Veri madenciliği için veri ambarı kullanılmıyorsa, veri temizleme, bütünleştirme indirgeme ve dönüştürme aşamalarının ayrıca yerine getirilmesi gerekmektedir (Dondurmacı, 2011).

2.2. Veri Hazırlama

Veritabanı içerisinde yer alan verinin ve bu verinin sonuçları olarak elde edilen veri madenciliği sonuçlarının kalitesinin artırılması, veriyi analize hazırlarken dikkat edilmesi gereken en önemli unsurdur. Veri madenciliği işlemlerini kolaylaştırmak ve verimliliği arttırmak için veriler, bir “ön işleme” aşamasından geçirilir. Verinin hazırlanması süreci olarak da kabul edilen bu işlemler öncelikli olarak veri tabanındaki bozuk değerleri düzeltmeyi ve veriler arasındaki tutarsızlıkları kaldırmayı amaçlamaktadır (Dondurmacı, 2011).

2.2.1. Veri temizleme

Veri temizleme, farklı alanlar da kullanılacak olan veri madenciliği modelinde veri kalitesini artırmak üzere yapılır (Gülçe, 2010). Genel olarak veri temizleme yöntemi, uyumsuzluk tespiti ve verilerdeki tutarsızlıkları düzeltmeyi amaçlamaktadır (Gülpınar, 2008).

Veri temizleme aşaması, eksik değerleri tanımlamak, aykırı değerleri belirleyerek verideki tutarsızlıkları gidermek ve gürültüyü azaltmak için kullanılan birçok tekniği kapsamaktadır.

Veri tabanındaki kayıp verilerin telafi edilebilmesi için bazı yollara başvurulur. Kayıp veri sorununun nasıl halledileceği konusuna geçmeden önce kayıp veri sorununu yaratan nedenler üzerinde durulmasında yarar vardır.

Kayıp veri birçok nedenden ötürü oluşmaktadır. Örneğin; müşteri bilgileri her zaman mevcut olmayabilir veya bu bilgilere ulaşamayabilir. Bazı veriler önemsiz görüldüğü için veri girişi sürecinde göz ardı edilmiş olabilir. Bazı veriler ise kayıtlı verilerle uyumlu olmadıkları düşünülerek silinmiş olabilir. Veri kümesini kayıp veri değerlerinden arındırmak için kullanılan yöntemlerden en önemlileri ve yaygın biçimde kullanılanları şunlardır (Dondurmacı, 2011):

- Kayıp verilerin yer aldığı kayıtlar tümüyle yok edilir. Bu şekilde yok edilen sorunlu kayıtların sayıları fazla ise bu yöntem kullanışlı olmaz.
- Kayıp değerler için genel bir sabit terim kullanılır. Bu yöntemde “BİLİNMIYOR” veya “\$” gibi sabit bir değer tanımlanabilir. Veri madenciliği programları bu veri değerini ortak bir değer gibi algılayabilir. Bu nedenle çok tercih edilen bir yaklaşım olmamıştır.
- Sayısal değerlere sahip nitelikler için hesaplanacak olan ortalama değer kayıp değerlerin yerine kullanılabilir.
- Kayıp değer yerine aynı sınıfa ait tüm örnekler için değişkenin ortalaması kullanılabilir.
- Kayıp değer, mevcut verilere dayanarak en uygun değer kullanılarak tanımlanabilir. En uygun değer belirlenebilmesi için regresyon ve karar ağacı gibi teknikler kullanılabilir. Örneğin veri setindeki diğer müşteri verilerini kullanarak gelir ile ilgili kayıp değerleri öngörmek için bir karar ağacı oluşturmak gibi.

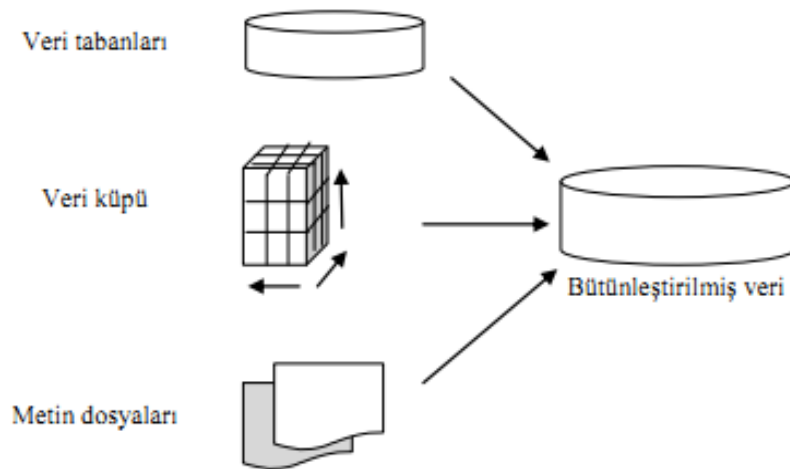
2.2.2. Veri bütünleştirme

Farklı kaynaklardan alınan verileri birlikte analiz edebilmek üzere tutarlı bir veri deposunda bir araya getirme işlemi veri bütünleştirme olarak adlandırılır (Gülpınar, 2008).

Veri madenciliği uygulamalarının kullanıldığı veri ambarının oluşturulması esnasında değişik kaynaklardan (veri küpleri, metin dosyaları, Veritabanı vb.) elde edilen veriler arasında uyum sağlanabilmesi için verilerin bütünleştirilmesi işleminin yerine getirilmesi beklenir. Bu adım, bütün veri kaynakları karakterlerini, özelliklerini ve toplama seviyelerini eşleştirerek tutarlılık sağlayarak bütünleştirilir.

Farklı kaynaklarda bulunan verilerin bütünleştirilmesi, örneğin bir kaynaktan gelen müşteri numarasının başka bir kaynaktan gelen müşteri numarası ile aynı tanıma sahip olması amacıyla yapılır. Veri tabanları ve veri ambarlarında veri türlerinin tanımlandığı meta data bu sorunu çözmekte yardımcı olur. Farklı bir örnekte de, bir bankanın farklı veritabanları içerisinde yer alan kredi kartları ile ilgili hesaplara sahip olması üzerine verilebilir. Bu hesaplarda müşteri isimleri farklı biçimlerde kaydedilmiş olabilir. Özetle ifade etmek gerekirse veri bütünleştirilmesi veri kalitesi ile ilgili sorunların çözülmesi için kullanılan bir yöntemdir.

Veri bütünleştirme aşamaları Şekil 2.3.'de gösterilmiştir. Farklı veri tabanlarındaki veri küpü ve kurumun ürettiği metin dosyaları veri bütünleştirme işlemlerinden geçerek veri madenciliği için hazır hale getirilir. Veri küpleri çok boyutlu birer veritabanıdır (Dondurmacı, 2011).



Şekil 2.3. Veri bütünleştirme

2.2.3. Veri indirgeme

Bilindiği üzere veri madenciliği uygulamalarının bazılarında milyonlarca kayıt ve binlerce nitelik ile karşılaşmak mümkündür. Bu durumda verinin temel özelliklerini kaybetmeden, bir kısmını atarak azaltma yoluna gidip gidilmeyeceği sorunu ile karşılaşılır. Bu sorunları azaltmanın yolu olan birleştirme işleminde önemli noktalardan biri de veri indirgemedir. Veri indirgeme; veriyi birleştirme, boyut indirgeme, örnekleme ve genelleme biçiminde yapılabilmektedir (Gülpınar, 2008).

Değişkenlerde meydana gelen tutarsızlıklar elde edilen veri setlerinde fazlalıklara neden olabilmektedir. Bu fazlalıkların tespiti için korelasyon analizinden yararlanılmaktadır. Örneğin korelasyon analizi ile müşteri numarası ile müşteri kimliği arasında yüksek düzeyde bir ilişki bulunursa değişkenlerden biri veri deposundan çıkarılarak indirgeme yapılabilir. Veri madenciliği yöntemleri hesaplama ağırlıklı olduğundan, veri indirgemesiyle doğal olarak maliyetlerin azalması sağlanabilmektedir. Veri indirgeme işlemleri sonucunda aşağıda belirtilen yararlar elde edilir (Dondurmacı, 2011):

- Öğrenme (model) maliyetleri azalır.
- Öğrenme (model) performansı artar.
- İlgisiz boyut sayısı azalır.
- Gereksiz boyut sayısının azalması

2.2.4. Veri dönüştürme

Veri madenciliğinde verileri aynen işleme katmak kurulan sistem için bazen uygun olmayabilir. Bazı değişkenlerin ortalaması ve varyansları diğer değişkenlerden farklı olabilir. Bu varyansların çok büyük veya çok küçük olması durumunda bu büyük fark yaratan değişkenlerin diğerleri üzerinde analiz aşamasında etkisi daha çok olur ve onların rolleri önemli ölçüde azalır. Ayrıca bu çok büyük ve çok küçük değerlerde çözümlenmenin sağlıklı bir şekilde yapılmasını engeller. Bu durumda sağlıklı çözümleme yapılabilmesi için verilerin standartlaştırılması ya da normalleştirilmesi gerekir. Normalleştirme işlemi verileri 0-1 aralığında sayısal değerlere dönüştürme işlemi iken, standartlaştırma işlemi ise verilerin ortalaması ve standart hatası göz önüne alınarak yeni değerlere dönüştürülmesi işlemidir (Gülpınar, 2008). Verinin standartlaştırılması için min-max modelleştirme veya Z-score standartlaştırma

yöntemleri kullanılabilir. Ya da veri madenciliği modeline göre değişkenlerin aralık değerlerini değiştirip gruplama yapılabilir (Güçle, 2010).

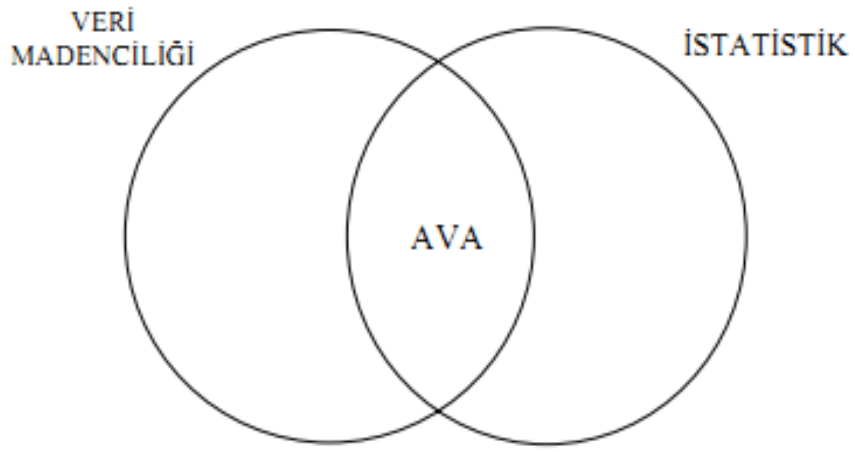
Veri dönüştürme de önemli olan diğer nokta ise ölçekleme ve kodlamadaki farklılıklardır. Örneğin uzunluk bir sistemde metre ile kaydedilmişken başka bir sistemde inç ile kaydedilmiş olabilir. Verilerin heterojen bir yapı oluşturması veri bütünlüğü açısından büyük tehlikeler oluşturabilmektedir (Dondurmacı, 2011).

2.3. Veri Madenciliği ve İstatistik

İstatistik ve veri madenciliği arasında önemli farklılıklar göze çarpsa da pek çok noktada bu ikisi örtüşmektedir. Bu farklılıklardan birincisi kullanılan veri tiplerinden kaynaklanmaktadır. İstatistik geleneksel şekilde, birinci elden elde edilen veri ile çalışırken veri madenciliği ise çoğu kez farklı kaynaklardan toplanan ikinci elden veri ile çalışmaktadır. Başka bir fark ise, veri madenciliğinde toplanan bilgilerin saklanması için kullanılan geniş veritabanında yatan önemli bilgi parçalarının önceden bilinemeyeceği ve bu nedenden ötürü hipotezlerin formüle edilmesinde kullanılamayacağıdır. Bir diğer fark ise, istatistiğin ana kütle hakkındaki çıkarsamalarını, ana kütleden çekilen temsili örnekten yapmasıdır (Terlemez, 2008).

Veri madenciliğinin istatistikçiler tarafından değil ticari kullanıcılar tarafından kullanılması veri madenciliği ile istatistik arasındaki temel farklardandır. Veri madenciliği istatistikteki birçok metottan yararlanmaktadır (Gülpınar, 2008).

Veri madenciliği ve istatistik arasındaki en büyük fark, veri madenciliği tarafından araştırılan veritabanlarının tablo, grafik ve diğer insan gücü görevlerle çalışmak gibi geleneksel veri analizi yaklaşımlarını imkânsız hale getirmesidir. Veri madenciliği bilgisayar gücü görevlere, istatistik düzensizlikleri bulan ve potansiyel olağanüstü ilginçlikleri bulmayı amaçlayan algoritmalara başvurulur. Yani veri madenciliği ikinci elden, çok büyük ve düşük kalite veriye uygulanmış oldukça otomatikleşmiş bir açıklayıcı veri analizidir. İstatistik ile veri madenciliği arasındaki farklar Çizelge 2.2.'de gösterilmiştir. Veri madenciliğinde bu durum tartışma konusu olabilmekte, üzerinde çalışılan veri tabanı örneklendiğinde, küçük fakat aşırı derecede önemli bilginin gözden kaçırılabilceği söylenebilmektedir. Veri madenciliği ile istatistik arasındaki ilişki olan açıklayıcı veri analizi (AVA) Şekil 2.4'de gösterilmiştir (Terlemez, 2008).



Şekil 2.4. Disiplinler arası ilişki

Çizelge 2.2. İstatistik ile veri madenciliği arasındaki farklar (Terlemez, 2008)

İSTATİSTİK		VERİ MADENCİLİĞİ
Orta derecede küçük n=onlar veya yüzler p=kullanışlı , nadiren birkaç onluktan fazla	Veri Hacmi	Geniş n=yüzlerce binlik- milyonluk p=binlerce- milyonlarca
- Veri bir modeli test etmek için veya belirli soruları cevaplamak için toplanır veya bir araya konulur(birinci elden veri). - Durum- kontrol çalışmaları - Fikir anket araştırmaları - Gözlemsel çalışmalar	Veri Tipi	- Veri elektronik olarak elde edilir ve gelecekte olası kullanım için depoda tutulur(ikinci elden veri). - Süper market işlemleri verisi - Telefon aramaları verisi - Hava kontrol verisi - Hisse senedi Pazar verisi - İnternet trafiği verisi
İnsan elinden izlek “Veri bilgisayar yardımıyla insan tarafından işlenir”.	Veri İşleme	Oldukça otomatikleşmiş izlek “veri insan yardımıyla bilgisayar algoritmaları tarafından işlenir”.
- Model uydurma - Model test etme - Güven ve kestirim aralıkları	Araştırma Amaçları	- Örüntü arama ve tanımlama - Sınıflandırma - gruplandırma
- Daha iyi istatistiksel ana görüş geliştirme - Metotların istatistiksel ve matematiksel özelliklerini çalışma		- Veri madenciliği görevlerini yerine getirmek için daha iyi/ hızlı algoritmalar geliştirme - Madencilik algoritmalarının deneysel performansını çalışma

Çizelge 2.2.'de gösterilen istatistik ve veri madenciliği arasındaki farklar, veri hacmi, veri tipi, veri işleme, yerine getirilen görevler ve araştırma amaçlarından kaynaklanmaktadır.

2.4. Veri Madenciliği yöntemleri

Veri madenciliği istatistiksel tabanlı olmak üzere çok sayıda yöntem ve algoritma içermektedir. Veri madenciliği yöntemleri farklı kaynaklarda farklı şekilde sınıflandırılmıştır. Tez kapsamında veri madenciliği 4 grupta toplanmıştır (Gülpınar, 2008).

- Sınıflandırma
- Kümeleme
- Tahmin-Öntahmin
- Benzer gruplama

2.4.1. Sınıflandırma

Veri madenciliğinin en çok kullanıldığı alan sınıflandırmadır. Veri madenciliğinde sıkça kullanılan yöntemlerden biri olan sınıflandırma, veri tabanındaki gizli örüntüleri ortaya çıkarmakta kullanılır (Gülçe, 2010). Yeni bir nesnenin özelliklerini açıklamak ve bu yeni nesnenin daha önceden tanımlanmış sınıf setlerinden birine atamasını yapmak sınıflandırma alanının amacıdır. Sınıflandırma yapabilmek için veriye uygun bir model uygulanmaktadır. Veri madenciliğinin sınıflandırma alanında en çok kullanılan teknik karar ağacıdır. Bunun dışında lojistik regresyon, diskriminant analizi, fuzzy setleri ve sinir ağları kullanılmaktadır. İnsanlar verileri sınıflandırdıkları, derecelendirdikleri ve kategorize ettikleri için sınıflandırma, veri madenciliğinin temeli ve hazırlama aracı olarak kullanılabilir (Koldere, 2008). Nesnelerin ait olduğu grupların bilinip bilinmemesine göre sınıflandırma eğitici ve eğitici-siz sınıflandırma olmak üzere ikiye ayrılır. “i” sınıf sayısı her bir grup örneğin hangi sınıfa ait olduğu bilinirse eğitici sınıflandırma eğer hangi nesnenin hangi sınıfa ait olduğu ve grup sayısı belirsiz ise eğitici-siz sınıflandırma kullanılır (Gülpınar, 2008).

2.4.2. Kümeleme

Kümeleme, bir nesne kümesini gruptama yani nesnelerin kendilerini ya da diğer nesnelerle olan ilişkilerini tarif eden bilgileri kullanarak nesneleri gruplara ayırma işlemidir (Gülçe, 2010).

Kümelerin amacı heterojen yapıda olan ana kütleli homojen gruplara veya kümelerle ayırmaktır. Kümeleme önceden tanımlanmış sınıfları temel alarak gruptama işlemi yapmaz, gruptamada kümeleme benzerlileri kullanılırken, sınıflandırma önceden tanımlanmış sınıflar modelini temel almaktadır. Kümeleme işlemi, veri madenciliği sürecinden önce veri hazırlama esnasında da kullanılabilir. K-ortalamlar ve toplayıcı algoritma teknikleri ile onların varyasyonları kümeleme yöntemleri olarak kullanılmaktadır (Koldere, 2008).

2.4.3. Tahmin-ön tahmin

Tahmin, bir olayın veya durumun sonucu hakkında bazı bilgi ve deneyimlere dayanılarak yapılan kestirmeleri ifade eder. Ön tahmin ise beş duyu organımız kullanılmadan, hafıza, muhakeme veya tahmin gibi zihinsel işlemler sonucunda elde edilmemiş henüz gerçekleşmemiş bir olayın önceden algılanmasıdır.

Tahmin ve ön tahmin birbirlerine benzeyen süreçlerdir. Sonucunda genellikle bir sayısal değer vermektedir. Verilmiş olan bir grup veri girdileri ile tahmin ve ön tahmin gelir, ihtimal ve yaşam süresi gibi bilinmeyen değişkenler için ortaya bir değer koymaktadır. İstatistikteki regresyon analizi, pazar sepet analizi, sinir ağları ve sınıflandırma da söz konusu olan bazı teknikler burada da uygulanmaktadır (Koldere, 2008).

2.4.4. Benzer gruptama

Amacı, birlikte olabilecek ya da gerçekleşebilecek olaylar hakkında karar vermektir. Benzer gruptamaya örnek olarak, süpermarketlerde yapılan alışverişlerde ne ile neyin alınacağına karar vermek gösterilebilir. Benzer gruptamaya ‘Pazar Sepet Analizi’ de denilmektedir. Pazar sepet analizi birliktelik kuralları ile aynı amaç için kullanılmaktadır. Bunun sebebi ise her iki analiz yapıldığında ikisinde de sonuçların aynı çıkmasıdır (Koldere, 2008).

2.5. Veri Madenciliği Metotları ve Algoritmalar

Farklı görevleri yerine getirmek için birçok farklı algoritmaya ihtiyaç duyan veri madenciliği, bu algoritmaların tümü ile veriyi inceler ve verinin karakteristiğine uygun en yakın olan modeli belirler.

Veri madenciliğinde İstatistiksel Metotlar, Karar Ağacı, Birliktelik Kuralları, Yapay Sinir Ağları, Genetik Algoritma, Bulanık Kümeleme ve Bulanık Mantık metotları kullanılmaktadır (Terlemez, 2008):

- İstatistiksel Metotlar

- Regresyon analizi
- Varyans analizi
- Lojistik Regresyon
- Log-Lineer Modeller
- Doğrusal Ayırma Analizi
- Kümeleme Analizi

-Karar Ağaçları

- ID3 Algoritması
- C4.5 Algoritması
- C5.0 Algoritması
- Sınıflandırma Ve Regresyon Ağaçları (CART)
- Ölçeklendirilebilir Karar Ağacı Teknikleri

-Birliktelik Kuralları

- Büyük Öğe Setleri(LargeItemsets)
- Apriori Algoritması
- Örneklem Algoritması
- Bölümlendirme

- Yapay Sinir Ağları

- Genetik Algoritma

- Bulanık Kümeleme ve Bulanık Mantık

2.5.1 Karar ağacı ile sınıflandırma

Karar ağacı ile sınıflandırma yöntemi, sınıflama teknikleri içerisinde en sık kullanılan, ilgi çekici ve anlaşılması kolay bir yöntemdir (Karaibrahimoğlu, 2014).

Karar ağaçları tekniği, basit kural çıkartması, görsel sonuç sunması ve yorumlanmasının kolay olması nedeni ile son yıllarda yaygın olarak uygulanan veri madenciliği tekniklerinden biridir (Dondurmacı, 2011).

Karar ağaçları (KA), bağımlı değişken üzerindeki farklılıkların maksimize edilmesi için veri setinin sıralı olarak bölünmesini ifade eder. Verileri belirli değişken değerlerine göre sınıflandırmaya yarayan KA' da kullanılan algoritmalarda girdi ve çıktılar verilerin belirlenen değişkenleridir. KA algoritması çıktı veri değişkenleri için girdi ve değişkenlerinin veri yapılarını keşfeder.

KA, verilerden sınıflandırıcılar üretmek için kullanılan etkili yöntemlerden birisidir. KA sunumu, en çok kullanılan mantık yöntemidir. Makine öğreniminde ve uygulamalı istatistik literatürün de tanımlanan çok fazla sayıda tümevarım KA algoritması vardır. Bu algoritmalar, bir girdi-çıktı kümesinden KA oluşturan denetlenmiş öğrenme yöntemidir. Bir KA öğrenme sistemi, araştırma alanının bir kısmında çözüm arayan, yukarıdan aşağıya yöntemini benimsemektedir. Bu yöntem ile basit bir ağacın bulunabileceğini garanti eder. KA, değişkenlerin test edildiği yerlerde düğümler içermektedir. Düğümden dışarı açılan dallar, düğümdeki testin bütün olası sonuçlarına karşılık verir (Gülpınar, 2008).

2.5.1.1 Karar ağaçlarının elde edilme süreci

Karar ağaçları, sınıfları bilinen örnek veriden tümevarım yöntemiyle öğrenilen ağaç şekilli bir karar yapısı çeşididir. Bir karar ağacı, basit karar verme adımları uygulanarak, büyük miktardaki kayıtları, çok küçük kayıt gruplarına bölerek kullanılan bir yapıdır. Her başarılı bölme işlemiyle, sonuç gruplarının üyeleri bir diğeriyle çok daha benzer hale gelmektedir.

Karar ağaçları, değişkenleri parçalayarak bir ağaç oluşturmaya dayanmaktadır. Karar ağaçlarında, tanımlanan sorunların cevapları gruplara ayrılmaktadırlar. Sorulara verilecek ölçütler belirlendikten sonra kümeler arasındaki riski maksimize edecek şekilde cevaplar bölünmektedir. En iyi bölünmeyi bulabilmek için her soruda bu işlemin tekrar edilmesi gerekmektedir. Bir soru için grup oluşturulup gruplar arasındaki riski maksimize ettikten sonra oluşan iki grup için aynı işlemler devam ettirilir. Bu işlemler, istatistiksel olarak anlamlı bir fark ortaya çıkana kadar devam ettirilip istatistiksel olarak anlamlı farkın oluşmadığı durumlarda ise son vermektedir. Ayırıştırma işlemi tamamlandıktan sonra ise o grup içerisinde yer alan gözlemlerin oranına göre grup değerlendirilmektedir.

Yüksek hacimli verileri kullanarak bunlara uygun bir karar ağacının elde edilmesi amacıyla birçok yöntem geliştirilmiştir. Bunlardan biri de regresyon ağaçlarıdır. Bağımsız değişkenler ile bağımlı değişken arasındaki ilişkiyi inceleyen regresyon analizi, doğrusallık, normallik, homojenlik, toplanabilirlik gibi varsayımların yerine getirilmesinden sonra uygulanabilen bir istatistiksel yöntemdir. Varsayımların gerçekleştirilememesi durumunda, bilinen bir grup dönüştürme işlemine tabi tutulacak olan veri kümesi, uygun hale getirilmeye çalışılır. Bu da veri setindeki orijinal değerlerin dönüşümü ya karekök ya da logaritmik dönüşüm yöntemi ile yapılabilir. Parametrik yöntemlerde, dönüştürme metotları kullanılarak ön koşul varsayımlar yerine getirilmeye çalışılsa da yapılan analiz dâhilinde veri setine ilişkin yanlış sonuçlar elde edilmesi söz konusu olabilmektedir. Alternatifin olmadığı durumlarda verileri dönüştürme yoluna gidilmesi istatistikî açıdan daha doğru bulunmaktadır. Bahsedilen avantajlı yönleriyle çoklu regresyon analizine alternatif sayılabilecek ve çoklu regresyon ağacının kullanılması doğru olur.

Regresyon ağacı ve sınıflama, sadece bağımsız değişken ile bağımlı değişken arasındaki ilişkinin yapısını açıklamakla kalmaz; aynı zamanda bağımsız değişkenlerin birbirleri ile olan etkileşimlerini de ortaya koymaya çalışır. Regresyon ağacı ve sınıflama analizinin kullanıldığı algoritma, bağımlı değişkenle bağımsız değişkenlerin ilişkisini değerlendirmede ve model içindeki etkileşim yapısının çözümlenmesinde önemli avantajları bulunmaktadır. Regresyon ağacı ve sınıflandırmanın sahip oldukları algoritma, benzerlik gösteren değişkenlerin aynı ağaç düğümünde toplanmasıyla alakalı olup, bütün oluşturduğu alt dalları (sub-branches) bağımlı değişken olan düğüme bağlamasıyla sonuca ulaşmaktadır.

Kullanılacak veri çeşidine göre parametrik veya parametrik olmayan yöntemler grubunda yer alma özelliğine sahip olan CART yönteminde, bağımsız değişkenlerin sürekli veya kesikli olması önem arz etmediği gibi, parametrik testlerin önem arz ettiği normallik, homojenlik ve doğrusallık gibi varsayımlar dikkate alınmamaktadır. Araştırmacılara sağladığı bu avantajla birlikte elde edilen sonuçların diyagram halindeki sunumu, tabloyla gösterimin ötesinde diğer araştırmacılar için açıklayıcıdır (Dondurmacı, 2011).

2.5.1.2. Örneklerin kullanımı

Sınıflandırma ağaçlarının oluşturulabilmesi için girdi verilerine, yani örneklerle ihtiyaç vardır. Örnekler kategorik veya sayısal niteliklerden oluşabilir. Bunlar dışında bir adet hedef niteliğe ihtiyaç duyulur. Hedef nitelik çıktı değerlerini içermektedir.

Sınıflandırma sürecinde burada kullanılacak verinin seçilmesi gerekmektedir. Temizlenerek kullanıma hazır hale getirilen bütün veri ağacın eğitilmesinde kullanılmamaktadır. Aksi takdirde elde edilecek ağacın doğruluk oranını test etme olanağı yoktur. Bundan dolayı, ilgili veri kümesi ikiye bölünür. Bu kümelerden birincisi (eğitim kümesi) ile ağaç oluşturulurken, ikincisi (test kümesi) de ağacın test edilmesinde kullanılır. Böylelikle karar verme mekanizmasının ne kadar güvenileceği bilgisine ulaşılmış olur. Bu ikiye bölme yöntemi dışında başka yöntemlerin araştırılmasının sebebi ise, her veri kümesinin bölünecek kadar büyük olmamasından kaynaklanmaktadır. Bu tarz durumlarda karar ağacını test etmede önemli sorunlar ortaya çıkmaktadır.

Büyük çaplı veri kümesinin alt kümesi olduğu kabul edilen Çizelge 2.3.'teki veri, eğitim kümesi olarak kabul edilebilir. Bu küme, kasko sigortası yaptıran bireylerin cinsiyetleri, meslekleri ve yaşları ile olan ilişkinin karar ağacı yoluyla çözümlemek için kullanılacaktır.

Sekiz adet gözlemden oluşan Çizelge 2.3.'teki bu veri kümesinde “Cinsiyet”, “Meslek” ve “Yaş” birer kategorik nitelik olarak yer almaktadır. Hedef ise, kategorik olan “KASKO isimli niteliktir. Kasko niteliğini kaskoyu yaptıran bireyler için “Evet”, yaptırmayan bireyler için “Hayır” olmak üzere iki farklı değeri vardır (Dondurmacı, 2011).

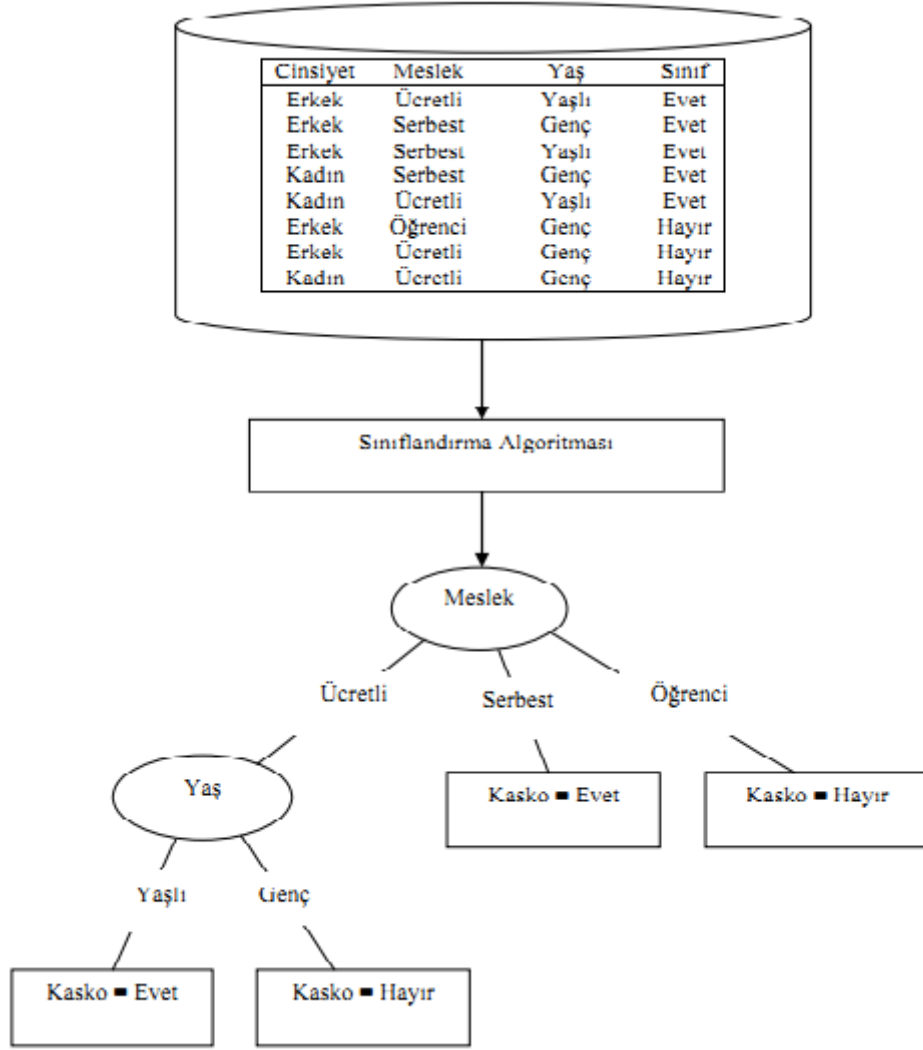
Çizelge 2.3. Cinsiyet, meslek ve yaş niteliklerine ait eğitim kümesi

Cinsiyet	Meslek	Yaş	Kasko
Erkek	Ücretli	Yaşlı	Evet
Erkek	Serbest	Genç	Evet
Erkek	Serbest	Yaşlı	Evet
Kadın	Serbest	Genç	Evet
Kadın	Ücretli	Yaşlı	Evet
Erkek	Öğrenci	Genç	Hayır
Erkek	Ücretli	Genç	Hayır
Kadın	Ücretli	Genç	Hayır

2.5.1.3. Öğrenme Süreci

Eğitim için ayrılan veri kümeleri ile önce sınıflandırma ağacı ve onun bir sonucu olarak karar kuralları üretilir. Karar ağacı eğitim verilerinden hareketle bir öğrenme olayı sonucunda oluşturulmaktadır.

Karar ağaçlarının oluşturulması için çok sayıda veri madenciliği algoritmaları (CHAID, C4.5, C5.0, ID3, QUEST, vb.) geliştirilmiştir. Karar ağacı oluşturulurken kullanılan algoritmanın hangisi olduğu önemlidir. Kullanılan algoritmaya göre ağacın tipi değişebilir. Değişik ağaç yapıları farklı sınıflandırma sonuçları oluşturacaktır. Kök adı verilen ilk düğümü oluşturan değişkenin farklı olması en uçta bulunan yaprağa ulaşırken izlenecek yolu dolayısıyla da sınıflandırmayı değiştirecektir. Gerek kök düğümünün gerekse de ondan sonraki her düğümün belirlenmesinde en büyük ölçüt o noktadan dallara ayrıldığında veritabanındaki verilerin kalan kısmı belli eşit parçalara bölünmüştür. Örneğin veritabanında bulunan alanın cevabı evet/hayır gibiyse iki eşit parçaya, evet/hayır/belki gibi üç değişkenliyse mümkün olduğu kadar üç eşit parçaya bölünmesi istenmektedir. Buradaki amaç, en kısa yoldan istenilen sonuca ya da sınıfa ulaşmaktır. Kullanılan bu algoritmalar yardımıyla bir karar ağacı oluşturulduğunda öğrenme olayı da gerçekleşmiş olur. Şekil 2.5 'te bu süreç gösterilmektedir.



Şekil 2.5. Sınıflandırma ağaçlarında öğrenme süreci

Karar kuralları, karar ağacının kökten başlayarak yapraklara doğru yorumlama yapılarak elde edilir. Elde edilen karar kuralları karar verme sürecinde kullanılmaktadır. Şekil 2.5.'de gösterildiği gibi, veri kümesine bir sınıflandırma algoritması uygulanarak karar ağacı elde edilmektedir. Eğitilmiş olan karar ağacı kullanılarak yeni bir örnek hakkında karar verilebilmektedir (Dondurmacı, 2011).

2.5.1.4 Karar ağaçlarının elde edilmesi

Karar verme sürecindeki seçenekler, olaylar ve bunlara bağlı elde edilecek olan sonuçların grafik biçiminde gösterilmesi sonucunda ortaya çıkan grafiğe “karar ağacı” denir. Veri madenciliği sınıflandırma uygulamalarında karar ağaçları yaygın

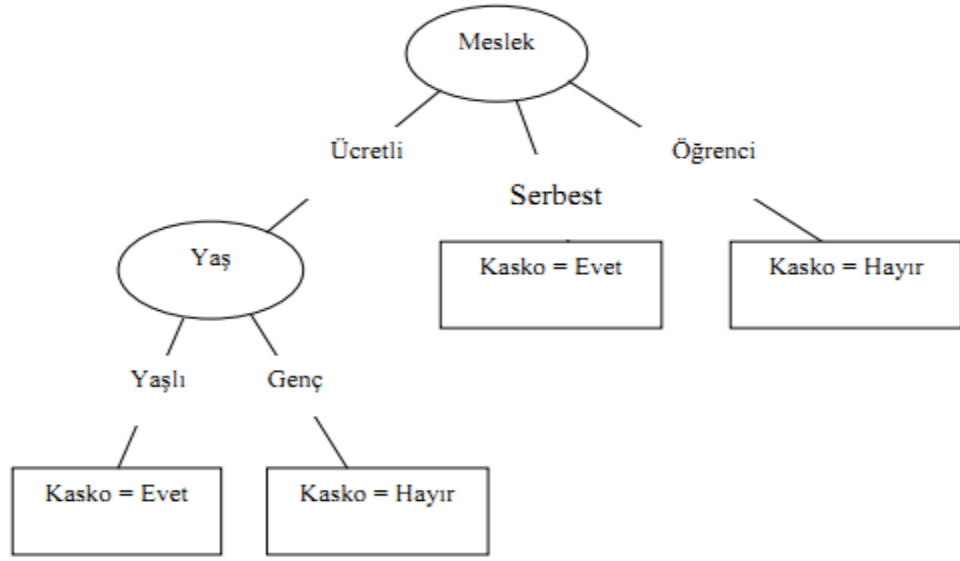
biçimde kullanılmaktadır. Adından da anlaşılacağı gibi, karar ağaçları kökten, dallardan ve yapraklardan oluşan bir ağaç görünümündedir.

Karar ağaçları yöntemi ile sınıflandırma süreci kök düğümden başlayıp, ağacın her bir düğümü bir karar ağacı niteliği tanımlar. Düğümlerden niteliğin her bir değerine göre dallara ayrılır. Bu süreç yapraklar elde edilinceye kadar devam eder. Karar ağaçları bilginin en iyi gösterim biçiminden biri olmasına rağmen bazı durumlarda karmaşık ve yorumlanması oldukça zor olan bir yapıya sahip olabilmektedir. Karar ağacında sadece bir kök düğümü bulunur. Dallar ve yapraklar için bir sayı sınırlaması yoktur.

Şekil 2.6.'daki karar ağacı Çizelge 2.3.'de verilen örneklerle dayalı olarak oluşturulmuştur. Karar ağacının kök düğümü bu örnekte “Meslek” olarak seçilmiştir. Bu seçimler rastgele yapılamaz. Uygun bir veri madenciliği algoritması ile seçim işlemleri yürütülür. Kök düğümden sonra söz konusu niteliğin düzeyleri olan “Ücretli”, “Serbest” ve “Öğrenci” dalları ayrılmaktadır.

Şekil 2.6.'da yer alan karar ağacında, kök düğüme bağlanan her bir dal için yeniden bir seçim işlemi uygulanarak uygun olan düğüm elde edilmektedir. Örneğin şekildeki karar ağacı üzerinde “Ücretli” dalına bağlı olarak “Yaş” düğümü elde edilmiştir. Aynı şekilde “Yaş” düğümünden itibaren bu niteliğin düzeyleri olan “Yaşlı” ve “Genç” dallarına ayrılmaktadır. Sonuçta dikdörtgen biçiminde ifade edilen yapraklara ulaşılmaktadır. Örnek veride yer alan “Cinsiyet” niteliği karar ağacında yer almamıştır.

Yapraklar karar sınıflarını belirlemektedir. Düğümler ise test işlemine tabi tutulan nitelikleri göstermektedir (Quinlan, 1993). Karar ağacının elde dilme süreci Şekil 2.6' da gösterilmektedir (Dondurmacı, 2011).



Şekil 2.6. Karar ağacının elde edilmesi

Karar ağaçlarında yer alan bazı bileşenlerin ne anlama geldiği Şekil 2.7.’de açıklanmıştır.

Bileşen	İşlevi
	Düğüm Yaprak Nitelik değerine göre bir dal

Şekil 2.7. Karar ağacında yer alan bileşenler

2.5.1.5. Karar ağacının kullanım alanları

Karar ağacı, uygulamalarının ucuz olması, yorumlanmalarının kolay olması, veri tabanı sistemleri ile kolayca bütünleştirilebildikleri ve güvenilirliklerinin iyi olması nedenleri ile veri madenciliğindeki sınıflama modelleri içerisinde en yaygın kullanıma sahip yapıdır. Genel olarak kullanım alanları şu şekilde sıralanabilir (Gülpınar, 2008):

- Belirli bir sınıfın muhtemel üyesi olacak elemanların belirlenmesi,
- Çeşitli durumların yüksek, orta, düşük risk grupları gibi kategorilere ayrılması,
- Gelecekteki olayların tahmin edilebilmesi için kurallar oluşturulması,
- Sadece belirli alt gruplara özgü olan ilişkilerin tanımlanması,
- Kategorilerin birleştirilmesi ve sürekli değişkenlerin kesikliye dönüştürülmesi
- Parametrik modellerin kurulmasında kullanılmak üzere çok miktardaki değişken ve veri kümesinden faydalı olacakların seçilmesi

2.5.1.6. Karar ağacının avantajları ve dezavantajları

Sınıflandırma işlemini yapmak için KA kullanmak analizciye pek çok avantaj sağlamaktadır. KA basit olduğundan çıkan sonuçların anlaşılması ve yorumlanması diğer yöntemlere göre çok kolaydır (Gülpınar, 2008).

Karar ağaçlarının kullanımı kolay ve etkin olması sınıflandırma açısından birçok avantaj sağlamaktadır. Karar ağaçları kolay yorumlanabilir ve anlaşılabilir kurallar oluşturabilir. Ağacın boyutu veri tabanından bağımsız olduğu için büyük veri tabanları için ölçeklenebilirler. Birçok özelliğe sahip veriler için ağaçlar oluşturulabilir. Özellikle değerlerinin dağılımı hakkındaki ve özelliklerin bağımsızlığı hakkındaki varsayımlardan bağımsızdır (Terlemez, 2008).

KA yönteminin diğer tekniklere göre önemli bir üstünlüğü sinir ağlarına karşıt olarak kuralları temsil etmesidir. Kurallar SQL gibi doğrudan bir veri tabanı dilinde özel kategoridekileri kayıtların geri çağırılabilmesi için kullanıldığı gibi insanların anlayabileceği şekilde de ifade edilebilmektedir. Kararları bulunan düğümleri tanımlamanın hem operatör hem de sezgisel olarak kolayca anlaşılabilmesi de KA yöntemini diğer sınıflandırma yöntemlerinden ayıran temel özelliklerdendir.

KA'nın avantajlarının yanı sıra dezavantajları da mevcuttur. Karar vermek genellikle KA yaklaşımı için ciddi bir sorun oluşturmaktadır. Sorunun kaynaklandığı

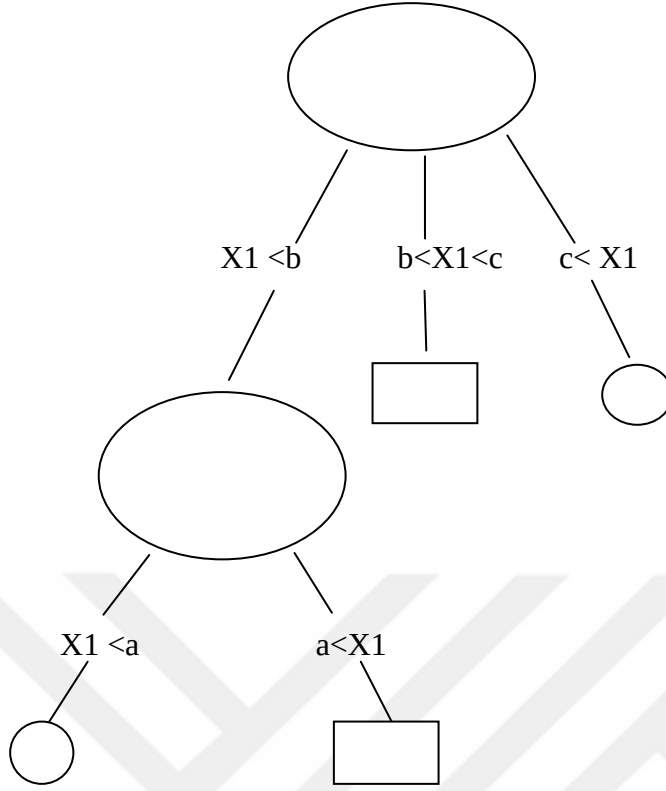
nokta, ağaç, ayrılan bölümleriyle genişledikçe, yapılan sınıflandırma sonucunda bölüm düğümlerinde daha az bilgi kalacaktır. KA veriyi çok küçük parçalara böler. Bu takımlar daha özel bir hal aldıkça küçülmeye başlar. İncelenmesi gereken durumların sayısı fazlalaştıkça, eğitim kümelerinin her biri daha da küçüktür. Rakamlardaki bu azalma nedeniyle sınıflandırmanın doğru bir şekilde gösterilmesinde daha az güvenilirlik söz konusu olmaktadır. KA çok sayıda küçük dallardan oluşturulursa bu düğümlerin içerisinde herhangi doğru olabilecek istatistiksel incelemelerden başarı ile çıkabilecek kuralların bulunması iyimser bir olasılıktır. Genellikle bu dallardan çıkacak olan her bir düğüm, bütün sınıfların küçük oranlarını içermektedir. Bu uygulamalarda sorunlara yol açmaktadır (Gülpınar, 2008).

Çok fazla değişkenin yer aldığı büyük hacimli veri gruplarında, karar ağaçlarının elde edilmesi sınıflama yöntemlerine göre daha fazla hesaplama maliyeti gerektirebilir (Dondurmacı, 2011).

KA farklı dallarda belli özellikler açısından veri kümelerini birbirleriyle karşılaştırılabilmektedir. Bu nedenle birden fazla değişkenin söz konusu olduğu durumlarda değişkenler arasındaki ilişkiyi saptama konusunda yetersizlik göstermektedir. KA modeli sayısal değerleri kategorik olarak değerlendirmektedir. Bunun dışında kategorilerde yer alan veri sayısında azalma oldukça KA'nın hatalı sonuç verme olasılığı artmaktadır. Verinin karmaşıklığı arttıkça KA'da oluşturulacak dalların sayısı artmakta ve bu sayede KA' nı yönetmek zorlaşmaktadır (Gülpınar, 2008).

2.5.2. Karar ağacı yapısı

Ağaç üç tür düğüme sahiptir. Birinci düğüm, giriş olan bir kök düğüm; ikinci düğüm, her biri tamamen bir giriş kenarına bağlı ve iki veya daha fazla çıkış kenarına sahip olan iç düğümler, üçüncü düğüm ise, her biri tamamen bir giriş kenarına sahip olan ve hiç çıkış kenarına sahip olmayan yaprak veya uç düğümlerdir. Şekil 2.8'de ikili dallanma için karar ağacı örneği verilmiştir.



Şekil 2.8. İkili dallanma için karar ağacı örneği

Kök düğüm ve iç düğümler sorunları ifade etmektedir. Dallar bu soruların cevaplarını temsil etmektedir. Yapraklar ise kararın verildiği sınıfı temsil etmektedir. Ağacın ilk düğümünden başlanarak sorular sorulmaya başlanır ve bu işlem yapraklara ulaşıncaya kadar devam eder. Yaprak adedi kadar kural koşulu olur.

Düğümlerin yapıları bakımından *KA* 'lar üçe ayrılır.

- 1) Tek değişkenli *KA* 'lar
- 2) Çok değişkenli *KA* 'lar
- 3) Melez *KA* 'lar

Tek değişkenli yapıya sahip *KA* 'lar da düğümlere sorulan soruların ilgili olayın tek bir değişkene bakılarak yanıtlanmasıyla oluşur. Bu durum aslında uzayı dikine bölme işlemidir. Çok değişkenli yapıya sahip *KA* 'lar da ise, düğümlerdeki sorular birden fazla değişkendin oluşturduğu uzayda ifade edilmektedir. Melez yapıya sahip *KA* 'lar da ise bu iki test biçimi birlikte kullanılmaktadır.

Şekil 2.8' deki *KA*, tek değişkenli karar ağacıdır. Burada olayların tek bir özelliği göz önüne alınarak testler yapılmıştır. Eğer düğümdeki sorular X_1 ve X_2 'yi

aynı anda içerseydi çok değişkenli KA' ların yapısı oluşturulmuş olurdu (Gülpınar, 2008).

2.5.2.1. Karar ağacı oluşturma aşamaları

Karar ağacı oluşturma aşamaları aşağıda verilmiştir (Gülpınar, 2008).

- 1. Aşama: Başlangıç olarak bağımlı değişken ve bağımlı değişkene etki eden bağımsız değişkenler belirlenir. Başlangıç noktasını oluşturan düğüm bağımlı değişken olup, başka bir değişle kök değişkenini (root) ifade etmektedir.
- 2. Aşama: Bu aşamada bağımlı değişkene etki ettiği düşünülen bağımsız değişkenlerin her biri incelenir. Bu durumda da KA, tekrarlanan bu bölümlere (recursivepartitioning) dayalı bir süreç takip eder. Bu işlemler incelenen bölümler arasında en anlamlı bölümlendirmeye ulaşıncaya kadar devam edecektir. Bölümlendirme kriterleri bazı durumlarda araştırmacı tarafından da belirlenebilmektedir. Genellikle KA' da bölümlendirmenin istatistiksel olarak anlamlı olup olmadığının tespit için ki – kare analizi kullanılmaktadır. KA'da oluşturulan ikili bölümlendirmeler arasında seçim yapılırken sınıflar arasındaki varyasyonun maksimize olduğu, sınıf içi varyasyonun ise minimize olduğu bölümlendirme tercih edilmektedir.
- 3. Aşama: Yukarıdaki işlemler her değişken için gerçekleştirildikten sonra tahmin düzeyi en yüksek olan değişken seçilir ve bu değişken yaprak düğümlerini oluşturmak için kullanılır.

2.5.2.2. Karar ağaçlarının karşılaştırılması

KA yapısı verilerin ve değişkenlerin özelliklerine göre çeşitlilik gösterebilir (Gülpınar, 2008).

Ayrılma değişkenlerinin sıralanması: Ayrılma kriterinde seçilen değişkenin sıralaması yani hangi niteliğin önce seçileceği çok önemlidir. Bu niteliğin seçimi bazı modellerde rassal olarak seçilirken bazı modellerde entropi yöntemi ile seçilir.

Ayrılmalar: Ayrılmaların sayısı değişkenlerin sıralanmasına bağlı olarak belirlenir. Değişkenlerin bazılarının etki kümesi küçüktür, bu nedenle ayrılmaların sayısı etki kümesine göre artar veya azalır. Etki kümesi eğer devamlı ya da çok fazla sayıda değere sahipse, bu durumda ayrılmaların sayısına kolayca karar verilemez.

Ağaç Yapısı: Sınıflandırma için ağaç uygulama gösterimini geliştirmede, en az aşamalı, yorumlanması kolay, dengeli bir ağaç istenebilir. Bu durumda, çoklu dallarda daha karmaşık karşılaştırmalara ihtiyaç duyulabilir. Bazı algoritmalar sadece ikili ağaçlar inşa eder.

Ölçütü Durdurma: Veri kümesi iyi bir şekilde sınıflandırıldığında dallanma kriteri kesinlikle son bulur. Daha büyük ağaçların ölçütlerini önlemek için bazı durumlarda önceden bir durdurma yapılabilir. Ayrıca önceden durdurma işlemi uygulamada bazı sorunları gidermek için de yapılır. Eğer veri kümesinde gösterilmeyen veri bölünmelerinin olduğu biliniyorsa, ihtiyaç duyulandan daha fazla veri kullanılması akla uygundur.

Eğitim Verisi: KA, üzerinde çalıştığımız eğitim verisine göre oluşturulur. Bu veri kümesi eğer çok küçükse oluşan ağaç, daha fazla veriyle çalışmaya yeterli olacak kadar spesifik olmayabilir. Tam tersi veri çok fazlaysa, ağacı uygulamada sorun yaşanabilir.

Budama: Ağaç oluşturulunca, ağacın gösterimini geliştirmek için sınıflandırma sürecinde bazı değişimlere ihtiyaç duyulabilir. Budama evresi, daha iyi bir sunuma ulaşmak için gereksiz karşılaştırmaları ve alt ağaçları ortadan kaldırabilir.

2.5.3 Karar ağacı algoritmaları

Bir sınıflandırma modelinde KA algoritması tahmin görevinde kullanılabilir. Karar ağacı algoritmasında durumlar farklı gruplara bölünerek sorunların en iyi şekilde çözümlenebilmesi amaçlanmıştır.

Karar ağacı algoritmalarında önemli olan sınıflandırma veya tahminin iyi bir şekilde yapılmasıdır, bundan dolayı bazı durumlarda modellerin nasıl çalıştığı ile pek ilgilenilmez. Bazı uygulamalarda ise kararın sebebini açıklayabilmek önemli olmaktadır. Bu gibi durumlar da verilerin kullanılabilmesi için tanımlamaların çok iyi yapılması gerekmektedir. İstenen kalitede ve yorumlanabilirlikte KA'lar oluşturmada çeşitli algoritmalar kullanılır.

Prensip olarak, verilen bir dizi değişkenden oluşturulabilecek pek çok KA bulunmaktadır. Oluşturulan karar ağaçlarından bazıları diğerlerine göre daha doğru olurken, en uygun ağacın bulunması araştırma alanının gittikçe büyüyen boyutu nedeniyle hesaplanamamaktadır. Yine de makul bir süre içinde, makul ölçüde doğru

KA 'nın indüklenmesi amacıyla etkin algoritmalar geliştirilmiştir. Bu algoritmalarından biri, ID3, C4.5 ve CART dâhil olmak üzere, mevcut pek çok KA endüksiyon algoritmasının temeli olan Hunt'un algoritmasıdır (Gülpınar, 2008).

KA algoritmalarının önemli özellikleri aşağıda özetlenmiştir.

- KA algoritması, sınıflandırma modellerinin oluşturulmasında kullanılan parametrik olmayan bir yöntemdir.
- Uygun bir KA 'nın bulunması, parametrik olmayan eksiksiz bir sorundur. Birçok KA algoritması, geniş hipotez alanlarındaki arayışları yönlendirmek için buluşsal yöntemlere dayalı bir yaklaşımdan yararlanmaktadır.
- KA oluşturmak için geliştirilen yöntemler, hesap olarak oldukça basittir ve düzenleme boyutu ne kadar geniş olursa olsun model oluşturmayı mümkün hale getirmektedir.
- KA yorumlanması kolaydır fakat dallanmanın artması boyutun büyümesine neden olacağı için, yorum yapmayı güçleştiren etki yapar.
- Bir alt KA 'nın birçok kez tekrarlanması KA 'yı gereğinden daha karmaşık hale getirmekte ve yorumlanmasını güçleştirmektedir. Böyle bir durum her bir dâhili düğümde tek bir değişken testi koşuluna dayanan KA uygulamalarından doğabilir. KA algoritmalarının çoğu bir 'böl ve yönet' bölünme stratejisi kullanıldığı için aynı test koşulu, değişken alanındaki farklı bölümlere uygulanabilir ve bu da alt ağaçların tekrarlanması (çoğalması) problemini beraberinde getirir.

2.5.3.1. Hunt algoritması

Hunt'un algoritmasında KA , eğitim kayıtlarını sırasıyla daha saf alt gruplara ayırarak tekrarlı bir şekilde büyümektedir. D_t , t düğümü ile birlikte oluşturulan eğitim kayıtları kümesi ve $Y = y_1, y_2, \dots, y_c$ ise sınıf etiketleri olsun. Hunt'un algoritmasının tekrarlanan bir tanımı aşağıda verildiği gibidir.

Adım1: D_t içindeki bütün kayıtların y_i sınıfına ait olması durumunda t , y_i olarak etiketlenen bir yaprak düğümüdür.

Adım2: D_t 'nin birden fazla sınıfa ait olan kayıtlar içermesi durumunda, kayıtların daha küçük alt kümeler ayrılması için değişken testi koşulu seçilmektedir.

Hunt'un algoritmasının çalışması, eğitim verilerindeki her değişken değerlerinin kombinasyonu mevcut olması ve her bir kombinasyonun tek bir sınıf

etiketi taşıması durumunda gerçekleşir. Pek çok uygulamada bu varsayımlar sağlanamamaktadır.

KA yöntemi, ID3, C4.5 ve geliştirilmiş versiyonu C5.0, Sınıflandırma ve Regresyon Ağaçları (Classification and Regression Trees - Cart), Ki-kare Otomatik Etkileşim Tespiti (Chi-Square Automatic Interaction Detection, CHAID) gibi pek çok sayıda özel algoritmayı kapsar. 1970'lerin sonunda J.R. Quinlan, ID3 (Tekrarlanan Çatallaşma – Iterative Dichotomiser) algoritmasının geliştirmiştir. ID3 geliştirilen ilk KA algoritmasıdır (Gülpınar, 2008).

2.5.3.2. ID3 algoritması

ID3 KA algoritmasını 1970'lerin sonunda, J.Ross Quinlan, Hunt'ın 'Böl ve Elde Et' (*Divide and Conquer*) algoritmasını geliştirerek oluşturmuştur. Hunt'ın kullanmış olduğu metotta değişken seçimi rastgele yapılırken Quinlan, değişken seçimi için entropi yöntemini kullanmış ve böylelikle Hunt metodunun en önemli eksikliğini gidermiştir. ID3, başlangıçta satranç oyununda etkin stratejiler öğrenme gibi amaçlara yönelik olarak kullanılmıştır. Günümüzde ID3 algoritması hem sanayi, hem de akademik alanda pek çok sorunu çözme amaçlı kullanılmış, daha sonra geliştirilerek zaman içinde yaygın kullanım alanı bulmuştur (Gülpınar, 2008).

ID3 tekniği karar ağacı kurmak için bilgi teorisi üzerine kurulmuştur ve karşılaştırmaların beklenen sayısı, minimize etmeye çalışır. En yüksek bilgiye sahip bölme özelliklerini seçmek ID3'ün kullandığı temel stratejidir. Bu özellik değeri ile ilişkilendirilmiş bilginin miktarını ortaya çıkma olasılığına ilişkindir. Bilgiyi nitелеmek için kullanılan kavram ise entropidir. Bir veri setindeki düzensizliğin veya rassallığın miktarını ölçmek için entropi kullanılır (Terlemez, 2008).

ID3 algoritması ilk olarak ağacın kök düğümündeki düzeltme örnekleriyle başlar. Bu düzeltme örnekleri bölümlenmek için bir değişken seçilir. Her bir değişken için bir dal oluşturulur ve dal tarafından kendisine yeni bir özellik kazandırılmış olan örnek alt kümelerde yeni oluşturulan alt düğüme yerleştirilir. Bir düğümdeki bütün örnekler tek bir sınıfa ait olana kadar algoritma her bir alt düğüme tekrar tekrar uygulanır. KA yaprağındaki her yol, bir sınıflandırma kuralını ortaya koyar. Böyle bir tepeden aşağı KA çıkarım algoritmasında önemli bir husus, düğümdeki niteliğin seçimidir. ID3 ve C4.5 algoritmalarındaki değişken seçimi için düğümdeki örneklere uygulanan entropi ölçüt bilgisini en aza indirme temelini dayanır.

Ağaç yapısında yaprağı olmayan bir düğümde örnekler dallara bölünür ve her bir alt düğüm örneklerin karşılığı olan alt kümeleri elde eder. Tek değişkenli bölümlerde kullanılan karar ağaçlarının basit bir sunum şekli vardır. Bu da kullanıcı için, ifade edilen modelin daha iyi anlaşılmasını kolaylaştırır. Aynı zamanda modelin ifade etme yetersizliğine de bir sınırlandırma getirir. Genellikle, özel bir ağaç sunumundaki herhangi bir sınırlama, önemli ölçüde işlev şeklini ve dolayısıyla da modelin yaklaşım gücünü de sınırlayabilir. Tek değişkenli bölümleri temel alan karar ağacı oluşturmak için kullanılan ağaç geliştirme algoritmalarında kullanılır. Bilgi teorisine dayanan yaklaşım, bir veri tabanında, örneklerin sınıflandırılmasını sağlayan testlerin sayısının en aza indirilmesinde ısrarlıdır. *KA* 'nın karmaşıklığı büyük ölçüde, verilerin özelliğın değeriyle yer değıştiren bilgi miktarına bağılıdır. Bilgi esaslı iz sürüm bilimi, en yüksek bilgi kazancını sağlayan niteliğı sınıflandırmak için sonuç alt-ağacında gerek duyulan bilgiyi en aza indiren niteliğı seçer. C4.5 algoritmasında sınıflandırma alanı kategorik özelliklerden sayısal özelliklere kadar uzanır. Ölçüt; içindeki örneklerin çoğunluğu tek bir sınıfa ait olduğunda, düşük sınıf entropisine sahip olan alt kümelere bölünmesiyle ortaya çıkan değışkenleri destekler. Algoritma temel olarak, sınıflar arasında yerel olarak en fazla ayırım derecesine sahip özellikleri seçer. Tümevarım algoritmasının temel fikri, cevapları en çok bilgi sağlayan soruları sormaktır (Gülpınar, 2008).

2.5.3.2.1. Karar ağacında entropi

Sistem belirsizliğinin ölçüsüne entropi denir. *KA* ' da, veri kümesindeki belirsizlik, sürpriz ya da rastgele miktarını ölçmek için entropi kullanılır. Eğer verilerin tümü belirli bir sınıfa aitse belirsizlik söz konusu olmaz bu durumda entropi sıfırdır. *KA* sınıflandırmasının asıl amacı, bir veri kümesini, sonuncu alt kümenin elemanlarının tümünü aynı sınıfa ait olacak şekilde tekrar tekrar bölümlemektir.

Karar ağacı oluştururken dikkat edilecek en önemli nokta ağaca hangi düğümden başlanacağıdır. Hunt algoritmasının en önemli eksiğı ağaca hangi düğümden başlanılacağıının rastgele seçilmesidir. Kök düğümün rassal bir seçimle belirlenmesi olası tüm ağaçları ortaya çıkarmak için büyük bir emek ve zaman kaybına sebep olacaktır.

Bu nedenle ağacın hangi düğümlle başlayacağına (kök düğümün belirlenmesi) entropi yöntemiyle karar verilir. Bir alanın entropi ölçüsü ne kadar fazla ise, o alan

kullanılarak ortaya konulan sonuçlar da o oranda belirsiz ve kararsızdır. Bu nedenle KA algoritmasının kökünde entropi ölçüsü en az olan alanlar kullanılır.

Entropi hesabı veri tabanından eğitim için elde edilen kayıt kümesinden hareketle hesaplanır. Eğitim kümesinin sınıf niteliğinin (bağımlı değişken) alacağı değerlere göre $\{C_1, C_2, \dots, C_k\}$ olmak üzere k tane sınıfa ayrıldığı varsayılın. Bu sınıflarla ilgili olarak ortalama bilgi miktarı ise T tane sınıf değerini içeren P_T (2.1)'deki şekilde hesaplanır.

$$P_T = \left(\frac{|C_1|}{|T|}, \frac{|C_2|}{|T|}, \dots, \frac{|C_k|}{|T|} \right) \quad (i=\overline{1, k}) \quad (2.1)$$

$|C_i|$ ifadesi C_i kümesindeki elemanların sayısını vermektedir. Örneğin $\frac{|C_j|}{|T|}$, P_j olasılığı ifade etmektedir. Buna göre T için ortalama bilgi miktarı yani entropi eşitlik (2.2)'deki gibi hesaplanır (Gülpınar, 2008).

$$H(T) = - \sum_{i=1}^n p_i \log_2(p_i) \quad (i=1,2,\dots,n) \quad (2.2)$$

2.5.3.2.2. Bölünme bilgi ve kazanç ölçütü

Hedef niteliğini ifade eden T, hedef niteliği olmayan (yani sınıf niteliği olmayan) bir X niteliğinin değerine bağlı olarak $T_1, T_2, \dots, T_n$ alt kümelerine ayrılın, T'nin bir elemanının sınıfını belirlemek için gerekli olan bilginin ağırlık ortalamasıdır. O halde, bu tanıma bağlı olarak T'nin bir elemanının sınıfını belirlemek için gerekli bilgi eşitlik (2.3)'deki gibi hesaplanır:

$$H(X, T) = \sum_{i=1}^n \frac{T_i}{T} H(T_i) \quad (i=1,2,\dots,n) \quad (2.3)$$

T veri tabanını X testine göre bölmekle elde edilen bilgileri ölçmek için 'kazanç ölçütü' adı verilen bir ölçüte başvurulur. Bu ölçütü eşitlik (2.4)'teki şekilde hesaplanır.

$$Kazanç(X, T) = H(T) - H(X, T) \quad (2.4)$$

Burada ayırma işlemi yapılırken $Kazanç(X, T)$ değeri en yüksek olan değişken seçilerek kazancı maksimize edilecek X testinin seçilmesi amaçlanmıştır (Gülpınar, 2008).

2.5.3.2.3. Kazanç oranı

Bir kategorik özelliğin olası değer çeşitliliği ne kadar yüksek olursa o özelliğin bilgi kazancı gereksiz bir şekilde yüksek çıkar ve bu durum ağacın doğruluğunu olumsuz etki eder. Bu tip özellikler işe yaramadıkları gibi kazancı yüksek özelliklerin de önüne geçip veride gizlenmiş kuralların çıkarılmasına engel teşkil eder. Örneğin, insanlarda şeker hastalığının olup olmadığına karar veren bir KA oluşturulacaktır. Bu konuda şeker hastalarından bazı veriler toplanır. Ancak bu veriler içinde hastaların etiket numaraları da bulunur. Bu durumda hasta numarası en yüksek bilgi kazancına sahip değişken olacaktır. Şeker hastalığının olup olmadığına etiket numarası ile karar verilmiş olacaktır. Bu olumsuz durumu telafi edebilmek için Quinlan, C4.5 algoritmasında değer çeşitliliği fazla olan özelliklerin bilgi kazancını azaltarak algoritmanın gereksiz bazı çıkarımlar yapmasını engellemektedir (Gülpınar, 2008).

T kümesi için X niteliğinin değerini belirlemek için gereken bilgi miktarını ölçmede, (bilgi miktarı $H(P_{X,T})$ miktarı) T kümesindeki X niteliği için ‘bölünme bilgisi miktarı $H(P_{X,T})$ olmak üzere Eşiklik (2.5)’deki gibi hesaplanır:

$$P_{X,T} = \left(\frac{|T_1|}{|T|}, \frac{|T_2|}{|T|}, \dots, \frac{|T_k|}{|T|} \right) \quad (i=1, k) \quad (2.5)$$

Bölünme bilgisine ait kazanç oranı ise Eşitlik (2.6)’daki gibidir.

$$Kazanç\ oranı(X,T) = \frac{Kazanç(X,T)}{H(P_{X,T})} \quad (2.6)$$

Burada kazanç oranı en yüksek olan değişken seçilir.

2.5.3.3. C4.5 algoritması

C4.5 algoritması ID3 algoritmasının geliştirilmiş halidir. Karar ağacı her düğümden çıkan çoklu dallar ile oluşturulur. Oluşturulan dalların sayısı ise tahmin edicinin kategori sayısına eşitti. Birden çok karar ağacını tek bir sınıflayıcı birleştirebilir. Bilgi kazancı yöntemi ile karar ağacında bölünme işlemi yapılır. Budama işlemi ise her yapraktaki hata oranına dayanır (Yılmaz, 2014). Bu algoritma Quinlan tarafından geliştirilmiştir. Sayısal değerlere sahip niteliklerinde karar ağaçlarını oluşturma olanağı sağlamıştır. C4.5 algoritması kayıp veri, sürekli veri, budama ,kurallar, bölümlleme yollarıyla ID3 algoritmasını iyileştirir (Terlemez, 2008).

C4.5 algoritması ID3’ten farklı olarak eksik ve sürekli değişken değerlerini ele alabilmekte, KA ’nın budanması ve kural çıkarımı gibi işlemler yapabilmektedir.

Yakın zamanda kural türetme hızının ve kalitesinin kendinden önceki versiyonu olan C4.5'ten daha iyi seviyede olan C5.0 geliştirilmiştir. C5.0 bunlara ek olarak çoklu KA'ları tek bir sınıflandırıcı bünyesinde birleştiren destekleme (boosting) tekniğini de uygulamaya koymuştur. Destekleme, farklı sınıflandırıcıları birlikte kullanma yaklaşımıdır. Destekleme normalde belirli bir sınıflandırıcıyı çalıştırmak için daha fazla zaman harcarken doğruluk oranını arttırmaktadır. Bazı veri kümelerinde hata oranının, C4.5 ile bulunanın yarısından daha az olduğu görülmüştür. Eğitim verisi çok gürültü içerdiğinde destekleme her zaman etkili olmaz. Desteklemenin çalışma prensibi, bir eğitim kümesinden birden fazla eğitim kümesinin oluşturulmasıdır. Eğitim kümesindeki her kaleme ağırlık tayin edilir. Ağırlık, söz konusu kalemin sınıflandırma açısından önemini temsil eder. Kullanılan her ağırlıklar kombinasyonu için sınıflandırıcı oluşturulur. Böylece aslında çok sayıda sınıflandırıcı oluşturulmuş olur. C5.0 ile sınıflandırma yapıldığında her sınıflandırıcıya oy tayin edilir, oylama yapılır ve hedef değişkenler grubu, çok az oy alan guruba tahsis edilir.

C4.5 'in karar ağaçları ve kural kümelerini türetmek üzere geliştirilmiş programları tek bir programda birleştirilmiştir. Bununla birlikte, pek çok nümerik özniteliğe sahip büyük veritabanlarında C5.0'ın, C4.5'e karşı üstünlüğü daha da belirgindir.

Ayrıca C5.0 algoritması bir karar ağacı veya bir kural kümesi oluşturur. Her bir yaprak modülü eğitim verilerine ait belirli bir alt kümeyi tanımlar ve eğitim verilerindeki her bir durum tam olarak ağaçta bulunan bir yaprak nodülüne aittir. Bir başka deyişle, karar ağacına sunuları belli bir veri kaydı için tam olarak bir tahmin mümkün olmaktadır (Gülpınar, 2008).

C4.5 aşağıdaki hususlarda ID3' e göre üstündür:

- Eksik veri: Eksik veri KA kurulduğunda basitçe ihmal edilir. Yani kazanç oranı yalnızca, söz konusu parametre için bir değere sahip diğer kayıtlara bakılarak hesaplanır. Eksik bir parametre değeri olan kaydı sınıflandırmak için söz konusu kalemin değeri, diğer kayıtlar için parametre değerine ilişkin bilgiler kullanılarak tahmin edilebilir.
- Sürekli veri: Sayısal değerlere sahip değişken içerisinde uygun eşik değeri bulunduktan sonra ikili ya da daha çok bölünme ile veri kümesi bölünebilir.
- Budama: C4.5 algoritmasında önerilen iki ana budama mevcuttur:

- ‘Alt ağaç değiştirmeli’ adı verilen stratejide değiştirme sonucunda, başlangıçtaki ağacınkine yakın hata oranı elde edilebiliyorsa alt ağaç, yaprak düğümle değiştirilir. Alt ağaç değiştirme, ağacın altında yukarıdaki köküne kadar uygulanır.

- ‘Alt ağaç yükseltme’ adı verilen diğer budama stratejisiyle alt ağaç, en çok kullanılan kendi alt ağacıyla değiştirilir. Böylece alt ağaç, mevcut konumundan ağacın daha üst noktasındaki düğüme yükseltilir. Yine de bu değiştirme için hata oranındaki artışı tespit etmemiz gerekmektedir.

- Ayırma: Aşırı uyum (overfitting) sebebiyle oluşan hata, C4.5 tarafından geliştirilen yöntemle telafi edilmeye çalışılmaktadır. C4.5’ te budama konusunda ayrıntılı olarak değinilecektir.

Sayısal değerlere sahip nitelikler:

Sayısal niteliklere ilişkin testlerin formüle edilmesinde bazı zorluklar görülebilir. Diğerleri iki aralığı bölmek için gelişmiş güzel eşikler bulunmaktadır. Ancak en uygun eşik değerini hesaplamak için çeşitli yollar bulunmaktadır. Eşik değerlerinin belirlenmesi amacıyla, en büyük bilgi kazancını sağlayacak biçimde bir eşik değeri belirlenir. Bunun için nitelik değerleri sıralanır ve $\{V_1, V_2, \dots, V_n\}$ biçimini alır. Bu eşik değeri kullanılarak nitelik değeri iki parçaya ayrılır. Eşik değeri olarak eşitlik(2.7)’ deki $[V_i, V_{i+1}]$ aralığın orta noktası alınabilir.

$$t_i = \frac{V_i + V_{i+1}}{2} \quad (i=1,2,\dots,n) \quad (2.7)$$

C4.5’ te eşik değeri olarak $[V_i, V_{i+1}]$ aralığının en büyük değeri eşik olarak alınır. Bu veri tabanında bir niteliğin sürekli, yani sayısal değerlere sahip ise genellikle iki test uygulanır. Bu karşılaştırmaya göre nitelik değerleri iki ayrı sayısal olmayan değere dönüştürülür. Böylece sayısal niteliklerin oluşturulduğu sorun çözülmüş olur.

Bilinmeyen nitelik değeri

ID3 algoritması eksik olmayan veri yığınlarını dikkate alarak analiz yapmaktadır. Veri tabanında bazı bilgilerin yer almaması ya da veri toplanırken bazı değerlerin kaydedilmemiş olması gerçek dünya verilerinde sıkça rastladığımız durumlardandır. Çünkü değer belirli bir örneğe bağlı değildir. Örneğin, veri tabanına veri aktarılırken hatalı bilgi girilmiş olabilir.

Bilinmeyen değerler sorununu çözmek için iki seçenek vardır:

- 1) Bilinmeyen veriyle birlikte veri tabanındaki bütün örnekleri yok saymak

2) Yeni bir algoritma tanımlamak ya da bilinmeyen veriyle işe yarayacak, var olan bir algorithmada değişkenler yapmak

İlk çözüm basittir ama örnekler takımında çok fazla değer atlanmışsa bu kabul edilemez. İkinci seçenek için ise bazı soruların cevaplandırılması gereklidir.

- Bilgi kazancı ve bilgi oranı değerleri olmayan özellikler için nasıl hesaplanır?
- KA'nı oluştururken alt ağaç yaratma işlemi sırasında özellik değeri olmayan satırlar hangi alt ağaca eklenecektir?

- KA oluşturulduktan sonra, kayıp değerleri olan yeni bir olay sınıflandırılırken nasıl bir test yapılacaktır?

Bütün bu ve buna benzer sorular bilinmeyen veriyle ilgili bir çözüm bulmada her zaman karşımıza çıkacaktır. Bilinmeyen veriyle çalışırken işe yarayan birkaç sınıflandırma algoritması, genellikle bilinmeyen değerlerin yerini olası bir değerle doldurma temeline dayanır. Hata oranlarının çok yüksek olduğunu bu yöntem bilinmeyen verilerin yüksek olduğu uygulamalarda KA'nın güvenilirliğini azaltır. Diğer bir yaklaşım verilen değişken için bütün değerlerin olası dağılımına bakmaktadır. Bu yaklaşımlardan hiç biri çok iyi değildir. Bu sebeple C4.5 algoritması bilinmeyen verilerle çalışmada farklı çözümler üretmiştir (Gülpınar, 2008).

2.5.3.3.1. Bilgi kazancı

Bir önceki bölümde bilgi kazancı formülleri, tüm verileri eksiksiz olan veri kümeleri için gösterilmiştir. C4.5'te geçerli prensip, bilinmeyen değerli örnekleri, bilinen değerlerin birbirleriyle bağıntılı frekanslarına ve olası durumlara göre dağıtmaktır. T çalışkan veri kümesi, Bilgi (T) genel bilgi kazancı, X veri kümesinin herhangi bir özelliği ve $Bilgi_x(T)$ X özelliğinin bilgi kazancı olmak üzere, Bilgi(T) ve $Bilgi_x(T)$ 'nin daha önce belirtilen biçimde hesaplandığı varsayılır. Yalnızca değerleri bilinen değişkenler hesaplamalara dâhil edilecektir. Böylelikle bilinmeyen değerleri içeren değişkenler eğitim kümesinden çıkartılmış olur. Bu durumda kazanç ölçütü, bir F faktörüyle düzeltilebilir. F faktörü, bilinen belirli bir niteliğin olasılığını ifade eder. Dolayısıyla F değeri eşitlik (2.8)'deki şekilde bulunur:

$$F = \frac{\text{Veri tabanında bilinen niteliğe sahip örneklerin sayısı}}{\text{Veri tabanındaki tüm örneklerin sayısı}} \quad (2.8)$$

Yeni kazanç ölçütü şu şekilde olacaktır:

$$Kazanç(x) = F(\text{Bilgi}(T) - \text{Bilgi}_x(T)) \quad (2.9)$$

Benzer şekilde değerleri bilinmeyen örnekler de bir ek gruba ayrılarak $\text{Bilgi}(T)$ bilgi bölmesi değeri hesaplanabilmektedir. X testinin bu test veri kümesini $n+1$ alt kümeye ayırdığını varsayarak $\text{Bilgi}_x(T)$ değeri hesaplanır. Bu kazanç oranının değerini doğrudan etkiler (Gülpınar, 2008).

2.5.3.3.2 C4.5 algoritmasında budama

Basit bir veri yığınının çok büyük bir ağacın elde edilmesine aşırı uyum ya da şişme (overfitting) denir. Aşırı uyum, her veri yığını için karşılaşılması mümkün bir sorundur. Aşırı uyum, veri yığınınındaki gürültüden kaynaklanabileceği gibi seçilen veri kümesinin o olayı temsil edebilme yeteneğinin olmamasından da kaynaklanabilir. Aşırı uyum ya da farklı sorunların bir sonucu olarak KA 'nın çok büyük çıkması, anlamsız sonuçlar oluşturabilir.

KA budanmasında yapılması gerek asıl görev, bir ya da daha fazla alt ağacı çıkarıp bunları yapraklarla değiştirerek KA 'nı sadeleştirmektedir. Alt ağacın bir yaprakla değiştirilmesinde algoritmanın, öngörülen hata oranını düşürmesi ve sınıflandırma modelinin kalitesini yükseltmesi beklenir. Fakat hata oranının hesaplanması kolay değildir. Sadece bir eğitim kümesine dayalı hata oranı uygun bir tahmin sağlanamaz. Öngörülen hata oranını tahmin etmenin bir yolu da, varsa yeni ve ilave test örneklerinin ya da çapraz geçerlilik sıralaması tekniklerinin kullanılmasıdır. Bu teknik başlangıçta mevcut olan örnekleri eşit boyutlu bloklara böler ve ağaç, her bir blok için, bu blok hariç olmak üzere, bütün örneklerden faydalanarak kurulup verilen bir örnekler bloğuyla test edilir. KA 'nın budanmasındaki temel fikir, daha az karmaşık ve böylelikle daha kapsamlı bir ağaç oluşturmak için, mevcut eğitim ve test örnekleriyle, görülmeyen test örneklerinin doğru bir şekilde sınıflandırılmasına katkıda bulunmayan alt ağaçların çıkarılmasıdır. Yinelemeli ayırma metodunun değiştirilebileceği iki yol vardır (Gülpınar, 2008).

1. Bazı koşullar altında bir örnekler kümesinin daha fazla bölünmemesine karar verilmesi: Durma kriteri genellikle ki-kare (χ^2) testi gibi bazı istatistiksel testlere dayanmaktadır: Bölünmeden önce ve sonra sınıflandırma doğruluğunda belirgin farklar olmaması halinde, bir akım düğümü bir yaprak

olarak gösterilir. Karar ayırım yapılmadan önce verilir, bu yüzden bu yaklaşım ön budama olarak adlandırılır.

2. Ağaç yağısının bir kısmının, seçilen doğruluk kriterlerinin kullanılmasıyla, geriye dönük olarak kaldırılması: Bu son budama prosesindeki karar, ağaç oluşturulduktan sonra verilir.

C4.5 son budama yaklaşımını izlemekte, fakat öngörülen hata oranını tahmin etmek için özel bir teknik kullanmaktadır. Bu metot kötümser budama olarak adlandırılmaktadır. Bir ağaçtaki her düğüm için, tahmin edilen üst güven limiti U_{cf} binom dağılımı istatistik tablosu kullanılarak hesaplanır. U_{cf} parametresi verilen bir düğüm için T_i ve E 'nin bir fonksiyonudur. C4.5, %25 varsayılan güven sınırını kullanır ve verilen her bir düğümdeki T_i için $U_{25\%}(\frac{|T_i|}{E})$, düğüm yapraklarının güven aralığı ile karşılaştırılır. Her bir yaprakta ağırlıklar olayların toplam sayısıdır. Bir alt ağaçtaki kök düğümün beklenen hatası, yapraklardaki için $U_{25\%}$ toplam ağırlıktan (alt ağaç için öngörülen hata) az olması halinde alt ağaç, budanan ağaçta yeni bir yaprak haline gelen kök düğümüyle değiştirilecektir.

C4.5'te budama yöntemini daha net açıklayabilmek için Quinlain (1993) tarafından yapılan bir çalışma da 10 niteliğe sahip bir veri kümesi oluşturulmuştur. Değişkenlerin her biri 1 veya 0'lar dan oluşan iki veri kümesi rastgele yaratılır. Sınıf, 'Evet' ve 'Hayır' değerlerinden oluşmak üzere ikili dallanma oluşturur. 'Evet' 0.25 ağırlığına, hayır ise 0.75 ağırlığına sahiptir. Rastgele oluşturulan bir adet olay, 500 satırlık bir test kümesi oluşturulacak şekilde ayrılmıştır. Bu verilerde C4.5' in başlangıçtaki ağaç oluşum rutiniyle, test olaylarında, %35 hata oranına sahip, 119 düğümlük mantıksız bir ağaç meydana gelir. Bulunan hata oranı ağaçta oluşan tahmini hata oranının bile altındadır.

Karar ağaçlarının başlangıcında oluşan dallanma genellikle aşırı karmaşıktır ve basit bir ağaca göre daha büyük bir hata oranına sahip olabilir. Yukarıda belirtilen rastgele seçilen veri için, sadece 'hayır' yaprağı içeren bir ağaç, görülmeyen durumlarda %25 hata oranı verir. Ancak her iki şıkkı içeren ağaç daha büyük bir hata oranı vermiştir.

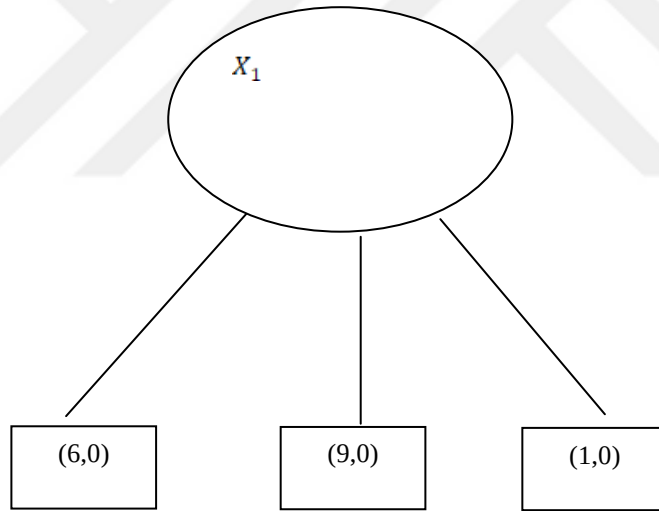
Çoğunlukla sınıfına ait olan durumların oranı, $p \geq 0.5$ burada 'hayır' dır. Eğer bir sınıflandırıcı tüm bu olayları bu çoğunluk sınıfına atarsa, sınıfın beklenen hata oranı $1-p$ olur. Diğer taraftan sınıflandırıcı bir olayı, p olasılıklı bir çoğunluk sınıfına

ve $1-p$ olasılıklı diğer sınıflara atarsa, beklenen hata oranı aşağıdaki olasılıkların toplamı olur.

- Çoğunluk sınıfına ait bir olayın diğer sınıfa atanması olayı, $p(1-p)$
- Diğer sınıfa ait bir olayın çoğunluk sınıfına atanması olasılığı, $(1-p)p$

Bu değerlerin toplamı $2p(1-p)$ olur. p en az 0.5 olduğu için, ki bu genellikle $1-p$ den daha büyüktür, ikinci sınıflandırıcı daha büyük bir hata oranına sahip olacaktır. Söz konusu veri tabanı için tahmini hata oranı $2 \cdot (0.75) \cdot (0.25) = 0.375$ (%37.5) sonucuna ulaşılır. KA bu tahmini hata oranlarının çok altında bir performansa sahip olabilmektedir. Oysaki rastgele seçilen bir veri yığınının oluşturduğu ağacın hata oranı bu tahmini hata oranına çok yakındır. Bu ise aşırı uyumun en büyük sorununun verinin söz konusu olayla ilgili olmaması veya bir şey ifade etmemesi olduğunu göstermiştir.

Karar ağaçlarında bu işlem yine Quinlain'ın verdiği bir örnekle açıklanacaktır. Şekil 2.9.'da karar ağacı budama örneği gösterilmektedir.



Şekil 2.9. Karar ağacında budama örneği

Eğitim harcamaları= n: demokrat (6)

Eğitim harcamaları= y: demokrat (9)

Eğitim harcamaları= u: cumhuriyetçi (1)

Alt ağacın, eğitim kümesinde hiçbir ilişki hatası bulanmamaktadır. İlk yaprak için $N=6, E=0$ ve (C4.5'in varsayılan %25'lik güven seviyesinin kullanarak),

$U_{\%25}(0,9) = 0.43$ ve $U_{\%25}(0,1) = 0.75$ olmak üzere bu alt ağaç için tahmini hataların sayısı aşağıda hesaplandığı gibi 3.273 olacaktır.

$$6.(0.206)+9.(0.143)+1.(0.750)=3.273$$

Eğer alt demokrat yaprağının yerini alsaydı, bir adet hata 16 aynı olayı kapsayacaktı. Dolayısıyla ilişkili tahmin edilen hatalar

$$16.U_{\%25}(1.16) = 16.(0.157) = 2.512$$

olacaktır.

Mevcut alt ağaç daha yüksek sayıda tahmin edilen hatalara sahip olduğu için bir yaprağa kadar budanıyor.

Yukarıdaki alt ağaç şu şekilde görünüyor:

Bütçe çözümünün kabulü = y: demokrat(151)

Bütçe çözümünün kabulü = u: demokrat(1)

Bütçe çözümünün kabulü = n: demokrat(16/1)

Bu alt ağaç için tahmin edilen hataların sayısı:

$$151.U_{\%25}(0,151) + 1.U_{\%25}(0,1) + 2.512$$

olup bu da 4.642'ye denk gelir. Eğer bu alt ağacın yerini demokrat yaprağı alırsa tahmin edilen hatalar $168.U_{\%25}(1,168) = 2.610$ olacaktır. Yaprak için tahmini hata oranı alt ağacinkine göre daha düşüktür. Dolayısıyla bu alt ağaç bir yaprak kalacak şeklinde budanır.

Bu durumda kriterin olmaması durumunda, bir ağaç modeli her bir düğüm bağımlı değişken değeri veya seviyeleri bağlamında özdeş gözlemler içerene kadar büyümeye devam eder. Bu yüzden ağacın büyümesi makul bir seviyede durmalıdır. KA hem çok büyük olmamalı hem de isabetli olmalıdır. İlk özellik ağacın yaprak sayısının az olmasına ikinci özellik ise yaprakların azami saflıkta (net bir sonuç – değer vermesi) olmasını gerektirir. Nihai karar iki muhalif strateji arasında bir uzlaşma ile elde edilir. Bazı karar ağaçları yaprak sayısına (mesela yaprak sayısı 5 i geçmeyecek dediğinde eşik=5 oluyor) göre, bazıları ise işlem sayısına göre durdurma kurallarına sahiptir. Diğer bazı algoritmalarındaki budama işlemi ise veriler üzerinde ihtimal hesaplarına dayanır. İhtimal hesaplarının yoğunluğundan ağacın saflığında azalma çok küçük olduğunda algoritma durur. KA sonuçları durdurma kuralı seçiminden son derece etkilenir (Gülpınar, 2008).

Karar kurallarını oluřturmak:

Ağaçlar budandıklarında orijinallerinden daha basit olmalarına rağmen hala çok karmaşık olabilirler. Büyük karar ağaçlarında her bir düğüm önceki düğümlerdeki test sonuçlarıyla oluřturulan spesifik bir bağlama sahip olduğundan, zor anlaşılabilir. Bir KA modelini daha okunaklı yapabilmek için, her bir yaprağı olan yol EĞER – İSE-DEĞİLSE (*IF-THEN-ELSE*) üretim kuralına çevrilebilir. EĞER (*IF*) kısmı bir yoldaki bütün testlerden oluşur ise İSE (*THEN*) kısmı nihai sınıflandırmalıdır. Bu şekildeki kurallar karar kuralları olarak adlandırılır ve bütün yaprak düğümleri için karar verme kurallarının bütünü, örnekleri tıpkı ağacın yaptığı gibi sınıflandırılır. Ağaç kuralların sırası önemli değildir

Karar verme kuralları ve KA'nın karmaşıklığını azaltmanın diğeri bir yolu da kategorik veriler için değişken değerlerinin gruplandırılması sürecidir. Çok sayıda değer büyük veri hacmine yol açar. Eğitim verilerinin yetersizliğinden dolayı kullanışlı örüntülerin tespit edilemeyeceği ya da örüntülerin tespit edilmesine rağmen modelin fazla karmaşık olacağı konusunda endişe yoktur. Değişken değerleri sayısını düşürmek için uygun gruplar tanımlanmasını gerekmektedir. Muhtemel ayırım sayısı büyüktür: n kadar değer için $2^{n-1} - 1$ kadar aşikâr olmayan ikili bölme mevcuttur. Değerlerin sıralanması halinde bile, ikili bölünme için $n-1$ kadar 'kesik değer' mevcuttur.

C4.5 metodu yalnızca ikili kategorik veriler değil, aynı zamanda n 'li bölünmeleri de içerdiğinden gruplama kombinasyonlarının sayısını artırır. Süreç tekrarlı olup, her değer için ayrı bir gruba temsil ettiği bir başlangıç dağılımıyla başlar ve sonrasında her yeni tekrar için öncesinde yer alan iki gruba, tek grup olacak şekilde birleştirilmesi olasılığını analiz eder. Birleştirme bilgi kazanç oranının azalmaması halinde kabul edilir. Nihai sonuç KA'lar ve karar kurallarına dayalı sınıflandırma modelini birleştirerek bir ya da iki grup şeklinde olabilir (Gülpınar, 2008).

2.5.3.4. Sınıflandırma ve regresyon ağaçları (CART)

Sınıflandırma ve Regresyon Ağaçları olarak bilinen CART kısaca (*Classification and Regression Trees*), Breiman, Friedman, Olshen ve Stone tarafından 1984 yılında yazılmış olan kitapta tanımlanmıştır. Bu ağaçların İstatistik dünyasında kullanılması, 1963 yılında Morgan ve Sonquist tarafından yazılan Otomatik Etkileşim Tespiti

(AID) ve daha sonra 1970'lerin başında Morgan ve Messenger tarafından yapılan THAID çalışmalarına kadar uzanmaktadır.

CART algoritması, bağımlı değişkenin kategorik veya nümerik olmasına bağlı olarak ikili KA oluşturan sınıflandırma ya da regresyon ağaçları üreten parametrik olmayan bir tekniktir (Gülpınar, 2008). CART algoritması kümeleme ve diskriminant analizlerinden oldukça farklı bir sınıflandırma yapısına sahiptir. Karmaşık yapıli veri setlerinde oldukça elverişli olan CART algoritmasında sınıf sayısı önceden belirlenir ve fonksiyonel bir form tanımlamayı gerektirir. Sapan değerlere oldukça duyarlı olan CART algoritması karışık, kategorik ve sürekli değişkenlerin herhangi bir birleşiminden oluşan veri setlerinin kullanabilirler (Yılmaz, 2014).

ID3 algoritmasının da her sayıda düğüm oluşturulabilirken CART 'ta yalnızca iki adet dal oluşturulur (Gülpınar, 2008). Bu nedenden dolayı ID3 ten farklıdır ama ID3 algoritmasında olduğu gibi en iyi bölme özelliğini belirlemek için entropi kullanır (Terlemez, 2008).

Kolay yorumlanabilir ağaç çıkarması CARTt analizinin son yıllarda kullanımın da artış sağlamıştır. Ağaca dayalı olan farklı yapısından dolayı kabul görmesi uzun zaman almıştır (Oğuzlar, 2004)

Maksimum ağacı oluşturmadaki temel hedef en saf, en iyi bölme durumunu elde etmektir. Bunun için en iyi kestirici değişken seçimi, kategorik bağımlı değişkenler için Gini ya da Twoing, sürekli bağımlı değişkenler için En Küçük Kareler sapması indeksi hesaplamalarına göre yapılmaktadır. Burada amaç, bağımlı değişkenle ilgili verinin, mümkün olduğunca homojen alt kümelerin meydana getirilmesidir.

C4.5 algoritmasının da kullanılan 'böl ve yönet' yönteminin temel metodolojisi aynı zamanda CART' ta da kullanılmaktadır. Aralarında ki fark, sadece ağaç yapısı, ayırma kriterleri, budama metotları ve eksik değerlerle alakalı uygulanan yöntemlerdir.

CART algoritmasının da hedef değişkeni nominal, ordinal veya sürekli olabilir. Karar ağaçlarında modelleme veri kümesinde belli değişkenlerin değerlerine dayanan kuralların toplanması ile oluşur. Kurallar, değişkenlerin değerine bağlı gözlemlerine kadar iyi farklılaştırılabileceğine bağlı olarak seçilir.

- Bir kuralın seçilip bir düğümün ikiye bölünmesinde kullanılan mantık her bir 'alt' düğüme uygulanır. Yani tekrarlı bir prosedür gerçekleştirilir.

- CART' ta daha fazla kazanç elde edilebileceği düşünüldüğünde veya bazı önceden ayarlanmış durma kurallarıyla karşılandığında, dallanma durur.
- Her ağaç bir terminal düğümde biter.
- Her gözlem sadece bir terminal düğüme düşer.
- Her terminal düğümü bir kurallar seti ile diğerlerinden ayrı olarak tanımlanmıştır (Gülpınar, 2008).

2.5.3.4.1. Ağaç yapısı

Ağaç yapısına dayalı olarak sınıflandırma ve regresyon modellerinin türetilmesi için kullanılan CART algoritması istatistiksel bir prosedürdür. CART ağaç modeli, tek değişkenli ikili kararların hiyerarşisini içerir. CART algoritmasında homojen ağaç yapısı elde edilmek istenir. Bunun için CART verileri iki alt kümeye ayırır ve her bir alt küme içindeki durumlar, bir önceki alt kümeden daha homojen olur. Homojenlik kriterine ulaşıncaya veya diğer bazı durma kriterleri sağlanıncaya kadar bu ardışık süreç kendini tekrar eder. Burada ağacın yapısı önceden belirlenmemekte, verilerden türetilmektedir.

CART algoritması çoklu ayırma işlemleri için bir alternatif sunmadığı için bu sınıflandırma ayırma kriterlerini basitleştirmektedir. Ayrıca, eğer bağımlı değişken sistemdeyse, ikili ayırma sınırlaması CART'ın kategorik özelliği (konkav - içbükey ayırma kriterlerini minimize ederek), zamanla özellik değerleri sayısı bakımından lineer zaman alan iki değer alt kümesini optimal olarak ayırmayı sağlamaktadır. Ama bu sınıflandırmanın da dezavantajları vardır. Çoklu olarak iyi ayrılabilen bir özellikte iyi bir ikili ayırma olması durumundaysa düşük kalitede ağaçlar elde edilebilir (Gülpınar, 2008).

2.5.3.4.2. Twoing algoritması

Twoing algoritmasında da ID3'de olduğu gibi, en iyi bölünme parametresi ve kriterinin seçiminde entropi kullanılmaktadır. Ancak, ID3 algoritması her alt kategori için bir alt dal oluştururken bu algoritma ID3'ten farklı olarak yalnızca iki dal oluşturulur. Bölünme ise en iyi bölünme noktası belirlenerek tespit edilen bu nokta etrafında yapılır. En iyi bölünmeyi tespit etmek için her adımda geniş kapsamlı bir arama yapılır (Gülpınar, 2008). Bu algoritma eğitim kümesinin her adımda iki parçaya ayrılması esasına dayanır. T eğitim kümesinin tümünü göstermek üzere, küme

iki parçaya bölündüğünde, ayrılan parçalardan t_{sol} sol taraftakini $t_{sağ}$ ise sağ taraftakini ifade eder.

Aday bölünmelerin her biri için P_{sol} ve $P(j/t_{sol})$ olasılıkları hesaplanır. Söz konusu olasılıklar eşitlik (2.10) da verilmektedir. $P(j/t_{sol})$ koşullu olasılık ifadesi ise bir j sınıf değerinin sol taraftaki bölünmede yer alma olasılığını ifade etmektedir.

Eğitim kümesindeki kayıtların sayısı N olmak üzere sol taraf için P_{sol} ve $P(j/t_{sol})$ olasılıkları şu şekil de hesaplanır.

$$P_{sol} = \frac{t_{sol} \text{ daki her bir nitelik değerlerinin ilgili nitelik sınıfındaki tekrar sayısı}}{\text{Eğitim kümesindeki kayıtların sayısı}}$$

$$P(j/t_{sol}) = \frac{t_{sol} \text{ daki kayıtların } j \text{ sınıfların sayısı}}{t_{sol} \text{ daki her bir nitelik değerlerinin ilgili nitelik sütünündaki tekrar sayısı}}$$

(2.10)

Aday bölünmelerin her biri için $P_{sağ}$ ve $P(j/t_{sağ})$ olasılıkları hesaplanır. Söz konusu olasılıklar eşitlik (2.11) da verilmektedir. $P(j/t_{sağ})$ koşullu olasılık ifadesi ise bir j sınıf değerinin sağ taraftaki bölünmede yer alma olasılığını ifade etmektedir.

$$P_{sağ} = \frac{t_{sağ} \text{ daki her bir nitelik değerlerinin ilgili nitelik sınıfındaki tekrar sayısı}}{\text{Eğitim kümesindeki kayıtların sayısı}}$$

$$P(j/t_{sağ}) = \frac{t_{sağ} \text{ daki kayıtların } j \text{ sınıfların sayısı}}{t_{sağ} \text{ daki her bir nitelik değerlerinin ilgili nitelik sütünündaki tekrar sayısı}}$$

(2.11)

$\emptyset(s/t)$, t düğümdeki s aday bölünmelerinin uygunluk ölçütü olsun. Söz konusu uygunluk ölçütünün nasıl hesaplandığı eşitlik (2.12)' de görülmektedir.

$$\emptyset(s/t) = 2P_{sol}P_{sağ} \sum_{j=1}^n |P(j/t_{sol}) - P(j/t_{sağ})| \quad (2.12)$$

şeklinde hesaplanır.

$\emptyset(s/t)$ değeri hesaplandıktan sonra bunlardan en büyük olanı ($\max \emptyset(s/t)$) seçilir. En büyük uygunluk değerine sahip olan satır, bölünmenin başlatılacağı niteliği belirler (Uysal vd, 2014).

2.5.3.4.3. Gini algoritması

CART kapsamında ele alınan sınıflandırma yöntemlerinden birisi olan gini algoritması da karar ağacı oluşturulmasında kullanılır. Karar ağacında dallanma ikiye

ayırma ilkesine dayanmaktadır. Karar ağacının ikili dallandırma oluşturmada aşamasında ilk hangi nitelikten bölüneceği ve bölünme değeri gini indeks değerine bakılarak hesaplanır. Veri setindeki varlıkların oranı gini indeksi olarak adlandırılır. İki varlığın gini değerlerinin aynı çıkması sonuç dağılımlarının aynı olduğunu gösterir.

Bir niteliğin gini değeri hesaplanmadan önce her nitelik değerleri ikili olacak biçimde gruplanır ve bu yolla elde edilen sol ve sağ bölümlere karşılık gelen sınıf değerleri gruplandırılır. Veri kümesindeki her bir nitelik ile ilgili sol ve sağ taraftaki bölünmeler için $gini_{sol}$ ve $gini_{sağ}$ değerleri eşitlik (2.13) ve (2.14)'daki gibi hesaplanır.

$$gini_{sol} = 1 - \sum_{i=1}^k \left(\frac{L_i}{|T_{sol}|} \right)^2 \quad (i=\overline{1, k}) \quad (2.13)$$

$$gini_{sağ} = 1 - \sum_{i=1}^k \left(\frac{R_i}{|T_{sağ}|} \right)^2 \quad (i=\overline{1, k}) \quad (2.14)$$

Burada; k sınıfların sayısı, T bir düğümdeki örneklerin sayısını, $|T_{sol}|$ sol taraftaki örneklerin sayısını, $|T_{sağ}|$ sağ taraftaki örneklerin sayısını, L_i sol tarafta i kategorisinde ki örneklerin sayısını ve R_i sağ tarafta i kategorisindeki örneklerin sayısını göstermektedir.

Gini algoritmasında $gini_{sol}$ ve $gini_{sağ}$ değerleri hesaplandıktan sonra bölünmenin başlayacağı niteliği seçmek üzere gini değeri hesaplanır. Bir niteliğin gini değeri eşitlik (2.15) 'deki denklem kullanılarak hesaplanır.

$$gini_j = \frac{1}{n} (|T_{sol}| gini_{sol} + |T_{sağ}| gini_{sağ}) \quad (2.15)$$

Hesaplanan tüm ölçütler arasından en küçük olan seçilir ve bölme bu nitelik üzerinden gerçekleşir. Geride kalan veri seti üzerinde aynı adımlar tekrar uygulanır ve diğer bölünme hesaplanır. seçilir ve bu değere sahip olan nitelik bölünmenin başlayacağı düğüm olarak belirlenir (Adak ve Yurtay, 2013).

2.5.3.4.4. Regresyon ağaçları

CART algoritması adından da anlaşıldığı gibi aynı zamanda regresyon ağaçlarını da destekler. Regresyon ağaçlarında CART'ta kullanılan budama ve büyüme kriterleri aynı olduğu için, regresyon ağaçları sınıflandırma ağaçlarından bir anlamda daha

basittir. Regresyon ağaçlarının yapısı, her bir yaprağın bir reel sayıyı işaret etmesi dışında regresyon ağaçları sınıflandırma ağaçlarına benzer (Gülpınar, 2008).

Yeni değer ikamesi, ortalama karesel hata olup Eşitlik (2.16)'deki gibi hesaplanır.

$$R(T) = \frac{1}{n} \sum_{i=1}^n (y_i - h(t_i))^2 \quad (i=1,2,3,\dots,n) \quad (2.16)$$

Formülde yer alan y_i değeri t_i için gerçek değerli etikettir. Formüldeki $h(t_i)$ de gerçek değerli tahmindir. Yeni değer ikamesi tahminini minimize etmek için ayırma kriteri seçilir (Gülpınar, 2008).

2.5.3.6. CHAID algoritması

Kass tarafından 1980' de bulunan CHİ-kare Otomatik Etkileşim Tespiti (CHAID – Chi-squared Automatic Interaction Detector), en eski en iyi bilinen ağaç sınıflandırma yöntemlerinden biridir. Bu yöntem değişkenlerdeki etkileşim ile bulunmuştur. Bu yöntem de sürekli bağımsız değişkenler, otomatik olarak sıralı alt sınıflara bölünür. CHAID, bir popülasyonu, bağımlı değişkendeki varyasyonu gruplar içi minimum ve gruplar arası maksimum olacak şekilde farklı alt gruplara veya bölümlere tekrarlı olarak ayıran bir tekniktir (Gülpınar, 2008).

CHAID algoritması CART'ın dışında en çok kullanılan karara ağacı algoritmalarından biridir. Bölümlendirme amaçlı kullanılan ve etkili bir yöntem olan CHAID optimal bölümlerin tespiti için ki-kare istatistiğini kullanır. Bir istatistiksel testin anlamlılığını kriter olarak kabul eder ve bu kriteri kullanarak, potansiyel ön kestirici değişkeninin tüm değerlerinin değerlendirir. Bağımlı değişkene veya aynı anlama gelmek üzere hedef değişkene göre istatistiksel olarak homojen bir şekilde değerlendirilir. Daha sonra ön kestirici değişkenin seçilmesiyle her bir düğümün seçilen değişkenin homojen değerlerinin bir gurubu oluşturmasını sağlar. Hedef değişkeninin ölçüm düzeyi kullanılan test istatistiğini belirler. Hedef değişkeni sürekli olduğu durumda F testi, hedef değişkeni kategorik olduğu durumda ise ki – kare testi kullanılmaktadır (Oğuzlar, 2004). Ki-kare testinde CHAID algoritması bölme işlemini gerçekleştirir. Dalların sayısı iki ile tahmin edicinin kategori sayısı arasında değişir (Yılmaz, 2014).

CHAID algoritması bağımlı değişken nominal, kategorik, ordinal kategorik veya sürekli, bağımsız değişkenler ise sürekli, kategorik veya nominal kategorik olduğunda uygulanabilen bir yöntemdir.

Bir regresyon denklemi eğer CHAID algoritması ile elde ediliyorsa bu denklem bilinen klasik varsayımlardan muaf tutulmaktadır. Çünkü güçlü bir öteleme algoritması (iteration algorithm) ile bütün olan evren kararlı alt düğümlere (node) bölünebilmektedir. Bu işlem beraberinde, verilen dağılımda beraberliğini ve normalliğini sağlayabilmektedir. Ayrıca hem kategorik hem de sürekli verileri, aynı anda modele alabilmeye olanak tanımaktadır. CHAID algoritması bu nedenle parametrik ve parametrik olmayan ayrımını ortadan kaldırmakta, yöntem algoritmasında istatistiksel olarak yarı parametrik (semi-parametric) özellik taşımaktadır.

Kategorik verileri sınıflandırma da CHAID, açıklayıcı bir yöntemdir. CHAID algoritması optimal dağılımları tanımlamak için ki-kare istatistiğini kullanır. CHAID algoritması daha büyük veri kümeleri için uygun olan, nispeten basit bir algoritmayı temel alan ve ikili olmayan ağaçlar oluşturur.

İstatistiksel bir işlem olarak, değişken değerleri kategorik olduğunda ve bağımsız değişkenler ile kategorik olarak ölçülebilen sonuçlar arasında ilişki arandığında oldukça yararlı olan sonuçlar üreten CHAID, kullanıcılara aşağıda sıralanan kolaylıkları sağlamaktadır (Gülpınar, 2008).

- CHAID algoritması bir ağaç yapılandırma yöntemi olduğundan varsayımlara dayanmaz, bu nedenle sıralı en küçük kareler (OLS–Ordinary Least Square) regresyonuna alternatif bir modelleme sağlar. OLS’ nin varsayımlarını dikkate almaksızın güvenilir tahminler temin ederek sürekli bir bağımlı değişken için regresyon ağacı oluşturulur.
- Modelin gerçek yapısal formunda belirlenen varsayımları dikkate almadığı için ikili (binary) ve multi nominal lojistik regresyon modellerine alternatif bir parametrik olmayan ağaç diyagramı olarak kullanılabilir.
- CHAID algoritması geniş örnek büyüklüklerinden yararlanmasından dolayı potansiyel olarak çok güvenilir tahminler sunar ve bağımsız değişkenlerdeki kayıp gözlemleri tahmin edebilir.

CHAID algoritmasının sağladığı avantajlar aşağıdaki şekilde sıralanabilir.

- Bağımlı ve bağımsız değişkenler için ölçü tipi nominal (kategorik), ordinal veya aralıklıdır.
- Bağımsız değişkenlerdeki tamamının aynı düzeyde ölçülmesine gerek yoktur.
- Bağımsız değişkenlerdeki kayıp değerler sabit olmayan kategori (floating category) olarak muhafaza edilebilir,

- Uygun bir istatistiksel kriter kullanılırsa, sonuçlandırılan modelden şansa bağlı olmaksızın çok güçlü sonuçlar elde edilmesini sağlar.

CHAID analizi, kategorik değişkenlere ilişkin veri kümesini ve bağımlı değişkeni en iyi açıklayabilecek şekilde ayrıntılı homojen alt gruplara böler. Bu alt kümeler, küçük tahmin edici gruplardan oluşur. En iyi tahmin sonucunu elde etmek için başlangıç değişkenleri bağımsız olarak yeniden kategorileştirilir. Bunun için ki-kare analizi kullanılır. Adımsal olarak uygulanan benzer kategorileri birleştirme işlemi, değişkenler arasında daha fazla birleştirme sağlanamayacağına istatistiksel olarak karar verilinceye kadar devam eder. Değişkenlerin bölünmeye uygun olup olmadığına Benferroni düzeltilmiş p değeri kullanılarak karar verilir.

Benferroni yaklaşımı, her bir grubun ortalama vektörlerinin genel ortalama vektöründen farkları bulunduktan sonra bu farkların sıfır olup olmadığını araştırmaya dayanır. Her bir değişken için, gruplar ikiye ayrılarak dikkate alınır ve eşlik kullanılarak i. değişken için elde edilen aralığın sıfır değerini içerip içermediği kontrol edilir. Eğer sıfır değeri belirlenen aralıkta yer alıyorsa, ilgili gruplar arasında anlamlı bir farklılık olmadığı, aksi durumda grupların farklı oldukları şeklinde yorumlanır.

Adım 1: Her bir bağımsız değişken için bağımlı değişkenin kategorileri ile bağımsız değişkenin kategorileri arasında çapraz tablo oluşturulur.

Adım 2: 2xd alt tablosunda bağımsız değişkene ait anlamlılığı en düşük olan kategori çiftleri bulunur. Birleşmeleri anlamlı bulunan iki kategori birleştirilir. Bu birleşme bir bileşik kategori olarak düşünülür ve bu adım bağımsız değişkenin kendi içindeki birleşmeleri anlamsız oluncaya kadar devam eder.

Adım 3: Üç ya da daha çok sayıda orijinal kategori içeren bileşik kategorilerin her biri için birleşmenin tekrar çözümlendiği en önemli iki bölünme bulunur. Eğer anlamlılık bir kritik değer altındaysa bölünme tamamlanarak ikinci adıma dönülür.

Adım 4: Optimum düzeyde birleştirilen bağımsız değişkenlerin her birinin anlamlılığı hesaplanır, en çok anlamlı olan ayrılır. Eğer bu anlamlılık kritik bir değerden büyükse seçilen bağımsız değişkenin birleştirilen kategorilerine göre veri alt gruplara bölünür.

Adım 5: Henüz analiz edilmemiş veri için birinci adıma gidilir.

Her bir bağımsız değişken için, kendi içinde kategorileri en anlamlı bir şekilde birleştirilip en iyi bölünme bulunduktan sonra, bağımlı değişkene göre olumsuzluk tablosu oluşturulur.

CHAID algoritmaları kesin bir durumda kriteri olarak ki-kare testi uygular, bu testte homojenlik hipotezinin belirginliği test edilir. Eğer homojenlik belli bir düğümde reddedilirse ayırma devam eder, aksi halde düğüm son-terminal düğüm olur. CART algoritmalarından farklı olarak, CHAID ağacın büyümesini, budama yerine ki-kare testinin belirginlik esasında durdurma kriteri ile durdurulur (Gülpınar, 2008).

2.5.3.7. Ayrıntılı CHAID

Orijinal CHAID yönteminin bazı zayıflıklarını telafi etmek için Biggs, de Villeve Suen (1991) tarafından Ayrıntılı CHAID geliştirilmiştir. Ayrıntılı CHAID algoritması, CHAID algoritmasının biraz daha geliştirilmiş halidir ve her bir tahmin değişkeni için mümkün olan tüm dağılımları daha ayrıntılı bir şekilde incelediği için bu şekilde hesaplama yapmak daha uzun zaman almaktadır. Ayrıntılı CHAID algoritmasında hedef değişken nominal, ordinal ya da sürekli olabilir (Gülpınar, 2008)

Ayrıntılı CHAID algoritmasının CHAID algoritmasına benzer yanları vardır. İstatistiksel testler ve kayıp değerlerin değerlendirilmesi bakımından benzerlik gösterir fakat Ayrıntılı CHAID algoritmasının hesaplanması uzun zaman almaktadır. Ayrıntılı bölmeler bulan CHAID algoritması zaman sorunu olmadığı durumlarda kullanıldığında daha faydalı sonuçlar verebilmektedir. Her ne kadar verilere bağlı olsalar da CHAID ile Ayrıntılı CHAID algoritması arasında farklılık bulunmamaktadır (Oğuzlar, 2004)

Uygulamada genelde, müşteri gruplarını seçmek ve bazı değişkenleri etkileyen diğer değişkenlere nasıl yanıt verdiklerini tahmin etmek için doğrudan pazarlamanın konusu olarak kullanılmaktadır. Diğer karar ağaçları gibi Ayrıntılı CHAID algoritmasının da avantajları çıktısının görsel olması ve kolay yorumlanabilmesidir. Varsayılan olarak çok yollu dalları kullandığı için, verimli çalışabilmesi ve kolay yorumlanabilmesidir. Varsayılan olarak çok yönlü dalları kullandığı için, verimli çalışabilmesi için oldukça büyük gruplara ihtiyaç duyar zira küçük gruplarla çalışıldığında yanıt gruplarının ebadı güvenilir analiz için çok küçük kalır (Gülpınar, 2008).

2.5.3.8 QUEST algoritması

QUEST (Çabuk Objektif Etkili İstatistik Ağacı) algoritması (1997) Wei-yin Loh ve Yu-Shan Shih tarafından geliştirilmiştir (Gülpınar, 2008). QUEST algoritması en son geliştirilmiş karar ağacı olma özelliği taşıyan ve çok sayıda kategoriye sahip ön kestiricileri destekleyen, diğer yöntemlerin yanlışlıklarından kaçınılmasını sağlayan ve hızlı hesaplanabilen bir yöntemdir (Oğuzlar, 2004). QUEST algoritmasının en önemli özelliği bağımlı değişkenin nominal olması gerektiğidir (Gülpınar, 2008).

QUEST algoritması ikili karar ağacı yapısı kullanan bir sınıflandırma algoritmasıdır. Budama ve doğrudan durma kuralı gibi tekniklerin kullanılabilmesinden dolayı ikili ağaç yapısını kullanmaktadır. CHAID ve CART algoritması ağacın oluşturulmasında değişken seçimi ve bölünmeyi eş zamanlı yaparken QUEST algoritması aksine hepsi ile ayrı ayrı ilgilenir. Ağacın dallanması sırasında önyargılı seçimin daha genel hala getirilmesi ve hesaplama maliyetinin düşürülmesi amacıyla QUEST algoritması geliştirilmiştir (Çalış, 2014).

CART detaylı araştırma yöntemleri, ağaç oluşturma sürecinde daha çok dağılımlar meydana götürebilen münferit değişkenleri seçme eğilimindedirler. Bu durum, modele ilişkin olarak sonuçların genelleştirilebildiklerini azaltan bir önyargıyı da ifade etmektedir. CART'ın geliştirmiş olduğu bir başka sınırlama da dağılımları araştırırken hesaplamaya çok fazla yatırım yapılıyor olmasıdır. QUEST algoritması işte bu sorunlara çözüm getirmek için tasarlanmıştır. QUEST'in, değişken seçimi önyargısı ve hesaplama maliyeti açısından detaylı araştırma yöntemlerinden çok daha uygun olduğu gözlemlenmiştir. Ancak sınıflandırma doğruluğu, dağılım noktalarının değişkenliği ve ağaç boyutu açısından değiştirilmemiş dağılımlar kullanıldığında halen net bir şekilde bir 'kazanan' ortaya çıkamamıştır (Gülpınar, 2008).

2.5.4. CART, CHAID ve QUEST karar ağacı algoritmalarının karşılaştırılması

CART, CHAID ve QUEST algoritmalarının üçü de kullandıkları karar kurallarıyla setleri alt gruplara ve bölümlere ayırarak bu değişkenler arasındaki ilişkilerin istatistiksel anlamlılığa bakar. Bu yöntemlerin tümünde amaç, istatistiksel olarak önemli olan modeli bularak karar kurallarını vermektir.

CART, CHAID ve QUEST algoritmalarının bu ortak noktalarına rağmen sahip oldukları değişkenin özelliklerine ve kullanım şekilleri ne göre önemli farklılıkları vardır. Aşağıda Çizelge 2.5.' te bu farklar bir arada verilmektedir (Gülpınar, 2008).



Çizelge 2.4. Karar ağacı algoritmalarının karşılaştırılması (Gülpınar, 2008)

Algoritma Türü	Analiz Bileşenleri	Analiz Kullanımı
CHAID	Bağımlı değişken: Nominal, kategorik, ordinal veya sürekli Bağımsız değişken: Sürekli, kategorik veya nominal	İkili olmayan bölünmeyi kullanır. Sınıflandırma modeli, diğerlerine göre daha güçlü sonuçlar vermektedir.
CART	Bağımlı değişken: Sürekli, ordinal veya nominal Bağımsız değişken: Sürekli, ordinal veya nominal	İkili bölünmeye olanak sağlayan ağaç oluşturur. Maliyet matrislerinin değişken seçiminde dikkate alınmasını isterse uygun sonuçlar verir. Maliyet karmaşıklığının indirgenmesi veya doğrudan durdurma kuralları gereklidir.
QUEST	Bağımlı değişken: Nominal Bağımsız değişken: Sürekli, ordinal veya nominal	Analistin elinde büyük veya karmaşık bir veri kümesi bulunduğunda ve ağacı tahmin etmek için etkin bir algoritmaya ihtiyaç duyduğunda kullanılır.

3. BENLİK SAYGISI, PROBLEM ÇÖZME VE KAYGI DÜZEYİ İLE İLGİLİ KAVRAMLAR

3.1. Benlik

Kendi kişiliğimize ilişkin tanımlarımız kendi kendimizi görüş ve düşünüş tarzımız benliğimizi oluşturur. Bu bakımdan benlik kişiliğin öznel yanı olarak tanımlanabilir (Taşgit, 2012).

Psikolojik bakımdan benlik, kişinin çevresinde olup bitenleri algılama da, değerlendirme de ve çevresine karşı tepkide bulunmasında en önemli dayanaktır (Yiğit, 2010).

Benlik, sosyal bilimlerin değişik alanlarında ele alınan ve tartışılan konuların başında gelmektedir. Benlik kavramı, yaklaşık son yarım asırdır en çok tartışılan ve üzerinde durulan psikoloji kavramlarından biridir. Bu yönüyle benlik, sosyal bilimlerde ve pek çok psikolojik alanda önemle üzerinde durulan bir kavram olmuştur. Psikoloji bilminde benlik, kişilik konusu içerisinde ele alınmaktadır. Bireyin kişiliğini ve kimliğini oluşturan temel yapılardan biri olarak görülmüştür. Bu nedenden dolayı benlikle kişilik arasında doğrudan bir ilişki vardır.

Benlik ve kişilik iç içe olmasına rağmen kişilikten farklı özellikler de taşımaktadır. Benlik kişinin kendini tanıması baskıya karşı olumlu ya da olumsuz direnç göstermesi ve kişinin kendisiyle ilgili verdiği özel hükümlerdir. Kendisini değil, kendisi dışındakileri de algılayıp, tavır sergilemesidir. Benlik nesnel değil, öznelidir. Kişiye göre değişir. Benlik bireyin kendine ait özellikleri, yetenekleri, değer yargıları ve ideallerine ilişkin yargılarının dinamik örüntüsü, kendi kişiliğine yönelik düşünceleri ve kendi kendini görüş tarzıdır. İnsanın kendisi hakkında sahip olduğu özel algıların çok boyutlu bir birleşimidir. Bu yönüyle benlik, kişilik gibi anlaşılması güç, karmaşık bir kavramdır (Çevik, 2013).

3.1.1. Benlik kavramı

Benlik konusunda pek çok alanda çalışma yapılmasına rağmen, gerek tanımlı gerekse yapısı konusunda fikir birliğine ulaşılamamıştır. Benlikle ilgili pek çok değişik terimlerin kullanıldığı görülmektedir. Bu terimler içerisinde benliği ifade etmek için, ben, benlik, kendilik, benlik kavramı, kendilik kavramı, öz kavramı, benlik saygısı, benlik tasarımı, benlik algısı, kendini değerlendirme, kendini kabul v.b. değişik

kavramlar kullanılmıştır. Bunlar içinde en yaygın olarak kullanılan benlik kavramıdır (Çevik, 2013). Benlik kavramı, bireyin kendi kişiliğini değerlendirme biçiminden oluşmaktadır. Bireyin kendi kişiliğine ilişkin kanaatlerinin, düşüncelerinin ve kendisi hakkındaki yargılarının değerlendirme biçimidir (Yiğit, 2010). Bireyin fiziksel, duygusal, zihinsel v.b. yönlerden kendisi hakkında geliştirdiği görüş ve kendisine ilişkin kabul edebileceği, farkında olduğu kanıların tümü benlik kavramını oluşturmaktadır. Yani benlik kavramı genel olarak kişinin kendisi hakkında sahip olduğu düşünceleri ve algıları ifade etmektedir. Ya da bir başka ifade ile, kişinin bilinçli bir şekilde kendi varoluşu olarak adlandırabildiklerinin toplamıdır. Bireyin kendisi ile ilgili olarak kafasında çizdiği görünümüdür.

Bilişsel süreçleri tarafından geliştirilen ve kontrol edilen bir yapıya sahip olan benlik kavramı insanın benliğini algılama, kavrama biçimi olarak tanımlanan pek çok kuram gibi psikoloji tarihinde de önemli bir yere sahiptir (Taşgit, 2012).

Sonuç olarak benlik, bireyin kişiliğini ve kimliğini oluşturan tüm özelliklerin karmaşık bir örüntüsü iken; benlik kavramı, kişinin kendisi hakkındaki görüşünü içermektedir (Çevik, 2013).

3.1.2. Benlik saygısı

Kişinin kendisini benimsemesi, değer vermesi, kendine güven ve saygı duyması anlamlarına gelen benlik saygısı bir anlamda, bireyin kendisi ile sıcak ve samimi bir tutum içerisinde olmasıdır.

Benlik saygısı bireyin kendisine yaptığı duygusal değerlendirmedir. Kendisini önemli, değerli, başarılı ve yetenekli algılama derecesidir. Benlik saygısı benlik kavramında bulunan, kimlik oluşturma, kendini kabul etme ve başkaları tarafından kabul edilme duygularının değerlendirilmesidir (Yiğit, 2010).

Benlik saygısı psikoloji ve psikiyatri literatüründe öz saygı, kendilik saygısı, öz değerlilik duygusu gibi farklı terimlerle ifade edilmektedir. Genel olarak benlik saygısı bireyin kendini değerlendirmesi sonucu, ulaştığı olumlu ya da olumsuz duygu durumudur. Kişinin kendisini değerlendirmede pek çok faktör etkili olabilir (Çevik, 2013).

Kişinin kendi benliğine saygı duyması ve kendini beğenmesi için üstün niteliklere sahip olması gerekmez. Çünkü benlik saygısı, bireyin kendisini olduğundan aşağı ya da üstün görmeden kendisinden memnun olması durumudur.

Kendini değerli, olumlu, beğenilmeye ve sevmeye değer bulmaktır. Benlik saygısı bu yönüyle, kendine güvenmeyi ve kendini olduğu gibi kabullenmeyi sağlayan olumlu bir ruh halidir.

Rosenberg (1965), benlik saygısını kişinin kendine karşı pozitif veya negatif tavrı olarak ele alır. Buna göre kişinin benlik saygısının yüksek olması kendini değerlendirmede olumlu bir tutum içinde olduğunu gösterirken; benlik saygısının düşük olması olumsuz bir tutum içinde olduğunu göstermektedir. Benlik saygısı yüksek olan bir birey, kendine saygı duymakta ve kendini toplumda değerli bir kişilik olarak görmektedir. Benlik saygısı düşük birey ise, genel olarak kalıcı ve sürekli bir şekilde kendini olumsuz değerlendirmektedir (Taşgit, 2012).

3.1.3. Benlik saygısını etkileyen faktörler

Benlik, yaşantımızda bir takım gelişmeler sonucu oluşan bir olgudur. Doğduğumuz andan itibaren çevremizdeki insanlarla kurduğumuz ilişkiler ve karşılaştığımız olaylar sonucu şekillenir. Benlik gelişimi çocuklukta, kişinin fiziksel sınırlılıklarını fark etmesiyle başlar, psiko-sosyal uyumun karmaşık örüntüsü içinde gelişir. Benliğin gelişimin de çevrenin ve kişinin diğer insanlarla kurmuş olduğu ilişkilerin büyük önemi vardır. Aile içinde yaşanan sıkıntılar, boşanma veya ayrı yaşama gibi durumlar, anne babadan birinin ölümü, çocukların olumsuz benlik kavramı geliştirmelerine sebep olmaktadır. Ayrıca, sosyo-ekonomik ve sosyo-kültürel düzey, anne babanın çocuklarıyla ilgilenme düzeyi, anne babanın birbirlerine karşı ilişki durumu, akademik başarı, ailedeki kardeş sayısı, doğum sırası, anne babanın mesleği, anne babanın eğitim durumu, sosyal statüsü, psikolojik bir rahatsızlığın olup olmaması, cinsiyet türü ve etnik kimlik değişkenleri benlik ve benlik saygısı gelişimini etkileyen önemli faktörlerdendir (Çevik, 2013).

3.2. Problem ve problem çözme

Yaşamsal süreç de sıklıkla karşılaşılan sıkıntılı, karmaşık ve istenmeyen durumlar problem olarak ifade edilir (Elkin ve Karadağlı, 2015).

Problem en genel ifade ile, bir kişinin istenilen bir amaca ulaşmak için o süreçte karşısına çıkan engel olarak tanımlanmaktadır. Engellenme bireyin hedefe ulaşmasını güçleştirmektedir. Böyle bir durumda problem çözme, engeli aşmanın en iyi yolunu bulmaktır (Vuslat ve Köksal, 2015).

Bir durumun kişinin kendisi tarafından problem olarak görülebilmesi için öncelikle problem olarak algılanması gerekmektedir. Yani bireyin karşılaştığı olumsuz bir durum bir başkası için olumsuz olmayabilir, dolayısıyla bu durum onu rahatsız etmeyecek ve bu durumu çözmek için çaba harcamayacaktır. Kısacası her bireye göre problem algılama durumu değişiklik göstereceğinden, problem çözme davranışları da buna göre değişecektir.

Problem çözme ilk olarak fen bilimleri alanında ele alınmıştır. Daha sonra 60'lı ve 70'li yıllarda, bireylerin psikolojik uyumlarıyla problem çözme becerileri arasındaki ilişkiler incelemek için sosyal bilimler alanında da uygulanmıştır. Zamanla ilişsel yaklaşımın yaygınlaşmasıyla birlikte problem çözme, farklı bilişsel ve düşünsel süreçlerin toplamı olarak tanımlanmıştır (Hatay, 2008).

3.2.1. Problem çözmeyi etkileyen faktörler

Problem çözme bireyin amaç, değer, ihtiyaç ve alışkanlık tutumları ile ilgilidir. Problem çözme, istenilen bir amaca ulaşmada karşılaşılan güçlüklerin üstesinden gelmedir. Bu nedenle bireyin problem çözme konusundaki başarısı problemin özelliğinden çok bazı kişisel faktörlere bağlıdır. Zekâ bu faktörlerden biridir. Birey karşılaştığı problem karşısında ne kadar zeki olursa, problem çözmedeki başarısı da o derece fazla olacaktır. (Hatay, 2008).

Çocukların okul öncesi dönemden itibaren problem çözme becerilerinin geliştirilmesi onların gerçek yaşama uyum sağlamaları açısından çok önemlidir (Vuslat ve Köksal, 2015). Bu nedenle “problem çözme” eğitim kurumlarında çocuklara kazandırılmalı ve onların birey olarak bu yönde yetişmeleri sağlanmalıdır çünkü problem çözme becerisi çocuklarda öğrenmenin temelini oluşturur. Problem çözme davranışları ile çocuklar bir konuda düşüncelerini yoğunlaştırmayı, alternatif çözümler üretmeyi, neden sonuç ilişkilerini kavramayı ve sonuçları tahmin etmeyi öğrenirler (Hatay, 2008).

Bireysel Farklılıklar

Bireysel farklılıklar, diğer alanlar da olduğu gibi problem çözme sürecinde de bireye etki eden önemli değişkenlerden biridir. Bu farklılıkların problem tanımlama sürecine olumlu etkileri olabileceği gibi olumsuz etkileri de olabilmektedir. Burada dikkat edilmesi gereken karşılaşılan problem ile problem çözücünün probleme yaklaşma tutumu karşısında izlediği yoldur.

Tanışık Olma

Daha önce benzer bir problemle karşı karşıya kalmış olmaktır. Problem çözme becerisini etkileyen en büyük etkenlerden biridir. Çünkü benzer problemler çözmüş bir problem çözücünün, otomatik olarak uygulayabileceği problem haritalarına sahip olur ve karşılaştığı problem karşısında nasıl bir yol izleyeceğinin bilincinde olur.

Bilişsel Faktörler

Biliş, bilme, anlama, tanıma, kavrama, düşünme, akıl yürütme, anımsama, unutma ve problem çözme gibi zihinsel işlevleri ifade eder. İnsanların dünyayı tanımalarına, anlamalarına ve öğrenmelerine yol açan zihinsel faaliyetlerdir. Başka bir deyişle, düşünme, öğrenme ve hatırlama süreçlerinin genel adıdır. Bilişsel gelişim ise bireyin anlamasını ve öğrenmesini sağlayan, zihinsel faaliyetlerdeki gelişimdir. Bilişsel faktörler, tahmin edileceği üzere problem çözümünde kullanılacak olan bilimsel yöntem bilimlerini yansıtmaktadır. Bu bilimsel yöntemler, her bireyin elde ettiği bilgiyi işleme biçimi, düşünme biçimi ve kontrol etme biçimi farklılık gösterebilmektedir. Bilişsel esnekli yüksek olan bir birey daha fazla alternatif üretime yetisine sahip olduğu gibi probleme çok daha analitik yaklaşabilmektedirler. Bu da daha iyi bir problem çözme becerisini beraberinde getirmektedir.

Tecrübe Faktörü

Adından da anlaşılacağı üzere, belli problemler ile karşılaşmış ve bu problemler üzerinde belli stratejiler kullanarak çözüm üretmiş problem çözücülerini kapsamaktadır. Problem çözme sürecinde tecrübe ne kadar çok önemli bir etken olsa da, problemin türü kapsamı ve rutin olup olmama durumuna göre de tecrübe farklılık gösterebilmektedir.

Demografik Özellikler

Demografik özelliklerin bireysel farklılık açısından problem çözme üzerinde yadsınamaz etkileri vardır. Demografik özellikler tahmin edilebileceği üzere, yaş, cinsiyet, nüfus yoğunluğu, yakın çevre, etki alanı, gelir düzeyi, meslek, eğitim düzeyi, etnik altyapı, aile yapısı, bölge gibi unsurlardır. Bu unsurların her biri problem çözümünde önemli rol oynamaktadır. Bireyin yetiştiği çevre, yaşı, cinsiyeti, aile yapısı, eğitim düzeyi ve mesleği problem çözümüne yönelik alternatiflerin geliştirilmesinde farklılıklar göstermektedir.

Duyuşsal Faktörler

Duyuşsal faktörler Ortaya çıkan problemle kişinin aniden ortaya çıkan duyuşsal durumudur ve bu faktörler olumlu ya da olumsuz kaygı durumlarıdır. Problem çözme sürecinde istek, özgüven, sabır ve azim gibi bazı olumlu duyuşsal faktörler bireyi motive eder ve problem çözmede başarılı olma arzusunu birlikte getirir. Problem çözme sürecinde stres, kaygı, belirsizlik, acelecilik, sabırsızlık vb. olumsuz duyuşsal faktörler ise bireylerin motivasyonunu düşürerek onları problem çözümünde olumsuz etkiler (Çam ve Tümkeya, 2007).

3.3. Kaygı

Tehdit ve tehlike durumunda kişilerde ortaya çıkan, gelecekle ilgili endişeleri içeren ve bedensel reaksiyonlarla da (kasların gerilmesi, terleme vb.) kendini belli eden hoş olmayan bir duygu durumu olarak tanımlanmıştır. Kaygı, içinde bulunduğu durumu tehdit edici olarak algılayan bireyin geliştirdiği temel duygulardan biridir. Korku, endişe ve gerginlik gibi subjektif olarak hissedilir. Genel anlamda ise kaygı bireyin çevresinde yaşanan ve psikolojik olaylara karşı göstermiş olduğu bir duygu olarak tanımlanır, belirli şartlarda evrensel ve olağan kabul edilir.

Türkiye’de kaygı, vesvese, kuruntu ya da evham olarak da adlandırılmaktadır. Kullanılan bu kavramlar insanların kendisinde oluşturduğu korku ve kötü bir şey olacağı düşüncesi ile kendi bedeninde oluşan titreme, stres, uyuşma, çarpıntı gibi durumları yoğun olarak yaşaması anlamına gelmektedir.

Kaygı durumu yalnızca bireylerde gizli kalan bir duygu durumu değildir. Kaygı aynı zamanda bireylerin bulunduğu ortamlarda rolleri gereği davranışsal bütünlüklerinde etki edebilecek çok yönlü etkileri olan bir durumu ifade etmektedir (Kara ve Acet, 2012).

Kaygı, hem normal hem de patolojik insan davranışlarında önemli bir yere sahip olması nedeniyle psikolojide de yaygın olarak incelenen kavramlardan biri haline gelmiştir. Bu nedenle de diğer duyguların tanımında olduğu gibi, kaygının da tanımını yapmak zordur. Kaygı bireyin, üzüntü, sıkıntı, korku, başarısızlık duygusu, çaresizlik, sonucu bilememe ve yargılanma duygularından birini veya çoğunu içerebilir. Kaygının temel nedeni kişiye rahatsızlık veren her hangi bir olay değil, olayın kişi için taşıdığı önemdir (Yüksel, 2014).

3.3.1. Durumluk ve sürekli kaygı kavramları

Durumluk kaygı, tehlikeli, istenmeyen bir durumla karşılaşıldığında bireyin içinde bulunduğu baskılı ve stresli durumdan dolayı hissettiği subjektif korkudur (Kaya ve Varol, 2004). İnsanlar tehlikeli gördüğü durumlarda kaygı duyarlar. Bu tehlikeli koşulların oluşturduğu korku ve tedirginlik bireyin geçirdiği normal kaygı olarak ifade edilir (Kara ve Acet, 2012). Otonom sinir sisteminde meydana gelen bir uyarılma sonucu terleme, sararma, kızarma ve titreme gibi fiziksel değişimler, bireyin gerilim ve huzursuzluk duygularının göstergeleridir. Stresin yoğun olduğu zamanlarda durumluk kaygı seviyesinde yükselme, stres azalınca ya da ortadan kalkınca durumluk kaygı seviyesinde düşme meydana gelmiştir. (Macila, 2013).

Sürekli kaygı, geleceğe yönelik olarak hissedilen ve bireyin kaygı düzeyi açısından kaygı yaşantısına genel yatkınlığına işaret eder. Buna kişinin içinde bulunduğu durumları genellikle stresli olarak algılama ya da stres olarak yorumlama eğilimi de denebilir. Dolayısıyla, sürekli kaygı kişisel bir özelliktir ve süreklilik içerir. Sürekli kaygı seviyesi yüksek olan bu tür kişilerin kolayca incindikleri ve karamsarlığa büründükleri görülmüştür (Kaya ve Varol, 2004).

Kişinin genel yatkınlığını yansıtır yansıtmaması açısından, durumluk ve sürekli kaygı duyguları birbirlerinden farklılık gösterir. Ayrıca sürekli kaygı seviyesinin yüksek olması kişinin durumluk kaygıya olan yatkınlığının daha da yüksek olmasına etki edebilir. Sürekli kaygının, fiziksel bir tehlikeden ziyade genelde özsaygıya yönelik psikolojik bir tehdit içeren durumlarda ortaya çıkması, durumluk kaygının da yüksek seviyelerde olmasına neden olduğu öne sürülmüştür (Yüksel, 2014).

3.3.2. Kaygının nedenleri

Huzurlu ve kendini güvende hissettiği bir insanda korku ve kaygı olmaz. Diğer yandan aynı çevredeki başka biri çevresindeki bazı durumları tehlikeli olarak değerlendirebilir ve bu algılamayla ilgili heyecanlar yaşayabilir. Bu nedenle hangi sosyal ortamı nasıl algılayacağımızı içinde yetiştiğimiz çevre belirler (Kaya ve Varol, 2004).

Kaygı nedeni olarak şunları sayabiliriz:

- Bireyin alışmış olduğu çevredeki desteğin çekilmesi durumunda insanlar kaygı duymaya başlarlar.

- Olumsuz sonuçların doğacağı bir durumda da insanlar kaygı duymaya başlarlar.
- Bireyin inandığı ve önem verdiği bir inancı ile göstermiş olduğu davranış arasında çelişki olması bireyde gerginlik oluşmasına ve kaygı duymasına neden olur.
- İnsanlar için en belli başlı kaygı nedeni ise gelecekte ne olacağını bilmemek yani gelecekle ilgili belirsizliklerdir.



4. UYGULAMA

4.1. Araştırmanın Amacı

Bu çalışma Ondokuz Mayıs Üniversitesi Fen Edebiyat, Eğitim, Mühendislik, Sağlık Bilimleri, İİBF, Spor Bilimleri, Hukuk, İlahiyat Fakülteleri ve Sağlık MYO’da öğrenim gören bireylerin belirli nitelik ve yetenekleri bakımından kendini nasıl algıladığı ya da nasıl olması gerektiği gibi davranışlarını belirlemek için Rosenberg Benlik Saygısı envanteri, bireyin içinde bulunduğu stresli (baskılı) durumdan dolayı hissettiği subjektif korku ve bireyin belli bir anda ve belirli koşullarda kendini nasıl hissettiğini belirlemek için Durumluk Kaygı Ölçeği, bireyin kaygı yaşantısına yatkınlığı, bireyin içinde bulunduğu durum ve koşullardan bağımsız olarak kendini nasıl hissettiğini belirlemek için Sürekli Kaygı Ölçeği bireylerin bir hedefe ulaşmada karşılaştıkları engelleri ve bireylerin kişilerarası ilişkilerde yaşadıkları problemleri çözme konusunda kendini nasıl değerlendirdiklerini belirlemek amaçlı Problem Çözme Envanteri kullanılmış ve bir veri madenciliği sistemi oluşturmak amaçlanmıştır. Çalışmada “Rosenberg Benlik Saygısı Envanteri”, “Problem Çözme Envanteri” hedef değişken olarak ele alınmış ve bu değişkenler üzerinde hangi demografik değişkenlerin etkili olduğu belirlenmek istenmiştir. Uygulamada SPSS CLMENTINE paket programı kullanılmıştır.

4.2. Yöntem

4.2.1. Araştırma modeli

Araştırma, üniversite öğrencilerinde benlik saygısı ve problem çözme becerileri bağımlı değişkenlerinin öğrencilerin demografik bilgileri bağımsız değişkenlerinden ne oranda etkilendiğinin incelenmesine yönelik bir çalışmadır. Bu çalışmada, seçilen örneklem grubundan bilgi toplamak için “anket tekniği” kullanılmıştır. Analizde karar ağacı algoritmalarından biri olan CART algoritması kullanılmıştır. CART ikili ağaç olarak büyüyen bir algoritmadır. CART veriyi iki alt kümeye ayırır. Böylece bir sonraki adımda oluşacak alt kümenin, bir öncekinden daha homojen olmasını sağlar. Bu sonuç bulunana kadar devam eden kendini tekrarlayan bir süreçtir.

4.2.2. Araştırmanın Evreni ve örneklemi

Araştırmanın çalışma evrenini, 2015–2016 Eğitim-Öğretim yılında Ondokuz Mayıs Üniversitesi, Fen Edebiyat, Eğitim, Mühendislik, Sağlık Bilimler, İİBF, Spor Bilimleri, Hukuk, İlahiyat Fakülteleri ve Sağlık MYO’da öğrenim gören öğrenciler oluşturmaktadır.

Araştırmanın örneklemini, çalışmanın evrenini oluşturan Ondokuz Mayıs Üniversitesinden tesadüfi olarak seçilen 1426 (742 kadın, 684 erkek) öğrenciden oluşturulmuştur.

Öğrencilerin fakültelere göre dağılımı Tablo 4.1’de verilmiştir. Araştırmaya katılan öğrencilerin, %22.1’i Eğitim Fakültesinde, %13.2’si Fen Edebiyat Fakültesinde, %22.2’si Mühendislik Fakültesinde, %16.3’ü Sağlık Bilimleri Fakültesinde, %4.2’si Sağlık MYO’da, %6.4’ü İİBF’de, %6.3’ü Spor Bilimleri Fakültesinde ve % 9.3’ü İlahiyat Fakültesinde öğrenim görmektedir.

Çizelge 4.1. Çalışmaya katılan öğrencilerin fakültelere göre dağılımı

	Frekans	Yüzde
Eğitim Fakültesi	315	22.1
Fen Edebiyat Fakültesi	188	13.2
Mühendislik Fakültesi	316	22.2
Sağlık Bilimleri Fakültesi	233	16.3
Sağlık MYO	60	4.2
İİBF	91	6.4
Spor Bilimleri Fakültesi	90	6.3
İlahiyat Fakültesi	133	9.3
Toplam	1426	100.

Ankete katılan 1426 öğrencinin demografik özellikleri verilmiştir. Bu öğrencilerin, % 48’i erkek, %52’si kızdır. Çalışmaya katılan öğrencilerin, % 14.1’i 18-19 yaş arasında, %75.8’i 20-23 yaş arasında ve % 10.1’i 24 yaş ve üstündedir. Bununla beraber öğrencilerin, %2.5’i birinci sınıf, %38.1’i ikinci sınıf, % 32.8’i üçüncü sınıf, %22.7’si dördüncü sınıf, %2.7’si beşinci sınıf ve % 1.2’si altıncı sınıfta okumaktadır.

Çalışmaya katılan öğrencilerin, % 63.5’inin ailesi ev sahibi, % 27.3’ünün ailesi kiracı, %3.7’sinin ailesi ev sahibi değil kira ödemiyor ve % 3.4’ünün ailesi ise lojmanda yaşamaktadır. Bu öğrencilerin, % 96.3’ünün anne babası birlikte yaşamakta, % 2.5’inin anne- babası boşanmış ve % 1.3’ü de ayrı yaşamaktadır.

Öğrencilerin, %4.8'inin anne eğitim durumu Okuma yazma bilmiyor, % 2.7'sinin anne eğitim durumu Okur yazar, % 50.1'inin anne eğitim durumu İlköğretim, % 14.2'sinin anne eğitim durumu Ortaokul, % 16.1'inin anne eğitim durumu Lise, % 1.2'sinin anne eğitim durumu Meslek veya Teknik Lise, %2.5'inin anne eğitim durumu Ön lisans,% 7.1'inin anne eğitim durumu Lisans ve %1.3'ünün anne eğitim durumu Lisansüstüdür. %0.9'unun baba eğitim durumu Okuma yazma bilmiyor, %0.9'unun baba eğitim durumu Okuryazar, %29.7'sinin baba eğitim durumu İlköğretim, %15.6'sinin baba eğitim durumu Ortaokul, %26.3'ünün baba eğitim durumu Lise, %2.2'sinin baba eğitim durumu Meslek veya Teknik Lise, %4.9'unun baba eğitim durumu Ön lisans,% 16.3'ünün baba eğitim durumu Lisans ve %3.2'sinin baba eğitim durumu Lisansüstüdür. Araştırmaya katılan öğrencilerin,% 18.7'sinin annesi çalışırken, %81.3'ünün annesi çalışmamaktadır. Bu öğrencilerin, % 77.4'ünün babası çalışırken, %22.6'sının babası çalışmamaktadır. Araştırmaya katılan öğrencilerin, %22.3'ünün toplam geliri 1500 TL den az, %24.5 'inin toplam geliri 1500-2000 TL arasında, %13.3'ünün toplam geliri 2001-2500 TL arasında, %12.8'inin toplam geliri 2501-3000 TL arasında, % 9.0'ının toplam geliri 3001-3500 TL arasında, %5.3'ünün toplam geliri 3501-4000 TL arasında, %2.9'unun toplam geliri 4001-4500 TL arasında, %2.2'sinin toplam geliri 4501-5000 TL arasında, %2.3'ünün toplam geliri 5001-5500 TL arasında, %2.0'sının toplam geliri 5501-6000 TL arasında, % 1.2'sinin toplam geliri 6001-6500 TL arasında ve %2.2'inin toplam geliri ise 6501TL ve üzerindedir. Bu öğrencilerin, %66'sı burs alırken, %34'ü burs alamamaktadır. Öğrencilerin, %17.7'si günde 1 saat internet kullanırken, %39.4'ü günde 1 ile 3 saat internet kullanmakta, %41.0'ı 3 saatten fazla internet kullanmakta ve %2'si hiç internet kullanmamaktadır. Ayrıca bu öğrencilerin, % 26.1'i günde ortalama 1 saat TV izlerken, % 16.7'si 1ile 3 saat arası TV izlemekte, %3.9'u 3 saatten fazla TV izlemekte ve %52.3'ü hiç TV izlememektedir. Çalışmaya katılan öğrencilerin, %73.2'si bilgisayara sahipken ,% 26.8'i bilgisayara sahip değildir. Öğrencilerin,%24.1'i her gün kitap okurken, %18.3'ü haftada bir defa kitap okumakta, %27.3'ü ayda bir kitap okumakta, %6.9'u yılda bir kitap okurken %12.1'i hiç kitap okumamaktadır.

Çizelge 4.2. Çalışmaya katılan öğrencilerin demografik özellikleri

	Frekans	Yüzde
Cinsiyet		
Erkek	684	48,0
Kız	742	52,0
Yaş		
18-19 yaş	201	14,1
20-23 yaş	1081	75,8
24 yaş ve üzeri	144	10,1
Sınıf		
1	23	2,5
2	178	38,1
3	395	32,8
4	349	22,7
5	202	2,7
6	135	1,2
Kardeş Sayısı		
1	40	1,8
2	442	30,0
3	499	34,0
4	215	14,1
5	100	7,0
6	130	8,1
7	32	2,2
8	21	1,3
9 ve üzeri	27	1,5
Konut ve Mülkiyet Durumu		
Ev Sahibi	906	63,5
Kıracı	389	27,3
Ev sahibi değil, kira ödemiyor	53	3,7
Lojman	30	2,1
Diğer	48	3,4
Anne –Baba Birliktelik Durumu		
Birlikteler	1373	96,3
Boşandılar	35	2,5
Ayrı Yaşıyorlar	18	1,2

Çizelge 4.2. Çalışmaya katılan öğrencilerin demografik özellikleri (devam)

	Frekans	Yüzde
Anne Eğitim Durumu		
Okuma Yazma Bilmiyor	69	4.8
Okuryazar	39	2.7
İlköğretim	715	50.1
Ortaokul	202	14.2
Lise (Fen-Anadolu)	230	16.1
Meslek veya Teknik Lise (Kız Meslek- Endüstri Meslek-İmam Hatip)	17	1.2
Ön lisans	35	2.5
Lisans	101	7.1
Lisansüstü	18	1.3
Baba Eğitim Durumu		
Okuma Yazma Bilmiyor	13	0.9
Okuryazar	13	0.9
İlköğretim	423	29.7
Ortaokul	222	15.6
Lise (Fen-Anadolu)	375	26.3
Meslek veya Teknik Lise (Kız Meslek- Endüstri Meslek-İmam Hatip)	31	2.2
Ön lisans	70	4.9
Lisans	233	16.3
Lisansüstü	46	3.2
Anneniz Çalışıyor mu?		
Evet	266	18.7
Hayır	1160	81.3
Babanız Çalışıyor Mu?		
Evet	1104	77.4
Hayır	322	22.6
Ailede Başka Çalışan Var Mı?		
Evet	450	31.6
Hayır	976	68.4

Çizelge 4.2. Çalışmaya katılan öğrencilerin demografik özellikleri (devam)

	Frekans	Yüzde
Ailenizin Toplam Geliri		
1500 TL den az	318	22,3
1500-2000 TL	350	24,5
2001-2500 TL	189	13,3
2501-3000 TL	182	12,8
3001-3500 TL	129	9,0
3501-4000 TL	76	5,3
4001-4500 TL	42	2,9
4501-5000 TL	32	2,2
5001-5500 TL	33	2,3
5501-6000 TL	28	2,0
6001-6500 TL	17	1,2
6501 TL den fazla	30	2,2
Burs/kredi alıyor musunuz?		
Evet	941	66,0
Hayır	485	34,0
Günlük internet kullanımı		
1 saat	252	16,7
1 ile 3 arası saat	562	39,4
3 saatten fazla	584	41,0
hiç kullanmıyorum	28	2,0
Günde ortalama TV izleme		
1 saat	372	26,1
1 ile 3 arası saat	252	17,7
3 saatten fazla	56	3,9
hiç kullanmıyorum	746	52,3
Bilgisayarınız var mı		
Evet	1044	73,2
Hayır	382	26,8
Hangi sıklıkta kitap okuyorsunuz		
Her gün	343	24,1
Haftada 1	261	18,3
Ayda 1	390	27,3
6 ayda 1	161	11,3
Yılda 1	99	6,9
Hiç	172	12,1

4.2.3. Veri toplama araçları

Araştırma verileri, demografik bilgileri içeren öğrenci bilgi formu, bireyin belirli nitelik ve yetenekleri bakımından kendini nasıl algıladığını belirlemek için Rosenberg Benlik Saygısı Envanteri, bireyin ani değişiklik gösteren heyecansal reaksiyonları değerlendirmede oldukça duyarlı bir araç olan Durumluk Kaygı Ölçeği, kişinin genelde, yaşama eğilimi gösterdiği kaygının sürekliliğini ölçmeyi amaçlayan Sürekli Kaygı Ölçeği ve problem çözme becerilerini belirlemek için Problem Çözme Envanteri kullanılarak toplanmıştır.

4.2.3.1. Sosyodemografik anket formu

Bu form öğrencilerin kişisel özelliklerini (yaş, cinsiyet, aile gelir düzeyi, anne ve babanın eğitim durumu, anne-babanın birliktelik durumu, aile içi ilişkileri, anne babanın çalışma durumu, ailede başka çalışan sayısı, oturulan konutun mülkiyet durumu, kardeş sayısı), okul yaşamlarını (bölüm, sınıf, bölüm tercihi, mezun olduğu lisenin türü, akademik başarı durumu, bilgisayar kullanma durumu, internet kullanım durumu, TV izleme durumu, Samsundaki ortalama gideri, burs/öğrenim kredisi alma durumu, kitap okuma durumu) sorgulayan soruları içermektedir.

4.2.3.2. Rosenberg benlik saygısı ölçeği

Araştırmada, benlik saygısı ölçümü için kullanılan bu ölçek, 1963 yılında Morris Rosenberg tarafından geliştirilmiştir. Amerika Birleşik Devletleri'nde güvenilirlik 19 geçerlilik çalışması yapıldıktan sonra birçok araştırmada ölçüm aracı olarak kullanılmıştır. Ülkemizde ölçeğin geçerlilik ve güvenilirlik çalışmaları Çuhadaroğlu tarafından yapılmış olup, geçerlilik katsayısı $r = 0,71$ olarak bulunmuştur. Test-tekrar test güvenilirlik yöntemi kullanılarak da güvenilirlik katsayısı $r = 0,75$ olarak saptanmıştır. Rosenberg Benlik Saygısı Ölçeği, çoktan seçmeli 63 sorudan oluşan bir öz bildirim ölçeğidir. Ölçek, on iki alt kategoriden oluşmaktadır. Araştırmanın amacı doğrultusunda benlik saygısını ölçmeye yönelik olarak, ölçeğin 1 (Benlik saygısı), 10 (ana-baba ilgisi), 11 (babayla ilişki) ve 12 (psşik izolasyon) alt kategorileri kullanılmıştır(Yılmaz, 2015).

D -1 (BENLİK SAYGISI)

- 1....(C)=0,17 (D)=0,34 6....(C)=0,50 (D)=1
2....(C)=0,16 (D)=0,33 7....(C)=0,50 (D)=1
3....(A)=0,17 (B)=0,33 8....(A)=1 (B)=0,50
4....(C)=0,25 (D)=0,50 9....(A)=0,50 (B)=0,25
5....(C)=0,25 (D)=0,50 10....(A)=0,50 (B)=0,25

D -10 (ANA –BABA İLGİSİ)

- 49....(A)=0 (B)=0 (C)=1 (D)=1
50....(A)=0 (B)=0 (C)=1 (D)=1
51....(A)=1 (B)=0
52....(A)=1 (B)=0
53....(A)=1 (B)=0
54....(A)=1 (B)=0
55....(A)=0 (B)=0 (C)=1

D -11 (BABAYLA İLİŞKİ)

- 56....(A)=1 (B)=0 (C)=0 (D)=0 (E)=0 (F)=0 (G)=0
57....(A)=1 (B)=1 (C)=0 (D)=0
58....(A)=1 (B)=1 (C)=0 (D)=0 (E)=0
59....(A)=1 (B)=1 (C)=0 (D)=0 (E)=0
60....(A)=1 (B)=1 (C)=0 (D)=0 (E)=0
61....(A)=1 (B)=1 (C)=0 (D)=0 (E)=0

D -12 (PSİŞİK İZOLASYON)

- 62....(A)=1 (B)=0
63....(A)=0 (B)=1

“Benlik Saygısı” alt testinde **0 –1** puan alanların **“yüksek”**; **2–4** puan alanların **“orta”** ve **5–6** puan alanların **“düşük”** benlik saygısına sahip oldukları kabul edilir.

“Ana–Baba İlgisi” alt testinde **“yüksek puan”** , ilgi **“azlığına”**; **“0-2”** puan ilginin çok olduğunu, **“3-4”** puan orta düzeyde ilgi olduğunu ve **“5-7”** puan ise ana-baba ilgisinin az olduğunu gösterir.

“Babayla İlişki” alt ölçeğinde **“yüksek puan”** , babayla ilişkinin **“fazla”** oluşuna; **“0-2”** puan babayla ilişki azlığına, **“3-4”** puan orta düzeyde bir ilişkiye, **“5-6”** puan da babayla ilişkinin fazla olduğuna ve

“Psşik İzolasyon” alt testinde ise **“yüksek puan”** , psşik izolasyonun **“fazlalığına”**; “1” puan izolasyonun az olduğuna, “2” puan izolasyonun çok olduğuna işaret eder.

4.2.3.3. Problem çözme envanteri

Heppner ve Peterson (1982) tarafından geliştirilen bu envanter; problem çözme süreci aşamalarını belirleyen, bireylerin kendi problem çözme davranışları ve yaklaşımları hakkında ne düşündüklerini değerlendiren, 35 maddeden oluşmuş 6’lı Likert tipi (1,“her zaman böyle davranırım.”, “2, çoğunlukla böyle davranırım.”, “3, sık sık böyle davranırım.”, “4, arada sırada böyle davranırım.”, “5, ender olarak böyle davranırım.”, 6, hiçbir zaman böyle davranmam.”) bir ölçektir. Problem Çözme Envanteri ergenlere ve yetişkinlere uygulanabilen, bireylerin kendi kendilerine cevaplandırabilecekleri, kolay uygulanabilen bir ölçektir. Ölçek maddelerine verilen cevaplar 1 ile 6 arasında değişen puanlarla puanlanmaktadır. Ölçek puanlanırken 9, 22. ve 29. maddeler puanlama dışı tutulur. 1, 2, 3, 4, 11, 13, 14, 15, 17, 21, 25, 26, 30. ve 34. maddeler ters olarak puanlanan maddelerdir. Bu maddelerin yeterli problem çözme becerilerini temsil ettiği varsayılır. Ölçek puanları 32-192 arasında değişmektedir.

Heppner ve Petersen (1982) tarafından geliştirilen Problem Çözme Envanteri’nin Türkçeye uyarlanması Nail Şahin ve Nesrin Hisli Şahin (1993) tarafından yapılmıştır. Ölçekten alınan toplam puanların yüksekliği bireyin problem çözme becerileri konusunda kendini yetersiz olarak algıladığını göstermektedir. Bu envanterin amacı, günlük yaşantınızdaki problemlerinize (sorunlarınıza) genel olarak nasıl tepki gösterdiğinizi belirlemeye çalışmaktır. Sözünü ettiğimiz bu problemler, matematik ya da fen derslerindeki alışmış olduğumuz problemlerden farklıdır. Bunlar, kendini karamsar hissetme, arkadaşlarla geçinmeme, bir mesleğe yönelme konusunda yaşanan belirsizlikler ya da boşanıp boşanmama gibi karar verilmesi zor konularda ve hepimizin başına gelebilecek türden sorunlardır (Hatay, 2008).

4.2.3.4. Sürekli-durum kaygı envanteri

Durumluk Kaygı (A – State) ; bireyin içinde bulunduğu stresli (baskılı) durumdan dolayı hissettiği sübjektif korkudur. Fizyolojik olarak da otonom sinir sisteminde meydana gelen bir uyarılma sonucu terleme, sararma, kızarma ve titreme gibi fiziksel değişmeler, bireyin gerilim ve huzursuzluk duygularının göstergeleridir. Stres’in

yoğun olduđu zamanlar durumluk kaygı seviyesinde yükselme stres ortadan kalkınca düşme olur.

Sürekli Kaygı (A- Trait) ; bireyin kaygı yaşantısına yatkınlığıdır. Buna, kişinin içinde bulunduđu durumları genellikle stresli olarak algılama ya da stres olarak yorumlama eğilimi de denilebilir. Bu tür kaygı seviyesi yüksek olan bireylerin kolaylıkla incindikleri ve karamsarlığa büründükleri görülür. Bu bireyler durumluk kaygıyı da diğerlerinden daha sık ve yoğun bir şekilde yaşarlar.

Durumluk Sürekli Kaygı Ölçeđi Spielberger ve arkadaşları (1970) tarafından geliştirilmiş, Öner ve Le Compte tarafından (1985) Türk toplumuna uyarlaması yapılmış, durumluk ve sürekli kaygı düzeylerini 20 soru ile ayrı ayrı ölçen likert tipi bir ölçektir. Yüksek puanlar yüksek kaygı seviyelerini, düşük puanlar düşük kaygı seviyelerini gösterir 1975 yılında Türkçeye çevrilerek geçerlik ve güvenirlik çalışması yapılmış olan ölçek yirmişer maddelik durumluk kaygı ve sürekli kaygı ölçeklerinden oluşmaktadır. Her iki ölçekten elde edilen toplam puan değeri 20-80 arasında değişir. Büyük puan yüksek kaygı seviyesini, küçük puan ise düşük kaygı seviyesini belirtir. “Hiç” ile “Tamamıyla” arasında değışen dört derecelik bir ölçektir. Envanterin Türkiye’deki geçerlilik ve güvenirliliđi N. Öner tarafından (1977) yapılmıştır. Durumluk - Sürekli Kaygı Envanterlerinde iki tür ifade vardır. Doğrudan ifadeler olumsuz duyguları, tersine dönmüş ifadeler ise olumlu duyguları dile getirir. Durumluk Kaygı Envanterindeki tersine dönmüş ifadeler 1,2,5,8,10,11,15,16,19 ve 20. maddelerdir. Sürekli Kaygı Envanterindeki tersine dönmüş ifadeler ise 21,26,27,30,33,36 ve 39 uncu maddeleri oluşturur. Doğrudan ve tersine dönmüş ifadelerin ayrı ayrı toplam ağırlıkları bulunduktan sonra doğrudan ifadeler için elde edilen toplam ağırlık puanından, ters ifadelerin toplam ağırlık puanı çıkarılır. Bu sayıya, önceden saptanmış ve değışmeyen bir değeri eklenir. Durumluk Kaygı Envanteri için bu değışmeyen değeri 50, Sürekli Kaygı Envanteri için 35’dir. En son elde edilen değeri bireyin kaygı puanıdır. Durumluk Kaygı Ölçeđi (DKÖ), ani değışiklik gösteren heyecansal reaksiyonları değeriendirmede oldukça duyarlı bir araçtır. Envanterin ikinci bölümünde yer alan yine 20 maddeden oluşan Sürekli Kaygı Ölçeđi (SKÖ), kişinin genelde, yaşama eğilimi gösterdiği kaygının sürekliliğini ölçmeyi amaçlamaktadır. Skorlar 20 (düşük anksiyete) ile 80 (yüksek anksiyete) arasındadır (Kaya ve Varol, 2004)

Durumluluk ve sürekli kaygı ölçeği puanlaması:

Üçten fazla ifadeye cevap verilmemişse, doldurulan form geçersiz sayılır ve puanlanmaz.

Durumluk kaygı ölçeğinde dört sınıfta toplanan cevap seçenekleri: (1) Hiç, (2) Biraz, (3) Çok ve (4) Tamamıyla şeklinde; Sürekli Kaygı Ölçeğindeki seçenekler ise (1) Hemen hiçbir zaman, (2) Bazen, (3) Çok zaman ve (4) Hemen her zaman şeklindedir.

Ölçeklerde iki türlü ifade bulunur. Bunlara (1) doğrudan ya da düz ve (2) tersine dönmüş ifadeler diyebiliriz. Doğrudan ifadeler, olumsuz duyguları; tersine dönmüş ifadeler ise olumlu duyguları dile getirir. Bu ikinci tür ifadeler puanlanırken 1 ağırlık değerinde olanlar 4 'e, 4 ağırlık değerinde olanlar ise 1'e dönüşür. Doğrudan ifadelerde 4 değerindeki cevaplar kaygının yüksek olduğunu gösterir. Tersine dönmüş ifadelerde ise 1 değerindeki cevaplar yüksek kaygıyı, 4 değerindekiler düşük kaygıyı gösterir. "Huzursuzum" ifadesi doğrudan, "kendimi sakin hissediyorum" ifadesi de tersine dönmüş ifadelere örnek olarak gösterilebilir. Bu durumda "huzursuzum" ifadesi için 4 ağırlıklı bir seçenek, "kendimi sakin hissediyorum" ifadesi için 1 ağırlıklı seçenek işaretlenmişse, bu cevaplar yüksek kaygıyı yansıtmış olur. Durumluk kaygı ölçeğinde on tane tersine dönmüş ifade vardır. Bunlar 1, 2, 5, 8, 10, 11, 15, 16, 19 ve 20. Maddelerdir. Sürekli kaygı ölçeğinde ise tersine dönmüş ifadelerin sayısı yedidir ve bunlar 21, 26, 27, 30, 33, 36 ve 39. Maddeleri oluşturur.

Puanlama iki şekilde olur: Elle ya da bilgisayarla puanlama. Elle Puanlama: Doğrudan ve tersine dönmüş ifadelerin her biri için iki ayrı anahtar hazırlanır. Böylece bir anahtarla doğrudan ifadelerin, ikinci anahtarla tersine dönmüş ifadelerin toplam ağırlıkları hesaplanır. Doğrudan ifadeler için elde edilen toplam ağırlıklı puandan ters ifadelerin toplam ağırlıklı puanı çıkarılır. Bu sayıya önceden tespit edilmiş ve değişmeyen bir değer eklenir. Durumluk kaygı ölçeği için bu değişmeyen değer 50, Sürekli kaygı ölçeği için ise 35' tir. En son elde edilen değer bireyin kaygı puanıdır.

Puanların yorumlanması:

Her iki ölçekten elde edilen puanlar kuramsal olarak 20 ile 80 arasında değişir. Büyük puan yüksek kaygı seviyesini, küçük puan düşük kaygı seviyesini ifade eder. Uygulamalarda belirlenen ortalama puan seviyesi 36 ile 41 arasında değişmektedir.

5. BULGULAR VE TARTIŞMA

Bu bölümde araştırma amacının gerçekleştirilebilmesi için elde edilen verilerin istatistiksel olarak çözümlenmesine ve $\alpha=0.05$ önem seviyesinde elde edilen bulgulara yer verilmiştir. Veri madenciliğinde karar ağacı algoritmalarından biri olan CART algoritması tekniğinden yararlanılarak sonuçlar elde edilmiştir. Çalışmada bütünsel görsel modelleme gereci olan SPSS CLEMENTINE12.0 paket programı kullanılmıştır.

Çizelge 5.1’de Benlik saygısı, problem çözme, sürekli ve durum kaygı değişkenlerinin cinsiyet açısından değerlendirilmesine ilişkin sonuçlar verilmiştir. Sonuçlara bakıldığında kızlar ve erkeklerin durum kaygıları ve benlik saygısı puanları arasında istatistiksel olarak anlamlı bir fark olmamasına rağmen ($p>0.05$); sürekli kaygı ve problem çözme puanları arasında istatistiksel olarak önemli bir farklılık elde edilmiştir ($p<0.05$). Kızların hem sürekli kaygı puanları hem de problem çözme puanları erkeklerden daha yüksektir. Yani kızların yüksek sürekli kaygı puanlarına sahip olması, problem çözme becerilerini de etkileyerek erkeklerden daha düşük hale gelmesine sebep olabilir.

Çizelge 5.1. Benlik saygısı, problem çözme, sürekli ve durum kaygı değişkenlerinin cinsiyet açısından değerlendirilmesi

Cinsiyet		N	ortalama	Std.sapma	p-değeri
Benlik Saygısı Puanı	Erkek	684	1.0024	0.64416	0.819
	Kız	742	0.9947	0.62737	
Problem Çözme Puanı	Erkek	684	95.1170	14.76099	0.000*
	Kız	742	98.0593	14.37325	
Sürekli Kaygı Puanı	Erkek	684	46.6871	5.08710	0.000*
	Kız	742	49.0404	5.24267	
Durum Kaygı Puanı	Erkek	684	48.9225	3.71867	0.364
	Kız	742	48.7480	3.54609	

* $p<0.05$

Çizelge 5.2’de benlik saygısı, problem çözme, sürekli ve durum kaygı değişkenlerinin bilgisayara sahip olup-olmama açısından değerlendirilmesine ilişkin sonuçlar verilmiştir. Buna göre; bilgisayarı olan ile bilgisayarı olmayan öğrencilerin

problem çözme, sürekli kaygı ve durum kaygı puanları arasında istatistiksel olarak anlamlı bir fark olmamasına rağmen ($p>0.05$); benlik saygısı puanları arasında istatistiksel olarak önemli bir farklılık elde edilmiştir ($p<0.05$). Kendine ait bilgisayarı olan öğrencilerin benlik saygısı puanları daha düşüktür.

Çizelge 5.2. Benlik saygısı, problem çözme, sürekli ve durum kaygı değişkenlerinin bilgisayara sahip olup-olmama açısından değerlendirilmesi

Kendine ait bilgisayar var mı		N	ortalama	Std.sapma	p-değeri
Benlik Saygısı Puanı	Evet	1044	0.9554	0.61216	0.000*
	Hayır	382	1.1159	0.68163	
Problem Çözme Puanı	Evet	1044	96.3238	14.66934	0.167
	Hayır	382	97.5340	14.50215	
Sürekli Kaygı Puanı	Evet	1044	47.8151	5.36504	0.256
	Hayır	382	48.1754	5.11168	
Durum Kaygı Puanı	Evet	1044	48.9272	3.56859	0.100
	Hayır	382	48.5707	3.78406	

* $p<0.05$

Çizelge 5.3’de benlik saygısı, problem çözme, sürekli ve durum kaygı değişkenlerinin okunan Fakülte ve MYO’a göre değerlendirilmesine ilişki sonuçları verilmiştir. Buna göre; farklı fakültelerde okuyan öğrencilerin sürekli kaygı puanları arasında istatistiksel olarak anlamlı bir fark olmamasına rağmen ($p>0.05$); benlik saygısı puanı, problem çözme puanı ve durum kaygı puanları arasında istatistiksel olarak önemli bir farklılık elde edilmiştir($p<0.05$).

Benlik saygısı puanı en yüksek grup Spor Bilimleri Fakültesi öğrencileri iken en düşük grup ise fen edebiyat fakültesi öğrencileridir.

Problem çözme puanı en yüksek grup Sağlık Bilimler öğrencileri iken en düşük grup ise Sağlık MYO öğrencileridir. Örnekleme yer alan sağlık bilimleri öğrencileri genel itibariyle 2,3,4. sınıf öğrencileri iken Sağlık MYO öğrencilerinin tamamına yakını 2. sınıf öğrencilerinden oluşması Sağlık MYO öğrencilerinin problem çözme becerilerinin sağlık bilimleri öğrencilerine göre daha iyi olmasına sebep olmuş olabilir. Durum kaygı puanı da bunu desteklemektedir.

Durum kaygısı en yüksek grup Fen Edebiyat öğrencileri iken en düşük puanlı grup ise MYO öğrencileridir. Sağlık MYO öğrencileri meslek edinme sürecinin en

sonunda olmalarından dolayı mesleki problemlere daha serinkanlı ve akılcı yaklaşmayı öğrenmişlerdir.

Çizelge 5.3. Benlik saygısı, problem çözme, sürekli ve durum kaygı değişkenlerinin okunan fakülteye göre değerlendirilmesi

Fakülte		N	Ortalama	Std.sapma	p-değeri
Benlik Saygısı Puanı	Eğitim	315	0.9708	0.61038	0.025*
	Fen Edebiyat	188	0.8971	0.62457	
	Mühendislik	316	1.0080	0.69244	
	Sağlık Bilimleri	233	0.9743	0.58604	
	Sağlık MYO	60	1.1208	0.66234	
	İİBF	91	1.0962	0.71200	
	Spor Bilimleri	90	1.1598	0.63180	
	İlahiyat	133	0.9948	0.55699	
Problem Çözme Puanı	Eğitim	315	97.7841	13.33394	0.028*
	Fen Edebiyat	188	96.3457	14.54452	
	Mühendislik	316	94.5728	13.38031	
	Sağlık Bilimleri	233	98.2403	14.25212	
	Sağlık MYO	60	93.4500	19.05249	
	İİBF	91	95.9451	15.50868	
	Spor Bilimleri	90	97.1000	18.57330	
	İlahiyat	133	98.1429	14.81203	
Durum Kaygı Puanı	Eğitim	315	48.9397	3.48877	0.000*
	Fen Edebiyat	188	49.8191	3.33500	
	Mühendislik	316	48.4905	3.60818	
	Sağlık Bilimleri	233	48.9056	3.64593	
	Sağlık MYO	60	48.0167	4.10660	
	İİBF	91	48.2637	3.66920	
	Spor Bilimleri	90	48.1000	4.21087	
	İlahiyat	133	49.1128	3.42596	
Sürekli Kaygı Puanı	Eğitim	315	48.1429	5.01163	0.157
	Fen Edebiyat	188	47.1383	5.56363	
	Mühendislik	316	47.6646	5.28984	
	Sağlık Bilimleri	233	48.1545	5.67832	
	Sağlık MYO	60	48.7833	4.83662	
	İİBF	91	47.2527	5.45607	
	Spor Bilimleri	90	48.2111	5.38515	
	İlahiyat	133	48.4737	4.85015	

*p<0.05

Çizelge 5.4’de benlik saygısı, problem çözme, sürekli ve durum kaygı değişkenlerinin sınıflara göre değerlendirilmesine ilişkin sonuçları verilmiştir. Sonuçlara bakıldığında problem çözme puanı ve durum kaygı puanının sınıflara göre farklılaşmadığı ($p>0.05$); ancak benlik saygısı puanı ve sürekli kaygı puanında sınıflara göre önemli bir farklılık olduğu görülmüştür ($p<0.05$). Bu farklılığı yaratan grubu bulmak için çoklu karşılaştırma testlerinden LSD ve Tukey testi yapılmıştır. Benlik saygısı puanı en yüksek grup birinci sınıf öğrencileri iken en düşük grup ise 6. sınıf öğrencileridir. Sürekli kaygısı en yüksek grup birinci sınıf öğrencileri iken en düşük puanlı grup ise altıncı sınıf öğrencileridir. Son sınıf öğrencileri çevreyi daha iyi tanıdığı ve en kıdemli oldukları için psikolojik olarak daha rahattırlar.



Çizelge 5.4. Benlik saygısı, problem çözme, sürekli ve durum kaygı değişkenlerinin sınıflara göre değerlendirilmesi

Sınıf		N	Ortalama	Std.sapma	p-değeri
Benlik Saygısı Puanı	1	36	1.2242	0.77151	0.002*
	2	543	1.0391	0.61386	
	3	468	1.0189	0.65809	
	4	323	0.8971	0.62650	
	5	39	0.9338	0.53004	
	6	17	0.7259	0.43484	
Problem Çözme Puanı	1	36	96.3333	13.89758	0.649
	2	543	96.4438	15.56868	
	3	468	96.1496	14.46161	
	4	323	97.3467	13.36296	
	5	39	98.0256	14.63321	
	6	17	101.1176	13.30358	
Durum Kaygı Puanı	1	36	48.3611	4.84121	0.449
	2	543	48.7017	3.56998	
	3	468	48.7607	3.79936	
	4	323	49.1115	3.39711	
	5	39	49.2821	3.06893	
	6	17	49.5882	3.24150	
Sürekli Kaygı Puanı	1	36	48.8889	5.39194	0.011*
	2	543	48.0295	5.30048	
	3	468	48.3547	5.44057	
	4	323	47.2105	5.14319	
	5	39	46.7692	4.11967	
	6	17	45.8235	4.77278	

*p<0.05

Çizelge 5.5’de Benlik saygısı, problem çözme, sürekli ve durum kaygı değişkenlerinin anne eğitim durumuna göre değerlendirilmesine ilişkin sonuçları verilmiştir.

Buna göre problem çözme puanı, durum kaygı puanı ve sürekli kaygı puanlarına göre anne eğitim düzeylerine göre istatistiksel olarak birbirlerinden farklılaşmamasına rağmen ($p>0.05$); benlik saygısı puanının anne eğitim düzeyleri arasında istatistiksel olarak önemli bir farklılık elde edilmiştir ($p<0.05$).

Benlik saygısı puanı en yüksek olanlar sadece okur-yazar anneye sahip olanlar iken puanı en düşük olanlar ise anne eğitimi ilköğretim düzeyinde olan gruptur. Kendileri okuyamamış annelerin çocuklarını cesaretlendirdiği ve eğitim süreçlerine katkı verdikleri düşünülebilir.

Çizelge 5.5. Benlik saygısı, problem çözme, sürekli ve durum kaygı değişkenlerinin anne eğitim durumu düzeylerine göre değerlendirilmesi

Anne Eğitim Durumu			N	Ortalama	Std.sapma	p-değeri
Benlik Saygısı Puanı	Okuma Yazma	Bilmiyor	69	1.1649	0.59891	0.024*
		Okuryazar	39	1.1900	0.75805	
		İlköğretim	715	1.7171	1.07499	
		Ortaokul	202	1.0341	0.60463	
		Lise (Fen-Anadolu)	230	0.9591	0.62405	
		Meslek veya Teknik Lise	17	1.1453	0.61684	
		Ön lisans	35	0.8997	0.59853	
		Lisans	101	0.9968	0.65612	
		Lisansüstü	18	1.0794	0.66066	
Problem Çözme Puanı	Okuma Yazma	Bilmiyor	69	95.0290	15.35748	0.694
		Okuryazar	39	98.0256	13.61208	
		İlköğretim	715	195.003	28.29114	
		Ortaokul	202	96.1931	13.36429	
		Lise (Fen-Anadolu)	230	96.0739	16.58770	
		Meslek veya Teknik Lise	17	95.2353	11.31111	
		Ön lisans	35	100.3714	14.03585	
		Lisans	101	96.6040	13.72667	
		Lisansüstü	18	92.2222	17.72853	

*p<0.05

Çizelge 5.5. Benlik saygısı, problem çözme, sürekli ve durum kaygı değişkenlerinin anne eğitim durumu düzeylerine göre değerlendirilmesi (devam)

Anne Eğitim Durumu		N	Ortalama	Std.sapma	p-değeri
Durum Puanı	Kaygı	Okuma Yazma Bilmiyor	69	49.0290	3.72579
		Okuryazar	39	48.3590	3.48294
		İlköğretim	715	96.4433	7.07762
		Ortaokul	202	48.7376	3.58902
		Lise (Fen-Anadolu)	230	48.7391	3.94302
		Meslek veya Teknik Lise	17	48.8824	4.85904
		Ön lisans	35	49.2286	3.81241
		Lisans	101	49.270	3.28190
		Lisansüstü	18	49.7778	2.15722
		Okuma Yazma Bilmiyor	69	48.8841	5.64003
Sürekli Puanı	Kaygı	Okuryazar	39	47.6923	4.73038
		İlköğretim	715	94.7534	10.09273
		Ortaokul	202	48.0891	5.27757
		Lise (Fen-Anadolu)	230	47.6478	5.90036
		Meslek veya Teknik Lise	17	46.0000	5.73367
		Ön lisans	35	48.4286	4.08934
		Lisans	101	47.4455	5.50177
		Lisansüstü	18	47.3889	5.56395

*p<0.05

Çizelge 5.6'da benlik saygısı, problem çözme, sürekli ve durum kaygı değişkenlerinin öğrencilerin günlük internette geçirdikleri zamana göre değerlendirilmesine ilişki sonuçları verilmiştir.

Tabloda problem çözme puanı, durum kaygı puanı ve sürekli kaygı puanı öğrencilerin günlük internette geçirdikleri zaman düzeyleri açısından istatistiksel olarak anlamlı bir farklılık oluşturmamasına rağmen ($p>0.05$); benlik saygısı puanı öğrencilerin internette geçirdikleri zaman düzeylerine göre istatistiksel olarak önemli bir farklılık göstermektedir ($p<0.05$).

İnternet kullanmayan öğrencilerin benlik saygısı puanı en yüksek iken, interneti günde 1 ila 3 saat arası kullananların benlik saygı puanı en düşüktür. Öğrencilerin internet yerine kitap, dergi vb. yazılı eser okuması bilgi ihtiyacını karşılarken benlik saygısı üzerinde de olumlu etki bırakıyor olabilir.

Çizelge 5.6. Benlik saygısı, problem çözme, sürekli ve durum kaygı değişkenlerinin öğrencilerin günde kullandıkları internet miktarına göre değerlendirilmesi

Günlük internet kullanımı	N	Ortalama	Std.sapma	p-değeri
Benlik Saygısı Puanı	1 saat	252	0.9725	0.010*
	1 ile 3 arası saat	562	0.9476	
	3 saatten fazla	584	1.0471	
	hiç kullanmıyorum	28	1.2343	
Problem Çözme Puanı	1 saat	252	96.5516	0.309
	1 ile 3 arası saat	562	97.5107	
	3 saatten fazla	584	95.9058	
	hiç kullanmıyorum	28	95.6786	
Durum Kaygı Puanı	1 saat	252	48.8333	0.587
	1 ile 3 arası saat	562	48.9520	
	3 saatten fazla	584	48.7483	
	hiç kullanmıyorum	28	48.1429	
Sürekli Kaygı Puanı	1 saat	252	47.9683	0.110
	1 ile 3 arası saat	562	47.5552	
	3 saatten fazla	584	48.1592	
	hiç kullanmıyorum	28	49.3929	

*p<0.05

Çizelge 5.7’de benlik saygısı, problem çözme, sürekli ve durum kaygı değişkenlerinin aile ilişkisine göre değerlendirilmesine ilişki sonuçları verilmiştir.

Tabloda problem çözme puan ortalamaları aile ilişkisi düzeylerine göre farklılaşmamaktadır ($p>0.05$). Benlik saygısı puanı durum kaygı puanı ve sürekli kaygı puanı ise aile ilişkisi düzeylerine göre farklılaşmaktadır ($p<0.05$).

Ailesinin kendisini anlamadığını söyleyen öğrencilerin benlik saygı puanı en yüksek iken, ailesi ile çok iyi anlaşılan öğrencilerin benlik saygısı puanı düşüktür.

Ailesi ile çok iyi anlaşılan öğrencilerin durum kaygı puanı en yüksek iken, beni anlamıyorlar cevabını veren öğrencilerin durum kaygı puanı en düşüktür.

Beni anlamıyorlar cevabını veren öğrencilerin sürekli kaygı puanı en yüksektir.

Çizelge 5.7. Benlik saygısı, problem çözme, sürekli ve durum kaygı değişkenlerin aile ilişki düzeylerine göre değerlendirilmesi

Aile ile ilişki düzeyleri		N	Ortalama	Std.sapma	p-değeri
Benlik Puanı	Saygısı	Çok iyi anlaşıyoruz	1180	0.9657	0.62360
		Beni anlamıyorlar	50	1.4094	0.65912
		Kuşak çatışması var	80	1.1933	0.66779
		Diğer	116	1.0196	0.64119
Problem Puanı	Çözme	Çok iyi anlaşıyoruz	1180	96.6093	14.88055
		Beni anlamıyorlar	50	94.5800	15.25978
		Kuşak çatışması var	80	97.1125	13.23497
		Diğer	116	97.6121	12.62349
Durum Puanı	Kaygı	Çok iyi anlaşıyoruz	1180	48.9915	3.59293
		Beni anlamıyorlar	50	47.6400	4.79694
		Kuşak çatışması var	80	48.6000	3.19652
		Diğer	116	47.8793	3.49945
Sürekli Puanı	Kaygı	Çok iyi anlaşıyoruz	1180	47.7178	5.20105
		Beni anlamıyorlar	50	50.3400	6.22572
		Kuşak çatışması var	80	49.9625	5.68018
		Diğer	116	47.4224	5.08048

*p<0.05

Çizelge 5.8’de benlik saygısı, problem çözme, sürekli ve durum kaygı değişkenlerinin günde TV başında geçirilen zamana göre değerlendirilmesine ilişkin sonuçlar verilmiştir. Buna göre benlik saygı puanı, durum kaygı puanı ve sürekli kaygı puan ortalamaları öğrencilerin günde TV başında geçirdikleri zamana göre farklılaşmamaktadır ($p>0.05$). Problem çözme puan ortalamaları ise öğrencilerin günde TV başında geçirilen zaman düzeylerine göre farklılık göstermektedir ($p<0.05$).

Günde 1 ile 3 saat arası TV izleyen öğrencilerin problem çözme puanı en yüksek iken 3 saatten fazla TV izleyen öğrencilerin problem çözme puanı en

düşüktür. Yani 3 saatten fazla TV seyreden öğrenciler problem çözme yeteneklerini izledikleri programlara göre daha fazla geliştirmektedir.

Çizelge 5.8. Benlik saygısı, problem çözme, sürekli ve durum kaygı değişkenlerinin günde izlenen TV miktarına göre değerlendirilmesi

TV		N	Ortalama	Std.sapma	p-değeri
Benlik Saygısı Puanı	1 saat	372	0.9823	0.64682	0.465
	1ile 3 arası saat	252	0.9621	0.58251	
	3 saatten fazla	56	1.0902	0.65475	
	Hiç kullanmıyorum	746	1.0118	0.64515	
Problem Çözme Puanı	1 saat	372	96.8091	14.49765	0.008*
	1ile 3 arası saat	252	98.6032	16.08051	
	3 saatten fazla	56	91.5536	17.33931	
	Hiç kullanmıyorum	746	96.2895	13.85648	
Durum Kaygı Puanı	1 saat	372	48.9489	3.58945	0.092
	1ile 3 arası saat	252	49.0913	3.42064	
	3 saatten fazla	56	49.5893	4.34753	
	Hiç kullanmıyorum	746	48.6287	3.65127	
Sürekli Kaygı Puanı	1 saat	372	47.8172	5.39980	0.663
	1ile 3 arası saat	252	47.6270	5.10407	
	3 saatten fazla	56	48.3750	6.18374	
	Hiç kullanmıyorum	746	48.0201	5.24676	

*p<0.05

BEHİSAĞBİSİM

Node 0

Category	%	n
benlik saygı puanı yüksek	61.651	802
benlik saygı puanı orta	38.149	544
benlik saygı puanı düşük	0.000	0
Total	100.000	1428

annebabılgönd

İyi çok

Node 1

Category	%	n
benlik saygı puanı yüksek	83.250	833
benlik saygı puanı orta	16.750	494
benlik saygı puanı düşük	0.000	0
Total	82.558	1317

surekikaygılgönd

İyi orta, iyi az

Node 2

Category	%	n
benlik saygı puanı yüksek	44.954	49
benlik saygı puanı orta	55.046	60
benlik saygı puanı düşük	0.000	0
Total	7.844	109

otayyet

orta

Node 3

Category	%	n
benlik saygı puanı yüksek	77.868	108
benlik saygı puanı orta	22.132	31
benlik saygı puanı düşük	0.000	0
Total	9.746	139

yüksek

Node 4

Category	%	n
benlik saygı puanı yüksek	81.545	725
benlik saygı puanı orta	18.455	463
benlik saygı puanı düşük	0.000	0
Total	82.809	1178

faikute

etnik

Node 5

Category	%	n
benlik saygı puanı yüksek	37.500	24
benlik saygı puanı orta	62.500	40
benlik saygı puanı düşük	0.000	0
Total	4.489	64

faikute

İyi

Node 6

Category	%	n
benlik saygı puanı yüksek	55.556	25
benlik saygı puanı orta	44.444	20
benlik saygı puanı düşük	0.000	0
Total	3.156	41

faikute

EĞİTİM, FEN EDEBİYAT, MÜHENDİSLİK, SAĞLIK

Node 7

Category	%	n
benlik saygı puanı yüksek	64.084	650
benlik saygı puanı orta	35.916	313
benlik saygı puanı düşük	0.000	0
Total	61.080	871

anneveababılgöndürümü

MYD, İBF, SPOR FAKÜLTESİ, HUKUK, İLAHİYAT

Node 8

Category	%	n
benlik saygı puanı yüksek	64.387	187
benlik saygı puanı orta	35.613	140
benlik saygı puanı düşük	0.000	0
Total	21.539	307

babılgönd

EĞİTİM, FEN EDEBİYAT

Node 9

Category	%	n
benlik saygı puanı yüksek	61.111	11
benlik saygı puanı orta	38.889	7
benlik saygı puanı düşük	0.000	0
Total	1.282	18

MÜHENDİSLİK, SAĞLIK, MYD, İBF, SPOR FAKÜLTESİ, HUKUK, İLAHİYAT

Node 10

Category	%	n
benlik saygı puanı yüksek	20.261	13
benlik saygı puanı orta	71.739	33
benlik saygı puanı düşük	0.000	0
Total	3.228	46

EĞİTİM, FEN EDEBİYAT

Node 11

Category	%	n
benlik saygı puanı yüksek	42.105	6
benlik saygı puanı orta	57.895	11
benlik saygı puanı düşük	0.000	0
Total	1.332	19

MÜHENDİSLİK, SAĞLIK, MYD, İBF

Node 12

Category	%	n
benlik saygı puanı yüksek	0.000	0
benlik saygı puanı orta	0.000	0
benlik saygı puanı düşük	0.000	0
Total	0.000	0

Node 13

Category	%	n
benlik saygı puanı yüksek	0.000	0
benlik saygı puanı orta	0.000	0
benlik saygı puanı düşük	0.000	0
Total	0.000	0

Node 14

Category	%	n
benlik saygı puanı yüksek	0.000	0
benlik saygı puanı orta	0.000	0
benlik saygı puanı düşük	0.000	0
Total	0.000	0

benlikler

Node 15

Category	%	n
benlik saygı puanı yüksek	64.782	544
benlik saygı puanı orta	35.218	298
benlik saygı puanı düşük	0.000	0
Total	50.808	842

İyandılar, aynı yapıyorlar

Node 16

Category	%	n
benlik saygı puanı yüksek	45.161	14
benlik saygı puanı orta	54.839	17
benlik saygı puanı düşük	0.000	0
Total	2.174	31

İyi çok

Node 17

Category	%	n
benlik saygı puanı yüksek	59.489	146
benlik saygı puanı orta	41.511	100
benlik saygı puanı düşük	0.000	0
Total	17.381	246

İyi orta, iyi az

Node 18

Category	%	n
benlik saygı puanı yüksek	37.288	22
benlik saygı puanı orta	62.712	37
benlik saygı puanı düşük	0.000	0
Total	4.137	59

durumkaygılgönd

orta

Node 19

Category	%	n
benlik saygı puanı yüksek	45.000	9
benlik saygı puanı orta	55.000	11
benlik saygı puanı düşük	0.000	0
Total	1.403	20

yüksek

Node 20

Category	%	n
benlik saygı puanı yüksek	55.244	535
benlik saygı puanı orta	34.756	283
benlik saygı puanı düşük	0.000	0
Total	57.594	820

Benlik saygısı değişkeni için ağacın dallanmasında en etkili değişkenin Şekil 5.1’de görüldüğü üzere anne baba ilgisi olduğu görülmektedir. Çalışmaya katılan 1426 kişiden % 61.851 kişinin benlik saygısı yüksek, % 38.149 kişinin ise benlik saygı düzeyi orta ve düşük düzeyde olduğu söylenebilir. Anne baba ilgisi yüksek olanlarda benlik saygı düzeyi yüksek olduğu görülmektedir. Yani bu kişiler içerisinde benlik saygı düzeyi yüksek olanlar %63.250 ile çoğunluktadır. Benlik saygısı düzeyi orta ve düşük düzeyde olan kişilerin %36.750’sinin anne baba ilgisi yüksek düzeydedir. Anne baba ilgisinin orta veya az düzeyde olduğu durumda %44.954’ünün benlik saygısı yüksek,%55.046’sinin ise benlik saygı orta ve düşük düzeydedir. Öğrencilerin %92.356’sının anne baba ilgi düzeyi yüksektir.

Anne baba ilgisinin yüksek olduğu durumda benlik saygıyı etkileyen en önemli etken sürekli kaygı olduğundan karar ağacı dallanmaya buradan devam ederken, ilginin orta ve az olması durumunda benlik saygısını etkileyen en önemli etken cinsiyet olduğundan karar ağacı dallanmaya buradan devam edecektir.

Ağacın ilk dalında sürekli kaygısı yüksek olanlar içerisinde %61.545’inin benlik saygı düzeyi yüksekken, % 38.455’inin benlik saygısı orta ve düşük düzeydedir. Sürekli kaygı düzeyi yüksek olanlarda benlik saygı düzeyini etkileyen en önemli etken fakülte olduğundan karar ağacı dallanmaya buradan devam edecektir.

Eğitim, Fen Edebiyat, Sağlık Bilimleri ve Mühendislik Fakültesinde okuyanlar içerisinde %64.064’ünün benlik saygısı yüksek iken ,%35.936’sının benlik saygısı orta ve düşük düzeydedir.

İİBF, Spor Bilimleri, Hukuk, İlahiyat Fakülteleri ve Sağlık MYO’da okuyanlar içerisinde %54.397’sinin benlik saygısı yüksekken, %45.603’ünün benlik saygısı orta ve düşük düzeydedir.

Eğitim, Fen Edebiyat, Sağlık Bilimleri ve Mühendislik Fakültelerinde okuyanlarda benlik saygı düzeyini etkileyen en önemli etken anne babanın birliktelik durumu olduğundan karar ağacı dallanmaya buradan devam ederken, İİBF, Spor Bilimleri, Hukuk ve İlahiyat Fakülteleri ve Sağlık MYO’da okuyanlarda benlik saygı düzeyini etkileyen en önemli etken baba ilişkisi olduğundan karar ağacı dallanmaya buradan devam edecektir.

Anne babası birlikte olanların büyük çoğunluğunda yani % 64.762’sinin benlik saygısı yüksekken, %35.238’inin benlik saygısı orta ve düşük düzeydedir.

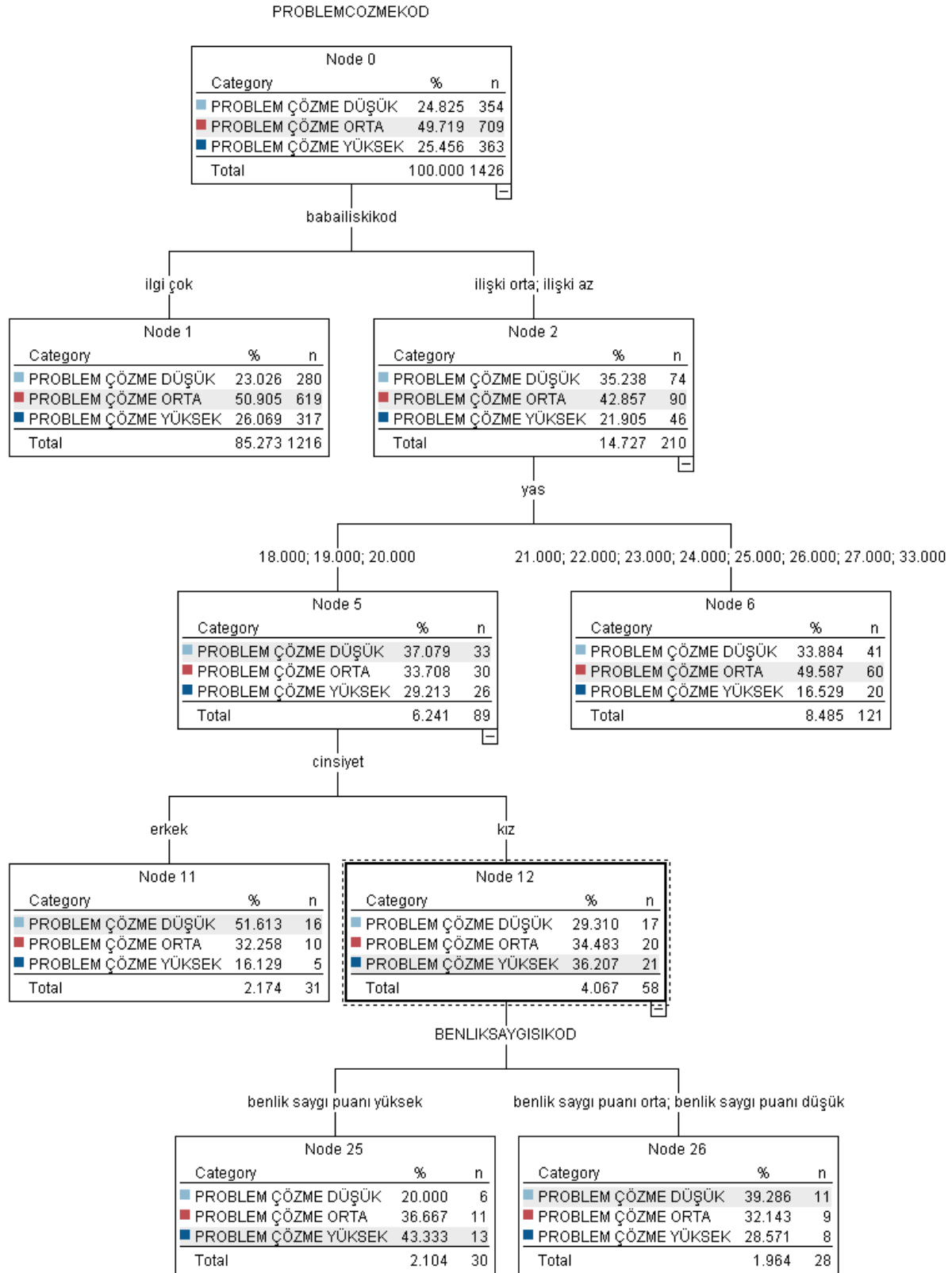
Anne baba birlikte olduğu durumda benlik saygıyı etkileyen en önemli etken durum kaygı olduğundan karar ağacı dallanmaya buradan devam edecektir. Durum kaygı

puanı orta olanlar içerisinde %45'inin benlik saygısı yüksek iken çoğunluğunun yanı %55'inin benlik saygısı orta ve düşük düzeydedir. Durum kaygısı yüksek olanlar içerisinde %65.244'ünün benlik saygısı yüksek iken , %34.756'sinin benlik saygısı orta ve düşük düzeydedir.

İİBF, Spor Bilimleri, Hukuk ve İlahiyat Fakülteleri ve Sağlık MYO'da okuyanlarda benlik saygısı değişkenini etkileyen en önemli etken olan baba ile ilişki durumunda ilişkinin yüksek düzeyde olduğu durumda %58.468'inin benlik saygısı yüksek iken,%41.532'sinin benlik saygısı orta ve düşük düzeydedir.

Ağacın ikinci dalında anne baba ilgisi orta ve az düzeyde olanlar içerisinde %44.954'ünün benlik saygısı yüksek iken ,% 55.046'sının benlik saygısı orta ve düşük düzeydedir. Anne baba ilgisinde ilginin orta ve az olduğu durumda benlik saygısını etkileyen en önemli etken cinsiyet olduğundan karar ağacı dallanmaya buradan devam edecektir. Erkekler içerisinde % 37.5'inin benlik saygısı yüksek iken, büyük çoğunluğunun yani % 62.5'inin benlik saygısı orta ve düşük düzeydedir. Erkeklerde benlik saygısını etkileyen en önemli etken fakülte olduğundan karar ağacı dallanmaya buradan devam edecektir. Eğitim ve Fen Edebiyatı okuyanlar içerisinde % 61.111'inin benlik saygısı yüksekken, %38.889'unun benlik saygısı orta ve düşük düzeydedir.

Kızların içerisinde % 55.556'inin benlik saygısı yüksekken, % 44.444'ünün benlik saygısı orta ve düşük düzeydedir. Kızlarda benlik saygıyı etkileyen en önemli etken fakülte olduğundan karar ağacı dallanamaya buradan devam edecektir. Mühendislik, Sağlık Bilimleri, İİBF, Spor Bilimleri, Hukuk ve İlahiyat Fakültesinde okuyanlar ile Sağlık MYO da okuyanlar içerisinde % 65.385'inin benlik saygısı yüksek iken % 34.615'inin benlik saygısı orta düzeydedir.



Şekil 5.2. Bağımlı değişken olarak problem çözme alındığında CART algoritması Clementine ekran çıktısı

Problem çözme değişkeni için ağacın dallanmasında en etkili değişkenin Şekil 5.2’de görüldüğü üzere baba ile ilişki olduğu görülmektedir. Problem çözme envanterinde puanı yüksek olan öğrencilerin problem çözme becerisi düşüktür. Tersine olarak da problem çözme envanterinde puanı düşük olan öğrencilerin problem çözme becerisi yüksektir. Buna göre çalışmaya katılan 1426 kişiden % 24.825 kişinin problem çözme becerisi yüksek, % 49.719 kişinin problem çözme becerisi orta düzeyde ve % 25.456’sının ise problem çözme becerisi düşüktür.

Baba ile ilişki düzeyi yüksek olanların çoğunda problem çözme becerisinin orta düzeyde olduğunu söyleyebilir. Yani bu kişiler içerisinde problem çözme becerisi orta düzeyde olanlar %50.905 ile çoğunluktadır. Problem çözme becerileri yüksek düzeyde olan kişilerin %23.026’sının babası ile ilişkisi yüksek düzeydedir.

Baba ile ilişkisi orta ve az düzeyde olan kişilerin %35.238’inin problem çözme becerisi yüksek,% 42.857’sinin problem çözme becerisi orta düzeyde ve %21.905’inin ise problem çözme becerisi düşüktür. Öğrencilerin %85.273’ünün babası ile olan ilişkileri yüksek düzeydedir.

Baba ile ilişkinin orta ve az düzeyde olmasını etkileyen en önemli etken yaş olduğundan karar ağacı dallanmaya buradan devam edecektir. 18- 20 yaş arasında olan bireyler içerisinde %37.079’unun problem çözme becerisi yüksek, %33.708’inin problem çözme becerisi orta düzeyde ve%29.213’ünün problem çözme becerisi düşüktür. 21-33 yaş arasında olan bireyler içerisinde %33.884’unun problem çözme becerisi yüksek, %49.587’inin problem çözme becerisi orta düzeyde ve %16.529’inin ise problem çözme becerisi düşüktür.

18-20 yaş arasında olan bireylerde problem çözme becerisini etkileyen en önemli etken cinsiyet olduğundan karar ağacı dallanmaya buradan devam edecektir. Kızlar içerisinde %29.310’unun problem çözme becerisi yüksek, %34.483’ünün problem çözme becerisi orta düzeyde ve %36.207’sinin problem çözme becerisi düşüktür. Kızların problem çözme becerilerini etkileyen en önemli etken benlik saygısı olduğundan karar ağacı dallanmaya buradan devam edecektir. Benlik saygısı yüksek olanlar içerisinde %20’sinin problem çözme becerisi yüksek,%36.667’sinin problem çözme becerisi orta düzeyde ve %43.333’ünün problem çözme becerisi düşüktür.

Anne baba ilgisi çok, sürekli kaygısı orta ve düşük olan öğrencilerin benlik saygısı puanı yüksek olması olasılığı 0.07573 iken orta ve düşük olması olasılığı ise 0.02174 tür.

Anne baba ilgisi çok, sürekli kaygısı yüksek, Eğitim, Fen Edebiyat, Mühendislik ve Sağlık Bilimleri Fakültelerinde okuyan, anne babası birlikte yaşayan, durum kaygısı orta ve düşük olan öğrencilerin benlik saygısı puanı yüksek olması olasılığı 0.0063 iken orta ve düşük olması olasılığı ise 0.007714'dür.

Anne baba ilgisi çok, sürekli kaygısı yüksek, Eğitim, Fen Edebiyat, Mühendislik ve Sağlık Bilimleri Fakültelerinde okuyan, anne babası boşanmış veya ayrı yaşayan öğrencilerin benlik saygısı puanı yüksek olması olasılığı 0.00982 iken; orta ve düşük olması olasılığı ise 0.01192'dir.

Anne baba ilgisi çok, sürekli kaygısı yüksek, Eğitim, Fen Edebiyat, Mühendislik ve Sağlık Bilimleri Fakültelerinde okuyan, anne babası birlikte olan, durum kaygısı yüksek olan öğrencilerin benlik saygısı puanı yüksek olması olasılığı 0.3752 iken; orta ve düşük olması olasılığı ise 0.1999'dur.

Anne baba ilgisi çok, sürekli kaygısı yüksek, İİBF, Spor Bilimleri, Hukuk, İlahiyat Fakültesi ve Sağlık MYO'da okuyan, baba ilişkisi yüksek olan öğrencilerin benlik saygısı puanı yüksek olması olasılığı 0.10168 iken; orta ve düşük olması olasılığı ise 0.07223'dür.

Anne baba ilgisi çok, sürekli kaygısı yüksek, İİBF, Spor Bilimleri, Hukuk, İlahiyat Fakültesi ve Sağlık MYO'da okuyan, baba ilişkisi orta ya da az düzeyde olan öğrencilerin benlik saygısı puanı yüksek olması olasılığı 0.01543 iken; orta ve düşük olması olasılığı ise 0.02595'tir.

Anne baba ilgisi orta ya da az, erkek, Eğitim ve Fen Edebiyat Fakültelerinde okuyan öğrencilerin benlik saygısı puanı yüksek olması olasılığı 0.007714 iken; orta ve düşük olması olasılığı ise 0.00491'dir.

Anne baba ilgisi orta ya da az, erkek, Mühendislik, Sağlık Bilimleri, Sağlık MYO, İİBF, Spor Bilimleri, Hukuk ve İlahiyat Fakültelerinde okuyan öğrencilerin benlik saygısı puanı yüksek olması olasılığı 0.00912 iken; orta ve düşük olması olasılığı ise 0.02314'dür.

Anne baba ilgisi orta ya da az, kız, Eğitim, Fen Edebiyat Fakültelerinde okuyan öğrencilerin benlik saygısı puanı yüksek olması olasılığı 0.00561 iken; orta ve düşük olması olasılığı ise 0,007714'dür.

Anne baba ilgisi orta ya da az, kız, Mühendislik, Sağlık Bilimleri, Sağlık MYO, İİBF, Spor Bilimleri, Hukuk ve İlahiyat Fakültelerinde okuyan öğrencilerin benlik saygısı puanı yüksek olması olasılığı 0.0119; orta ve düşük olması olasılığı ise 0.0063'tür.

Problem çözme envanterinde puanı yüksek olan öğrencilerin problem çözme becerisi düşüktür. Tersine olarak da problem çözme envanterinde puanı düşük olan öğrencilerin problem çözme becerisi yüksektir. Bu bilgiye göre:

Baba ile ilişki düzeyi yüksek olan öğrencilerin problem çözme becerisinin yüksek olması olasılığı 0.19635 iken; orta düzeyde olma olasılığı 0.43408 düşük olma olasılığı 0.2223'tür.

Baba ile ilişki düzeyi orta veya az, 18-20 yaş arası, erkek öğrencilerin problem çözme becerisinin yüksek olması olasılığı 0.01122; orta düzeyde olması olasılığı 0.007013 iken düşük olması olasılığı 0.003506'dır.

Baba ile ilişki düzeyi orta veya az, 18-20 yaş arasında, kız, benlik saygısı puanı yüksek olan öğrencilerin problem çözme becerisinin yüksek olma olasılığı 0.004208, orta düzeyde olma olasılığı 0.007714 iken düşük olması olasılığı 0.009116'dır

Baba ile ilişki düzeyi orta veya az, 18-20 yaş arasında, kız, benlik saygısı puanı orta düzeyde veya düşük olan öğrencilerin problem çözme becerisinin yüksek olma olasılığı 0.007714; orta düzeyde olma olasılığı 0.007103 iken düşük olması olasılığı 0.00561'dir.

Baba ile ilişki düzeyi orta veya az, 21-33 yaş arasında olan öğrencilerin problem çözme becerisinin yüksek olma olasılığı 0.02875; orta düzeyde olma olasılığı 0.04208(iken düşük olması olasılığı % 0.01402'dir.

6. SONUÇ VE ÖNERİLER

Sosyal Bilimler alanında bireylerin doğumdan ölüme kadar geçen yaşam sürelerinde toplum içinde kendini bilmesi, kabul ettirmesi, başarılı olması ve yaşam standartlarını belirlemesi kendi özünden gelen genetik faktörler ile çevresel etkenlere bağlıdır. Bu bağlamda problem çözme, benlik saygısı, süreklilik ve durumluluk kaygı değişkenleri bu süreçlere etki eden etkenler olarak karşımıza çıkmaktadır.

Bu tez çalışmasında ele alınan konunun önemine binaen klasik istatistiksel yöntemlerin yanı sıra daha ileri düzeyde sosyal bilimlerdeki literatür taramasında rastlanamayan veri madenciliği algoritmalarının uygulaması çalışmanın özgün değerini oluşturmuştur.

Erkekler kızlara göre daha pratik çözüm üreten problem çözme becerisi yüksek, sürekli kaygısı düşük olan gruptur. Bu ise istatistiksel olarak önemli olmasına rağmen erkeklerin benlik saygısı puanının yüksek olmasına katkı vermektedir.

Bilgisayar sahibi olmayan öğrencilerin benlik saygısı puanı daha yüksektir. Bu öğrenciler muhtemeldir ki interneti daha az kullanan, daha çok gazete, dergi, kitap okuyan kesimdir.

Sporla uğraşan öğrencilerin daima antrenman yapmaları, başarı hislerinin yüksek olması ve beraberinde benlik saygısı puanlarının da yüksek olmasına sebep olmaktadır. Fen Edebiyat öğrencileri ise akademik olarak daha düşük puanlarla okudukları bölümlere geldikleri ve iş imkânları kısıtlı olması sebebiyle başarı hissi gelişmemiş bireylerden oluşmaktadır. Bu ise benlik saygısı puanlarının düşük olmasına sebeptir.

Problem çözme becerisi en yüksek olan Sağlık MYO öğrencileri lisede de mesleki eğitimin içindedir. Dolayısıyla uygulamalı olarak edindikleri mesleki tecrübe problem çözüm becerilerini de yükseltmektedir.

Durum kaygıları en yüksek olan öğrenci grubu Fen Edebiyat Fakültesinde okuyanlardır. Yine akademik olarak belli bir meslek edinmemiş olma, iş bulmadaki sıkıntılar bu grubun kaygını yükseltmektedir.

Benlik saygısı puanı ve sürekli kaygı puanı üst sınıflara gittikçe almaktadır. Öğrenciler yıllar geçtikçe mesleğe adım atacakları ilk heyecanları yaşarken başarı noktasındaki mesleki tecrübesizlikten kaynaklanan korkuyu da daha fazla hissetmektedir.

Kendi okuyamamış ve akademik mesleki hayatındaki anneler kendi çocuklarını destekleyerek, psikolojik olarak öğrencilerin kendilerine güven duygularını yetiştirmişlerdir.

Bilgisayar sahibi olmayan dolayısıyla internete bağımlı olmayan öğrencilerin daha çok yazılı kaynaklarda bilgi ihtiyaçlarını gidermesi benlik saygısı üzerine olumlu etki yapmaktadır.

Öğrencilerin aile ilişki düzeylerinin çok iyi olmaması bireyin kendini yetiştirmede daha çok çaba sarf etmesine ve kendi başına öğrenme yetisinin gelişimine dolayısıyla benlik saygısının daha yüksek olmasına sebep olmaktadır.

Benzer olarak aileleri ile çok iyi anlaşılamayanlar problem çözme becerilerinin yüksek olmasında etkilidir. Bunun sebebi ise desteksiz bir şekilde her şeyi kendileri öğrenmiş ve hayata uygulayarak tecrübe edinmişlerdir.

Aileleri ile çok iyi anlaşılanların durum kaygıları yüksekliği kendilerine her zaman destek olmuş ailelerin karşısında başarıma ve mahcup olmama zorunluluğunu beraberinde getirmektedir.

Aileleriyle hiç anlaşılamayanların ise sürekli kaygıları yüksektir. Ailesiyle anlaşılamayanlarında kendilerini ispat etmesi “kendi ayakları üzerinde durma” gibi bir zorunlulukları vardır.

Öğrencilerin benlik saygısı puanını etkileyen en önemli faktör anne ve babanın ilişkisidir. Çünkü öğrenci kendini birey hissedene kadar anne ve babasının tecrübelerinden faydalanacak ve kendi hayat kurallarını oluşturacaktır. Anne- baba ilgisi yüksek olan öğrencilerin benlik saygısı da yüksektir.

Sürekli kaygısı yüksek öğrenciler eğitim sürecinin farkında olan, farkındalık düzeyi yüksek öğrencilerdir. Sorumluluk bilinci ve başarıma zorunluluğu öğrencilerin sürekli kaygıları da yükseltirken benlik saygısı puanını da yükseltmektedir. Anne ve babası beraber olanlar da yine ilişki kaynaktan beslendikleri için kendilerini bilen, anlayan ve kendileri ile barışık kişilerdir.

Aile olma bilinci ve farkındalığı yüksek ailelerin çocuklarının benlik saygısı puanının da yüksek olması topluma faydalı bireyler yetiştirmede ailenin ne denli önemli olduğunu bir kez daha istatistiksel olarak da göstermektedir.

Benlik saygısına en önemli katkıyı yapan anne- baba ilgisinin yanı sıra problem çözmeye önemli katkıyı yapan baba ile ilişki düzeyinin olması “birey” yetiştirme ve “birey olabilme” nin yolunun aileden geçtiğini bir kez daha göstermektedir. İstatistiksel olarak da kanıtlanan bu gerçeğin üzerine, baba ile ilişki düzeyinin az

olmasını öğrencinin yaşı etkilemektedir. Kızlarda ise baba ilişki düzeyini etkileyen en önemli etkenin benlik saygısı olduğu görülmektedir. Bu bağlamda gerek benlik saygısı gerek problem çözme becerilerinin gelişmesinde demografik değişkenlerden çok bireyin babasıyla olan ilişkisi ve anne-babasının bireye olan ilgisi çok önemli bir rol oynamaktadır. Bu sebepten dolayı sağlıklı toplumların oluşturulması için sağlıklı bireylerin yetiştirilmesi elzemdir. Bu ise bireyin topluma kazandırılması, kendine ve çevresine faydalı olmasını sağlayacak süreçlerin aile ile birlikte üstlenilmesi, koordineli bir şekilde eğitim öğretim programlarının planlanarak uygulanması gerekliliğini ortaya koymaktadır. Ancak maddi, manevi, sosyal ve kültürel olarak aile kurumunu koruyup, yücelterek benlik saygısı ve problem çözme becerisi yüksek, bilinçli toplumlar oluşabilir. Bu ise geleceğimizin teminatıdır.



KAYNAKLAR

- Adak, M. F. ve Yurtay, N. 2013. Gini Algoritmasını Kullanarak Karar Ağacı Oluşturmayı Sağlayan Bir Yazılım Geliştirilmesi. *Sakarya Üniversitesi Bilişim Teknolojiler Dergisi*, 6:3.
- Akpınar, H. 2000. Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği. *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, 29: 1, 1-22.
- Aldana, W. A. (2000). Data Mining Industry: Emerging Trends and New Opportunities. Master of Engineering in Electrical Engineering and Computer Science at the Massachusetts Institute of Technology ,B.S. Electrical Engineering and Computer Science, 182, Cambridge, ABD.
- Bacanlı, H., İlhan, T. ve Aslan, S. 2009. Beş Faktör Kuramına Dayalı Bir Kişilik Ölçeğinin Geliştirilmesi: Sıfatlara Dayalı Kişilik Testi (SDKT). *Türk Eğitim Bilimleri Dergisi*, 7: 2, 261-279.
- Başaran, M. H., Taşgın, Ö., Sanioglu, A. ve Taşkın, A. K. 2009. Sporcularda Durumluluk ve Sürekli Kaygı Düzeylerinin Bazı Değişkenlere Göre İncelenmesi. *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 21, 533-542.
- Baykasoğlu, A. (2005). Veri Madenciliği ve Çimento Sektöründe Bir Uygulama. Akademik Bilişim '05-VII. Akademik Bilişim Konferansı, Gaziantep, Türkiye.
- Bidgoli, B. M. (2004). Data Mining For a Web Based Educational System, Doctoral Dissertation, Department of Computer Science and Engineering, Michigan State University, 208, Michigan, ABD.
- Biggs, D., De Ville, B. ve Suen, E. 1991. A method of choosing multi- way partitions for classication and decision trees. *Journal of Applied Statistics*, 18:1, 49-62.
- Bindak, R. 2005. İlköğretim Öğrencileri için Matematik Kaygı Ölçeği. *Fırat Üniversitesi Fen ve Mühendislik Bilimleri Dergisi* , 2: 17, 442-448.
- Breiman, L., Friedman, J., Olshen, R. ve Stone, C. 1984. *Classification and Regression Trees*. Belmont, California: Wadsworth International Group.
- Bounsaythip, C. ve Rinta-Runsala, E. (2001). Overview of Data Miningfor Customer Behaviour Modeling, VTT Information Technology Research Report TTEI, 29 June 2001, Espoo, Finland.
- Bölükbaş, N., Karabulut, N. ve Özer, H. 1998. Ameliyathane Ortamının Çalışan Ekip Üzerine Yarattığı Anksiyete Düzeyinin İncelenmesi. *Düşünen Adam: Psikiyatri ve Nörolojik Bilimler Dergisi*, 1: 11, 27-31.
- Chakrabarti, S. 2003. *Miningthe Web: Discovering Knowledge from Hypertext Data*, Morgan Kaufmann Publishers, San Francisco.
- Çalış, A., Kayapınar, S. ve Çetinyokuş, T. 2014. Veri Madenciliğinde Karar Ağacı Algoritmaları İle Bilgisayar ve İnternet Güvenliği Üzerine Bir Uygulama. *Gazi Üniversitesi Endüstri Mühendisliği Dergisi*, 25:3-4, 2-19
- Çam, S. ve Tümkaya, S. 2007. Kişiler Arası Problem Çözme Envanteri'nin (KPÇE) geliştirilmesi Geçerlilik ve Güvenirlilik Çalışması. *Türk Psikolojik Danışma ve Rehberlik Dergisi*, 3: 28.

- Çevik Demir, Ş. (2013). Ergenlerde Benlik Saygısı ve Dini Başa Çıkma. Doktora Tezi, Uludağ Üniversitesi, Sosyal Bilimler Enstitüsü Felsefe ve Din Bilimleri Anabilim Dalı, Bursa.
- Çuhadaroglu, F. (1985). Gençlerde benlik saygısı ile ilgili bir araştırma. XXI. Ulusal Psikiyatri ve Nörolojik Bilimler Kongresi Bilimsel Çalışmaları, 107-108.
- Doğan, K. ve Arslantekin, S. 2016. Büyük Veri: Önemi, Yapısı ve Günümüzdeki Durumu. *DTCF dergisi*, 56:1, 15-36
- Doğan, N. ve Özdamar, K. 2003. Chaid Analizi ve Aile Planlaması ile İlgili Bir Uygulama. *Türkiye Klinikleri Tıp Bilimleri Dergisi*, 23: 5, 392-397.
- Doğan, Ş., ve Türkoğlu, İ. 2007. Hypothyroidi and Hyperthyroidi Detection from Thyroid Hormone Parameters by Using Decision Trees. *Doğu Anadolu Bölgesi Araştırmaları Dergisi*, Cilt 5, No 2, 163-169.
- Dolgun, M. Ö., Özdemir, T. G. ve Oguz, D. 2009. Veri Madenciliği'nde Yapısal Olmayan Verinin Analizi: Metin ve Web Madenciliği. *İstatistikçiler Dergisi*, 2, 48-58.
- Dondurmacı G. (2011). Veri Madenciliğinde Regresyon Ağaçları ile Sınıflandırma ve Bir Uygulama, Yayınlanmamış Doktora Tezi. Mimar Sinan Güzel Sanatlar Üniversitesi Fen bilimleri Enstitüsü, 166, İstanbul.
- Elkin, N. ve Karadağlı, F. 2015. Üniversite Öğrencilerinin Problem çözme Becerilerinin Değerlendirilmesi. *Adıyaman Üniversitesi Sağlık Bilimleri Dergisi*, 1:1.
- Erduran, G. Y. (2017). Online Müşteri Şikayetlerinin Veri Madenciliği İle İncelenmesi. Doktora Tezi, Trakya Üniversitesi Sosyal Bilimler Enstitüsü İşletme Anabilim Dalı, 205, Edirne.
- Ertuğrul, İ., Organ, A. ve Şavlı, A. 2013. Veri Madenciliği uygulamasına İlişkin Paü Hastanesinde Hasta Profiline Belirlenmesi. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 19:2, 97-103.
- Fieldman, R. ve Sanger, J. 2006. *The Text Minig Hand Book – Advanced Approaches in Analyzing Unstructured Data*. UK: Cambridge University Press.
- Giudici, P. 2003. *Applied Data Mining: Statistical Methods for Business and Industry*. Faculty of Economics University of Pavia, John Wiley & Sons Ltd, Italy.
- Gülçe, G. (2010). Veri Ambarı ve Veri Madenciliği Teknikleri Kullanarak Öğrenci Karar Destek Sistemi Oluşturma. Yayınlanmamış Yüksek Lisans Tezi, Pamukkale Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, 109, Denizli.
- Gülpınar, V. (2008). Avrupa Birliği Ülkeleri ile Göstergelerin Karar Ağacı Yöntemi ile Karşılaştırılması. Yüksek Lisans Tezi, Marmara Üniversitesi Sosyal Bilimler Enstitüsü Ekonometri Anabilim Dalı, 152, İstanbul.
- Han, J. ve Kamber, M. 2001. *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers, San Francisco.
- Hand, D.J., 1998. *Data Mining: Statistics and More?*. The American Statistician, Cilt 52, 112-118.

- Hatay Polat, R. (2008). Sınıf Öğretmenliği Öğrencilerinin Bazı Sosyo- Demografik Özellikleri ve Düşünme İhtiyacına Göre Problem Çözme Becerilerinin İncelenmesi. Yayınlanmamış Yüksek Lisans Tezi, Çukurova Üniversitesi Sosyal Bilimler Enstitüsü İlköğretim Ana Bilim Dalı, 97, Adana.
- Heppner, P.P. ve Petersen, C.H. 1982. The development and implications of a personal problem - solving inventory. *Journal of Counseling Psychology*, 29, 66-75.
- Jacobs, P. 1999. Data Mining: What General Managers Need to Know. *Harvard Management Update*, Cilt 4, No 10, 8.
- Kara, H. ve Acet, M. 2012.Spor Yöneticilerinde Durumluluk Kaygısının Otomatik Düşünceler Üzerindeki Etkisinin İncelenmesi. *e-Journal of New World Sciences Academy*, 7:4.
- Karaibrahimoğlu, A. (2014). Veri Madenciliğinde Birliktelik Kuralı İle Onkoloji Verilerinin Analiz Edilmesi: Meram Tıp Fakültesi Onkoloji Örneği. Yayınlanmamış Doktora Tezi, Selçuk Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı, 126, Konya.
- Kass, G.V. 1980. An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Applied Statistics*, 29:2, 119-127.
- Kavzoğlu T. ve Çölkesen, İ. 2010. Karar Ağaçları İle Uydu Görüntülerinin Sınıflandırılması: Kocaeli Örneği. *Harita Teknolojileri Elektronik Dergisi*, 1, 36-45.
- Kaya, M. ve Varol, K. 2004. İlahiyat Fakültesi Öğrencilerinin Durumluluk-Sürekli Kaygı Düzeyleri ve Kaygı Nedenleri (Samsun Örneği). *Ondokuz Mayıs Üniversitesi İlahiyat Fakültesi Dergisi*, 17:12
- Koldere Akın, Y. (2008).Veri Madenciliğinde Kümeleme Algoritmaları ve Kümeleme Analizi. Yayınlanmamış Doktora Tezi, Marmara Üniversitesi Sosyal Bilimler Enstitüsü Ekonometri Anabilim dalı, 189, İstanbul.
- Koyuncuğil, A. S. ve Özgülbaş, N. 2006. İMKB’de İşlem Gören KOBİ’lerin Finansal Başarısızlığına Etki Eden Faktörlerin Veri Madenciliği İle Belirlenmesi. 3. KOBİ ve Verimlilik Kongresi, 17–18 Kasım, İstanbul.
- Koyuncuğil, A. S. (2007). *Borsa Şirketlerinin Sektörel Risk Profillerinin Veri Madenciliğiyle Belirlenmesi*. Sermaye Piyasası Kurulu Araştırma Raporu, Araştırma Dairesi, Ankara.
- Koyuncuğil, A. S. ve Özgülbaş, N. 2009. Veri Madenciliği: Tıp ve Sağlık Hizmetlerinde Kullanımı ve Uygulamaları. *Bilişim Teknolojileri dergisi*, 2: 2, 21.
- Köktürk ,F., Ankaralı, H. ve Sümbüloğlu, V. 2009. Veri Madenciliği Yöntemlerine Genel Bakış. *Journal of Biostatistics*, 1: 1, 20.
- Kurt, Ç. ve Erdem, A. 2012. Öğrenci Başarısını Etkileyen Faktörlerin Veri Madenciliği Yöntemiyle İncelenmesi. *Politeknik Dergisi*, 15:2, 111-116.
- Kusiak, A., Kernstine, K. H., Kern, J. A., Mc Laughlin, K. A. ve Tseng, T. L. (2000). Data Mining: Medicaland Engineering Case Studies. Proceedings of the Industrial Engineering Research 2000 Conference, May 21-23,1-7, Ohio, ABD.

- Macila, E. (2013). Basketbol Oyuncularının Durumluluk ve Süreklilik Kaygı Düzeylerinin Belirlenmesi. Yayınlanmamış Yüksek Lisans Tezi, Yakın Doğu Üniversitesi Sağlık Bilimleri Enstitüsü Beden Eğitimi ve Spor Programı Anabilim Dalı, 82, Lefkoşa.
- Mehta, M., Agrawal, R. ve Rissanen, J. (1996). SPRINT: A fast scalable classifier for data mining. InProc. Of the Fifth Int'l Conference on Extending Database Technology (EDBT), Avignon, France.
- Morgan, J.N. ve Sonquist, J.A. 1963. Problems in the analysis of survey data and a proposal. *JASA*, 58, 415-434. (Original AID).
- Morgan, J. N. ve Messenger, R. C. 1973. *THAID: a sequential search program for the analysis of nominal scale dependent variables*. University of Michigan, Institute for Social Research, ABD.
- Oğuzlar, A. 2004. CART Analizi İle Hane Halkı İşgücü Anketi Sonuçlarının Özetlenmesi. *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 18, 3-4.
- Öner, N. ve Le Compte, A., 1983. *Durumluluk-Süreklilik Kaygı Envanteri El Kitabı*. Boğaziçi Üniversitesi Yayınları, İstanbul, 1983.
- Öner, N. (1977). Durumluluk ve süreklilik kaygı envanterinin Türk toplumunda geçerliği Yayınlanmamış doktora tezi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Özçakır, F.C. ve Çamurcu, A.Y. 2007. Birliktelik Kuralı Yöntemi İçin Bir Veri Madenciliği Yazılımı Tasarımı ve Uygulama. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 6:12, 21-37.
- Quinlan, J.R. 1986. *Induction of decision trees*. Machine Learning, 1: 81, 106.
- Quinlan, J. R. 1993. *C4.5: Programs for Machine Learning*. San Mateo: Morgan Kaufmann Publishers, San Francisco.
- Pehlivan G. (2006). CHAID Analizi ve Bir Uygulama. Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, 69, İstanbul.
- Rosenberg, M. 1963. Parental Interest and Children's Self-Conceptions. *Sociometry*, 26, 35-49.
- Rosenberg, M. 1965. *Society and the adolescent self-image*. Princeton University Press, Princeton, N.J.
- Sahin, N., Sahin, N. ve Heppner, P. P. 1993. Psychometric properties of the Problem Solving Inventory in a group of Turkish university students. *Cognitive Therapy and Research*, 17, 379-396.
- Sarıcaoğlu, F., Akıncı, S.B., Gözaçan, A., Güner, B., Rezaki, M. ve Aypar, Ü. 2005. Gece ve Gündüz Vardiya Çalışmasının Bir Grup Anestezi Asistanının Dikkat ve Anksiyete Düzeyleri Üzerine Etkisi. *Türk Psikiyatri Dergisi*, 2: 16, 106-112.
- Savaş, S., Topaloğlu, N. ve Yılmaz, M. 2012. Veri Madenciliği ve Türkiye'deki Uygulama Örnekleri. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 21, 1-23
- Spielberger, C., Gorsuch, R. ve Lushene, R. 1970. *Manual for State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologist Press.

- Silahtaroğlu, G. 2013. *Kavram ve Algoritmalarıyla Temel Veri Madenciliği*. Papatya Yayıncılık, İstanbul.
- Taşgit, M.S. (2012). Üniversite Öğrencilerinin Benlik Saygısı ve Karar Verme Düzeylerinin İncelenmesi. Yayınlanmamış Yüksek Lisans Tezi, Karamanoğlu Mehmetbey Üniversitesi Sosyal Bilimler Enstitüsü Beden Eğitimi ve Spor Öğretmenliği Ana Bilim Dalı, 106, Karaman.
- Terlemez, L. (2008). Eş İşlem Stratejisi Yöntemiyle İMKB’ de Pörföy Oluşturmada Veri Madenciliği Uygulaması. Yayınlanmamış Doktora Tezi, Anadolu Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı, 152, Eskişehir.
- Tosun, T. (2006). Veri Madenciliği Teknikleriyle Kredi Kartlarında Müşteri Kaybetme Analizi. Yayınlanmamış Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, 51, İstanbul.
- Uysal, İ., Bilen, M. ve Ulukuş, S. (2014). Twoing Algoritması ile Sınıflandırma: Kalp Hastası Uygulaması. Akademik Bilişim’14 - XVI. Akademik Bilişim Konferansı Bildirileri 5 - 7 Şubat 2014 Mersin Üniversitesi ,127. Mersin, Türkiye.
- Vahaplar, A. ve İnceoğlu, M. (2001). Veri Madenciliği ve Elektronik Ticaret. VII. *Türkiye’de İnternet Konferansı*. 1-3 Kasım, Harbiye İstanbul.
- Vuslat, O. ve Köksal Akyol, A. 2015. Problem Çözme Becerisi Ölçeği (PÇBO) Geçerlilik ve Güvenirlilik Çalışması. *Çukurova Üniversitesi Eğitim Fakültesi Dergisi*, 44:1, 105-122.
- Yılmaz, H. (2014). Random Forests Yönteminde Kayıp Veri Probleminin İncelenmesi ve Sağlık Alanında Bir Uygulama. Yüksek Lisans Tezi, Eskişehir Osmangazi Üniversitesi Sağlık Bilimleri Enstitüsü, Biyoistatistik Anabilim Dalı, 92, Eskişehir.
- Yılmaz, N. (2015). Obez Bireylerde Benlik Saygısı, Sosyal Görüş Kaygısı ve Vücut Algısı. Tıpta Uzmanlık Tezi, Dokuz Eylül Üniversitesi, Aile Hekimliği Anabilim Dalı, 57, İzmir.
- Yiğit, H. (2010). Ergenlerin Benlik Saygılarının Yaşam Doyumu ve Bazı Özlük Nitelikleri Açısından İncelenmesi. Yüksek Lisans Tezi, Selçuk Üniversitesi Eğitim Bilimleri Enstitüsü Eğitim Bilimleri Anabilim Dalı, 112, Konya.
- Yüksel, B. (2014). Kaygı Belirtilerini Açıklamada Bağlama, Pozitif ve Negatif Duygu Düzenleme ve Belirsizlik Tahammülsüzlük Arasındaki İlişkiyi Bütünleyici Model Arayışı. Yüksek Lisans Tezi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü Psikoloji Anabilim Dalı, 136, Ankara.
- Loh, W. and Shih, Y. 1997. Split Slection Methods Classification Tress. *Statistica Sinica*, 7, 815-840.

EKLER

I.DEMOGRAFİK BİLGİLERİ

S.1. Cinsiyetiniz : 1 Erkek 2 Kadın

S.2. Yaşınızı belirtiniz.

S.3. Üniversiteye giriş puanınız?.....

S.4. Okuduğunuz bölüm?.....

S.5. Sınıf?..... Okul Numaranız:.....

S.6. Üniversite akademik ortalamanız GANO :.....

S.7. Siz dahil kaç kardeşiniz?.....

S.8. bitirdiğiniz okul aşağıdakilerden hangisidir?

Fen Lisesi İmam Hatip Lisesi Anadolu Lisesi Sosyal Bilimler Lisesi
Sağlık Meslek Lisesi Kız Meslek Lisesi Erkek Meslek Lisesi

S.9. Oturduğunuz konutun mülkiyet durumu nedir ?

Ev sahibi.....	1
Kiracı.....	2
Ev sahibi değil, kira ödemiyor..	3
Lojman.....	4
Diğer (belirtiniz).....	5

S.10. Anne ve Babanızın birliktelik durumu nedir?

Birlikteler1 ☐

Boşandılar 2 ☐

Ayrıyaşıyorlar3 ☐

S.9. Babanızın eğitim durumu :

İlkokul.....	1
Ortaokul.....	2
İlköğretim.....	3
Lise.....	4
Meslek veya teknik lise..	5
önlisans.....	6
lisans.....	7
Yüksek lisans..	8
Doktora	9
Okuma-yazma	
bilmiyor.....	10
Okur yazar...	11

S.11. Annenizin eğitim durumu :

İlkokul.....	1
Ortaokul.....	2
İlköğretim.....	3
Lise.....	4
Meslek veya teknik lise..	5
önlisans.....	6
lisans.....	7
Yüksek lisans..	8
Doktora	9
Okuma-yazma	
bilmiyor.....	10
Okur yazar...	11

S.12. Anneniz herhangi bir işte çalışıyor mu?

Evet 1 ☐ hayır 2 ☐

S.13. Babanız herhangi bir işte çalışıyor mu?

Evet 1 ☐ hayır 2 ☐

S.14. Anne/ Babanız dışında ailede başka çalışan var mı?

Evet 1 ☐ hayır 2 ☐

S.15. Anne/ Babanız dahil olmak üzere ailede kaç kişi çalışıyor?.....

S.16. Ailenizin aylık toplam geliri nedir ?

1500 TL'DEN AZ.....	☐	1
1500 - 2000.....	☐	2
2001 - 2500.....	☐	3
2501 - 3000.....	☐	4
3001 - 3500.....	☐	5
3501 - 4000.....	☐	6
4001 - 4500.....	☐	7
4501 - 5000.....	☐	8
5001 - 5500.....	☐	9
5501 - 6000.....	☐	10
6001 - 6500.....	☐	11
65001 +	☐	12

S.17. Ailenizin aylık toplam gideri nedir ?

1500 TL'DEN AZ.....	☐	1
1500 - 2000.....	☐	2
2001 - 2500.....	☐	3
2501 - 3000.....	☐	4
3001 - 3500.....	☐	5
3501 - 4000.....	☐	6
4001 - 4500.....	☐	7
4501 - 5000.....	☐	8
5001 - 5500.....	☐	9
5501 - 6000.....	☐	10
6001 - 6500.....	☐	11
65001 +	☐	12

S.18. Burs Öğrenim kredisi vb. ek geliriniz var mı?

Evet 1 ☐ hayır 2 ☐

S.19. Şuanda okumakta olduğunuz bölüm kaçınıcı tercihinizdi?

S.20. . Günde ortalama kaç saat internet kullanıyorsunuz?

1 saat ☐ 1 ile 3 saat arası ☐ 3 saatten fazla ☐ Hiç kullanmıyorum ☐

S.21. Günde ortalama kaç saat TV izliyorsunuz?

1 saat ☐ 1 ile 3 saat arası ☐ 3 saatten fazla ☐ Hiç seyretmiyorum ☐

S.22. Samsunda ortalama gideriniz ne kadardı?

150 den az....	☐	1
150 - 300	☐	2
301 - 450	☐	3
451 -600	☐	4
601 - 750	☐	5
751 - 900	☐	6
901 den fazla	☐	7

S.23. Kendinize ait bilgisayarınız var mı?

Evet 1 ☐ hayır 2 ☐

S.24. Hangi sıklıkla kitap okuyorsunuz?

Hergün ☐ haftada 1 ☐ ayda 1 ☐ 6 ayda 1 ☐ Yılda 1 ☐
hiç ☐

S.25. Aile içi ilişkileriniz nasıldır?

Çok iyi anlaşıyoruz ☐ Beni anlamıyorlar ☐ Kuşak çatışması var ☐
Diğer ☐

2. PROBLEM ÇÖZME ENVANTERİ(PÇE)

Yanıtlarınızı aşağıdaki ölçeğe göre değerlendirin:

1. Her zaman böyle davranırım
2. Çoğunlukla böyle davranırım
3. Sık sık böyle davranırım
4. Arada sırada böyle davranırım
5. Ender olarak böyle davranırım
6. Hiçbir zaman böyle davranmam

1. Bir sorunumu çözmek için kullandığım çözüm	(1)	(2)	(3)	(4)	(5)	(6)
yolları başarısız ise bunların neden başarısız olduğunu araştırmam.						
2. Zor bir sorunla karşılaştığımda ne olduğunu tam olarak belirleyebilmek için nasıl bilgi toplayacağımı uzun boylu düşünmem.	(1)	(2)	(3)	(4)	(5)	(6)
3. Bir sorunumu çözmek için gösterdiğim ilk çabalar başarısız olursa o sorun ile başa çıkabileceğimde şüpheyeye düşerim.	(1)	(2)	(3)	(4)	(5)	(6)
4. Bir sorunumu çözdükten sonra bu sorunu çözerken neyin işe yaradığını, neyin yaramadığını ayrıntılı olarak düşünmem.	(1)	(2)	(3)	(4)	(5)	(6)
5. Sorunlarımı çözmek konusunda genellikle yaratıcı ve etkili çözümler üretebilirim.	(1)	(2)	(3)	(4)	(5)	(6)
6. Bir sorunumu çözmek için belli bir yolu denedikten sonra durur ve ortaya çıkan sonuç ile olması gerektiğini düşündüğüm sonucu karşılaştırırım.	(1)	(2)	(3)	(4)	(5)	(6)
7. Bir sorunun olduğunda onu çözebilmek için başvurabileceğim yolların hepsini düşünmeye çalışırım.	(1)	(2)	(3)	(4)	(5)	(6)
8. Bir sorunla karşılaştığımda neler hissettiğimi anlamak için duygularımı incelerim.	(1)	(2)	(3)	(4)	(5)	(6)
9. Bir sorun kafamı karıştırdığında duygu ve düşüncelerimi somut ve açık seçik terimlerle ifade etmeye uğraşmam.	(1)	(2)	(3)	(4)	(5)	(6)
10. Başlangıçta çözümünü fark etmesem de sorunlarımın çoğunu çözmeye yeteneğim vardır.	(1)	(2)	(3)	(4)	(5)	(6)

11. Karşılaştığım sorunların çoğu, çözebileceğimden daha zor ve karmaşıktır.	(1)	(2)	(3)	(4)	(5)	(6)
12. Genellikle kendimle ilgili kararları verebilirim ve bu kararlardan hoşnut olurum.	(1)	(2)	(3)	(4)	(5)	(6)
13. Bir sorunla karşılaştığımda onu çözmek için genellikle aklıma gelen ilk yolu izlerim.	(1)	(2)	(3)	(4)	(5)	(6)
14. Bazen durup sorunlarım üzerinde düşünmek yerine gelişigüzel sürüklenip giderim	(1)	(2)	(3)	(4)	(5)	(6)
15. Bir sorunla ilgili olası bir çözüm yolu üzerinde karar vermeye çalışırken seçeneklerimin başarı olasılığını tek tek değerlendirmem.	(1)	(2)	(3)	(4)	(5)	(6)
16. Bir sorunla karşılaştığımda, başka konuya geçmede önce durur ve o sorun üzerinde düşünürüm.	(1)	(2)	(3)	(4)	(5)	(6)
17. Genellikle aklıma il gelen fikir doğrultusunda hareket ederim.	(1)	(2)	(3)	(4)	(5)	(6)
18. Bir karar vermeye çalışırken her seçeneğin sonuçlarını ölçer, tarar, birbirleriyle karşılaştırır, sonra karar veririm.	(1)	(2)	(3)	(4)	(5)	(6)
19. Bir sorunumu çözmek üzere plan yaparken o planı yürütebileceğime güvenirim.	(1)	(2)	(3)	(4)	(5)	(6)
20. Belli bir çözüm planını uygulamaya koymadan önce, nasıl bir sonuç vereceğini tahmin etmeye çalışırım	(1)	(2)	(3)	(4)	(5)	(6)
21. Bir soruna yönelik olası çözüm yollarını düşünürken çok fazla seçenek üretmem.	(1)	(2)	(3)	(4)	(5)	(6)
22. Bir sorunumu çözmeye çalışırken sıklıkla kullandığım bir yöntem; daha önce başıma gelmiş benzer sorunları düşünmektir.	(1)	(2)	(3)	(4)	(5)	(6)
23. Yeterince zamanım olur ve çaba gösterirsem karşılaştığım sorunların çoğunu çözebileceğime inanıyorum.	(1)	(2)	(3)	(4)	(5)	(6)
24. Yeni bir durumla karşılaştığımda ortaya çıkabilecek sorunları çözebileceğime inancım vardır.	(1)	(2)	(3)	(4)	(5)	(6)
25. Bazen bir sorunu çözmek için çabaladığım halde, bir türlü esas konuya giremediğim ve gereksiz ayrıntılarla uğraştığım duygusunu yaşıyorum.	(1)	(2)	(3)	(4)	(5)	(6)
26. Ani kararlar verir ve sonra pişmanlık duyarım.	(1)	(2)	(3)	(4)	(5)	(6)
27. Yeni ve zor sorunları çözebilme yeteneğine güveniyorum.	(1)	(2)	(3)	(4)	(5)	(6)
28. Elimdeki seçenekleri karşılaştırırken ve karar verirken kullandığım sistematik bir yöntem vardır.	(1)	(2)	(3)	(4)	(5)	(6)

29. Bir sorunla başa çıkma yollarını düşünürken çeşitli fikirleri birleştirmeye çalışmam.	(1)	(2)	(3)	(4)	(5)	(6)
30. Bir sorunla karşılaştığımda bu sorunun çıkmasında katkısı olabilecek benim dışımdaki etmenleri genellikle dikkate almam.	(1)	(2)	(3)	(4)	(5)	(6)
31. Bir konu ile karşılaştığımda, ilk yaptığım şeylerden biri, durumu gözden geçirmek ve konuyla ilgili olabilecek her türlü bilgiyi dikkate almaktır.	(1)	(2)	(3)	(4)	(5)	(6)
32. Bazen duygusal olarak öylesine etkilenirim ki, sorunumla başa çıkma yollarından pek çoğunu dikkate bile almam.	(1)	(2)	(3)	(4)	(5)	(6)
33. Bir karar verdikten sonra, ortaya çıkan sonuç genellikle benim beklediğim sonuca uyar.	(1)	(2)	(3)	(4)	(5)	(6)
34. Bir sorunla karşılaştığımda, o durumla başa çıkabileceğimden genellikle pek emin değilimdir.	(1)	(2)	(3)	(4)	(5)	(6)
35. Bir sorunun farkına vardığımda, ilk yaptığım şeylerden biri, sorunun tam olarak ne olduğunu anlamaya çalışmaktır.	(1)	(2)	(3)	(4)	(5)	(6)

3.ROSENBERG BENLİK SAYGISI ÖLÇEĞİ

- Kendimi en az diğer insanlar kadar değerli buluyorum.
a. ÇOK DOĞRU b. DOĞRU c. YANLIŞ d. ÇOK YANLIŞ
- Bazı olumlu özelliklerim olduğunu düşünüyorum.
a. ÇOK DOĞRU b. DOĞRU c. YANLIŞ d. ÇOK YANLIŞ
- Genelde kendimi başarısız bir kişi olarak görme eğilimindeyim.
a. ÇOK DOĞRU b. DOĞRU c. YANLIŞ d. ÇOK YANLIŞ
- Ben de diğer insanların birçoğunun yapabildiği kadar birşeyler yapabilirim.
a. ÇOK DOĞRU b. DOĞRU c. YANLIŞ d. ÇOK YANLIŞ
- Kendimde gurur duyacak fazla birşey bulamıyorum.
a. ÇOK DOĞRU b. DOĞRU c. YANLIŞ d. ÇOK YANLIŞ
- Kendime karşı olumlu bir tutum içindeyim.
a. ÇOK DOĞRU b. DOĞRU c. YANLIŞ d. ÇOK YANLIŞ
- Genel olarak kendimden memnunum.
a. ÇOK DOĞRU b. DOĞRU c. YANLIŞ d. ÇOK YANLIŞ
- Kendime karşı daha fazla saygı duyabilmeyi isterdim.
a. ÇOK DOĞRU b. DOĞRU c. YANLIŞ d. ÇOK YANLIŞ

9. Bazen kesinlikle kendimin bir işe yaramadığımı düşünüyorum.
a. ÇOK DOĞRU b. DOĞRU c. YANLIŞ d. ÇOK YANLIŞ
10. Bazen kendimin hiç de yeterli bir insan olmadığımı düşünüyorum.
a. ÇOK DOĞRU b. DOĞRU c. YANLIŞ d. ÇOK YANLIŞ
11. Siz 10 – 11 yaşlarınızdayken **anneniz** arkadaşlarınızı tanır mıydı ?
a. Hepsini tanırdı b. Çoğunu tanırdı
c. Bazılarını tanırdı d. Hemen hemen hiçbirini tanımazdı
12. Bu dönemde **babanız** arkadaşlarınızı tanır mıydı ?
a. Hepsini tanırdı b. Çoğunu tanırdı
c. Bazılarını tanırdı d. Hemen hemen hiçbirini tanımazdı
13. 5. – 6. sınıflardayken karneniz **iyi** olduğunda **anneniz** çoğu zaman ilgilenmezdi.
a. DOĞRU b. YANLIŞ
14. 5. – 6. sınıflardayken karneniz **iyi** olduğunda **babanız** çoğu zaman ilgilenmezdi.
a. DOĞRU b. YANLIŞ
15. 5. – 6. sınıflardayken karneniz **kötü** olduğunda **anneniz** çoğu zaman ilgilenmezdi.
a. DOĞRU b. YANLIŞ
16. 5. – 6. sınıflardayken karneniz **kötü** olduğunda **babanız** çoğu zaman ilgilenmezdi.
a. DOĞRU b. YANLIŞ
17. Sizce diğer aile bireyleri sizin söylediğiniz şeylerle ne kadar ilgilenirler ?
a. Çok ilgilenirler b. Oldukça ilgilenirler c. İlgilenmezler
18. Büyümekte olduğunuz dönemde babanızın en çok tuttuğu çocuğu kimdi ?
a. Ben b. Ağabeyim c. Ablam d. Erkek kardeşim
e. Kız kardeşim f. Bildiğim kadarıyla çok tuttuğu birisi yoktu
19. Bu dönemde babanız arkadaşlarınızı tanır mıydı ?
a. Hepsini tanırdı b. Çoğunu tanırdı
c. Bazılarını tanırdı d. Hiçbirini tanımazdı
20. Anne ve babanızın hangisi ile daha rahat konuşabiliyorsunuz ?
a. Babamla çok daha fazla b. Babamla biraz daha fazla
c. Her ikisi ile eşit oranda d. Annemle biraz daha fazla
e. Annemle çok daha fazla
21. Anne ve babanızın hangisi sizi daha çok över ?
a. Babam çok daha fazla b. Babam biraz daha fazla
c. Her ikisi eşit oranda d. Annem biraz daha fazla
e. Annem çok daha fazla
22. Anne ve babanızın hangisi size daha çok şefkat gösterir ?
a. Babam çok daha fazla b. Babam biraz daha fazla
c. Her ikisi eşit oranda d. Annem biraz daha fazla

e. Annem çok daha fazla

23. Anne ve babanız anlaşılmadıkları zaman siz genellikle hangisinden yana olursunuz ?

a. Çok daha fazla olarak babamdan yana b. Biraz fazla olarak babamdan yana

c. Eşit oranda her ikisinden yana d. Biraz fazla olarak annemden yana

e. Çok daha fazla olarak annemden yana

24. Yalnız bir insan olmaya eğilimli misinizdir ?

a. Evet b. Hayır

25. İnsanların çoğu sizin nasıl bir kişi olduğunuzu bilirler mi, yoksa çoğunun sizi gerçekten tanımadıklarını mı düşünürsünüz ?

a. Çoğu benim nasıl biri olduğumu bilir.

b. Çoğu gerçekten beni tanımaz.

4.SÜREKLİ-DURUM KAYGI ÖLÇEĞİ

		HİÇ	BİRAZ	ÇOK	TAMAMIYLA
1.	Şu anda sakinim	(1)	(2)	(3)	(4)
2.	Kendimi emniyette hissediyorum	(1)	(2)	(3)	(4)
3.	Su anda sinirlerim gergin	(1)	(2)	(3)	(4)
4.	Pişmanlık duygusu içindeyim	(1)	(2)	(3)	(4)
5.	Şu anda huzur içindeyim	(1)	(2)	(3)	(4)
6.	Şu anda hiç keyfim yok	(1)	(2)	(3)	(4)
7.	Başıma geleceklerden endişe ediyorum	(1)	(2)	(3)	(4)
8.	Kendimi dinlenmiş hissediyorum	(1)	(2)	(3)	(4)
9.	Şu anda kaygılıyım	(1)	(2)	(3)	(4)
10.	Kendimi rahat hissediyorum	(1)	(2)	(3)	(4)
11.	Kendime güvenim var	(1)	(2)	(3)	(4)
12.	Şu anda asabım bozuk	(1)	(2)	(3)	(4)
13.	Çok sinirliyim	(1)	(2)	(3)	(4)

14.	Sinirlerimin çok gergin olduğunu hissediyorum	(1)	(2)	(3)	(4)
15.	Kendimi rahatlamış hissediyorum	(1)	(2)	(3)	(4)
16.	Şu anda halimden memnunum	(1)	(2)	(3)	(4)
17.	Şu anda endişeliyim	(1)	(2)	(3)	(4)
18.	Heyecandan kendimi şaşkına dönmüş hissediyorum	(1)	(2)	(3)	(4)
19.	Şu anda sevinçliyim	(1)	(2)	(3)	(4)
20.	Şu anda keyfim yerinde.	(1)	(2)	(3)	(4)

		HİÇ	BİRAZ	ÇOK	TAMAMİYLE
1.	Şu anda sakinim	(1)	(2)	(3)	(4)
2.	Kendimi emniyette hissediyorum	(1)	(2)	(3)	(4)
3.	Su anda sinirlerim gergin	(1)	(2)	(3)	(4)
4.	Pişmanlık duygusu içindeyim	(1)	(2)	(3)	(4)
5.	Şu anda huzur içindeyim	(1)	(2)	(3)	(4)
6.	Şu anda hiç keyfim yok	(1)	(2)	(3)	(4)
7.	Başıma geleceklerden endişe ediyorum	(1)	(2)	(3)	(4)
8.	Kendimi dinlenmiş hissediyorum	(1)	(2)	(3)	(4)
9.	Şu anda kaygılıyım	(1)	(2)	(3)	(4)
10.	Kendimi rahat hissediyorum	(1)	(2)	(3)	(4)
11.	Kendime güvenim var	(1)	(2)	(3)	(4)
12.	Şu anda asabım bozuk	(1)	(2)	(3)	(4)
13.	Çok sinirliyim	(1)	(2)	(3)	(4)
14.	Sinirlerimin çok gergin olduğunu hissediyorum	(1)	(2)	(3)	(4)
15.	Kendimi rahatlamış hissediyorum	(1)	(2)	(3)	(4)
16.	Şu anda halimden memnunum	(1)	(2)	(3)	(4)
17.	Şu anda endişeliyim	(1)	(2)	(3)	(4)

18	Heyecandan kendimi şaşkına dönmüş hissediyorum	(1)	(2)	(3)	(4)
19.	Şu anda sevinçliyim	(1)	(2)	(3)	(4)
20.	Şu anda keyfim yerinde.	(1)	(2)	(3)	(4)



ÖZGEÇMİŞ

Adı ve Soyadı : Esmâ AKDEMRCİ

Doğum Yeri : Pazar/ Rize

Doğum Tarihi : 22.10.1990

Yabancı Dili : İngilizce

Eğitim Durumu

Lise : Pazar Atatürk Lisesi

Lisans : Sinop Üniversitesi / Fen-Edebiyat Fakültesi / İstatistik Bölümü (2012)

Yayınlar



