

**KIRIKKALE ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ**

**BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI  
YÜKSEK LİSANS TEZİ**

**Veri Madenciliği ile  
Lise Öğrenci Performanslarının  
Değerlendirilmesi**

**Semra YURDAKUL**

**HAZİRAN 2015**

**Bilgisayar Mühendisliđi Anabilim Dalında** Semra YURDAKUL tarafından hazırlanan VERİ MADENCİLİĐİ İLE LİSE ÖĐRENCİ PERFORMANSLARININ DEĐERLENDİRİLMESİ adlı Yüksek Lisans Tezinin Anabilim Dalı standartlarına uygun olduđunu onaylarım.

Prof. Dr. Hasan ERBAY  
Anabilim Dalı Başkanı

Bu tezi okuduđumu ve tezin **Yüksek Lisans Tezi** olarak bütün gereklilikleri yerine getirdiđini onaylarım.

Yrd. Doç. Dr. Taner TOPAL  
Danışman

Jüri Üyeleri

Başkan : Prof. Dr. Erdem Kamil YILDIRIM \_\_\_\_\_  
Üye (Danışman) : Yrd. Doç. Dr. Taner TOPAL \_\_\_\_\_  
Üye : Doç. Dr. Necaattin BARIŞCI \_\_\_\_\_

...../...../.....

Bu tez ile Kırıkkale Üniversitesi Fen Bilimleri Enstitüsü Yönetim Kurulu Yüksek Lisans derecesini onaylamıştır.

Prof. Dr. Mustafa YİĐİTOĐLU  
Fen Bilimleri Enstitüsü Müdürü

## ÖZET

### VERİ MADENCİLİĞİ İLE LİSE ÖĞRENCİ PERFORMANSLARININ DEĞERLENDİRİLMESİ

YURDAKUL, Semra

Kırıkkale Üniversitesi

Fen Bilimleri Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı, Yüksek Lisans Tezi

Danışman: Yrd. Doç. Dr. Taner TOPAL

Haziran 2015, 105 sayfa

Günümüzde bilim ve teknoloji hızla gelişirken, bu gelişimi takip etmede en önemli rol şüphesiz eğitime düşmektedir. Bu rolün gerçekleştirilebilmesi için eğitim etkinliklerinde teknolojiden en verimli şekilde yararlanmak gerekmektedir.

Bilgisayar ve iletişim teknolojilerindeki gelişmelere paralel olarak donanımın da ucuzlaması verilerin uzun süre depolanmasına dolayısıyla da büyük kapasiteli veri tabanlarının oluşmasına neden olmuştur. Büyük veri tabanlarının kullanılmaya başlanması ile istenilen anlamlı, kullanılabilir ve farklı bilgiye erişme gereksinimi ortaya çıkmış ve bu ihtiyaçlar yeni bir disiplinin doğmasına sebep olmuştur. Veriler çeşitli istatistiksel metotlarla analiz edilerek kurumların karar verme sürecinin etkinliğine ve yeni stratejiler geliştirmesine katkı sağlamaktadır. Büyük veri yığınları arasında bulunan, anlamı daha önce keşfedilmemiş, potansiyel olarak faydalı ve anlaşılır bilgilerin çıkarılmasında veri madenciliği teknikleri önemli yer tutmaktadır.

Bu çalışma, Kırıkkale İli Anadolu Liselerinde okuyan 11.sınıf öğrencilerine uyguladığımız anket sonucu elde edilen verilerden yararlanarak gerçekleştirilmiştir. Lise öğrencilerine ait 231 adet veri kullanılmıştır. Öğrenci performansına etki eden faktörler belirlenip, başarı ya da başarısızlığa etki eden faktörlerin birbiri ile olan ilişkisi araştırılmıştır. Ayrıca başarıyı artıracak bilgiler verilmesi amaçlanmıştır. Uygulama Weka 3.7 programı ile yapılmıştır.

Bu alıřma ile veri madenciliđinin eđitim ve đretimin kalitesini ve verimliliđini artırmak iin nasıl kullanılabilirleđi irdelenmiř ve yapılacak sonraki alıřmalar iin neriler sunulmuřtur.

**Anahtar Kelimeler:** Veri Madenciliđi, đrenci Performansı, Yapay Sinir Ađı, ok Katmanlı Algılayıcı

## **ABSTRACT**

### **ASSESSMENT OF HIGH SCHOOL STUDENTS' PERFORMANCE BY MEANS OF DATA MINING**

YURDAKUL, Semra

Kırıkkale University

Graduate School of Natural and Applied Sciences

Department of Computer Engineering, M.Sc. Thesis

Supervisor: Asst. Prof. Taner TOPAL

June 2015, 105 pages

Today, education should follow the rapid developments in science and technology. Therefore the educational activities it is necessary to take advantage of science and technology in an efficient way.

Parallel to the developments in science, communication technologies and decrease in hardware prices result in string data for much longer times capacity formation of very large databases. Along with the use of large databases need for reaching meaningful, useful and different information occurred. This in return caused a formation of a new discipline. Analysed data, by using various statistical data, contribute to firm's decision processes and development of new strategies. Data mining techniques are very important in finding unnoticed, understandable and potentially usefull information in very large pile of data.

Data mining techniques play an important role in extracting information which are held in large information volumes, not discovered before and potentially useful and understandable. This study was carried out according to data which are collected in the Anatolian High Schools by surveying with 11. grade students. 231 data were used related to high school students. The factors which influence the performance of the students were determined and the relationship between the factors which has an

impact on succes and failure were analyzed. Beside, it is aimed to give information about succes enhancement. Implementation is made by Weka 3.7 program.

In this study, the use of data mining for increasing quality and efficiency of education is examined and some proposals are presented for further research.

**Key Words:** Data Mining, Student Performance, Artificial Neural Network, Multilayer Perceptron

## TEŐEKKÜR

Tezimin hazırlanması esnasında yardımlarını esirgemeyen tez danışmanım Sayın Yrd. Doç. Dr. Taner TOPAL'a, bu çalışmanın taslađını oluştururken yardımcı olan en değerli arkadaşım Fevzi GÖNÜLTAŐ'a, fedakârlıklarından dolayı hakkını ödeyemeyeceđim ablam Sevil ÖZARSLAN'a, son düzenlemelerimi yaparken yardımcı olan Galip SARI ve Ođuzhan YILMAZ'a, manevi kardeŐlerim Melek ÇELİK ve Kübra USTA'ya, teŐekkürlerimi sunarım.

Varlıkları ile her zaman yanımda olan ve desteklerini hiçbir zaman esirgemeyen canım aileme, özellikle babam Haydar YURDAKUL'a teŐekkürü bir borç bilirim.

# İÇİNDEKİLER DİZİNİ

## Sayfa

<b>ÖZET</b> .....	i
<b>ABSTRACT</b> .....	iii
<b>TEŞEKKÜR</b> .....	v
<b>İÇİNDEKİLER DİZİNİ</b> .....	vi
<b>ŞEKİLLER DİZİNİ</b> .....	ix
<b>ÇİZELGELER DİZİNİ</b> .....	x
<b>SİMGELER DİZİNİ</b> .....	xi
<b>KISALTMALAR DİZİNİ</b> .....	xii
<b>1. GİRİŞ</b> .....	1
<b>2. MATERYAL VE YÖNTEM</b> .....	7
2.1. Veritabanı .....	7
2.1.1. Veritabanı Nedir? .....	7
2.1.2. Veritabanı Özellikleri.....	8
2.1.3. Veritabanı Nesneleri.....	8
2.1.4. Veritabanı Sorgulaması Nedir? .....	9
2.2. Veri Ambarı .....	9
2.2.1. Veri Ambarı Nedir? .....	9
2.2.2. Veri Ambarı Özellikleri .....	10
2.2.3. Veritabanı ile Veri Ambarı Arasındaki Farklar .....	11
2.3. Veri Madenciliği .....	12
2.3.1. Veri Madenciliğine Giriş.....	12
2.3.2. Veri Madenciliği Nedir? .....	13
2.3.3. Veri Madenciliği Ne Değildir?.....	16
2.3.4. Veri Madenciliğine Tarihçesi.....	16
2.3.5. Veri Madenciliği Yararları .....	18
2.3.6. Veri Madenciliği Kullanım Alanları .....	19
2.3.7. Veri Madenciliği Süreci .....	21
2.3.7.1. İş ve İş Ortamını Anlama .....	24
2.3.7.2. Veriyi Anlama.....	24



2.3.7.3. Veri Hazırlama .....	25
2.3.7.4. Modelleme .....	26
2.3.7.5. Değerlendirme .....	26
2.3.7.6. Yayıma .....	27
2.3.8. Veri Madenciliği Yöntem ve Teknikleri .....	27
2.3.8.1. Sınıflama ve Regresyon .....	28
2.3.8.2. Kümeleme .....	37
2.3.8.3. Birliktelik Kuralları .....	42
2.4. Öğrenci Performansını Etkileyen Faktörler .....	44
2.4.1. Yöneticiden Kaynaklı Nedenler .....	45
2.4.2. Öğretmenlerden Kaynaklı Nedenler .....	46
2.4.3. Aileden Kaynaklı Nedenler .....	46
2.4.4. Çevreden Kaynaklı Nedenler .....	48
2.4.5. Akademik Kaygıdan Kaynaklı Nedenler .....	48
2.4.6. Bireysel Nedenler .....	49
2.5. WEKA .....	51
2.6. Kullanılan Veri Madenciliği Sınıflama Algoritmaları .....	52
2.6.1. Çok Katmanlı Algılayıcı (Multilayer Perceptron) Algoritması .....	52
2.6.2. k- En Yakın Komşu (IBk) Algoritması .....	54
2.6.3. J48 Algoritması .....	55
2.6.4. JRIP Algoritması .....	57
2.6.5. Saf (Naive) Bayes .....	59
2.7. Sınıflama Algoritmalarının Başarısını Test Etme .....	60
<b>3. ARAŞTIRMA BULGULARI .....</b>	<b>62</b>
3.1. Veri Tanımlama .....	62
3.2. Modeli Kurma .....	68
3.3. Modeli Değerlendirme .....	69
3.3.1. Çok Katmanlı Algılayıcı İle Oluşturulan Veri Modelleme .....	70
3.3.2. k- En Yakın Komşu Algoritması İle Oluşturulan Veri Modelleme ...	72
3.3.3. J48 Algoritması İle Oluşturulan Veri Modelleme .....	73
3.3.4. JRIP Algoritması İle Oluşturulan Veri Modelleme .....	76
3.3.5. Saf Bayes Algoritması İle Oluşturulan Veri Modelleme .....	79
3.4. Weka İle Elde Edilen Görsel Sonuçlar .....	79

3.5. Birliktelik Kuralları (Associate) İle Elde Edilen Kurallar .....	88
<b>4. TARTIŞMA VE SONUÇ</b> .....	91
<b>KAYNAKLAR</b> .....	94
<b>EKLER</b> .....	102
<b>EK 1. ANKET UYGULAMASI İÇİN VALİLİK ONAYI</b> .....	102
<b>EK 2. ANKET ÖRNEĞİ</b> .....	103

## ŞEKİLLER DİZİNİ

<u>ŞEKİL</u>	<u>Sayfa</u>
2.1. Veri Madenciliğini Oluşturan Disiplinler .....	14
2.2. CRISP-DM Süreç Şeması .....	23
2.3. Sınıflandırma Model Kurma Süreci .....	29
2.4. Sınıflandırma Test Verileri .....	30
2.5. Karar Ağacı Yapısı.....	35
2.6. k-means Yöntemiyle Kümeleme Örneği.....	40
2.7. k-medoids Yöntemiyle Kümeleme Örneği .....	41
2.8. Hiyerarşik Kümeleme Örneği .....	42
2.9. Weka Arayüz Görünümü .....	52
2.10. JRip Algoritma Kuralları .....	58
3.1. Çalışmada Oluşturulan arff Dosyasının Başlık Kısmı .....	67
3.2. Çalışmada Oluşturulan arff Dosyasında Verilerin Bulunduğu Kısım.....	67
3.3. Weka Explorer Penceres .....	68
3.4. Weka Classify Paneli .....	69
3.5. Çok Katmanlı Algılayıcı ile Oluşturulmuş Modelin Sonuç Ekranı .....	70
3.6. J48 Algoritması İçin Karar Ağacı Sonuç Ekranı.....	75
3.7. Weka Görselleştirme (Visualize) Paneli .....	79
3.8. Kardeş Sayısına Göre Başarı Durumu .....	80
3.9. Sağlık Problemine Göre Başarı Durumu .....	81
3.10. Anne Eğitim Düzeyine Göre Başarı Durumu .....	82
3.11. Ekonomik Duruma Göre Başarı Durumu .....	83
3.12. Öğrencinin Kendi Odası Olup Olmamasına Göre Başarı Durumu.....	84
3.13. Öğrencinin Öğretmenlerle Olan İletişimine Göre Başarı Durumu .....	85
3.14. Öğrencinin Arkadaşlarıyla Olan İletişimine Göre Başarı Durumu.....	86
3.15. Anne Eğitim Düzeyi İle Öğrencinin Öğretmenlerle Olan İletişimi .....	87
3.16. Ailenin Hayatta Olma Durumu İle Öğrencinin Öğretmenleriyle İletişimi .	88
3.17. Weka Associate Paneli .....	89
3.18. Weka İle Elde Edilen Birliktelik Kuralları.....	90

## ÇİZELGELER DİZİNİ

### ÇİZELGE

	<u>Sayfa</u>
2.1. OLTP Veritabanı İle Veri Ambarı Karşılaştırma Tablosu.....	11
2.2. Veri Madenciliğinin Tarihsel Süreci.....	17
2.3. Sınıflama Model Gösterimi.....	29
2.4. Regresyon Model Gösterimi.....	31
2.5. Karar Ağacı Oluşturulacak Örnek Tablo.....	34
2.6. Apriori Algoritması Müşteri Alışverişleri Tablosu.....	43
2.7. Yaklaşımlara Göre Performans Tanımları.....	44
2.8. JRip Algoritması Kural Açıklamaları.....	58
2.9. İki Sınıflı Bir Model İçin Sınıflama Matrisi.....	61
3.1. Başarı Durumu.....	63
3.2. Kardeş Sayısı.....	63
3.3. Anne – Baba Hayatta Olma Durumu.....	64
3.4. Anne – Baba Yaşı.....	64
3.5. Veri Madenciliği Çalışması İçin Kullanılacak Verilerin Dağılımı.....	64
3.6. Veritabanı İstatistikleri 1.....	65
3.7. Veritabanı İstatistikleri 2.....	65
3.8. Çok Katmanlı Algılayıcı Algoritması İçin Düzensizlik Matrisi.....	71
3.9. Çok Katmanlı Algılayıcı Algoritması İçin Detaylı Doğruluk Tablosu.....	71
3.10. k- En Yakın Komşu Algoritması İçin Düzensizlik Matrisi.....	72
3.11. k- En Yakın Komşu Algoritması İçin Detaylı Doğruluk Tablosu.....	73
3.12. J48 Algoritması İçin Düzensizlik Matrisi.....	73
3.13. J48 Algoritması İçin Detaylı Doğruluk Tablosu.....	74
3.14. JRip Algoritması İçin Düzensizlik Matrisi.....	76
3.15. JRip Algoritması İçin Detaylı Doğruluk Tablosu.....	77
3.16. Saf Bayes Algoritması İçin Düzensizlik Matrisi.....	77
3.17. Saf Bayes Algoritması İçin Detaylı Doğruluk Tablosu.....	78
3.18. Seçilen Sınıflandırma Algoritmaları ve Doğruluk Yüzdeleri.....	79

## SİMGELER DİZİNİ

$F_k$	Giriş katmanındaki k. nöronun çıkışı
$G_k$	Giriş katmanına dış dünyadan gelen bilgi
$Net_j$	j. prosesin net girdisi
$F_i$	j. nörona bilgi gönderen nöronların çıkış bilgisi
$W_{ij}$	j. nörona bilgi gönderen i. nöron ile j. nöron arası ağırlık
$E_m$	Hata - Çıkış katmanında elde edilen çıkış bilgisi ile olması gereken çıkış bilgisi arası fark

## KISALTMALAR DİZİNİ

OLTP	Online Transaction Processing / Çevrimiçi İşlem Yürütme
VTYS	Veritabanı Yönetim Sistemi
SQL	Structured Query Language / Yapısal Sorgu Dili
CRISP-DM	Cross Industry Standart Process for Data Mining / Çapraz Endüstri Veri Madenciliği Standart Süreci
YSA	Yapay Sinir Ağı
WEKA	Waikato Environment for Knowledge Analysis / Waikato Bilgi Analiz İçin Ortam
ÇKA	Çok Katmanlı Algılayıcı
RIPPER	Repeated Incremental Pruning to Produce Error Reduction / Yinelenen Artımlı Budama İçin Hata Azaltmayı Üretme
IREP	Incremental Reduced Error Pruning / Artan Maliyetli Budama İçin İndirgenmiş Hata
DP	Doğru Pozitif Sayısı
YP	Yanlış Pozitif Sayısı
YN	Yanlış Negatif Sayısı
DN	Doğru Negatif Sayısı
ARFF	Attribute Relation File Format / Özellik İlişki Dosya Biçimi

## 1. GİRİŞ

Eđitim bir toplumun geliřimi iin en nemli unsurlardan biridir. đrenmeyi en st dzeye ıkartmak iin yıllarca birok yntem denenmiř ve bu yolda srekli geliřmeler elde edilmiřtir. Klasik eđitim yntemlerine her defasında yenilikler katılmıř ve gnn teknolojik geliřmelerinden yararlanılmıřtır.

Bilim ve teknolojide yařanan hızlı geliřmeler lkeleri kaınılmaz bir yarıřın iine sokmuř ve bu yarıř var olan teknolojik olanakların geliřtirilmesini bir ayrıcalık olmaktan ıkarıp zorunluluk haline getirmiřtir. ađın gereklerine ayak uydurmada ve geliřimi yakalamada en nemli rol řphesiz eđitime dřmektedir. Bu roln gerekleřtirilebilmesi iin ise eđitim etkinliklerinde teknolojiden en verimli řekilde yararlanmak gerekmektedir. Teknoloji bilimin retim, hizmet, ulařım gibi alanlardaki sorunlara uygulanması srecinde yararlanan ve bilim ile uygulama arasında kpr grevi gren makineler, iřlemler, sistemler, sreler, yntemler, ynetim ve kontrol mekanizmalarının tmdr. Eđitim teknolojisi ise đrenmenin tm ynlerini ieren sorunları sistemli bir řekilde inceleyen, bu sorunlara zmler bulmak amacıyla insan gc, bilgi, yntem, teknik, ara-gere ve dzenleme gibi đeleri iře kořarak uygun tasarımlar geliřtiren, uygulayan, deđerlendiren ve yneten karmařık bir sretir. Kısacası eđitim teknolojisi đrenme-đretme srelerinin tasarlanması, uygulanması ve geliřtirilmesi srecidir [1].

Son yıllarda, iletiřim araları ve biliřim teknolojisinde ok byk geliřmeler yařanmıřtır. Biliřim ve iletiřim aralarındaki retim maliyetlerinin dřrlmesiyle, insanların ve kurumların son teknolojik rnlere ulařması kolaylařmıřtır. Bu sayede, milyarlarca byte tutan veriler elektronik ortamlarda depolanmıř, insanlar ve kurumlar arasındaki her trl iřlem kayıt altına alınmıřtır. Buna paralel olarak, iletiřim teknolojilerinde ise adeta kablosuz bir dneme girilmiřtir. İnsanlar ve kurumlar iin zaman ve mekanın nemi giderek azalmıřtır. Bu sayede dnyada uzak mesafeler yakınlařmıř, daha nce ulařılmayan yerlere ulařılabilmıřtir. Tm bu geliřmeler, insanlar ve kurumlar arasında rekabeti de beraberinde getirmiřtir. Klen dnya ve artan rekabet kořulları altında kurumların yařamaları ve gelecekte var olmaları da

alacakları kararların doğruluđuna bađlıdır. Gnmzde, kurumların aldıđı yanlış kararların geri dnş hemen hemen kalmamıřtır. Birok kurum, aldıđı kararlar neticesinde faaliyetlerini durdurmak zorunda kalmıřtır. Bunun tam aksine, birok kurum da aldıđı dođru kararlar sonucunda bymř, geleceklerini garanti altına almıřtır. Dođru karar vermenin nemini anlayan kurumlar, yaptıkları tm iřlemleri, daha sonra kullanmak zere elektronik ortamlarda saklamaya bařlamıřlardır.

Bilgisayar ve iletiřim teknolojilerindeki geliřmelere paralel olarak donanımın ucuzlaması verilerin uzun sre depolanmasına dolayısıyla da byk kapasiteli veri tabanların oluřmasına neden olmuřtur. Bu nedenle byk veri tabanlarında istenilen anlamlı, kullanılabilir ve farklı bilgiye eriřmek yeni bir disiplinin dođmasına sebep olmuřtur. Veriler eřitli istatistiksel metotlarla analiz edilerek kurumların karar verme srecinin etkinliđine ve yeni stratejiler geliřtirmesine katkı sađlamaktır.

İlk bařlarda ama, verilere istendiđi anda ulařabilmek, verileri listeleyip raporlayabilmektir. Bu yapıldıktan sonra, depolanan verilerden analiz yaparak eřitli sonulara ulařmak sonraki ama oldu. nk son yıllarda kurumlar arası rekabet artıřı alınan karar sayısındaki artıřı da beraberinde getirmiřtir. Alınan kararların hızlı ve dođru bir řekilde alınması gerekmiřtir. Karar almayı etkileyen faktrler ve parametreler de gemiřle kıyaslanamayacak lde artmıř ve karmařıklařmıřtır.

zetle; gemiřte verilere ulařım daha uzun zaman alıyor ancak bu verilerden yapılan analizler daha kısa srede tamamlanıyordu. Bugn ise, verilere ulařım hızlanmıř ve kolaylařmıř ancak bu verilere dayalı yapılan analizler de o derece karmařıklařmıř ve zorlařmıřtır. Bundan dolayı, analiz yapma, sonuca ulařma ve karar verme srelerini de bilgisayarlara yaptırabilme ihtiyacı ve fikri dođmuřtur. Bunun sonucunda, eřitli matematiksel ve istatistiksel hesaplamalara dayanan algoritmalar geliřtirilmiř ve Veri Madenciliđi kavramı dođmuřtur [2].

Veri madenciliđi, diđer bir adla veri tabanında bilgi keřfi; ok byk veri hacimleri arasında tutulan, anlamı daha nce keřfedilmemiř potansiyel olarak faydalı ve anlaşılır bilgilerin ıkarıldıđı veri analiz teknikleridir. Arka planda veri tabanı



yönetim sistemleri, istatistik, yapay zekâ, makine öğrenme, paralel ve dağıtık işlemlerin bulunmaktadır [3].

Veri madenciliği yeni bir disiplin olmasına karşın oldukça geniş bir alanda uygulanmaktadır. İş dünyası ve bilimin bazı alanlarında birçok problemin çözülmesinde etkin rol oynamıştır. Veri madenciliği aracılığıyla finans ve ekonomi, sağlık hizmetleri, güvenlik hizmetleri, sosyal hizmetler, e-devlet, telekomünikasyon ve nakliye gibi alanlarda olduğu gibi eğitim alanında da başarılı bir şekilde gerçekleştirilmiş uygulamalar bulunmaktadır.

Veri madenciliğinin uygulandığı birçok alanda olduğu gibi eğitimde de anlamlı ilişkilerin araştırılabileceği ve faydalı bilginin türetilebileceği geniş veri tabanları mevcuttur. Eğitim alanındaki veri madenciliği çalışmaları eğitim sistemlerinde yer alan veri tabanlarında öğrencilere, akademik sorumlulara ve eğitimcilere faydalı olabilecek henüz keşfedilmemiş bilginin mevcut olduğu olgusundan yola çıkmaktadır [4].

Eğitim alanında, öğrencilerin

- Başarı veya başarısızlık nedenlerinin bulunması,
- Öğrenci başarısının artırılması için neler yapılabileceği,
- Anadolu Lisesi'ne yerleştirmede esas alınan giriş puanları ile öğrencinin okul başarısı arasında bir ilişkinin var olup olmadığı gibi soruların cevaplarının araştırılmasında veri madenciliği yöntemleri kullanılarak, eğitimin kalitesi ve performansı artırılabilir.

Günümüze kadar eğitim alanında yapılmış olan veri madenciliği çalışmaları aşağıda kısaca özetlenmiştir;

2003 yılında Konya Selçuk Üniversitesi'nde Onur İnan tarafından, hazırlık sınıfı, birinci sınıf ve mezun durumunda olan öğrenciler üzerinde, üniversite veri tabanındaki veriler kullanılarak; öğrencilerin başarılarını etkileyen etmenler, başarı düzeyleri, üniversiteyi kazanan öğrenci profilleri ve mezun olamayan öğrencilerin

okulu bitirmelerini engelleyen etmenler üzerinde çalışmalar gerçekleştirilmiş ve sonuçları yorumlanmıştır [5].

2004 yılında Murat Karabatak ve Melih Cevdet İnce tarafından yapılan çalışmada Veri Madenciliği teknikleri kullanılarak Fırat Üniversitesi Teknik Eğitim Fakültesi Bilgisayar Eğitimi bölümü öğrencilerinin notları kullanılmış ve öğrenci başarılarının analizi yapılmıştır. Bu analizi yapmak için Veri Madenciliğinde, birliktelik kuralı çıkarım algoritmalarından biri olan Apriori algoritması kullanılmıştır [6].

2005 yılında Şenol Zafer Erdoğan ve Mehpere Timor tarafından gerçekleştirilen çalışmada Maltepe Üniversitesi öğrencilerinin belirleyici özellikleri “K-Means” algoritması kullanılarak kümelendirilmiştir. 2003 yılına ait 722 öğrenciye ait verilerin kullanıldığı çalışmada öğrencilerin üniversiteye giriş sınavı sonuçları ile başarıları arasındaki ilişki kümeleme analizi ve K-Means algoritması teknikleri kullanılarak incelenmiştir [7].

2006 yılında yapılan çalışmanın KPSS’ye uygulanmış bir modeline benzeyen çalışmayı Hüseyin Özçınar gerçekleştirmiştir. Frekans analizi ve regresyon analizi yöntemleri kullanılarak derslere ve yıllara göre verinin özellikleri incelenmiştir. Oluşturulan regresyon modeli ile KPSS sonuçlarının değişimi üzerinde anlamlı katkısı olan değişkenler incelenmiş ve oluşturulan modellerin tahmin doğrulukları, ortalama mutlak hata ve ortalama hata karekök değerleri kullanılarak karşılaştırılmıştır [8].

2006 yılında Serdar Çiftci tarafından gerçekleştirilen çalışmada, uzaktan eğitime katılan öğrencilerin ders çalışma etkinliklerinin değerlendirilmesi için yapılan anketler ve log dosyaları karşılaştırılmış ve sonuçların farklı olup olmadığı incelenmiştir [9].

2007 yılında Y. Ziya Ayık, Abdülkadir Özdemir ve Uğur Yavuz tarafından yapılan çalışmada, Atatürk Üniversitesi öğrencilerinin mezun oldukları lise türleri ve lise mezuniyet dereceleri ile kazandıkları fakülteler arasındaki ilişki, veri madenciliği teknikleri kullanılarak incelenmiştir. Çalışma sonucunda, lise türünün arzu edilen bir

fakültenin kazanılmasında çok büyük öneminin olduğu, yine lise başarısının da aynı derecede önemli olduğu tespit edilmiştir [10].

2008 yılında Murat Kayri tarafından gerçekleştirilen bir çalışmada, öğrencilerin performans göstergelerinin sürekli izlenebilmesi ve ürünler arasındaki örüntünün bilgisayar sistemleri tarafından oldukça kolay yapılabildiği e-portfolio değerlendirmeleri için veri madenciliğinde kullanılan yöntemlerin alternatif bir ölçme yaklaşımı olarak kullanımı önerilmektedir [11].

2009 yılında Ahmet Selman Bozkır, Ebru Sezer ve Bilge Gök tarafından gerçekleştirilen bir çalışmada, ÖSYM tarafından 2008 ÖSS adayları için resmi internet sitesi üzerinden yapılan anket verileri üzerinde veri madenciliği yöntemleri kullanılarak, öğrencilerin başarılarını etkileyen faktörler araştırılmıştır. Bu çalışmada, veri madenciliği yöntemlerinden karar ağaçları ve kümeleme kullanılmıştır [12].

2006 yılında Serdar Çiftci tarafından gerçekleştirilen çalışmaya benzer olarak 2009 yılında Serdar Savaş ve Nursal Arıcı tarafından gerçekleştirilen bir çalışmada, web tabanlı uzaktan eğitim için video destekli ve animasyon destekli öğretim modeline uygun iki farklı öğretim materyali, bu materyallerin öğrenci başarısı üzerindeki etkilerinin incelenmesi için hazırlanmıştır. Analiz sonucunda video destekli öğretim materyallerinin animasyon destekli öğretim materyallerine göre öğrenci başarısını daha olumlu etkilediği belirlenmiştir [13]

2012 yılında Çağdaş Kurt, O. Ayhan Erdem tarafından yapılan çalışmada öğrencilerin başarılarına etki edebilecek faktörler farklı veri madenciliği algoritma ve modelleriyle incelenmiştir. Ekonomik, sosyal, kişisel, çevresel değişkenler üzerindeki yapılan uygulamada bazı sonuçlar saptanmış ve bunlara öneriler sunulmuştur. [14].

2014 yılında Sevil Özarslan ve Necaattin Barışçı tarafından yapılan çalışmada Kırıkkale Üniversitesi öğrenci bilgi sisteminden alınan; öğrencinin bölüme yerleştirmede esas alınan puan türü (sayısal, sözel, eşit ağırlık, yabancı dil, özel

yetenek, sınavsız geiř), ğrencinin eđitim grdđ akademik birim (faklter-yksekokul), đrencinin cinsiyeti (kız, erkek), đrencinin bařarı durumu (ok iyi, ortalama, bařarısız), đrencinin yař aralıkları, đrencinin dersi aldıđı dnem (gz, bahar), dersin verildiđi eđitim sistemi (yz yze eđitim, uzaktan eđitim) verileri kullanılmıřtır. J48 algoritması ile yapılan sınıflandırma sonucunda %82,22 ve ok Katmanlı Algılayıcı ile yapılan sınıflandırma sonucunda %80,74 bařarı yakalanmıřtır [15].

Arařtırma dahilinde 2012-2013 Eđitim-đretim Yılı'nda Kırıkkale İl'inde bulunan Anadolu Liselerinde okuyan 11. sınıf đrencilerine ait veriler Veri Madenciliđi yntemleri ile incelenmiřtir. Arařtırma dâhilinde Kırıkkale İl'inde bulunan 5 farklı Anadolu Lisesinde eđitim gren 231 đrenciyle yapılan anket sonucunda elde edilen veriler kullanılmıřtır.

Bu alıřma ile đrenci performansına etki eden faktrler belirlenip, bařarı ya da bařarısızlıđa etki eden faktrlerin birbiri ile olan iliřkisi arařtırılmıřtır. đrenci bařarısına etki eden faktrler belirlenirken daha nce eđitim alanında yapılan alıřmalar incelenmiřtir. Veri madenciliđinin eđitim ve đretimin kalitesini ve verimliliđini artırmak iin nasıl kullanılabileceđi irdelenmiř ve yapılacak sonraki alıřmalar iin neriler sunulması hedeflenmiřtir. Uygulama WEKA 3.7. programı ile gerekleřtirilmiřtir.

2015 yılında AB 2015 Akademik Biliřim Konferansı kapsamında yapılan alıřmanın anlatıldıđı bir bildiri yayınlanmıřtır [16].

## 2. MATERYAL VE YÖNTEM

Bu bölümde veri madenciliği hakkında genel bilgiler verilmiştir. Veri madenciliğini oluşturan disiplinlerden bahsedilip, veri madenciliği tekniklerine değinilmiştir.

### 2.1. Veritabanı

Veri madenciliği kavramının ve uygulama alanlarının anlaşılabilmesi için, çeşitli işlemlerden ve ortamlardan elde edilen bilgilerin elektronik ortamda saklandıkları mantıksal bölümler olan veritabanı kavramının ve özelliklerinin iyi bilinmesi gerekmektedir. Veri madenciliği teknikleri sonucunda ulaşılan sonuçların kaynağı veritabanlarında tutulan verilere dayanmaktadır.

#### 2.1.1. Veritabanı Nedir?

Kurumlarda, gerçek zamanlı operasyonel verilerin saklandığı ortamlardır. Önceleri, kurumlarda veri saklama ortamı olarak defterler, kartonlar kullanıldı. Kurumda yapılan her türlü işlem, işlenerek dosyalandı. Fihristler ve indexler oluşturularak çeşitli arama bulma yöntemleri geliştirildi. Ancak bütün bunlar insan eliyle manuel yapılmakta olduğundan, kayıt, arama, bulma, sorgulama işlemleri çok zaman alıyordu [17].

Bilişim teknolojilerindeki gelişmelere paralel olarak, kurumlardaki gerçek zamanlı işlevsel veriler elektronik ortamlara kayıt edilerek saklanmaya başlandı. Bilişim teknolojilerindeki yeni ürünlerin fiyatları ucuzladı ve birçok kurum tarafından ulaşılabilir hale geldi. Bunların sonucunda, hemen hemen tüm kurumlar işlevsel verilerini elektronik ortamda kayıt altına alıp saklamaya başladı [18].

### **2.1.2. Veritabanı Özellikleri**

Bilgisayar ortamında yapılan kayıt ve bilgiye sonradan erişim işlemleri geçmişle karşılaştırılmayacak kadar kolay ve hızlı yapılmaktadır. Kolaylık ve hız kadar diğer önemli kavramlar da doğruluk ve tutarlılıktır. Veritabanı sistemlerinin günümüzde popüler olarak kullanılmasının temel sebepleri de bunlardır.

Kurumların gerçek zamanlı işlevsel verilerinin tutulduğu veritabanlarına OLTP (Online Transaction Processing – Çevrimiçi işlem yürütme) veritabanları denir. Bu tür veritabanlarına, aynı anda birçok kullanıcı kayıt girebilir. Girilen kayıtlar üzerinde çok miktarda silme, güncelleme, sorgulama gibi temel veritabanı işlemleri yapılabilir [19].

### **2.1.3. Veritabanı Nesneleri**

OLTP veritabanları içerisinde birçok mantıksal nesne bulunur. Bunların temeli ve en önemlileri ise tablo ve indekslerdir. Veritabanına kayıt edilen bilgiler tablolarda tutulur. Bu tablolarda yer alan bilgilere daha hızlı erişim ve sorgulama yapılabilmesi için kullanılan nesnelere de indekslerdir. Bir veritabanında çok sayıda tablo olabileceği gibi, bir tabloda da farklı niteliklere göre hazırlanmış çok sayıda indeks olabilir. Her veritabanının içerisinde, kendi içinde oluşturulmuş olan nesnelere tutan ayrı birer sistem tabloları ve şemaları bulunmaktadır. Bunlara META-DATA denir. Tablo ve indeksler dışında, günümüz OLTP veritabanlarında birçok nesne bulunmaktadır. Bunların başlıcaları; görüntüler, saklanmış yordamlar, tetikleyiciler, kısıtlamalardır.

Günümüzde OLTP veritabanı sistemleri çok gelişmiştir. Tablolarda milyonlarca kayıt tutulabilir, bu kayıtlar üzerinde çok kısa sürelerde arama bulma işlemleri gerçekleştirilebilir. Ayrıca her OLTP veritabanı sistemi kendi VTYS (Veritabanı Yönetim Sistemi) sistemine sahiptir [20].

#### **2.1.4. Veritabanı Sorgulaması Nedir?**

OLTP veritabanlarında gerçek zamanlı operasyonel veriler tutulmaktadır. Değişik zamanlarda, bu veritabanları içerisinde yer alan veriler sorgulanarak, verilerden anlamlı bilgiler çıkarılmaya çalışılmaktadır. 1990'lı yılların ortalarına kadar, veriden anlamlı bilgi çıkarma işlemlerinin tamamına yakını veritabanı sorgulaması ile elde edilmiştir. Veritabanında, sorgulama ile elde edilen veriler çeşitli tablolama yazılımlarına aktarılmış ve ihtiyaç duyulan analizler o yazılımlar üzerinde yapılmıştır.

OLTP veritabanlarından sorgulama yapabilmek ve istenen kriterlere ait veri setlerini elde edebilmek için SQL (Structured Query Language – Yapısal Sorgu Dili) kullanılmaktadır.

## **2.2. Veri Ambarı**

### **2.2.1. Veri Ambarı Nedir?**

OLTP veritabanları belirli bir döneme aittirler. Bu veritabanları içerisinde ilgili döneme ait veriler tutulur. Kısa vadeli ihtiyaçlara yönelik sorgulama ve raporlamalar bu veritabanları üzerinden kolaylıkla yapılabilir. Ancak, uzun vadeli ve çok boyutlu karmaşık analizler yapılamamaktadır. Bunun 3 temel sebebi bulunmaktadır:

- Uzun vadeli stratejik kararların alınabilmesi için eldeki veri geniş bir aralığı kapsamalı, bu verinin geniş aralıkta değişimi analiz edilmelidir.
- OLTP veritabanlarındaki veri sürekli güncellenmektedir. Bundan dolayı, aynı geçmiş bir döneme ait farklı zamanlarda çalıştırılan iki sorgu farklı sonuç değerleri üretebilmektedir.
- OLTP veritabanları üzerinde kullanıcıların yapmış olduğu veritabanı işlemlerinden dolayı yoğunluk oluşmaktadır. Veritabanı sistemi, aynı anda hem INPUT (girdi) işlemlerine hem de yoğun OUTPUT (çıkıtı) işlemlerine yanıt vermekte zorlanmaktadır. Dolayısı ile veritabanı üzerinde çalıştırılan

karmaşık bir sorgunun sonuçlanması uzun zaman almakta bazen sistem kaynaklarının bitmesine sebep olmaktadır.

Bütün bu sebeplerden dolayı ortaya Veri Ambarı kavramı çıkmıştır. OLTP veritabanlarında tutulan veriler, konularına göre ayrılıp belirli zamanlarda zaman boyutuna göre farklı veritabanlarından arşivlenmektedir. Bu arşivlenmiş ve yapılmak istenen analizlere göre şeması yeniden yapılandırılmış olan veritabanlarına Veri Ambarı adı verilmektedir [21].

### 2.2.2. Veri Ambarı Özellikleri

Veri ambarlarına kısaca kurumlardaki veri deposu diyebiliriz. Kurum içerisindeki zaman içerisinde oluşan, gerçek zamanlı operasyonel verilerin konularına göre ayrıştırılıp, birleştirilip zaman boyutunda depolandığı ve çok boyutlu analizler için kullanıldığı ortamlardır. Veri ambarlarının ortak özellikleri ise şunlardır:

- a. **Konuya yöneliktir:** Kurumlarda veri ambarları belirli bir konuya göre belirlenirler. İşletmelerdeki veri ambarları düşünüldüğünde siparişler, satışlar, alışlar, müşteriler, ürünler birer veri ambarı konusudur.
- b. **Bütünleşiktir:** Veri ambarlarında bulunan veriler mutlaka bütünleştirilmiş olmalıdır. Farklı veritabanlarından gelen bilgilerde, aynı değeri ifade etmek için farklı semboller, kısaltmalar kullanılabilir. Bu türden farklılıklar yok edilmeli ve veriler alınmadan önce mutlaka dönüştürme ve standartlaştırma işlemi yapılmalıdır.
- c. **Zaman boyutu vardır:** Veri ambarında bulunan veriler, geçmişte belirli bir zaman dilimine aittirler. Ortalama zaman periyodu 5 yıldır.
- d. **Sadece okunabilir:** Veri ambarına yüklenen veriler sadece erişilebilirler.



### 2.2.3. Veritabanı ile Veri Ambarı Arasındaki Farklar

Öncelikle her iki sistemin kullanım amaçları farklıdır. OLTP veri tabanlarının kullanım amacı, kurumdaki faaliyetin devam etmesidir. Veri ambarlarının amacı ise, geçmiş verileri çok boyutlu olarak analiz etmek ve elde edilen sonuçları geleceğe dönük olarak alınan kararlar için kullanmaktır.

OLTP veritabanlarında gerçek zamanlı operasyonel veriler saklandığı için, giriş-çıkış odaklıdır. Veri ambarları ise, sorgulama ve çözümlene odaklı oldukları için belirli periyotlarda veri yüklemesinin dışında bir giriş-çıkış işlemi yapılmaz.

OLTP veritabanlarından ve veri ambarından bilgiye ulaşma yöntemleri farklıdır. OLTP veritabanlarında önceden tanımlanmış olan sabit SQL sorguları kullanılabilir. Ancak, veri ambarlarında böyle bir durum söz konusu değildir. Hangi verilerden ne tür bilgi elde edileceği önceden tanımlanmamıştır. OLTP veritabanlarında ilgili döneme ait veriler saklanmaktadır. Veri ambarlarında ise, geçmişe yönelik tüm veriler tutulmaktadır [22].

**Çizelge 2.1.** OLTP Veritabanı ile Veri Ambarları Karşılaştırma Tablosu [23]

	<b>OLTP Veritabanı Sistemi</b>	<b>Veri Ambarı</b>
Amaç	Günlük veri kayıt etmek	Çok boyutlu analiz yapmak
Veri Modeli	2 Boyutlu	Çok boyutlu
Tablo Yapısı	Az alanlı çok tablo	Çok alanlı az tablo
Güncelleme	Sürekli	Daha uzun aralıklarda
Veri Durumu	Dinamik	Statik
Sorgu Durumu	Statik	Dinamik

## **2.3. Veri Madenciliđi**

Veri setleri ierisinde ok fazla bilgi bulunmaktadır. Ancak, bu bilgilerin hepsi ilk bakışta grnmez. eřitli verilerin birbirleri arasındaki iliřkilerin eřitli algoritmalar ile incelenmesi neticesinde bu veriler anlam kazanarak bilgiye dnřler. Sistem kullanıcılarına bir Őeyler anlatmaya alıřırlar. Elbette, bu tr gizli kalmıř bilgileri ortaya ıkarmak kolay olmayacaktır. Bu konuda yođun alıřmalar yapılmakta olup eřitli yntemler ve teknikler geliřtirilmektedir.

### **2.3.1. Veri Madenciliđine Giriř**

Herhangi bir belge veya bilgiyi saklamaktaki temel amacımız, ileride bu belge ya da bilgiyi yeniden kullanabilmektir. Gemiř yıllarda insanlar bilgi ve tecrbelerini aktarmada kâđıt ortamlarını kullanmıřtır. Zamanla bu durum hem iř ykn arttırmıř hem de bilgiye ulařımı zorlařtırmıřtır. Bu durum verilerin bilgisayar ortamında tutulmasını zorunlu kılmıřtır.

Dijital verilerin gn getike artıř gstermesi ile birlikte bilgi miktarlarında byk artıřlar sz konusu olmaktadır. Gnmzde bilgi teknolojileri ok byk miktardaki verilerin toplanmasına, saklanmasına, iřlenmesine ve tekrar bilgiye dnřtrlmesine olanak sađlamaktadır.

Boyutları gn getike artıř gsteren veriler veri tabanlarında depolanmaktadır. Depolanan veriler genelde tek bařına bir anlam ifade etmemektedirler. Zamanla byk miktardaki eřitli veriler iinde sistemlerin ihtiya dođrultusunda anlamlı bilgilerin elde edilebilmesi gerekmektedir. Bundan dolayı byk miktardaki verilerden anlamlı bilgilerin ıkartılması iin veri inceleme ve analizi yapan eřitli teknolojiler geliřtirilmesine ihtiya duyulmuřtur. Dolayısı ile veri tabanlarından bilgi keřfi yapacađımız bir sre sz konusu olmuřtur.

Veri madenciliği dünyanın anlaşılabilirliğine önemli ölçüde destek olan bir kavramdır. Gelişen bilgi toplama, depolama ve işleme yetkinlikleri, giderek artan mevcut verilerin incelenerek anlamlı sonuçlar elde edilmesine olanak sağlamaktadır [24] .

### **2.3.2. Veri Madenciliği Nedir?**

Basit bir tanım yapmak gerekirse veri madenciliği, büyük ölçekli veriler arasından bilgiye ulaşma, bilgiyi işleme işidir. Ya da bir anlamda büyük veri yığınları içerisinde gelecekle ilgili tahminde bulunabilmemizi sağlayabilecek bağıntıların bilgisayar programı kullanarak aranmasıdır. Veri madenciliği geniş anlamda veri analiz teknikleri bütünüdür ve tek başına bir çözüm değildir. Mevcut problemleri çözmek, kritik kararları almak veya geleceğe yönelik tahminleri yapmak için gerekli olan bilgileri elde etmeye yarayan bir araçtır. Ortaya çıkarılması hedeflenen bilgiler; üstü kapalı, çok net olmayan, önceden bilinmeyen, daha önce keşfedilmemiş ancak potansiyel olarak kullanışlı anlamlı ve kritik bilgilerdir.

Veri madenciliği, büyük boyutlu veri ambarlarının oluşmasının bir sonucudur. Günümüzde kurumlar büyük miktarlarda veri üretmekte ancak bu veriler içerisinde anlamlı ve yararlı bilgiyi ortaya çıkarmakta zorluklar yaşamaktadırlar. Geleneksel istatistiksel yöntemlerle büyük boyuttaki veriyi çözümlenmek kolay değildir. Bu nedenle verileri işlemek ve çözümlenmek için özel yöntemlere gereksinim duyulmuştur. Veri madenciliği yöntemleri bu gereksinimi karşılamak üzere ortaya çıkmıştır [25].

1960'lerde veriler elektronik ortamda toplanmaya ve geçmiş veriler bilgisayarlar ile analiz edilmeye başlanmıştır. 1980'lerde bağıntılı veritabanları ve SQL ile verilerin dinamik ve anlık analiz edilmesine olanak sağlanmıştır. 1990'lara gelindiğinde toplanmakta olan verinin hacmi çok büyük boyutlara ulaşmış ve verilerin depolanması için veri ambarları kullanılmaya başlanmıştır. Veri madenciliği toplanan bu büyük veri kütlelerinin değerlendirilmesi için istatistik ve yapay zeka tekniklerinin kullanılması sonucunda ortaya çıkmıştır.

Veri madenciliği Şekil 2.1.'de görüldüğü gibi veri madenciliği yapay zekâ, istatistik ve karar ağaçları, raporlama gibi alanlar ile yakından ilişkili disiplinler arası bir alandır.



**Şekil 2.1.** Veri Madenciliğini Oluşturan Disiplinler [26]

Veri madenciliği kullanılan yaklaşıma bağlı olarak yapay sinir ağları, bulanık mantık, genetik algoritmalar, mantıksal programlama ya da makine öğrenmesi gibi diğer teknikler ile kullanılabilir. Veri madenciliği sistemleri analiz türüne ve verinin içeriğine bağlı olarak uzaysal veri analizi, örüntü tanımlama, görüntü analizi, sinyal işleme, bilgisayar grafikleri, web teknolojisi, ekonomi, iş dünyası, biyoinformatik veya fizyoloji alanlarına ilişkin teknikler ile entegre olabilir [22].

Genel olarak veri madenciliğinde vurgulanan unsurların istatistiğin tanımı içinde yer aldığı görülmektedir. Veri madenciliğinde ulaşılmak istenen amaç ile istatistik biliminin amacı; verilerden bilgiyi keşfetmektir. Her ikisinde de temel olan öğeler veri ve bilgidir. Bir tanım yapmak gerekirse veri madenciliği istatistik biliminin teknolojiyle bütünleşmesi sonucu oluşturulan bir araçtır [27].

Özetle, veri madenciliği iki gereksinimden ortaya çıkmaktadır:

- Toplanan çok miktarda verinin işleme ihtiyacı
- Artan rekabette doğru karar verebilme yetkinliğini artırmak ihtiyacı

Veri madenciliği ile ilgili bazı tanımlar şöyledir:

Veri madenciliği, hem yararlı hem de anlaşılabilir verilerle, alıılmamış yollarla verileri inceleyen ve gizli ilişkileri ortaya koyan bir analiz yöntemidir [28].

Veri madenciliği, öncelikle bilinmeyen desenlerin ortaya konması amacıyla bilimsel ve teknik yollarla veri araştıran, veri tabanındaki bilgi keşfi süreçlerinden biridir [29].

Gartner Group'a göre veri madenciliği, istatistiksel ve matematiksel yöntemler kadar desen tanıma teknolojilerinin de kullanılmasıyla, depolardaki muazzam miktardaki depolanmış verilerin elenmesi ile yeni anlamlı birliktelikler, desenler ve trendler keşfetme sürecidir [30].

Veri madenciliği büyük hacimli verilerden öz bilginin çıkarılması sürecidir [31].

Veri madenciliği büyük veri tabanlarındaki gizli bilgi ve yapıyı açıklamak için, çok sayıda veri analizi aracını kullanan bir süreçtir [32].

İş kararlarının alınabileceği doğru, alıılmamış, faydalı ve anlaşılabilir örüntüler veya modellerdir [33].

Bilgisayar teknolojilerinin sağlamış olduğu çok hızlı veri işleme ve yüksek hacimde veri depolama imkânları yardımıyla ve farklı disiplinlerin katkısıyla sağlanan araçlarla, sahip olunan çok büyük hacimlerdeki veriden, karar vericinin etkin ve daha fazla bilgiye dayalı karar vermesinde kullanabilmesi için önceden bilinmeyen, gizli, örtük, klasik metotlarla ortaya çıkarılması güç, faydalı, ilginç, anlaşılabilir; ilişki, örüntü ve bağıntıların otomatik veya yarı otomatik bir şekilde ortaya çıkarılması olarak tanımlanır [34].

Veri madenciliği genel anlamda, büyük miktarda veri içerisinde, gizli kalmış, değerli, kullanılabilir bilgilerin açığa çıkarılmasıdır [35].

Veri madenciliği ve öz bilgi keşfi, verilerde daha öncede bilinmeyen, anlamlı ve değerli bilgiler elde etme işlemidir [36].

Veri madenciliği, verinin bütünü kullanması bakımından diğer istatistiksel verilerden ayrılır. Bu yöntemle, geleneksel yollarla elde edilmiş küçük verilerle çalışma yerine daha kolay değerlendirme yapabilecek, yeni bağımsız veriler tercih edilebilmektedir [37].

### **2.3.3. Veri Madenciliği Ne Değildir?**

Veri madenciliği; veri toplamak, mevcut verilerden sorgulamalar yapmak veya gelişmiş analiz teknikleri kullanmanın ötesinde bir noktadır.

Bir satış şirketinde; hangi müşterilerin devamlılık gösterdikleri, hangi bölgelerde performans düşüklüğü yaşadıklarını belirlemek veri madenciliği değildir.

Gelir ile yaş ilişkisinin incelendiği bir değişken, bir sonuç ve az sayıda veriden oluşan bir modeli tanımlayarak yaşa göre gelir tahmini yapmak da veri madenciliği değildir.

### **2.3.4. Veri Madenciliğinin Tarihçesi**

Veri madenciliği teknikleri üzerine matematikçiler 1950'li yıllarda çalışmaya başlamışlar, mantık ve bilgisayar bilimleri alanlarında yapay zekâ ve makine öğrenme konularını geliştirmişlerdir. 1960'lı yıllarda ise istatistikçiler yeni bazı algoritmalar üzerinde çalışmışlardır. Örneğin regresyon analizi, en büyük olasılık kestirim, sinir ağları vb. yöntemler veri madenciliğinin ilk adımlarını oluşturmuştur.

Ayrıca veri tabanı sistemleri giderek gelişmiş ve büyük sayıda metin dokümanlarının saklanması ve bilginin geri kazanılması sağlanmıştır.

1970, 1980, 1990'lı yıllarda yeni programlama dilleri ve yeni bilgisayar tekniklerinin geliştirilmesi, genetik algoritmalar, kümeleme yöntemleri ve karar ağaçları gibi algoritmaları da içermiştir.

1990 yılıyla beraber veri tabanında bilgi keşfinin ilk adımları oluşturulmuş ve veri ambarı geliştirilmiştir. Ayrıca aynı zaman içerisinde yeni teknolojilerle beraber veri madenciliği değiştirilerek yaygın olarak kullanılan standart bir işin parçası olmuştur.

1990'lı yıllardan itibaren veri madenciliği konusu verilerin yoğun olduğu araştırma alanlarında bilgi keşfi ismiyle kullanılmaya başlanmıştır. İlk yıllar çoğunlukla veri tabanlarındaki veriler üzerinde yürütülen çalışmalar zamanla veri tabanında tutulmayan verileri de kapsayacak şekilde genişlemiştir. Geçmiş tüm bu çalışmaların değerlendirilmesi veri madenciliğinin geleceğini konusunda fikir vermesi açısından önem taşımaktadır.

## **Çizelge 2.2.** Veri Madenciliğinin Tarihsel Süreci [38]

1950'ler	<ul style="list-style-type: none"><li>• İlk bilgisayarlar (Sayım için)</li></ul>
1960'lar	<ul style="list-style-type: none"><li>• Veritabanı ve verilerin depolanması</li><li>• Perseptronlar (Algılayıcı, Fark edici)</li></ul>
1970'ler	<ul style="list-style-type: none"><li>• İlişkisel Veritabanı Yönetim Sistemleri</li><li>• Basit kurallara dayanan uzman sistemler ve Makine öğrenimi</li></ul>
1980'ler	<ul style="list-style-type: none"><li>• Büyük miktarda veri içeren veritabanları</li><li>• SQL sorgu dili</li></ul>
1990'lar	<ul style="list-style-type: none"><li>• Veritabanlarında Bilgi Keşfi Çalışma Grubu ve Sonuç bildirgesi</li><li>• Verimadenciliği için ilk yazılım</li></ul>
2000'ler	<ul style="list-style-type: none"><li>• Tüm alanlar için verimadenciliği uygulamaları</li></ul>

### 2.3.5. Veri Madenciliği Yararları

Veri madenciliğinin karar verici için olası yararları aşağıdaki gibi sıralanabilir:

- Mevcut müşterilerin karar verici tarafından daha iyi tanınmasını sağlayabilir.
- Özellikle finans sektöründe mevcut müşterileri bölümlere ayırıp, kredi risk davranış modelleri oluşturarak, yeni başvuruda bulunan müşterilere karşı riskin minimize edilmesini sağlayabilir.
- Finans ve borsa kuruluşlarında stok fiyatları tahminleri, portföy yönetimi yapılabilir.
- Mevcut müşterilerin ödeme performansları incelenip kötü ödeme performansı gösteren müşterilerin ortak özellikleri belirlenerek, benzer özelliklere sahip tüm müşteriler için yeni risk yönetim politikaları oluşturulabilir.
- En iyi müşteriler veya müşteri bölümlerinin bulunmasında kullanılabilir. Bulunan bu iyi müşteri bölümlerine yönelik yeni pazarlama stratejileri oluşturulabilir.
- Kuruluşlar tarafından düzenlenecek çeşitli kampanyalarda mevcut müşteri kitlesinin seçimi ve bu müşterilerin davranış özelliklerine yönelik kampanya şartlarının oluşturulması sağlanabilir.
- Bankacılık faaliyetlerinde, küçük işletmelere yönelik olarak makine ve ekipman satışı yapan dağıtıcı firmalarla ortak hareket ederek oluşturulacak satış paketleri ile pazarlama stratejileri geliştirilebilir.
- Veri madenciliği ile mevcut müşteriye tanıyarak kuruluşların müşteri ilişkileri yönetimlerinde düzenleme ve geliştirmeler yapılabilir.
- Günümüzde var olan yoğun rekabet ortamında kuruluşların hızlı ve kendisi için en doğru kararı almalarını sağlayabilir.
- Kuruluşlar veri analizi ile müşterilerini kişiselleştirilmiş ürün ve hizmetler hakkında bilgilendirebilirler.
- Veri madenciliği ile kuruluşların müşteriyle bütünleşmiş satış politikaları oluşturması sağlanabilir.
- Laboratuvar veya bilgisayar ortamında sistemlerin benzetimi ve analizi sürecinde elde edilen yüksek miktarda bilimsel veriler anlamlandırılabilir.



- Sağlık alanında tarama testlerinden elde edilen verileri kullanarak çeşitli kanserlerin ön tanısı, kalp verilerini kullanarak kalp krizi riskinin tespiti, acil servislerde hasta semptomlarına göre risk ve öncelikler tespit edilebilir.
- Öğrenci işlerinde veriler analiz edilerek öğrencilerin başarı ve başarısızlık nedenleri, başarının artırılması için hangi konulara ağırlık verilmesi gerektiği, üniversite giriş puanları ile okul başarısı arasındaki bir ilişkinin var olup olmadığı belirlenebilir.
- Birçok web sunucusu veya çevrimiçi servisten kullanıcı erişim desenlerinin analizi ve keşfi yapılabilir.
- Dokümanlar arasında elle bir tasnif gerektirmeden benzerlikler hesaplanabilir.

### 2.3.6. Veri Madenciliği Kullanım Alanları

Veri madenciliği günümüzde yaygın bir kullanım alanı bulmaktadır. Örneğin, pazarlama, eğitim, bankacılık ve sigortacılık gibi alanlarda ve elektronik ticaret ile ilgili alanlarda yaygın şekilde kullanılmaktadır. Bunlar kullanım yerlerine göre aşağıdaki gibi sınıflandırılabilir.

Pazarlama Alanında;

- Müşterilerin satın alma alışkanlıklarının belirlenmesi,
- Müşterilerin demografik özellikleri arasındaki bağlantıların ortaya konması,
- Mevcut müşterilerin elde tutulması, yeni müşterilerin kazanılması,
- Pazar sepeti analizi,
- Müşteri ilişkileri yönetimi,
- Müşteri değerlendirme,
- Benzer özellikleri olan müşterilerin bulunması,
- Satış tahmini.

Perakendecilik Alanında;

- Etkin ürünlerin benzerliklerini tespit etmek,
- Müşteriler için birçok ürün bulunması,

- Ürün satışları arasındaki ilişkiyi tespit etmek.

#### Bankacılık Alanında;

- Müşteri değerlerinin tanımlanması,
- Müşteriler arasındaki benzerliklerin tanımlanması,
- Aldıkları hizmeti iptal etme riski bulunan müşterileri gösteren raporlar oluşturulması,
- Farklı finansal göstergeler arasındaki gizli korelasyonların tespiti,
- Kredi kartı dolandırıcılıklarının tespiti,
- Kredi kartı harcamalarına göre müşteri gruplarının belirlenmesi,
- Kredi taleplerinin değerlendirilmesi.

#### Sigortacılık Alanında;

- Yeni poliçe talep edecek müşterinin tahmin edilmesi,
- Sigorta dolandırıcılıklarının tespit edilmesi,
- Riskli müşteri gruplarının belirlenmesi.

#### Elektronik Ticaret Alanında;

- Saldırıların çözümlenmesi,
- E-CRM uygulamalarının yönetimi,
- Web sayfalarına yapılan ziyaretlerin çözümlenmesi,
- Kullanıcı davranışlarına göre web sitesinin yenilenmesi.

#### Telekomünikasyon Alanında;

- İletişim ağlarında sorunlu bölgelerin tespiti,
- Kaçak hat kullanımlarının belirlenmesi,
- Kullanıcı davranışlarının belirlenmesi,
- Müşteri davranışlarına göre yeni hizmetlerin sunulması.

#### Tıp Alanında;

- DNA içerisindeki genlerin sıralarının belirlenmesi,
- Protein analizlerinin yapılması,
- Hastalık haritalarının hazırlanması,
- Hastalık tanıları,

- Sağlık politikalarına yön verilmesi,
- Hastalıkları etkileyen faktörlerin ortaya çıkartılması,
- Hastalıklara erken teşhis koyularak sağlığın korunması ve doğru tedavi yöntemlerinin seçilmesi,
- Sağlık hizmetlerinin kalitesinin artırılması ve geleceğe dönük doğru sağlık politikalarının oluşturulması,
- Koruyucu hekimliğin yaygınlaştırılması ve sağlık harcamalarının düşürülmesi,
- Salgın hastalıkların tespit edilmesi gerekli önlemlerin alınması.

Eğitim alanında;

- Öğrenci profillerine göre başarının tahmin edilmesinde,
- Benzer özellikleri gösteren öğrencilerin belirlenmesinde,
- Ölçme ve değerlendirme sistemlerini geliştirmede,
- Öğrenme ortamlarının geliştirilmesine yönelik araştırma-geliştirme çalışmalarının yapılmasında,
- Başarılı öğrenme ortamlarının oluşturulabilmesi için çeşitli uygulamalar geliştirilmesinde.

Bunların dışında veri madenciliğinin faydalı olabileceği ve kullanılabileceği alanlardan bazıları şunlardır;

- Taşımacılık ve ulaşım,
- Turizm ve otelcilik,
- Belediyeler,
- Bilim ve mühendislik.

### **2.3.7. Veri Madenciliği Süreci**

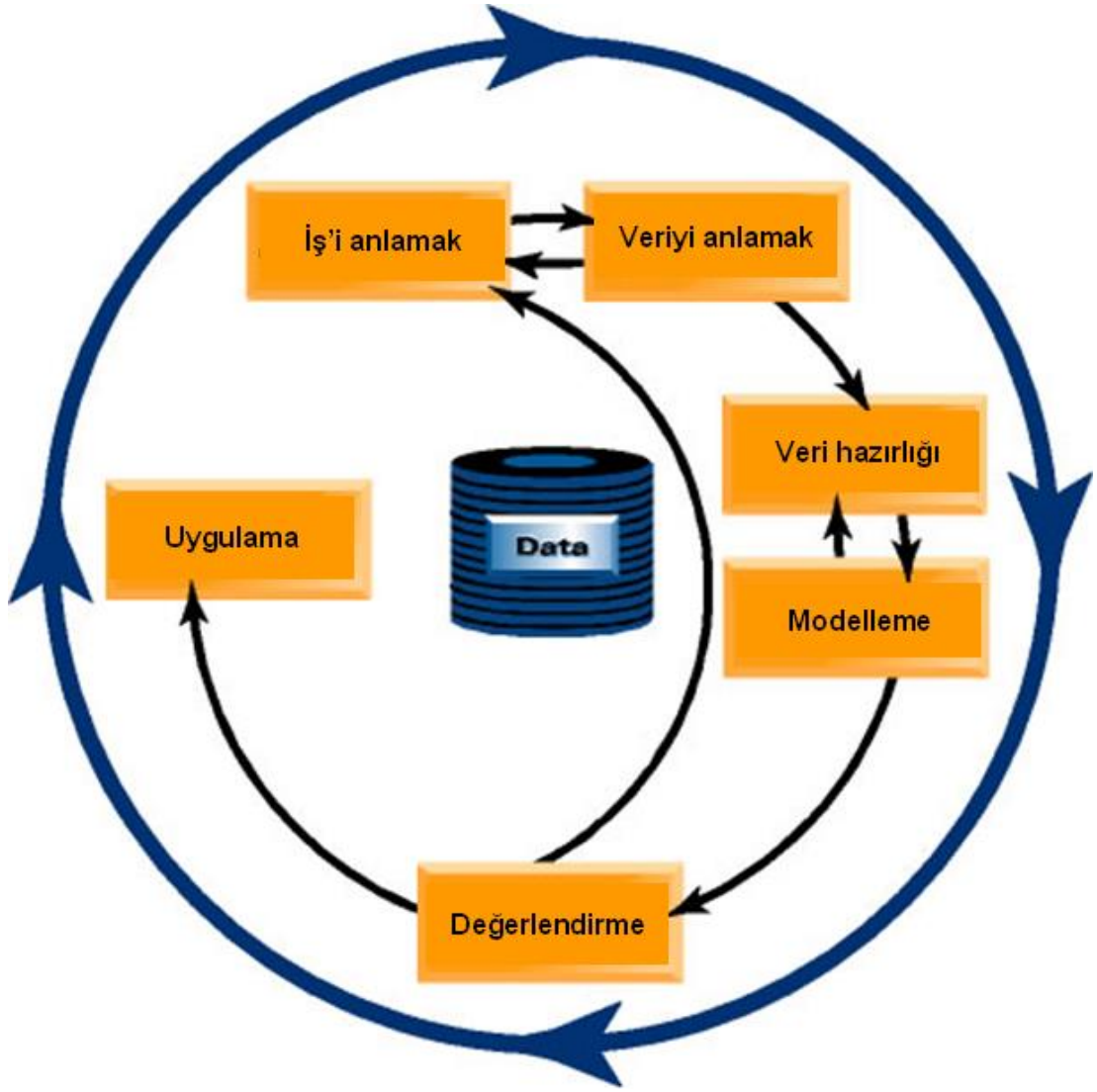
Veri madenciliği bir süreçtir. Bu süreçte ana unsur süreci gerçekleştiren uygulamacıdır. Süreçte bulunan adımlar doğru olarak yerine getirilmediği sürece istenilen sonuca ulaşılması mümkün değildir [39].

Veri madenciliği sureci en basit şekli ile dört adımdan oluşmaktadır.

- Veri Seçmek: Mevcut olan ve elde edilebilecek verilerin oluşturduğu veri havuzundan çalışma için kullanılacak verilerin seçilmesi
- Veri İşlemek: Farklı kaynaklardan alınan verilerin birleştirilmesi, hatalı verilerin çıkarılması gibi işlemlerle seçilmiş verilerin kullanılabilir hale getirilmesi
- Veri Eğilimlerini / Desenlerini Belirlemek: İşlenmiş verilerin veri madenciliği fonksiyonları ve algoritmaları ile değerlendirilerek verilerden anlamlı eğilimlerin, desenlerin çıkarılması
- Bilgiye Ulaşmak: Verilerden çıkarılan anlamlı eğilimler ve desenlerin yorumlanarak bilgi elde edilmesi

Veri madenciliği farklı disiplinlerden faydalanırken kendi içerisinde de süreçlere sahiptir. Bu süreçlerin doğru uygulanması elde edilen bilgilerin doğruluğunu, kalitesini artıracaktır. Bu süreçlerin standart hale getirilmesi için yapılan yoğun çalışmalar sonucunda veri madenciliği süreçlerinden en yaygın olarak kullanılanı, Veri madenciliği araçlarını satan firmalardan bazılarının ve veri madenciliği uygulamalarını işletme faaliyetlerine uyarlayan ilk kuruluşların 1996 yılında oluşturduğu bir konsorsiyum tarafından geliştirilen, CRISP-DM (Cross Industry Standard Process for Data Mining)/ (Çapraz Endüstri Veri Madenciliği Standart Süreci)'dir. CRISP-DM, bilgi keşfi için veri madenciliğinin temel adımlarını tanımlayan kademeli bir süreçtir [40]. Veri madenciliği projelerinin hızlı, daha verimli ve daha az maliyetli gerçekleştirilmesi için geliştirilmiş olan bu süreç altı adımdan oluşmaktadır.

Çapraz Endüstri Veri Madenciliği Standart Süreci süreç şeması Şekil 2.2. 'de gösterilmiştir.



**Şekil 2.2.** Çapraz Endüstri Veri Madenciliği Standart Süreci Süreç Şeması [41]

İlk adım “iş'i anlama” adımıdır. Bu adımda çalışmanın amaçları ve ihtiyaçları belirlenip problem tanımlanır. İkinci adım “veriyi anlama” adımıdır. Bu adım tanımlanan problemin çözümü için kullanılacak verilerin bir araya getirilmesi, verinin incelenmesi, veri kalite problemlerinin çözülmesi gibi işlem faaliyetlerini içermektedir. “Veri hazırlama” adımı ise başlangıç veri kümesinden modelde kullanılacak veri kümesini oluşturmak için dönüştürme ve temizleme işlemleri uygulanır. “Modelleme” adımı problem ve veri özelliklerine uygun modelleme teknikleri seçilir ve model parametrelerinin en iyi değerleri belirlenir. Son iki adımda modelin değerlendirilmesi ve uygulanmasına ilişkin görevler yer almaktadır [42].

### 2.3.7.1. İŖi ve İŖ Ortamını Anlama

İŖi anlama aŖamasında, veri madenciliđi uygulamasını yapacak olan kiŖi iŖletmenin geliŖtirmek istediđi amacı dikkate almalıdır. alıŖmanın temel amacının belirlenmesi ve bu amacın mmkn olduđunca ikincil amalardan ayrıŖtırılarak net olarak tanımlanması gerekir. Analizi yapan kiŖinin hedefi, veri madenciliđi uygulamasının sonularını etkileyebilecek nemli kriterleri ortaya ıkarmak olmalıdır [43].

### 2.3.7.2. Veriyi Anlama

Veriyi anlama adımı ilk verilerin toplanması, mevcut verilerin uygunluđunun deđerlendirilmesi, modeli oluŖturmak iin gerekli farklı veri ihtiyalarının tespit edilmesi, sahip olunan kayıt sayısının yeterliliđi gibi veri kalite ve yeterliliđine ynelik dŖnce szgecinden geirilmesi aŖamasıdır. Veriyi anlamak ile iŖi anlamak i ie gemiŖ alt srelerdir. İŖi anladıka farklı verilere bakmak veya verilerin gsterdiklerini anlamak, verilere baktıka iŖ ile ilgili farklı bakıŖ acıları kazanmak mmkndr. Bu dng kendi iinde devam ettike alıŖmada kullanılacak veriler netlik kazanır. Bu adımda Ŗu iŖlemler yapılır:

- BaŖlangı verilerini toplamak
- Veriyi tanımlamak
- Veriyi keŖfetmek
- Verinin kalitesini belirlemek

Byk hacimli verilerin bulunduđu veri ortamlarında byk sorunlar ortaya ıkabilir. Bu nedenle kk veri kmelerinde, benzetim ortamlarında hazırlanmıŖ veri madenciliđi sistemleri, byk hacimli, eksik, grltl, boŖ, atık, aykırı veya belirsiz veri kmelerinin bulunduđu ortamlarda yanlıŖ alıŖabilir. Bu nedenle veri madenciliđi sistemleri hazırlanırken bu sorunların zlmesi gerekmektedir.

Veri madenciliği uygulamalarında karşılaşılabilecek sorunlar şunlardır:

- Atık veri: Problemden istenilen sonucu elde etmek için kullanılan örneklem kümesindeki gereksiz niteliklerdir.
- Belirsizlik: Yanlılıkların şiddeti ve verideki gürültünün derecesi ile ilgilidir.
- Boş veri: Bir veri tabanında boş değer, birincil anahtarlar yer almayan herhangi bir niteliğin değeri olabilir.
- Dinamik veri: Kurumsal çevrim içi veri tabanları dinamiktir ve içeriği sürekli olarak değişir.
- Eksik veri: Veri kümesinin büyüklüğünden ya da doğasından kaynaklanmaktadır. Eksik veriler olduğunda eksik veri içeren kayıt veya kayıtlar çıkarılabilir, değişkenin ortalaması eksik verilerin yerine kullanılabilir.
- Farklı tipteki verileri ele alma: Gerçek hayattaki uygulamalar makine öğreniminde olduğu gibi yalnızca sembolik veya kategorik veri türleri değil, aynı zamanda tamsayı, kesirli sayılar, çoklu ortam verisi, coğrafi bilgi içeren veri gibi farklı tipteki veriler üzerinde işlem yapılmasını gerektirir.
- Gürültü ve kayıp değerler: Veri girişi veya veri toplanması esnasında oluşan sistem dışı hatalara gürültü denir.
- Sınırlı bilgi: Veri tabanları genel olarak basit öğrenme işlerini sağlayan özellik veya nitelikleri sunmak gibi veri madenciliği dışındaki amaçlar için hazırlanmışlardır. Bu yüzden, öğrenme görevini kolaylaştıracak bazı özellikler bulunmayabilir.
- Veri tabanı boyutu: Veri tabanı boyutları büyük bir hızla artmaktadır. Veri tabanı algoritması çok sayıda küçük örnekleme ele alabilecek biçimde geliştirilmiştir. Aynı algoritmaların yüzlerce kat büyük örneklemelerde kullanılabilmesi için çok dikkat gerekmektedir.

### 2.3.7.3. Veri Hazırlama

Bu aşama başlangıç verilerinin, çalışmalara temel oluşturacak final verilere dönüştürülmesi aşamasıdır. Bu çalışmanın adımlarının belirgin bir sırası veya tekrar

sayısı yoktur. Modelin kurulması aşamasında ortaya çıkacak sorunlar, bu aşamaya sık sık geri dönülmesine ve verilerin yeniden düzenlenmesine neden olacaktır. Bu durum verilerin hazırlanması ve modelin kurulması aşamaları için, bir analistin veri keşfi sürecinin toplamı içerisinde enerji ve zamanının %50 - %85'ini harcamasına neden olmaktadır [44]. Bu adımda şu işlemler yapılır:

- Veri setini tanımlamak
- Veriyi seçmek
- Veriyi temizlemek
- Veriyi kurmak
- Veriyi birleştirmek
- Veri formatlamak

#### **2.3.7.4. Modelleme**

Birbirine benzeyen veri madenciliği problemleri için birden çok çözüm tekniği olabilmektedir. Bazı teknikler verilerde özel ihtiyaçlar duyarlar. Bu nedenle veri hazırlama ve model kurma aşamaları, en iyi olduğu düşünülen modele varılıncaya kadar yinelenen bir süreçtir. Bu adımda şu işlemler yapılır:

- Model tekniğini seçmek
- Model test tasarımı yapmak
- Modeli kurmak

#### **2.3.7.5. Değerlendirme**

Bu aşamaya gelindiğinde kurulmuş bir model vardır. Bu aşama, modelin nihai olarak sunulmasından önce modelin yoğun olarak değerlendirilmesi ve iş hedefleri ile uyumlu olup olmadığının kontrol edilmesini amaçlar. Kapsanmamış konu olup olmadığı değerlendirilir. Bu adımda şu işlemler yapılır:



- Sonuçları değerlendirmek
- Süreci değerlendirmek
- Gelecek adımları planlamak

#### **2.3.7.6. Yayma**

Modelin tamamlanmış olması projenin nihai sonucu değildir. Modelin amacı veriler hakkında bilinenleri artırmak dahi olsa, elde edilen veri kullanılacak biçimde organize edilmeli ve sunulmalıdır. Genellikle gerçek verilerden örneklerin sunulması şeklinde olur. Bu adımda şu işlemler yapılır:

- Yayma planını oluşturmak
- Takip ve bakımı planlamak
- Final raporu hazırlamak
- Projeyi değerlendirmek

#### **2.3.8. Veri Madenciliği Yöntem ve Teknikleri**

Veri Madenciliği uygulamalarında, her biri farklı bir amaca hizmet eden çok sayıda yöntem ve teknik bulunmaktadır. Bu yöntemlerin hepsi çeşitli matematik ve istatistiksel hesaplamalara dayanmaktadır. Her projede, amaca ve eldeki veri setine en uygun yöntem seçilmelidir.

Veri madenciliğinde kullanılan modeller, tahmin edici (Predictive) ve tanımlayıcı (Descriptive) olmak üzere iki ana başlık altında incelenmektedir [45].

Tahmin edici modellerde, sonuçları bilinen verilerden hareket edilerek bir model geliştirilmesi ve kurulan bu modelden yararlanılarak sonuçlan bilinmeyen veri kümeleri için sonuç değerlerin tahmin edilmesi amaçlanmaktadır [46]. Örneğin; bir banka daha önceden verdiği kredilerin ödenme durumlarını inceleyerek, müşteri tiplerine göre vereceği kredilerin sonuçlarını tahmin edebilir. Bu sayede başvuru

yapan müşterilerin nasıl davranacakları tahmin edilebilir. Modelin bir diğer kullanım yeri de veri tabanındaki boş alanların diğer alanlar ve eski kayıtlar kullanılarak tahmin edilmesidir [47].

Tanımlayıcı modellerde ise karar vermeye rehberlik etmede kullanılacak mevcut verilerdeki örüntülerin tanımlanması sağlanmaktadır. X/Y aralığında geliri ve iki veya daha fazla arabası olan çocuklu aileler ile çocuğu olmayan ve geliri X/Y aralığından düşük olan ailelerin satın alma örüntülerinin birbirlerine benzerlik gösterdiğinin belirlenmesi tanımlayıcı modellere bir örnektir [18].

Veri madenciliği modellerini gördükleri işlevlere göre üç ana başlık altında incelemek mümkündür.

#### **2.3.8.1. Sınıflama ve Regresyon**

Sınıflama ve regresyon, önemli veri sınıflarını ortaya koyan veya gelecek veri eğilimlerini tahmin eden modelleri kurabilen veri analiz yöntemidir. Sınıflama kategorik değerleri tahmin ederken, regresyon süreklilik gösteren değerlerin tahmin edilmesinde kullanılır [22].

Sınıflama en temel veri madenciliği fonksiyonlarından biri olarak kategorik sonuçları tahmin etmek için kullanılır. Modeli kurabilmek için, sonuçları önceden bilinen durumlar ve bu durumlarda ilgili faktörlerin aldığı değerler gereklidir. Bu değerler eğitim verisi olarak adlandırılır. Örneğin satışlarını artırmak için kampanya düzenlemek isteyen bir otomobil firması, kampanyasına katılma ihtimali olan potansiyel alıcıları belirlemek için daha önceden satış yapmış olduğu müşterilerinin verilerini kullanarak, hangi özelliklere sahip adayların kampanyaya katılabileceğini belirli bir olasılık aralığında tahmin edebilir. Bu şekilde; ihtiyacı kadar veri satın alarak ve sadece alma potansiyeli yüksek olan adaylara ulaşmaya çalışarak tasarruf sağlamaktadır. Aşağıdaki örnekte adayın gelir düzeyi, mesleği, yaşı, çocuk sayısı, kullandığı mevcut aracın modeli, sınıfı, yaşı, gibi faktörler göz önüne alınarak bir model tasarlanmıştır.

**Çizelge 2.3.** Sınıflama Model Gösterimi [48]

Durum	Girdi Faktörleri							Sonuç
	Mevcut Aracın Markası	Mevcut Aracın Sınıfı	Mevcut Aracın Yaşı	Çocuk Sayısı	Gelir Düzeyi	Yaşı	Mesleği	
Aday 1	Ford	B	6	2	40.000	60	Emekli	Hayır
Aday 2	Renault	B	2	1	120.000	40	Serbest	Hayır
Aday 3		A	5	0	60.000	35	Muhasebe Uz.	Evet

Detaylı açıklayacak olursak verilerin sınıflandırılma süreci iki adımdan oluşur:

- Veri kümelerine uygun bir model ortaya konur. Söz konusu model veri tabanındaki alan isimleri kullanılarak gerçekleştirilir. Sınıflandırma modelinin elde edilmesi için veritabanından bir kısım eğitim verileri olarak kullanılır. Bu veriler veritabanından rastgele seçilir.

#### **Eğitim Verisi**

Müşteri	Borç	Gelir	Risk
Ali	Yüksek	Yüksek	Kötü
Ayşe	Yüksek	Yüksek	Kötü
Fatma	Yüksek	Düşük	Kötü
Fuat	Düşük	Yüksek	İyi
Ece	Düşük	Düşük	Kötü
Ayla	Düşük	Yüksek	İyi



**Sınıflandırma algoritması**

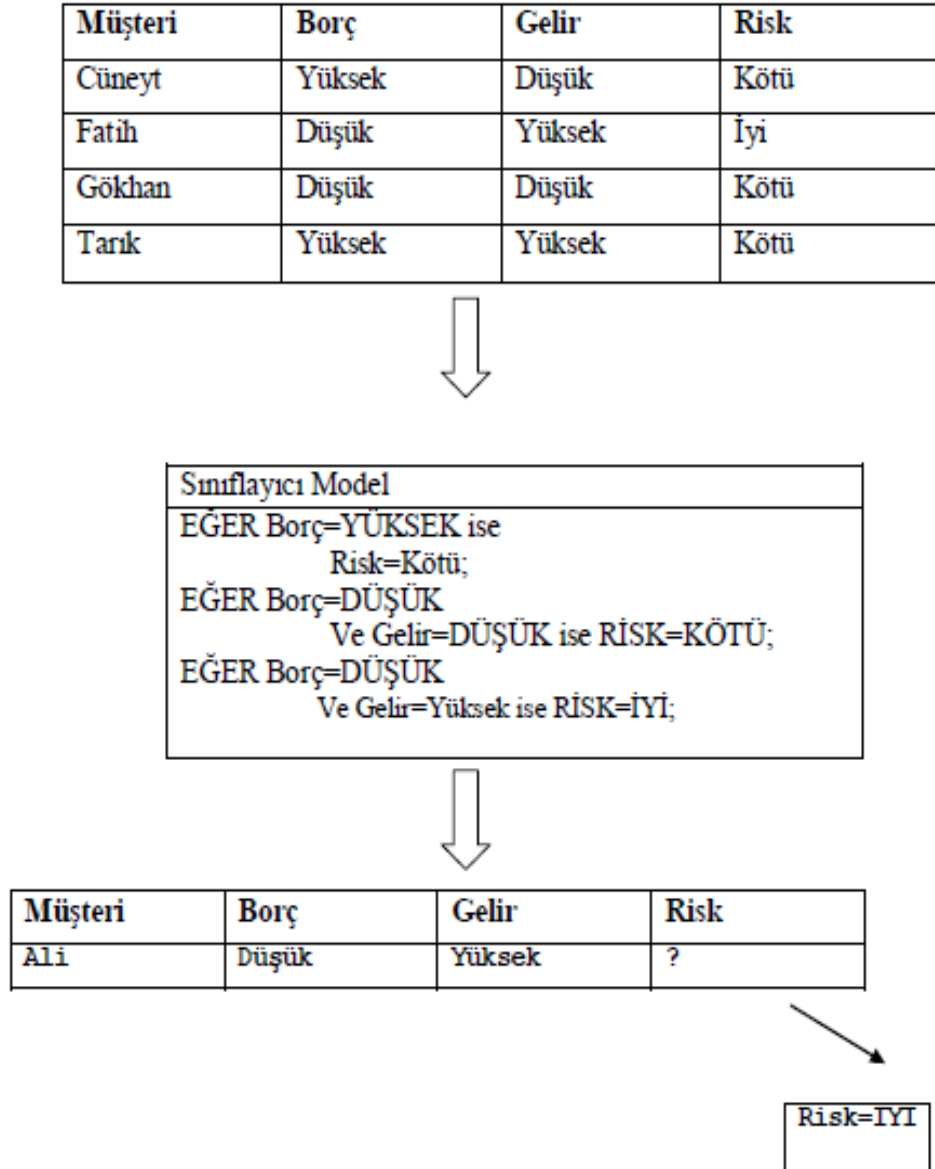


#### **Sınıflayıcı Model**

EĞER Borç=YÜKSEK ise  
Risk=Kötü;  
EĞER Borç=DÜŞÜK  
Ve Gelir=DÜŞÜK ise RİSK=KÖTÜ;  
EĞER Borç=DÜŞÜK  
Ve Gelir=Yüksek ise RİSK=İYİ;

**Şekil 2.3.** Sınıflandırma Model Kurma Süreci

- ii. Test verileri üzerinde sınıflandırma kuralları belirlenir. Ardından söz konusu kurallar bu kez test verilerine dayanarak sınanır. Örneğin Ali adlı yeni bir banka müşterisinin kredi talebinde bulunduğunu varsayalım. Bu müşterinin risk durumunu belirlemek için örnek verilerden elde edilen karar kuralı doğrudan uygulanır. Bu müşteri için Borç=Düşük, Gelir=Yüksek olduğu biliniyorsa risk durumunun Risk=İYİ olduğu hemen anlaşılır.



Şekil 2.4. Sınıflandırma Test Verileri

Yukarıdaki test sonucunda elde edilen modelin doğru olduğu kabul edilecek olursa, bu model diğer veriler üzerinde de uygulanır. Elde edilen sonuç model mevcut ya da olası müşterilerin gelecekteki kredi talep risklerini belirlemede kullanılır.

Regresyon süreklilik gösteren değerleri tahmin etmek için kullanılan fonksiyonlardır. Regresyon ile amaç girdiler ile çıktıyı ilişkilendirecek modeli oluşturup, en iyi tahmine ulaşmaktır. Sonuç ‘bağımlı değişken’, girdiler ‘bağımsız değişken’ olarak adlandırılır. Sonucun alacağı değer genellikle bir güvenlik aralığı içinde belirtilir. Girdiler, çözülecek probleme göre bir veya birden fazla olabilir. Gerçek hayatta çözülecek problemlerin hemen hepsinde doğru tahmine ulaşmak için birden fazla girdiden faydalanmak gereklidir. Bu noktada önemli olan konu girdilerin sonucun doğru tahmin edilmesine yaptıkları katkıdır. Bazı durumlarda sonuca katkısı limitli olan girdileri modelden çıkarmak, daha etkin bir model oluşturmak için önemli bir gerekliliktir.

Örneğin Bir dergiye ilk kez reklam vermeye başlayacak olan bir şirket daha önce reklam vermiş olduğu dergilerin sayfa maliyetlerini kullanarak, çalışılmaya başlanılacak olan derginin vermiş olduğu fiyatın uygunluk seviyesini belirli bir güven aralığı içinde değerlendirebilir. Ya da daha sonra yapacağı kampanyalarda çalışmakta olduğu dergilerin verecekleri fiyatların ne kadar makul olduğunu önceden öngörebilir. Aşağıdaki örnekte derginin okuyucu sayısı, bayan okuyucuların payı, okuyucuların ortalama yıllık kazancı, gibi faktörler göz önüne alınarak bir model tasarlanmıştır.

**Çizelge 2.4.** Regresyon Model Gösterimi [48]

Durum	Girdi Faktörleri			Sonuç
	Okuyucu Sayısı	Bayan Okur Sayısı	Ortalama Yıllık Gelir (YTL)	Sayfa Maliyeti (YTL)
Cosmopolitan	24.000	70%	100.000	10.000
Capital	20.000	30%	50.000	20.000
Esquire	9.000	5%	45.000	5.000

Mevcut verilerden hareket ederek geleceğin tahmin edilmesinde faydalanılan ve veri madenciliği teknikleri içerisinde en yaygın kullanıma sahip olan sınıflama ve regresyon modelleri arasındaki temel fark, tahmin edilen bağımlı değişkenin kategorik veya süreklilik gösteren bir değere sahip olmasıdır. Ancak çok terimli lojistik regresyon gibi kategorik değerlerin de tahmin edilmesine olanak sağlayan her iki model giderek birbirine yaklaşmakta ve bunun bir sonucu olarak aynı tekniklerden yararlanılması mümkün olmaktadır.

Sınıflama ve regresyon modellerinde kullanılan başlıca teknikler şunlardır [18]:

- Karar Ağaçları (Decision Trees)
- Yapay Sinir Ağları (Artificial Neural Networks)
- Genetik Algoritmalar (Genetic Algorithms)
- K-En Yakın Komsu (K-Nearest Neighbor)
- Bellek Temelli Nedenleme (Memory Based Reasoning)
- Saf Bayes

a. Karar Ağaçları

Karar ağaçları, veri madenciliğinde kuruluşlarının ucuz olması, yorumlanmalarının kolay olması, veri tabanı sistemleri ile kolayca entegre edilebilmeleri ve güvenilirliklerinin iyi olması nedenleri ile sınıflama modelleri arasında en yaygın kullanıma sahip tekniktir. Karar ağacı, adından da anlaşılacağı gibi bir ağaç görünümünde, tahmin edici bir tekniktir [49].

Karar ağacı karar düğümleri, dallar ve yapraklardan oluşur [22]. Karar düğümü, gerçekleştirilecek testi belirtir. Bu testin sonucu ağacın veri kaybetmeden dallara ayrılmasına neden olur. Her düğümde test ve dallara ayrılma işlemleri ardışık olarak gerçekleşir ve bu ayrılma işlemi üst seviyedeki ayrımlara bağımlıdır. Ağacın her bir dalı sınıflama işlemini tamamlamaya adaydır. Eğer bir dalın ucunda sınıflama işlemi gerçekleşemiyorsa, o dalın sonucunda bir karar düğümü oluşur. Ancak dalın sonunda belirli bir sınıf oluşuyorsa, o dalın sonunda yaprak

vardır. Bu yaprak, veri üzerinde belirlenmek istenen sınıflardan biridir. Karar ağacı işlemi kök düğümünden başlar ve yukarıdan aşağıya doğru yaprağa ulaşana kadar ardışık düğümleri takip ederek gerçekleşir.

Karar ağacı tekniğini kullanarak verinin sınıflanması iki basamaklı bir işlemdir [22]. İlk basamak öğrenme basamağıdır. Öğrenme basamağında önceden bilinen bir eğitim verisi, model oluşturmak amacıyla sınıflama algoritması tarafından analiz edilir. Öğrenilen model, sınıflama kuralları veya karar ağacı olarak gösterilir. İkinci basamak ise sınıflama basamağıdır. Sınıflama basamağında test verisi, sınıflama kurallarının veya karar ağacının doğruluğunu belirlemek amacıyla kullanılır. Eğer doğruluk kabul edilebilir oranda ise, kurallar yeni verilerin sınıflanması amacıyla kullanılır.

Örneğin, bir eğitim verisi incelenerek kredi duruma sınıfını tahmin edecek bir model oluşturuluyor. Bu modeli oluşturan bir sınıflama kuralı

IF yas = "41...50" AND gelir = yüksek THEN kredidurumu = mükemmel

şeklindedir. Bu kural gereğince yası "41...50" kategorisinde olan (yası 41 ile 50 arasında olan) ve gelir düzeyi yüksek bir kişinin kredi durumunun mükemmel olduğu görülür.

Oluşturulan bu modelin doğruluğu, bir test verisi aracılığı ile onaylandıktan sonra model, sınıfı belli olmayan yeni bir veriye uygulanabilir ve sınıflama kuralı gereği yeni verinin sınıfı "mükemmel" olarak belirlenebilir.

Tekrarlamak gerekirse bir karar ağacı, bir alandaki testi belirten karar düğümlerinden, testteki değerleri belirten dallardan ve sınıfı belirten yapraklardan oluşan akış diyagramı şeklindeki ağaç yapısıdır. Her bir nitelik bir düğüm tarafından temsil edilir. Dallar ve yapraklar ağaç yapısının elemanlarıdır. En son yapı "yaprak", en üst yapı "kök" ve bunların arasında kalan yapı ise "dal" olarak adlandırılır [50]. Karar ağaçları sınıflama algoritmasını uygulayabilmek için uygun bir alt yapı sağlamaktadır. Örneğin;

**Çizelge 2.5.** Karar Ağacı Oluşturulacak Örnek Tablo

<b>Borç</b>	<b>Gelir</b>	<b>Statü</b>	<b>Risk</b>
Yüksek	Yüksek	İşveren	Kötü
Yüksek	Yüksek	Ücretli	Kötü
Yüksek	Düşük	Ücretli	Kötü
Düşük	Düşük	Ücretli	İyi
Düşük	Düşük	İşveren	Kötü
Düşük	Yüksek	İşveren	İyi
Düşük	Yüksek	Ücretli	İyi
Düşük	Düşük	Ücretli	İyi
Düşük	Düşük	İşveren	Kötü
Düşük	Yüksek	İşveren	İyi

Tablodan yararlanılarak karar ağacı oluşturulur. Karar ağacı oluşturulduktan sonra karar kuralları oluşturulur.

Kurallar:

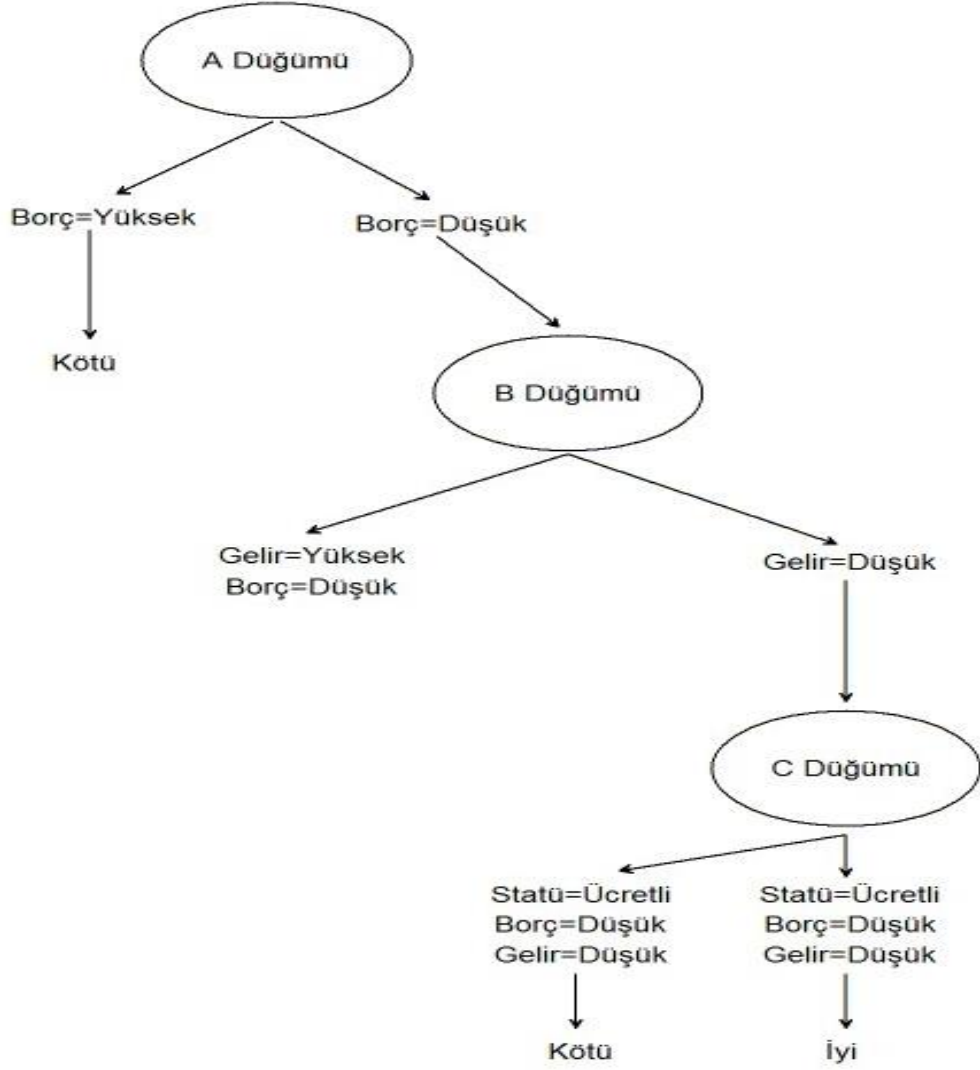
Kural.1: Borç Yüksek ise Risk Kötü

Kural.2: Borç Düşük ve Gelir=Yüksek ise Risk=İyi

Kural.3: Borç Düşük ve Gelir=Düşük ve Statü=İşveren ise Risk=Kötü

Kural.4: Borç Düşük ve Gelir=Düşük ve Statü=İşveren ise Risk=Kötü





**Şekil 2.5.** Karar Ağacı Yapısı

b. Yapay Sinir Ağları

Bu yöntem, belirli bir profile uyması için kalıp düzenlerini kontrol etmektedir ve bu süreç içerisinde belli bir öğrenme faaliyeti gerçekleştirerek sistemi geliştirmektedir. Yapay sinir ağlarında kullanılan öğrenme algoritmaları, veriden üniteler arasındaki bağlantı ağırlıklarını hesaplar. Yapay Sinir Ağları istatistiksel yöntemler gibi veri hakkında parametrik bir model varsaymaz yani uygulama alanı daha geniştir ve bellek tabanlı yöntemler kadar yüksek işlem ve bellek gerektirmez [51].

Yapay Sinir Ağı (YSA), insan beyninin sinir sistemi ve çalışma ilkelerinden ilham alınarak geliştirilmiş, ağırlıklı bağlantılar denilen tek yönlü iletişim kanalları vasıtası ile birbirleriyle haberleşen, her biri kendi hafızasına sahip birçok işlem elemanından oluşan paralel ve dağıtık bilgi işleme yapılarıdır. Bir anlamda insan beyninin ufak bir kopyası gibidir. İnsan beyninin öğrenme yoluyla yeni bilgiler üretebilme, keşfedebilme, düşünme ve gözlemlemeye yönelik yeteneklerini, yardım almadan yapabilen sistemler geliştirmek için tasarlanmışlardır.

Bir yapay sinir ağı belirli bir amaç için oluşturulur ve insanlar gibi örnekler sayesinde öğrenir. Yapay sinir ağları tekrarlanan girdiler sayesinde kendi yapısını ve ağırlığını değiştirir. Yapay sinir ağları aynen canlıların sinir sistemi gibi adapte olabilen bir yapıya sahiptir.

YSA'lar gerçek dünyaya ait ilişkileri tanıyabilir, sınıflandırma, kestirim ve işlev uydurma gibi görevleri yerine getirebilirler. Desen tanıma tekniğinin gerekliliği, gerçek dünya ile bilgisayar ilişkisinin başlaması ile ortaya çıkmıştır. Bu durum YSA'nın çok güçlü örnek tanıma tekniği olarak ortaya çıkmasına ve gelişmesine neden olmuştur [52].

#### c. Genetik Algoritmalar

Diğer veri madenciliği algoritmalarını geliştirmek için kullanılan optimizasyon teknikleridir. Sonuç model veriye uygulanarak gizli kalmış kalıpları ortaya çıkarılmakta ve bu sayede tahminler yapılabilmektedir. Doğrudan postalamaya, risk analizi ve perakende analizlerinde kullanılabilir. Genetik algoritmalar yapay zeka tekniklerinden doğal olayları modelleyen bir optimizasyon tekniğidir.

Genetik algoritmalar çözüme ulaşmak için rassal arama tekniklerini kullanır ve değişken kodlama esasına dayalı bir tekniktir. Genetik algoritmalar, farklı problemlere göre doğru parametreler ile çalışıldığı takdirde optimuma yakın çözümler verir [53].

#### d. k-En Yakın Komşu

En yakın komşu sınıflandırıcıları benzerlik yöntemi ile öğrenmeyi esas alır. Eğitim örnekleri n-boyutlu sayısal nitelik ile tanımlanırlar. Her bir örnek n-boyutlu uzayda bir noktayı temsil eder. Bu şekilde tüm eğitim örnekleri n-boyutlu uzayda depolanır. Bilinmeyen bir örnek geldiğinde, bir k-en yakın komşu sınıflandırıcısı bilinmeyen örneğe en yakın k eğitim örneğini bulmak için örüntü uzayını tarar. K eğitim örnekleri bilinmeyen örneğin k-en yakın komşularıdır.

Bilinmeyen örnek, örüntü uzayında kendisine en yakın eğitim örnekleri kümesine atanır. En yakın komşu sınıflandırıcıları tüm eğitim örneklerini depoladıkları için örnek tabanlıdır. Sınıflandırılmamış bir örnek karşılaştırılmak istendiğinde eğer olası komşularının sayısı fazlaysa hesaplama zamanı oldukça yüksektir. Bu durumda indeksleme tekniklerinin kullanılması gerekebilir. Karar ağacındaki tümevarım ve tümdengelim sürecinde uygulananın aksine, en yakın komşu sınıflandırıcıları her bir niteliğe eşit ağırlık verirler. Bu durum, veride çok fazla ilgisiz nitelik bulunduğunda karışıklığa sebep olabilir [22].

#### e. Bellek Temelli Nedenleme

Bellek tabanlı veya örnek tabanlı bu yöntemler istatistikte 1950'li yıllarda önerilmiş olmasına rağmen o yıllarda gerektirdiği hesaplama ve bellek yüzünden kullanılamamış ama günümüzde bilgisayarların ucuzlaması ve kapasitelerinin artmasıyla, özellikle de çok işlemcili sistemlerin yaygınlaşmasıyla, kullanılabilir olmuştur.

### 2.3.8.2. Kümeleme

Kümeleme, veriyi sınıflara veya kümelere ayırma işlemidir [54]. Aynı kümedeki elemanlar birbirleriyle benzerlik gösterirlerken, başka kümelerin elemanlarından farklıdırlar. Kümeleme veri madenciliği, istatistik, biyoloji ve makine öğrenimi gibi pek çok alanda kullanılır. Kümeleme modelinde, sınıflama modelinde olan veri

sınıfları yoktur [55]. Verilerin herhangi bir sınıfı bulunmamaktadır. Sınıflama modelinde, verilerin sınıfları bilinmekte ve yeni bir veri geldiğinde bu verinin hangi sınıftan olabileceği tahmin edilmektedir. Oysa kümeleme modelinde, sınıfları bulunmayan veriler gruplar halinde kümelere ayrılırlar. Bazı uygulamalarda kümeleme modeli, sınıflama modelinin bir önişlemi gibi görev alabilmektedir [55].

Marketlerde farklı müşteri gruplarının keşfedilmesi ve bu grupların alışveriş örüntülerinin ortaya konması, biyolojide bitki ve hayvan sınıflandırmaları ve işlevlerine göre benzer genlerin sınıflandırılması, şehir planlanmasında evlerin tiplerine, değerlerine ve coğrafik konumlarına göre gruplara ayılması gibi uygulamalar tipik kümeleme uygulamalarıdır. Kümeleme aynı zamanda Web üzerinde bilgi keşfi için dokümanların sınıflanması amacıyla da kullanılabilir [56].

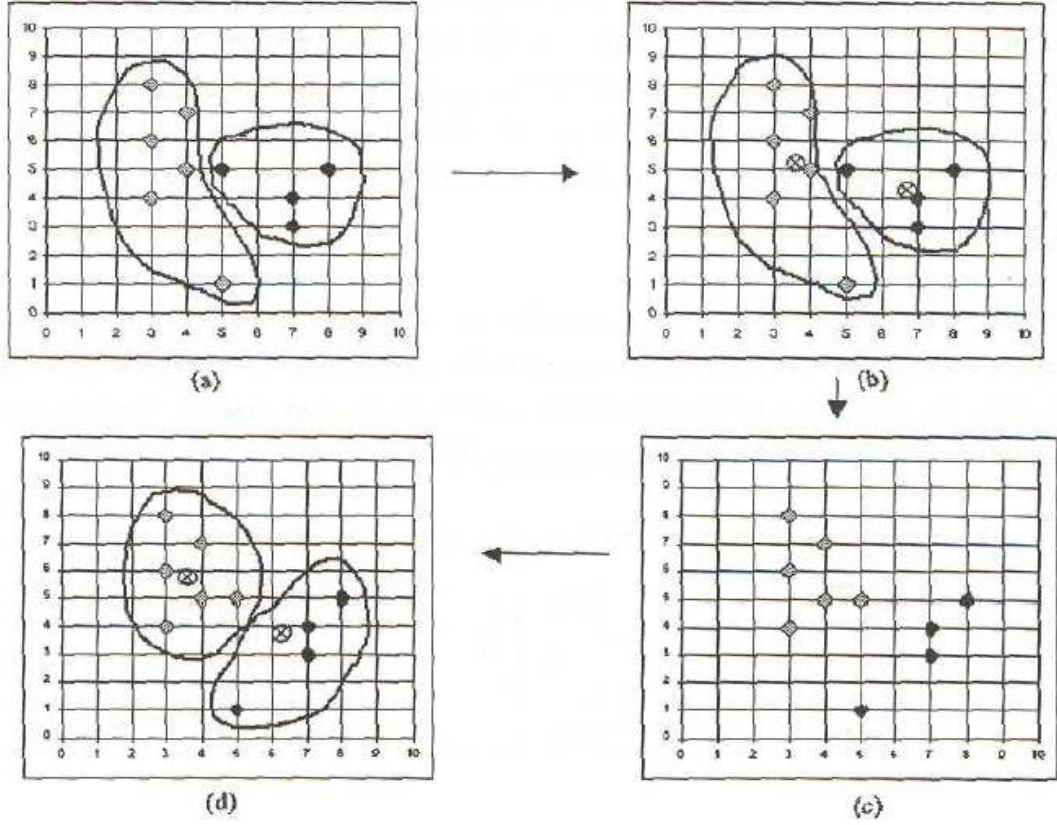
Veri kümeleme güçlü bir gelişme göstermektedir. Veri tabanlarında toplanan veri miktarının artmasıyla orantılı olarak, kümeleme analizi son zamanlarda veri madenciliği araştırmalarında aktif bir konu haline gelmiştir. Literatürde pek çok kümeleme algoritması bulunmaktadır. Kullanılacak olan kümeleme algoritmasının seçimi, veri tipine ve amaca bağlıdır. Genel olarak başlıca kümeleme yöntemleri şu şekilde sınıflandırılabilir [22]:

- Bölme yöntemleri
- Hiyerarşik yöntemler
- Yoğunluk tabanlı yöntemler
- Izgara tabanlı yöntemler
- Model tabanlı yöntemler

Bölme yöntemlerinde,  $n$  veri tabanındaki nesne sayısı ve  $k$  oluşturulacak küme sayısı olarak kabul edilir. Bölme algoritması  $n$  adet nesneyi,  $k$  adet kümeye böler. Kümeler tarafsız bölme kriteri olarak nitelendirilen kritere uygun oluşturulduğu için aynı kümedeki nesnelere birbirlerine benzerken, farklı kümedeki nesnelere farklıdır [22]. En iyi bilinen ve en çok kullanılan bölme yöntemleri  $k$ -means yöntemi,  $k$ -medoids yöntemi ve bunların varyasyonlarıdır [8].

k-means yöntemi, ilk önce  $n$  adet nesneden rasgele  $k$  adet nesne seçer ve bu nesnelerin her biri, bir kümenin merkezini veya orta noktasını temsil eder. Geriye kalan nesnelere her biri kendisine en yakın olan küme merkezine göre kümelere dağılırlar. Yani bir nesne hangi kümenin merkezine daha yakın ise o kümeye yerleşir. Ardından her küme için ortalama hesaplanır ve hesaplanan bu değer o kümenin yeni merkezi olur. Bu işlem tüm nesnelere kümelere yerleşinceye kadar devam eder [22].

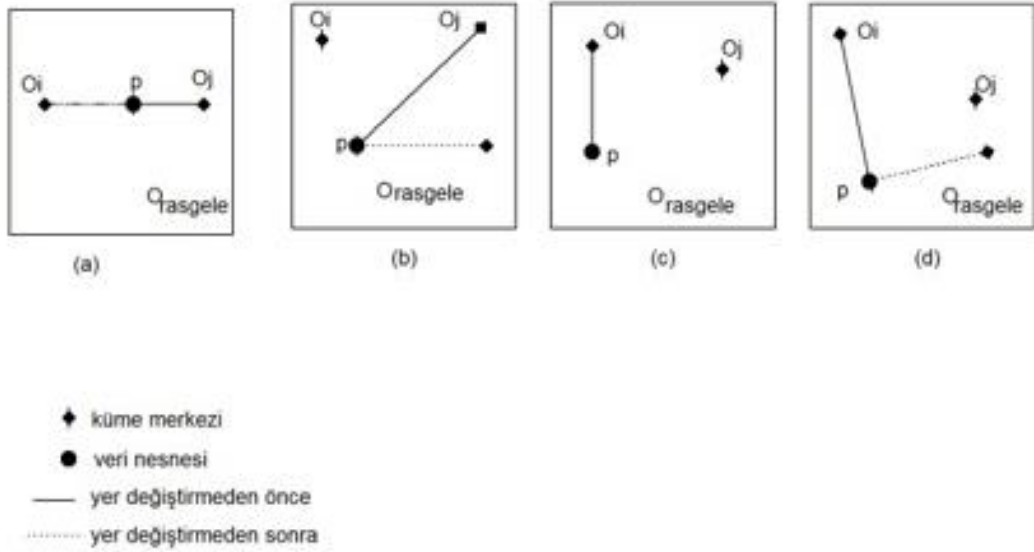
Bir nesne grubunun, Şekil 2.6.'da görüldüğü gibi uzayda konumlanmış olduğu varsayalım. Kullanıcının bu nesnelere iki kümeye ayırmak istediği varsayılırsa,  $k=2$  olur. İlk önce rasgele iki nesne, iki kümenin merkezi olarak seçilmiş ve diğer nesnelere de bu merkezlere olan yakınlıklarına göre iki kümeye ayrılmışlardır. Bu ayırma göre her iki kümenin nesnelere yeni ortalaması alınmış ve bu değer kümelerin yeni merkezleri olmuştur. Bu yeni merkezler Şekil 2.6.(b)'de üstünde çarpı işareti bulunan noktalarla gösterilmektedir. Bu yeni çarpı işaretli merkezlere göre, her iki kümede de birer nesne diğer kümenin merkezine daha yakın duruma gelmişlerdir. Bu durum Şekil 2.6.(c)'de görülmektedir. (5,1) koordinatındaki nesne ile (5,5) koordinatındaki nesne küme değiştirmişlerdir. Her iki kümedeki bu yeni katılımlar ile kümelerdeki nesnelere ortalama değerleri ve dolayısıyla merkezleri değişmiştir. Yeni hesaplanan merkezler Şekil 2.6.(d)'de üstünde çarpı işareti bulunan noktalarla gösterilmektedir. Artık açıkta bir nesne kalmadığı ve her nesne içinde bulunduğu kümenin merkezine en yakın durumda bulunduğu için k-means yöntemi ile kümelere bölünme işlemi Şekil 2.6.(d)'de görüldüğü gibi sonlanmıştır.



**Şekil 2.6.** k-means Yöntemiyle Kümeleme Örneği [57]

k-means yöntemi, sadece kümenin ortalaması tanımlanabildiği durumlarda kullanılabilir. Kullanıcıların k değerini, yani oluşacak küme sayısını belirtme gerekliliği bir dezavantaj olarak görülebilir. Esas önemli olan dezavantaj ise dışarıda kalanlar olarak adlandırılan nesnelere karşı olan duyarlılıktır. Değeri çok büyük olan bir nesne, dahil olacağı kümenin ortalamasını ve merkez noktasını büyük bir derecede değiştirebilir. Bu değişiklik kümenin hassasiyetini bozabilir. Bu sorunu gidermek için kümedeki nesnelere ortalamasını almak yerine, kümede ortaya en yakın noktada konumlanmış olan nesne anlamındaki medoid kullanılabilir. Bu işlem k-medoids yöntemi ile gerçekleştirilir. k-medoids kümeleme yönteminin temel stratejisi ilk olarak n adet nesnede, merkezi temsili bir medoid olan k adet küme bulmaktır. Geriye kalan nesnelere, kendilerine en yakın olan medoide göre k adet küme yerleşirler. Bu bölünmelerin ardından kümenin ortasına en yakın olan nesneyi bulmak için medoid, medoid olmayan her nesne ile yer değiştirir. Bu işlem

en verimli medoid bulunana kadar devam eder. Şekil 2.7.'de k-medoids yöntemiyle kümeleme örneği verilmiştir.



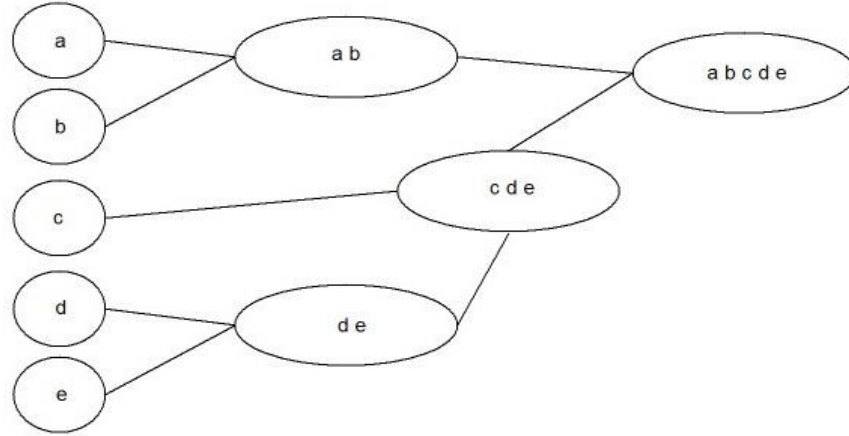
**Şekil 2.7.** k-medoids Yöntemiyle Kümeleme Örneği [3]

Kümeleme yöntemlerinden biri olan hiyerarşik yöntemler, veri nesnelerini kümeler ağacı şeklinde gruplara ayırma esasına dayanır. Hiyerarşik kümeleme yöntemleri, hiyerarşik ayrışmanın aşağıdan yukarıya veya yukarıdan aşağıya doğru olmasına göre agglomerative (kümelenmiş) ve divisive (bölünmeye neden olan) hiyerarşik kümeleme olarak sınıflandırılabilir. Kümelenmiş hiyerarşik kümelemede, Şekil 2.8.'de görüldüğü üzere hiyerarşik ayrışma aşağıdan yukarıya doğru olur. İlk olarak her nesne kendi kümesini oluşturur ve ardından bu atomik kümeler birleşerek, tüm nesnelere bir kümede toplanıncaya dek daha büyük kümeler oluştururlar.

Bölünmeye neden olan hiyerarşik kümelemede, Şekil 2.8.'de görüldüğü üzere hiyerarşik ayrışma yukarıdan aşağıya doğru olur. İlk olarak tüm nesnelere bir kümededir ve her nesne tek başına bir küme oluşturana dek, kümeler daha küçük parçalara bölünürler.

Basamaklar:

Agglomerative  
(AGNES)



Basamaklar:

Divisive  
(DIANA)

**Şekil 2.8.** Hiyerarşik Kümeleme Örneği [57]

### 2.3.8.3. Birliktelik Kuralları

Birliktelik kurallarının amacı, büyük veri kümeleri arasından birliktelik ilişkilerini bulmaktır. Depolanan verilerin sürekli artması nedeniyle şirketler, veritabanlarındaki birliktelik kurallarını ortaya çıkarmak isterler. Büyük miktarda depolanan verilerden değişik birliktelik ilişkileri bulmak, şirketlerin karar alma süreçlerini olumlu etkilemektedir.

Birliktelik kuralları için verilebilecek örnek market sepeti uygulamasıdır. Bu kural, müşterilerin satın alma alışkanlıklarını analiz etmek için, müşterilerin satın aldıkları ürünler arasındaki ilişkileri bulur. Bu tür ilişkilerin analizi sonucunda, müşterilerin satın alma davranışları öğrenilebilir ve yöneticiler de öğrenilen bilgiler sonucunda daha etkili satış stratejileri geliştirebilirler. Örneğin bir müşterinin süt ile birlikte ekmek satın alma olasılığı nedir? Elde edilen müşteri bilgilerine dayanarak rafları düzenleyen market yöneticileri ürünlerindeki satış oranlarını arttırabilirler. Örneğin bir marketin müşterilerinin sütün yanında ekmek satın alma oranı yüksekse, market yöneticileri süt ile ekmek raflarını yan yana koyarak ekmek satışlarını arttırabilirler [57].



Apriori algoritması birliktelik kurallarının üretilmesi için kullanılan en yaygın yöntemdir. Aşamaları:

- ✓ Destek ve güven ölçütlerini karşılaştırmak üzere eşik değerler belirlenir.
- ✓ Destek sayıları hesaplanır. Bu destek sayıları eşik destek sayısı ile karşılaştırılır. Eşik destek sayısından küçük değerlere sahip satırlar çözümlenmeden çıkarılır ve koşula uygun kayıtlar göz önüne alınır.
- ✓ Bu seçilen ürünler bu kez ikişerli gruplandırılarak bu grupların tekrar sayıları elde edilir. Bu sayılar eşik destek sayıları ile karşılaştırılır. Eşik değerden küçük değerlere sahip satırlar çözümlenmeden çıkarılır.
- ✓ Bu kez üçerli, dörderli vb. gruplandırmalar yapılarak bu grupların destek sayıları elde edilir ve eşik değeri ile karşılaştırılır, eşik değere uygun olduğu sürece işlemlere devam edilir.
- ✓ Ürün grubu belirlendikten sonra kural destek ölçütüne bakılarak birliktelik kuralları türetilir ve bu kuralların her birisiyle ilgili olarak güven ölçütleri hesaplanır.

**Çizelge 2.6.** Apriori Algoritması Müşteri Alışverişleri Tablosu

Müşteri	Aldığı Ürünler
1	Makarna, yağ, meyve suyu, peynir
2	Makarna, ketçap
3	Ketçap, yağ, meyve suyu, bira
4	Makarna, ketçap, yağ, meyve suyu
5	Makarna, ketçap, yağ, bira

Apriori algoritması uygulandığında şu sonuçlar elde edilir:

- {ketçap, meyve suyu}-{yağ} (s=0,4 c=1.0)
- {ketçap, yağ}-{meyve suyu} (s=0,4 c=0.67)
- {yağ, meyve suyu}-{ketçap} (s=0,4 c=0.67)
- {meyve suyu}-{ketçap, yağ} (s=0,4 c=0.67)

- {yağ}- {ketçap, meyve suyu} (s=0,4 c=0.5)
- {ketçap}- {yağ, meyve suyu} (s=0,4 c=0.5)

#### 2.4. Öğrenci Performansını Etkileyen Faktörler

Performans genel anlamda amaçlı ve planlanmış bir etkinlik sonucunda elde edileni, nicel ya da nitel olarak belirleyen bir kavramdır. Bu sonuç mutlak ya da görel olarak açıklanabilir. Bir iş sisteminin performansı, belirli bir zaman sonucundaki çıktısı ya da çalışma sonucudur. Bu sonuç işletme amacının ya da görevinin yerine getirilme derecesi olarak algılanmalıdır [58].

Performans kavramının tanımlanmasındaki değişiklikler, temelde bu kavrama hangi yaklaşımla bakıldığından kaynaklanmaktadır. Performansı irdelenecek olan örgütsel sistemin veya bireyin hangi işlevleriyle ilgilenildiği, değişik performans tanımlarını beraberinde getirmiştir. Farklı yaklaşımlara göre performans tanımları aşağıdaki tabloda gösterilmektedir [59].

**Çizelge 2.7. Yaklaşımlara Göre Performans Tanımları [60]**

<b>Yaklaşım</b>	<b>Performans Tanımı</b>
Amaç yaklaşımı	Bir organizasyon, ifade ettiği amaçlara ulaştığı derecede başarılıdır.
Sistem kaynakları yaklaşımı	Bir organizasyon, gereksiniz duyduğu kaynakları elde ettiği sürece başarılıdır.
İç süreç yaklaşımı	Bir organizasyon, iç bileşenleri arasında uyumluluk gösterdiği sürece başarılıdır.
Yüksek performanslı sistem yaklaşımı	Bir organizasyon, benzerlerine göreceli olarak üstün olduğu derecede başarılıdır.
Hata yaklaşımı	Bir organizasyon, hata yapmadığı ölçüde başarılıdır.
Rasyonel Sistem Yaklaşımı	Bir organizasyon, belirli bir zaman periyodunda yaptığı üretim miktarı kadar başarılıdır.

Bilgi ve teknoloji alanındaki hızlı gelişmeler, toplum yaşamını da büyük ölçüde değiştirmiştir. Bireyin değişen bu toplumsal yaşama uyum sağlayabilmesinde örgütlü bir sistem olarak okulun önemi tartışılmaz bir biçimde artmaktadır. Okul, eğitim amacıyla kurulmuş özel bir ortamdır. Kontrollü bir ortam olan okulda, öğrenciye kazandırılacak bilgi, beceri ve tutumlar önceden belirlenmiştir. Bunlar, bu konuların uzmanı olan öğretmenler tarafından planlı bir biçimde düzenlenen öğretim etkinlikleri ile öğrencilere kazandırılır. Öğrencilerin bu davranışları kazanma düzeyleri, akademik başarı ile ölçülmektedir [61].

Akademik başarı, öğrencilerin okul yaşamında amaçlanan davranışlara ulaşma düzeyi olarak tanımlanmaktadır. Öğrencilerin sergilemiş oldukları akademik başarı, istedik düzeyde davranışların oluşup oluşmadığının belirlenmesinde bir ölçüt olarak kullanıldığı gibi, üniversiteye giriş, is başvurusu gibi alanlarda da dikkate alınmaktadır. Okulda öğrencilerin akademik başarılarını ölçmek, bu performans sonuçlarına dayanarak öğrenci hakkında karar vermek, öğrenciyi ilgi ve yeteneklerine göre başarılı olacakları alanlara yönleltmek, genel olarak eğitim sisteminin görevidir. Bunun yanında öğrenci akademik başarısını olumsuz etkileyen etmenlerin belirlenmesi de büyük önem taşımaktadır [62].

Öğrenci akademik başarısını etkileyen etmenler genel olarak 6 boyutta ele alınabilir [63]:

- ✓ Yöneticilerden kaynaklanan nedenler,
- ✓ Öğretmenlerden kaynaklanan nedenler,
- ✓ Aileden kaynaklanan nedenler,
- ✓ Çevreden kaynaklanan nedenler,
- ✓ Akademik kaygıdan kaynaklanan nedenler,
- ✓ Bireysel nedenlerdir.

#### **2.4.1. Yöneticilerden Kaynaklı Nedenler**

Okul yöneticilerinin öğrenci üzerine doğrudan etkisi olmasa bile, dolaylı olarak etkisi bulunmaktadır. Okulda örgüt iklimini oluşturan disiplin politikaları, sosyal ve sportif

etkinlikleri örgütleme, giriş-çıkış ve teneffüs saatlerinin belirlenmesi gibi pek çok yönetim etkinlikleri öğrenci performansını etkileyen etmenleri oluşturabilir. Okul yönetiminin demokratik bir tutum sergilememesi istenmeyen davranışları artıracaktır. Okul yönetiminin demokratik bir hoşgörü düzeyi tutturması, öğrenciyi sıkı koşullarla sınırlandırmak yerine, esnek bir düzen kurması, bu düzenin belirlenmesinde öğrenci görüşlerini alması, onların istek ve beklentilerini okul amaçlarına zarar vermeden gerçekleştirmelerine yardım etmesi öğrencinin okula devamsızlık nedenlerini azaltacağından dolayı olarak performansına da etki edecektir [64].

#### **2.4.2. Öğretmenlerden Kaynaklı Nedenler**

Öğretmenlerin sınıf içinde olumlu tutum ve davranışları öğrencilerin okula bağlılığı üzerinde olumlu etki yaparken; öğretmenin sınıf içindeki otoriter tavrı, öğrencilerle olan iletişiminin yetersizliği, öğrenciden yetenekleri üzerinde performans beklemesi gibi etmenler, öğrencinin performansını etkilemektedir [65].

#### **2.4.3. Aileden Kaynaklı Nedenler**

Öğrencinin aile ortamı, onun davranışlarının temel kaynağıdır. Anne babanın çocuk yetiştirme tarzları, eğitim düzeyleri, boşanma gibi pek çok etken öğrenci davranışlarını etkilemektedir. Ailenin çocuğun davranışlarını aşırı derecede kontrol etmeye çalışması ya da aşırı ilgisiz davranması, sorunlarını paylaşmaması öğrencilerin performansına etki etmektedir [63]. Bu başlık altında bireyin performansını etkileyen faktörler şu şekilde açıklanabilir:

- Huzursuz ve kaygı verici bir ev ortamı,
- Anne babanın, kendi hayatlarındaki sıkıntılarından dolayı eleştirel ve sabırsız olması, çocuğun hatalarını tolere edememesi, baskıcı tutumu, çocuğu zorlamaları, çocuğun iyi yanlarından ziyade yetersiz yanlarına yoğunlaşması, çocukta kendine güvensizliğe ve kaygıya bu da başarısızlığa yol açmaktadır.

- Anne babanın çok kaygılı olması çocuğunda kaygılanmasına yol açar ve çocuğun gerçek performansını ortaya koymasını engelleyerek başarısızlığa sebep olur.
- Ebeveynlerin çocuğun potansiyelinin üzerinde olan beklentilerini çocuklara yansıtmaları çocukta kaygı ve başarısız olma korkusu geliştirir.
- Evde çocuğun kendine ait bir çalışma mekânının olmaması,
- Anne babanın zamanı etkili kullanma, okuma, sorumluluklarını yerine getirme konularında olumsuz model olup, bir taraftan çocuğun çok fazla televizyon izlemesine, gezmesine kızarken diğer taraftan zamanlarını hep bu şekilde geçirmeleri.
- Teknolojik olanakların kullanımına sınır getirilmemesi sonucu çocukta bağımlılık yaratması,
- Ödül verme yöntemini doğru kullanamayarak çocuğun, başarıyı başlı başına bir ödül olarak görmesinin engellenmesi,
- Çocuğun ders çalışmaktan başka sorumluluğu yokmuş gibi davranarak aslında gelişimi için gerekli olan etkinliklerin anne baba tarafından gereksiz görülmesi,
- Ailenin eğitim hataları, ana baba tutumundaki kararsızlık, anne babanın eğitim anlayışındaki farklılık başarıyı engelleyici olabilmektedir.
- Aile içi iletişimin gencin başarısını etkileyişine bakıldığında çocukların okul başarılarının huzurlu aile ortamlarında arttığı bulunmuştur.
- Benliğe saygı ile başarı arasındaki ilişki incelendiğinde yeteneği ölçüsünde başarı göstermeyen erkek öğrencilerin başarılı erkek arkadaşlarına kıyasla daha olumsuz benlik tasarımına sahip oldukları bulunmuştur.
- Farklı sosyoekonomik düzeydeki öğrencilerin problem alanları ve başarı düzeyleri arasındaki ilişkilere bakıldığında öğrencilerin derslerden aldıkları notların aritmetik ortalamaları alt sosyo-ekonomik düzeyden üst sosyo-ekonomik düzeye gidildikçe arttığı görülmektedir.
- Kardeş sayısı ile öğrencilerin okul başarısı arasında önemli fark bulunmuştur. Buna göre tek çocuk veya iki kardeş olan öğrencilerin okul başarıları, dört, beş ve daha fazla olan öğrencilerin okul başarılarından yüksek olmaktadır.

- Anne ve babanın eğitim düzeyi yükseldikçe öğrencinin başarı düzeyi yükselmektedir.
- Ebeveynler çocukların yakın çevrelerindeki modeller olarak ilk özdeşim kuracakları kişilerdir. Anne baba ve kardeşlerin eğitim düzeyi ve sosyal statüleri çocukların eğitiminde ve hedef belirlemelerinde önemlidir.
- Anne baba geçimsizliğinin olduğu ailelerde çocuk istismarı ve ihmalinin daha yüksek olduğu ortaya çıkmıştır. Çocuk istismarı ve ihmalinin okul başarısını doğrudan etkilediği ve başarısız öğrenciler arasında istismarın daha fazla olduğu görülmektedir.
- Aile bireylerinden birini hastalanması veya ölmesi, ana babanın ayrılması, yeni kardeşin doğması gibi değişiklikler başarı üzerinde etkili olabilmektedir.
- Alt sosyo-ekonomik aileden gelen çocukların çevrelerinin sınırlı olması başarılarını ve meslek sahibi olma olasılıklarını olumsuz etkilemektedir.
- Eğitimsel süreçlere değer veren, öğretmenin çabasına saygı duyan ebeveynler çocuğun okula karşı tutumlarını olumlu etkilerken, öğretmene saygısı olmayan aileler genellikle olumsuz etkilemektedir. Benzer şekilde eğitimin önemli olduğunu söylemesine rağmen okuma ve öğrenmeye hiçbir kişisel ilgi göstermeyen ebeveyn, çocuğun okula duyduğu ilgiye engel olabilmektedir.
- Aile çocuğun gelişiminde ilk eğitimcisi ve etkili olan ilk çevre olduğu kadar okul başarısında da çok önemlidir. Aile içi ilişkilerin dengeli ve düzenli olması çocuğun başarısını olumlu etkiler.
- Anne babanın okula karşı tutumları çocuğu etkilemektedir.

#### **2.4.4. Çevreden Kaynaklı Nedenler**

Öğrencinin çevresinde yer alan çeşitli arkadaş grupları, öğrenci davranışı üzerinde etkilidir. Öğrenciler, grup dinamiğinin etkisi ile bir grup içerisine girebilmek ya da grup içerisinde kalabilmek için de performansının düşmesine neden olabilirler. Bunun yanında, öğrencinin evinin okula uzak olması gibi fiziki nedenler de öğrenci performansına etki eder [66].

#### **2.4.5. Akademik Kaygıdan Kaynaklı Nedenler**

Öğrencilerin öğrenimleri süresince başarılı olmalarını sağlayan, akademik başarılarını olumlu şekilde etkileyen etmenlerden birisi, öğrencilerin sahip oldukları etkin ve verimli ders çalışma alışkanlıklarıdır. Öğrencilerin düzenli bir ders çalışma alışkanlıkları yoksa performansları bundan olumsuz yönde etkilenmektedir.

#### **2.4.6. Bireysel Nedenler**

Öğrencilerin; okul dışında çalışmak zorunda olması, psikolojik rahatsızlıkları, yeterince güdülenmediği için okulu sevmemesi ve gelecek kaygısı gibi pek çok bireysel neden öğrenci performansını etkilemektedir [63]. Bu başlık altında bireyin performansını etkileyen faktörler şu şekilde açıklanabilir:

- Bilişsel, fiziksel ve duygusal olgunluk açısından yetersizlik,
- Beden imajı düşük olan öğrencilerin özsaygıları ve akademik başarılarının da düşük olması,
- Başarısızlığın devamlılığına yol açmada gelişim görevlerini gerçekleştirememek de etkilidir.
- Ergenlik döneminde yoğunluk kazanan duygusal nedenler, ilgi alanlarının değişmesi ve çeşitlenmesi önemli başarısızlık nedenlerinden biridir.
- Ergenin soyut düşünce yeteneğinin üst düzeye gelmiş olması beklendiğinden müfredat programı da bu doğrultuda hazırlanır. Ancak bilişsel açıdan henüz bu olgunluğa ermemiş öğrenciler başarısız olabilmektedir.
- Yaştlarına oranla fiziksel olarak geç olgunlaşma da gencin kendine güvenini zedelemekte, kişisel ve sosyal uyumunu bozmaktadır. Bunun getirdiği özgüvensizlik başarıyı olumsuz etkilemektedir.
- Kaygının çok yüksek yada çok düşük olması gibi motivasyon eksikliği de başarısızlığa neden olabilmektedir.
- Çocuğun ön bilgilerinin yetersiz olması, diğer bir ifadeyle bulunduğu sınıf düzeyine gelinceye kadar almış olduğu eğitimle oluşturduğu akademik temelin gereken becerileri ortaya koymasına engel olması,

- Anlayamadığı konularda soru sormaktan çekinen, utangaç, kendine güveni düşük ve sınavlarda çok heyecanlandığı için bildiği soruları dahi yapamayan, kaygılı kişilik yapısı,
- Yüksek sınav kaygılı öğrencilerin başarıları düşük sınav kaygılı öğrencilerin başarısından daha düşüktür.
- Geçmişte aynı dersten başarısız olma veya o ders, konuyla tanışık olmama nedeniyle "Nasıl olsa başarısız olacağım" önyargısıyla çocuğun yeterince çalışmaması,
- Görme, işitme kaybı, bulaşıcı hastalıklar gibi nedenler ders çalışmayı engellediği ve dikkat dağılmasına neden olduğu için başarısızlığa zemin oluşturur.
- Öğrencinin kaldığı yerin okul başarısına etkisi incelendiğinde, ailesinin yanında kalan öğrencilerin yatılı okuyan ya da yurttan kalan öğrencilere göre daha başarılı oldukları bulunmuştur. Bununla birlikte ailedeki birey sayısı fazla olan öğrencilerin, birey sayısı az olan öğrencilere oranla başarı seviyeleri daha düşüktür.
- Ders çalışma alışkanlıkları ile akademik başarı karşılaştırıldığında, öğrencilerin ders çalışma alışkanlıklarının akademik başarıyı olumlu etkilediği görülmektedir.
- Çocukta öğrenme güçlüğü olması başarısızlığın çocuktan kaynaklanan en önemli nedenidir.
- Öğrencinin doğuştan getirdiği özelliklerin yanı sıra onun başarısını etkileyen bir faktörde öğrencinin sorumluluk alma gücünü kazanmış olmasıdır. Bunun diğer bir anlamı çalışma alışkanlığının olmasıdır.
- Bunalım ve endişe halleri çocuğun okul başarısını da engelleyici rol oynar.
- Okul çocuk için yeni bir sosyal çevredir. Okulun uyulması gereken kuralları çocuğun diğer çocuklar ve öğretmeni ile tanışıp anlaşması ve başarmak zorunda olduğu öğrenim görevleri vardır. Bütün bunlar çocuğun çevreye uyumunu güçleştirebilir. Okuldaki sosyal çevreye uymakta zorluk çeken çocuklar ise daha çok okul öncesinde aile dışına çok az çıkmış sosyal ilişkilerden mahrum bırakılan çocuklardır.
- Ergenlik döneminin özellikleri de başarıda önemli bir etkidir.



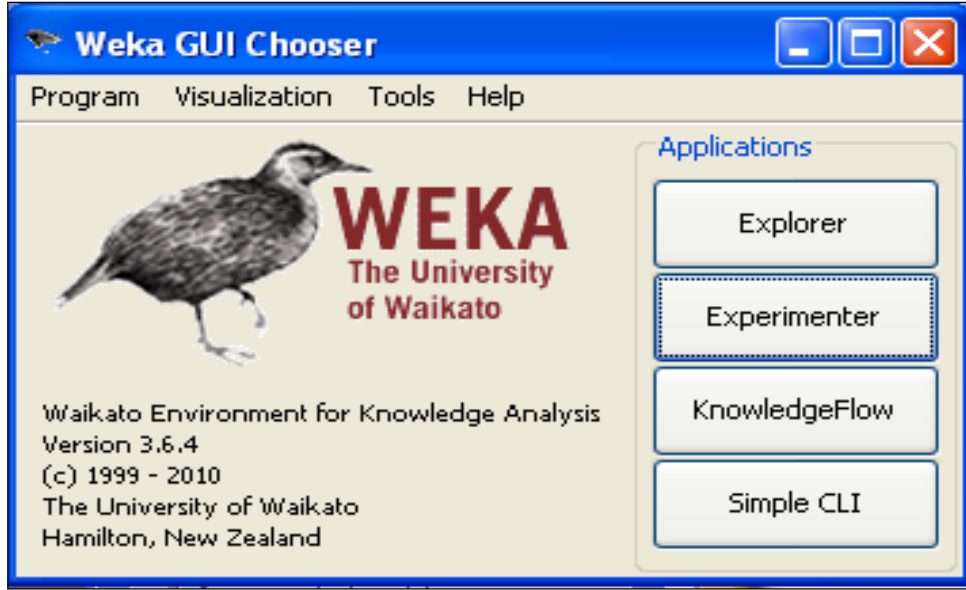
## 2.5. WEKA

Yeni Zelanda Waikato Üniversitesi'nde bir proje olarak başlayan Weka (Waikato Environment for Knowledge Analysis / Waikato Bilgi Analizi İçin Ortam), bugün dünya üzerinde birçok araştırmacı tarafından kullanılmaya başlanan bir Veri Madenciliği programıdır [67].

Alanında uzman biri tarafından, elle analiz edilemeyecek kadar büyük veri tabanlarından faydalı bilgi türetmek için kullanılmaktadır. Ayrıca eğitim amaçlı olarak da yaygın olarak kullanılmaktadır. Şekil 2.9.'da ara yüzü görünen Weka'nın, dünya çapındaki veri madenciliği problemlerini çözmek için geliştirilen tüm algoritmaların bir koleksiyonunu sunması ve kolay ulaşılabilirliği önemli avantajlarıdır [68].

Weka'daki birincil öğrenme metotları, sınıflandırıcılardır ve bu sınıflandırıcılar veriyi modelleyen kural kümeleri veya karar ağaçları üretirler. Weka aynı zamanda, birliktelik kuralları öğrenme ve veriyi kümeleme algoritmalarını da içerir. Tüm uygulamalar tek tip bir komut satır ara yüzüne sahiptir. Veriyi ön işlemek için araçlar veya filtreler, diğer önemli kaynaklardır. Öğrenme planları gibi, filtreler ortak bir komut satır seçimleri kümesiyle, standart ortak bir komut satır ara yüzüne sahiptir. Weka yazılımı, bilgisayar platformu ne olursa olsun veri madenciliği araçlarının kullanılabilirliğini kolaylaştırmak için tamamen Java'da yazılmıştır. Sistem, özette Java paketlerinin bir takımıdır [68].

Weka arayüz görünümü Şekil 2.9.'da gösterilmiştir.



Şekil 2.9. Weka Arayüz Görünümü

## 2.6. Kullanılan Veri Madenciliği Sınıflama Algoritmaları

### 2.6.1. Çok Katmanlı Algılayıcı (Multilayer Perceptron) Algoritması

Bu sınıflandırıcı, örnekleri sınıflandırmak için arka planda yayılma yöntemini kullanır. Bu ağ elle oluşturulabileceği gibi bir algoritmayla ya da her ikisiyle de oluşturulabilir. Eğitim zamanı sürecinde bu ağ görüntülenebilir ve değiştirilebilir. Düğümlerin eşik olmayan lineer birimler haline geldiği ve sınıfın nümerik (sayısal) olduğu durumlar dışında, ağdaki bu düğümler sigmamsı (kırımlı) biçimdedir. Çok Katmanlı Algılayıcı (ÇKA), yapay sinir ağları yapısının bir sınıfıdır. ÇKA, paralel bağlı ağ modeli kurarak insan beyninin öğrenme mekanizmasını taklit eden akıllı kodlar oluşturmak üzerine yoğunlaşır. Bir ÇKA modelinde önce sistem eğitilir ve ağ, en son güncellenen ağ parametreleri kullanılarak fonksiyonel haritalayıcı olarak çıktıları hesaplayabilir [69].

Çok Katmanlı Algılayıcı, standart geri yayılım algoritması ile sürekli geri beslenerek eğitilir. ÇKA denetimli (supervised) ağlardır ki bu yüzden ÇKA istenilen cevabın

eđitilmesi ihtiyaçı duyar. KA giriř verisinin istenilen cevap verisine nasıl dnüştürüleceđini ğrenir ve bu yüzden KA rüntü sınıflamasında yaygınca kullanılır. Bir veya iki gizli katmanla KA giriř ve ıkıř haritasını sanal olarak birbirine yaklařtırır. Birok yapay sinir ađları uygulaması KA ierir.

KA, algılayıcılar olarak adlandırılan basit sinir ađıdır. Algılayıcı, giriř ađlıklarına gre lineer kombinasyon formuna dnüştürlerek oklu gerek deđerli giriřlerden ve sonra da bazı lineer olmayan aktivasyon fonksiyonları iinde yer almıř mmkn ıkıřlardan tek bir ıkıř hesaplar [70].

Genellikle dıř dnya bilgileri analog bilgi olduđundan bu bilgilerin, sayısal hale dnüştürlmesi ve bu sayısal bilginin 0–1 arası skalaya indirgenmesi gerekmektedir. Bu durum eřitlik 2.1’ de gsterilmiřtir.

$$F_k = G_k \quad (2.1)$$

Burada;

$F_k$  : Giriř katmanındaki k. nronun ıkıřını ifade eder.

$G_k$  : Giriř katmanına dıř dnyadan gelen bilgiyi ifade eder.

Ara katmanda ve ıkıř katmanındaki nronların ıkıřının hesaplanabilmesi iin ilgili nrona gelen net girdinin hesaplanması gerekir. Bunun iin gelen bilgi ve ađlıklı arpımı kullanılır. Bu durum eřitlik 2.2’ de gsterilmiřtir.

$$Net_j = \sum_{i=1}^n W_{ij} F_i \quad (2.2)$$

Burada;

$Net_j$  : j. prosesin net girdisini ifade eder.

$F_i$  : j. nrona bilgi gnderen nronların ıkıř bilgisidir.

$W_{ij}$ :  $j$ . nörona bilgi gönderen  $i$ . nöron  $j$ . nöron arası ağırlığı ifade eder.

Ara katman ve çıkış katmanı nöronları için net bilginin sigmoid fonksiyonu kullanılarak nöron çıkışına dönüştürülmesi işlemi yapılır. Bu işlem eşitlik 2.3' de gösterilmiştir.

$$F_j = \frac{1}{1 + e^{(-Net_j + \beta)}} \quad (2.3)$$

Çıkış katmanında elde edilen çıkış bilgisi ile olması gereken çıkış bilgisi arası fark hatayı oluşturur. Hata kavramı eşitlik 2.4' de gösterilmiştir.

$$E_m = B_m - C_m \quad (2.4)$$

Ağın test başarısı aşağıdaki formülle hesaplanır. Ağın test başarısı eşitlik 2.5' de görüldüğü üzere hesaplanmaktadır.

$$P = \frac{D}{T} \times 100 \quad (2.5)$$

### 2.6.2. k- En Yakın Komşu (IBk) Algoritması

k- En Yakın Komşu algoritması basit bir örnek tabanlı öğrenme algoritmasıdır. Test örneklerinden kalan eğitim örneklerini bulmak amacıyla basit bir uzaklık ölçüsü kullanır ve eğitilen örnekle aynı sınıfı tahmin eder. Eğer birden çok örnek test örneğine aynı uzaklıkta ise, ilk bulunan örnek kullanılır [71].

k- En Yakın Komşu, özniteliklerin dağılımını normalize etmesinin dışında tipik bir en yakın komşuluk algoritması olarak ifade edilebilir. k- En Yakın Komşu, örnekleri giderek artan şekilde işler ve boş değerleri toleranse eden bir yol izler [72].

Söz konusu bu yöntem, örnek kümedeki gözlemlerin her birinin, sonradan belirlenen bir gözlem değerine olan uzaklıklarının hesaplanması ve en küçük uzaklığa sahip k sayıda gözlemin seçilmesi esasına dayanmaktadır. Uzaklıkların hesaplanmasında, i ve j noktaları için Öklit Uzaklık Formülü kullanılır.

$$d(i, j) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (2.6)$$

olarak ifade edilir.

k- En Yakın Komşu algoritması, gözlem değerlerinden oluşan bir küme için aşağıdaki adımların uygulanmasından ibarettir.

- K parametresi belirlenir. Bu parametre verilen bir noktaya en yakın komşuların sayısıdır.
- Bu algoritma verilen bir noktaya en yakın komşuları belirleyeceği için, söz konusu nokta ile diğer tüm noktalar arasındaki uzaklıklar tek tek hesaplanır.
- Hesaplanan uzaklıklara göre satırlar sıralanır ve bunlar arasından en küçük olan k tanesi seçilir.
- Seçilen satırların hangi kategoriye ait oldukları belirlenir ve en çok tekrarlanan kategori değeri seçilir.
- Seçilen kategori, tahmin edilmesi beklenen gözlem değerinin kategorisi olarak kabul edilir.

### 2.6.3. J48 Algoritması

Weka'nın sınıflandırma algoritmalarından birisi olan J48 algoritması, temel olarak C4.5 Karar Ağacı algoritmasını kullanır. Yani C4.5 algoritmasının Weka sınıflandırma aracındaki gerçekleşmesi J48 olarak bilinir. C4.5 karar ağacının oluşturulması aşağıdaki gibi özetlenebilir.

- Çıkış değerlerini en fazla farklılaştıran öznelik seçilir.
- Seçilen özneliğin her değeri için farklı bir dal oluşturulur..
- Seçilen düğümdeki öznelik değerlerini yansıtacak şekilde örnekler alt gruplara ayrılır.
- Her alt grup için öznelik seçimi durdurulur; Eğer
  - i. Alt gruptaki tüm üyeler aynı çıkış değerini üretiyorsa, ağacın ilerlemesi durdurulur ve çıkış değeri olarak son belirlenen değeri atanır.
  - ii. Alt grupta tek düğüm kaldıysa veya ayırt edici öznelikler belirlenemiyorsa ağacın ilerlemesi durdurulur.
- 3.aşamada belirlenen her alt grup için yukarıdaki işlem tekrarlanır.

C4.5 algoritması ile sayısal değerler içeren veri tabanları üzerinde karar ağaçlarının oluşturulma olanağı sağlamıştır.

Sayısal nitelikleri belirli aralıklara bölme konusunda bazı zorluklar görülebilir. Ancak en uygun t eşik değerini hesaplamak için çeşitli yöntemler bulunmaktadır. Nitelik değerleri sıralanır ve  $\{V_1, V_2, \dots, V_n\}$  şeklini alır. Nitelik değerler kümesi iki parçaya ayrılır ve Eşik değeri olarak  $[V_i, V_{i+1}]$  Aralığının orta noktası olarak alınabilir [25]:

$$t_i = (V_i + V_{i+1}) / 2 \quad (2.7)$$

C4.5 algoritması sınıflandırmada en ayırıcı özelliğe sahip değişkeni bulurken Entropi kavramından yararlanır. Entropi kavramı eldeki verinin sayısallaştırılmasıdır. Entropi bir veri kümesi içindeki belirsizliği ve rastgeleliği ölçmek için kullanılır.

Veri tabanının tamamının entropisi hesaplanır; eğer veri tabanı farklı bölümlere ayrılıyorsa her bir alt bölümün de entropisinin hesaplanması gerekir. C4.5 algoritması kullanılarak ağaç elde edilirken her bir alt ağaçlar yapraklara dönüştürülür. Ağaç yapısı oluşturmak için, her bir alt ağacın yaprağa dönüşümü kazanım ve ayırma oranları ile gerçekleştirilir [25].

#### 2.6.4. JRIP Algoritması

JRip, Repeated Incremental Pruning to Produce Error Reduction (RIPPER) - (Yinelenen Artımlı Budama İçin Hata Azaltmayı Üretme) algoritmasının WEKA uygulamasıdır [73].

RIPPER Algoritması “if .... then.... Kural tablosunu kullanan doğrudan kural tabanlı bir sınıflandırma tekniğidir. Amacı gürültülü veriler üzerinde etkili kural geliştirmektir ve bu doğrultuda C4.5 algoritması ile rekabet etmektedir.

Öğrenme algoritması IREP’in (Incremental Reduced Error Pruning – Artan maliyetli budama için indirgenmiş hata) gelişmiş bir versiyonu olan RIPPER kural öğrenme algoritması, tüm olumlu örneklerin kapsamakta ve algoritmanın, gürültülü veri setleri üzerine etkili performans gösterdiği bir kural setinden oluşmaktadır.

Bir kural oluşturulmadan önce mevcut eğitim örnekleri seti iki alt sete ayrılır. Bunlardan biri gelişen kurallar listesi (genellikle 2/3) diğeri de budama listesi (genellikle 1/3) dir. Kural gelişen kurallar listesindeki örneklerden oluşturulur. Bütün olumsuz örnekler belirlenene kadar kurallar bu listeye kurallar ilave edilir. Gelişen kurallar listesinde bir kural geliştirildikten sonra kural listesinin performansını arttırmak için budama (kural silme) yapılır. Bir kuralın budamasında RIPPER yalnızca bu kuraldan oluşan son koşulu göz önünde bulundurur [74].

Algoritmanın sonunda eğitim veri setine göre Eğer-O-Zaman kuralları listesi elde edildikten sonra yeni bir örneğin sınıflandırılmasında sırayla kullanılır. Eğer listedeki ilk kural örneği kapsamıyorsa, yani hem kural hem örnekteki nitelikler için eşleşen değerler yoksa o zaman bir sonraki kural denenir. Sırayla örnek bir kural tarafından sınıflandırılana kadar devam eder. Eğer kural hiçbir örneği kapsamıyorsa o zaman karar listesinin en altında varsayılan bir kural işletilir. Yani sınıflandırılmayan tüm örnekler bu sınıfta toplanır [75].

Karar setleri gürültülü eğitim veri setine uyma problemi ile karşı karşıyadır. Bu yüzden genellikle budama işlemi yapılır.

Bu tez çalışmasında JRip algoritması ile yapılan modellemede elde edilen kurallar Şekil 2.10’da görülmektedir.

```

JRIP rules:
=====

(OGRT_ILET = OGRT_ILET_KOTU) and (BABA_EGITIM = B_ILKOKUL) => BASARI_DURUM=BASARISIZ (13.0/3.0)
(OGRT_ILET = OGRT_ILET_KOTU) and (KARDES_SAYI = COK) and (ODA = ODA_YOK) => BASARI_DURUM=BASARISIZ (9.0/2.0)
(ANNE_EGITIM = A_LISE) and (SAGLIKPROB = SAGLIKLI) => BASARI_DURUM=BASARILI (48.0/5.0)
(ANNE_EGITIM = A_ORTAOKUL) and (ODA = ODA_VAR) and (KARDES_SAYI = AZ) => BASARI_DURUM=BASARILI (21.0/0.0)
(ANNE_EGITIM = A_UNI) and (ARK_ILIS = ARK_ILIS_IYI) => BASARI_DURUM=BASARILI (12.0/0.0)
=> BASARI_DURUM=ORTALAMA (107.0/12.0)

Number of Rules : 6

```

**Şekil 2.10.** JRip Algoritma Kuralları

Şekil 2.10.’da yer alan 6 adet JRip kuralının, kodları açık hale getirildiğinde oluşan açıklamalar Çizelge 2.8.’de yer almaktadır.

**Çizelge 2.8.** JRip Algoritması Kural Açıklamaları

KURAL	AÇIKLAMA
(OGRT_ILET = OGRT_ILET_KOTU) and (BABA_EGITIM = B_ILKOKUL) => BASARI_DURUM=BASARISIZ	Öğretmenle iletişimi kötü, babası ilkokul mezunu olan öğrencilerin başarı durumu başarısızdır.
(OGRT_ILET = OGRT_ILET_KOTU) and (KARDES_SAYI = COK) and (ODA = ODA_YOK) => BASARI_DURUM=BASARISIZ	Öğretmenle iletişimi kötü, kardeş sayısı çok, kendisine ait odası olmayan öğrencilerin başarı durumu başarısızdır.
(ANNE_EGITIM = A_LISE) and (SAGLIKPROB = SAGLIKLI) => BASARI_DURUM=BASARILI	Annesi lise mezunu olan sağlıklı öğrencilerin başarı durumu başarılıdır.
(ANNE_EGITIM = A_ORTAOKUL) and (ODA = ODA_VAR) and (KARDES_SAYI = AZ) => BASARI_DURUM=BASARILI	Annesi ortaokul mezunu, kendisine ait odası bulunan, kardeş sayısı az olan öğrencilerin başarı durumu başarılıdır.



### Çizelge 2.8. (devam)

(ANNE_EGITIM = A_UNI) and (ARK_ILIS = ARK_ILIS_IYI) => BASARI_DURUM=BASARILI	Annesi üniversite mezunu olan arkadaşlarıyla iletişimi iyi öğrencilerin başarı durumu başarılıdır.
BASARI_DURUM=ORTALAMA	Yukarıdaki 5 kurala uymayan öğrencilerin başarı durumu ortalamadır.

JRip algoritmasının avantajlı yönlerini aşağıdaki gibi sıralayabiliriz:

- Kural Kümesini yorumlamak diğerlerine göre daha kolaydır.
- Karar Ağacı öğrenmesine kıyasla daha iyi öğrenir.
- Birinci dereceden mantık gösterilemeyen uygulamalarda kolay uygulanabilir.

#### 2.6.5. Saf (Naive) Bayes Algoritması

Eldeki verilerin belirlenmiş olan sınıflara ait olma olasılıklarını öngören bir algoritmadır. Temeli, istatistikteki Bayes teoremine dayanır. Bu teorem; belirsizlik taşıyan herhangi bir durumun modelinin oluşturularak, bu durumla ilgili evrensel doğrular ve gerçekçi gözlemler doğrultusunda belli sonuçlar elde edilmesine olanak sağlar. Belirsizlik taşıyan durumlarda karar verme konusunda oldukça başarılıdır.

Genellikle belirsizlik durumlarında sınıflandırma ve tahmin yapmak için kullanılır. En önemli dezavantajı değişkenler arası ilişkinin modellenmemesi ve değişkenlerin birbirinden tamamen bağımsız olduğu varsayımdır [76].

Bayes Kuralı;

A ve B rastgele sayılar olsun;

$$P(A | B) = P(B | A)P(A) / P(B) \quad (2.8)$$

$P(A)$  : A olayının bağımsız olasılığı prior (öncül) olasılık

$P(B)$  : B olayının bağımsız olasılığı

$P(B|A)$  : A olayının olduğu bilindiğinde B olayının olasılığı likelihood (şartlı olasılık)

$P(A|B)$  : B olayının olduğu bilindiğinde A olayının olasılığı posterior (artçıl) olasılık

Bayes kuralına dayanarak  $P(A|B)$ ' yi maksimum yapan durumlar hesaplanabilir.

“E” A olayının bütün durumlarının kümesi;

$$A_{MAX} = \operatorname{argmax}_{A \in E} P(A|B)$$

$$= \operatorname{argmax}_{A \in E} \frac{P(B|A)P(A)}{P(B)}$$

$$= \operatorname{argmax}_{A \in E} P(B|A)P(A) \quad (2.9)$$

## 2.7. Sınıflandırma Algoritmalarının Başarısını Test Etme

Veri madenciliği modellerinin performansını hesaplamak için standart bir ölçütün kullanılması önemlidir. Veri madenciliğinde sınıflama modellerinin karşılaştırılması için en sık kullanılan yöntem hata oranını hesaplamaktır. İki sınıflı model için sınıflama matrisi Çizelge 2.9.'da verilmiştir.

**Çizelge 2.9.** İki Sınıflı Bir Model İçin Sınıflama Matrisi

	<b>Gerçek Sınıf</b>		
		<b>Pozitif</b>	<b>Negatif</b>
<b>Modelin Sınıf Tahmini</b>	<b>Pozitif</b>	Doğru Pozitif Sayısı (DP)	Yanlış Pozitif Sayısı (YP)
	<b>Negatif</b>	Yanlış Negatif Sayısı (YN)	Doğru Negatif Sayısı (DN)

Modeli oluşturan toplam örnek sayısı =  $N = DP + YP + YN + DN$

Modelin doğru sınıflama oranı (Doğruluk) =  $(DP + DN) / N$

Doğru Pozitif Oranı (Duyarlılık) =  $DP / (DP + YN)$

Doğru Negatif Oranı (Belirlilik) =  $DN / (DN + YP)$

Hassaslık (Precision) =  $DP / (DP + DN)$

### 3. ARAŞTIRMA BULGULARI

Tez uygulaması Kırıkkale ilinde bulunan 5 farklı Anadolu Lisesi'nde okuyan 11.sınıf öğrencilerine uygulanan anket uygulaması sonucunda elde edilen verilerden yararlanılarak gerçekleştirilmiştir. Bu çalışmada 210 adet öğrenciye ait veriler kullanılmıştır. Öğrencinin başarısına etki edebilecek etmenler ve bunların birbiri üzerindeki etkisi incelenmiştir. Öğrencilerin liselere yerleştirmede esas alınan puanın öğrencinin lise hayatında etkisinin olup olmadığı, ne düzeyde etkili olduğu araştırılmıştır.

Öğrencilerin cinsiyet, sağlık durumu, kardeş sayısı, anne baba eğitim durumu, anne baba yaşı, anne baba hayatta olma durumu, çocuğun ders dışı aktivitelere katılmalarını destekleme durumları, okul hayatıyla ilgilenme durumları, ailenin ekonomik durumu, çocuğun kendisine ait odasının olup olmadığı, öğrencinin öğretmenleriyle ve arkadaşlarıyla olan ilişkisinin durumuna göre başarı notları kıyaslanmıştır. Yapılan çalışma sonucunda öğrencilerin farklı kriterlere göre başarısızlıkları ve bu başarısızlıklarının nedenini bulup çözümlenmek hedeflenmiştir. Liselere yerleştirme puanlarının aynı düzeyde üniversiteye yerleştirmede etkili olup olmayacağı yordanmak istenmiştir. Farklı liselerde okuyan öğrencilerden alınan verilerin tamamı uygulamaya dahil edilerek incelenmiş ardından da tüm veriler içerisinden rasgele seçilen veriler ve aynı okulda okuyan öğrencilere ait veriler ayrı ayrı incelenmiştir. Uygulama Weka 3.7 programı yardımı ile yapılmıştır.

#### 3.1. Veri Tanımlama

Öncelikle öğrencinin başarısına etkisi muhtemel kriterler belirlenmiştir. Bunlar;

Öğrencinin cinsiyeti (kız, erkek)

Kardeş Sayısı (tek, az, çok)

Sağlık problemi (hasta, sağlıklı)

Anne-Baba eğitim durumu (ilkokul, ortaokul, lise, üniversite, yüksek lisans)

Anne-Baba hayatta olma durumu (sağ, ölü)  
Anne-Baba yaşı (genç, orta yaş, yaşlı)  
Ders dışı aktivitelere destek (var, yok)  
Anne-Babanın öğrencinin okul hayatıyla ilgisi (ilgili, ilgisiz)  
Ailenin ekonomik durumu (iyi, kötü)  
Öğrencinin kendisine ait odası (var, yok)  
Öğrencinin öğretmenleri ile iletişimi (iyi, kötü)  
Öğrencinin arkadaşları ile iletişimi (iyi, kötü)  
Öğrencinin başarı durumu (başarısız, ortalama, başarılı) şeklinde belirlenmiştir.

Yapılan anket uygulaması sonucunda elde edilen verilerden gereksiz alanlar temizlenmiştir.

Verilerin rahat modellenmesi için bazı alanların yapısı değiştirilmiştir.

**Çizelge 3.1.** Başarı Durumu

Not Ortalaması Aralığı	Başarı Durumu
0-2.49	BAŞARISIZ
2.50-3.99	ORTALAMA
4.00-5.00	BAŞARILI

**Çizelge 3.2.** Kardeş Sayısı

Kardeş Sayısı	Durumu
1	BAŞARISIZ
2 – 3	ORTALAMA
4 ve üzeri	BAŞARILI

**Çizelge 3.3.** Anne-Baba Hayatta Olma Durumu

Anne-Baba	Hayatta Olma Durumu
Sağ-Sağ	SAĞ
Sağ-Ölü	ÖLÜ
Ölü-Sağ	ÖLÜ

**Çizelge 3.4.** Anne-Baba Yaşı

Anne-Baba Yaş Ortalaması Aralığı	Anne Baba Yaşı
30-39	GENÇ
40-49	ORTA YAŞ
50 ve üzeri	YAŞLI

210 adet kayda ait veri madenciliği uygulamasında kullanılacak veri seti üzerindeki veri dağılımı Çizelge 3.5.'de verilmiştir.

**Çizelge 3.5.** Veri Madenciliği Çalışması İçin Kullanılacak Verilerin Dağılımı

Sınıf	Sayı
BAŞARILI	85
ORTALAMA	105
BAŞARISIZ	20

Verilere ait istatistiki bilgiler Çizelge 3.6.'da gösterilmiştir.

**Çizelge 3.6.** Veritabanı İstatistikleri 1

Niteliği	Özelliği	Kişi Sayısı
Cinsiyeti	Kız	110
	Erkek	100
Kardeş Sayısı	Tek	20
	Az	149
	Çok	41
Sağlık Problemi	Hasta	158
	Sağlıklı	52

**Çizelge 3.7.** Veritabanı İstatistikleri 2

Niteliği	Özelliği	Kişi Sayısı
Anne Eğitim Durumu	İlkokul	99
	Ortaokul	33
	Lise	63
	Üniversite	14
	Y.Lisans	1
Baba Eğitim Durumu	İlkokul	36
	Ortaokul	26
	Lise	81
	Üniversite	64
	Y.Lisans	3
Anne-Baba Hayatta Olma Durumu	Sağ	194
	Ölü	16
Anne-Baba Yaşı	Genç	38
	Orta Yaş	141
	Yaşlı	31

**Çizelge 3.7. (devam)**

Ders dışı Aktivitelere Destek	Var	149
	Yok	61
Anne-Babanın Öğrencinin Okul Hayatıyla İlgisi	İlgili	114
	İlgisiz	96
Ailenin Ekonomik Durumu	İyi	155
	Kötü	55
Öğrencinin Kendisine At Odası	Var	147
	Yok	63
Öğrenci-Öğretmen İlişkisi	İyi	147
	Kötü	63
Öğrenci-Arkadaş İletişimi	İyi	156
	Kötü	54

Verileri Weka programında kullanabilmek için öncelikle Attribute - Relation File Format (ARFF – Özellik İlişki Dosya Biçimi) formatına dönüştürülmesi gerekmektedir. Arff dosyaları, değişken tanımlamasına izin veren ASCII metin dosyalarıdır. Arff dosyasının başlık kısmında, değişkenler (veri tabanındaki her bir kolonun ismi), bunlar arasındaki ilişkiler ve her bir değişkenin türü ve alacağı değerler bulunmaktadır.

Arff dosyasını oluşturmak için Microsoft Excel 2010 programı kullanılmıştır. Çalışmada kullanılan verilere ilişkin tanımlanan değişkenler ve tipleri Şekil 3.1.'de gösterilmiştir.



```

@RELATION LISE_BASARI
@ATTRIBUTE CINSIYET {KIZ,ERKEK}
@ATTRIBUTE KARDES_SAYI {TEK,AZ,COK}
@ATTRIBUTE SAGLIKPROB {SAGLIKLI,HASTA}
@ATTRIBUTE ANNE_EGITIM {A_ILKOKUL,A_ORTAOKUL,A_LISE,A_UNI,A_YLISANS}
@ATTRIBUTE BABA_EGITIM {B_ILKOKUL,B_ORTAOKUL,B_LISE,B_UNI,B_YLISANS}
@ATTRIBUTE A_B_HAYAT {SAG,OLU}
@ATTRIBUTE A_B_YAS {GENC,ORTA_YAS,YASLI}
@ATTRIBUTE AKT_DES {AKT_DES_VAR,AKT_DES_YOK}
@ATTRIBUTE AILE_ILGI {ILGILI,ILGISIZ}
@ATTRIBUTE EKONOMIK_DUR {EKONOMIK_D_IYI,EKONOMIK_D_KOTU}
@ATTRIBUTE ODA {ODA_VAR,ODA_YOK}
@ATTRIBUTE OGRT_ILET {OGRT_ILET_IYI,OGRT_ILET_KOTU}
@ATTRIBUTE ARK_ILIS {ARK_ILIS_IYI,ARK_ILIS_KOTU}
@ATTRIBUTE BASARI_DURUM {BASARILI,ORTALAMA,BASARISIZ}

```

**Şekil 3.1.** Çalışmada Oluşturulan arff Dosyasının Başlık Kısmı

Verilerin bulunduğu kısım @DATA satırından sonra gelir. Şekil 3.2.de verilerin bulunduğu kısım gösterilmektedir.

```

@DATA
ERKEK,AZ,SAGLIKLI,A_LISE,B_UNI,SAG,ORTA_YAS,AKT_DES_VAR,ILGILI,
EKONOMIK_D_IYI,ODA_VAR,OGRT_ILET_IYI,ARK_ILIS_IYI,BASARILI
KIZ,AZ,SAGLIKLI,A_LISE,B_ILKOKUL,SAG,GENC,AKT_DES_VAR,ILGILI,
EKONOMIK_D_IYI,ODA_VAR,OGRT_ILET_IYI,ARK_ILIS_IYI,BASARILI
ERKEK,AZ,SAGLIKLI,A_ORTAOKUL,B_ORTAOKUL,SAG,GENC,AKT_DES_VAR,ILGISIZ,
EKONOMIK_D_IYI,ODA_VAR,OGRT_ILET_IYI,ARK_ILIS_IYI,BASARILI
ERKEK,COK,SAGLIKLI,A_ORTAOKUL,B_LISE,SAG,GENC,AKT_DES_VAR,ILGILI,
EKONOMIK_D_IYI,ODA_VAR,OGRT_ILET_IYI,ARK_ILIS_IYI,ORTALAMA
ERKEK,AZ,SAGLIKLI,A_ILKOKUL,B_LISE,SAG,ORTA_YAS,AKT_DES_VAR,ILGILI,
EKONOMIK_D_IYI,ODA_VAR,OGRT_ILET_KOTU,ARK_ILIS_KOTU,ORTALAMA
ERKEK,TEK,SAGLIKLI,A_ILKOKUL,B_ILKOKUL,SAG,ORTA_YAS,AKT_DES_YOK,ILGISIZ,
EKONOMIK_D_KOTU,ODA_YOK,OGRT_ILET_IYI,ARK_ILIS_IYI,ORTALAMA
ERKEK,AZ,SAGLIKLI,A_ILKOKUL,B_UNI,SAG,ORTA_YAS,AKT_DES_VAR,ILGILI,
EKONOMIK_D_IYI,ODA_VAR,OGRT_ILET_IYI,ARK_ILIS_IYI,BASARILI
ERKEK,TEK,SAGLIKLI,A_LISE,B_LISE,SAG,ORTA_YAS,AKT_DES_VAR,ILGILI,
EKONOMIK_D_IYI,ODA_VAR,OGRT_ILET_IYI,ARK_ILIS_IYI,BASARILI
KIZ,COK,SAGLIKLI,A_LISE,B_LISE,SAG,YASLI,AKT_DES_YOK,ILGISIZ,
EKONOMIK_D_IYI,ODA_VAR,OGRT_ILET_KOTU,ARK_ILIS_KOTU,ORTALAMA
KIZ,AZ,SAGLIKLI,A_ILKOKUL,B_ILKOKUL,SAG,ORTA_YAS,AKT_DES_VAR,ILGISIZ,
EKONOMIK_D_KOTU,ODA_YOK,OGRT_ILET_KOTU,ARK_ILIS_KOTU,ORTALAMA

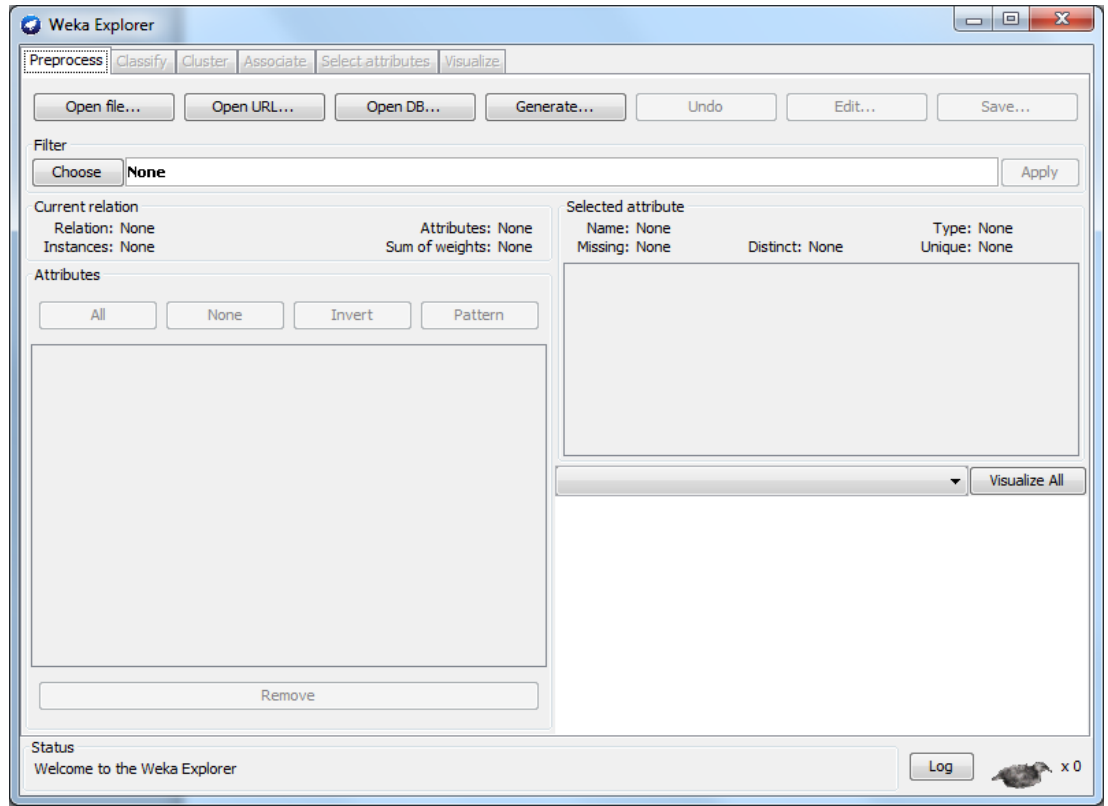
```

**Şekil 3.2.** Çalışmada Oluşturulan arff Dosyasında Verilerin Bulunduğu Kısım

### 3.2. Modeli Kurma

Arff formatındaki veri dosyası oluşturulduktan sonra Weka programı çalıştırılmıştır.

Weka ara yüzünde bulunan menüde Explorer butonuna basılarak Explorer penceresi açılmıştır. Verileri ön işlemde geçirmek için Preprocess sayfasında dosya aç seçeneği yardımı ile oluşturmuş olduğumuz arff dosyası açılmıştır. Şekil 3.3.'de Weka Explorer ara yüzü görülmektedir.



Şekil 3.3. Weka Explorer Penceresi

Model kurulurken verilerin belli bir kısmı eğitilmiş (training), eğitilen verilerin oluşturduğu örüntüler kullanılarak geri kalan veriler test edilmiştir. Örüntü bulma ve test veri sınıflarının tahmini işlemlerini Weka sınıflandırma algoritmaları yapmıştır.

Verileri işlemede tahmin edilecek alan için doğruluk performansı yüksek olan algoritma ele alınmıştır.

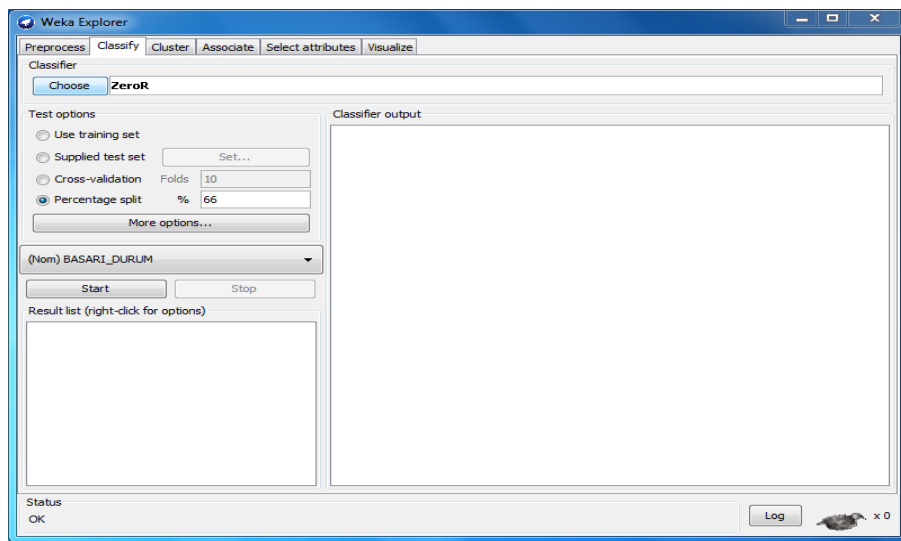
Sınıflandırma algoritmalarının performanslarını test etmek için Weka'daki sınıflandırma algoritmalarından Çok Katmanlı Algılayıcı, k- En Yakın Komşu, J48, JRIP ve Saf Bayes Algoritması kullanılmıştır.

Çalışmada Weka 3.7 sürümü kullanılmıştır. Veriler tüm sınıflandırıcılar ile test edilmiştir. Bu işlem yapılırken yüzde ayırma (percentage split) yöntemi kullanılmıştır.

Verilerin %66'sı (program varsayılan değer olarak da %66 kullanılmaktadır) yüzde ayırma yöntemi ile eğitim için, diğer %34'ü de test verisi olarak kullanılmıştır. Çeşitli algoritmalarla test verisi test edilerek algoritmaların doğruluğu belirlenmiştir.

### 3.3. Modeli Değerlendirme

Weka Explorer penceresinde "Sınıflama" panelinde bulunan çeşitli algoritmalar seçilerek modeller oluşturulmuştur. Şekil 3.4'te Weka Explorer Penceresinde Sınıflama paneli ekranı görülmektedir.



Şekil 3.4. Weka Sınıflama Paneli

### 3.3.1. Çok Katmanlı Algılayıcı İle Oluşturulan Veri Modelleme

Şekil 3.5.'de Uygulamada çok katmanlı algılayıcı algoritması ile oluşturulmuş veri modellemesi ve sonuç ekranı görülmektedir. Şekil 3.5.'deki çıkış ekranı incelendiğinde;

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a**

Test options

Use training set

Supplied test set Set...

Cross-validation Folds 10

Percentage split % 66

More options...

(Nom) BASARI\_DURUM

Start Stop

Result list (right-click for options)

17:57:59 - functions.MultilayerPerceptron

Classifier output

=== Evaluation on test split ===

=== Summary ===

Correctly Classified Instances	63	88.7324 %
Incorrectly Classified Instances	8	11.2676 %
Kappa statistic	0.8008	
Mean absolute error	0.0825	
Root mean squared error	0.244	
Relative absolute error	21.0608 %	
Root relative squared error	54.4523 %	
Coverage of cases (0.95 level)	95.7746 %	
Mean rel. region size (0.95 level)	39.4366 %	
Total Number of Instances	71	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.857	0.028	0.968	0.857	0.909	0.976	BASARILI
0.967	0.171	0.806	0.967	0.879	0.969	ORTALAMA
0.667	0	1	0.667	0.8	0.974	BASARISIZ
Weighted Avg.	0.887	0.086	0.902	0.887	0.887	0.973

=== Confusion Matrix ===

a	b	c	<-- classified as
30	5	0	a = BASARILI
1	29	0	b = ORTALAMA
0	2	4	c = BASARISIZ

Status

OK Log x 0

Şekil 3.5. Çok Katmanlı Algılayıcı ile Oluşturulan Modelin Sonuç Ekranı

210 kaydın %66'sı olan 139 adet kayıt modelin eğitimi için kullanılmıştır. Geriye kalan 71 kayıt ise test amaçlı kullanılmıştır. Çıkış ekranında görülen Hata Matrisine göre (Confusion Matrix) 71 adet kayıta ait hata matrisi Çizelge 3.8.'de görüldüğü gibidir.

**Çizelge 3.8.**Çok Katmanlı Algılayıcı Algoritması İçin Düzensizlik Matrisi

a	b	c	
30	5	0	a=BASARILI
1	29	0	b=ORTALAMA
0	2	4	c=BASARISIZ

Çizelge 3.8.'e göre;

35 adet başarı durumu “BASARILI” değerli test verisinin 30 tanesi “BASARILI”, 5 tanesi “ORTALAMA”,

30 adet başarı durumu "ORTALAMA" değerli test verisinin 1 tanesi “BASARILI” 29 tanesi “ORTALAMA”,

6 adet başarı durumu “BASARISIZ” test verisinin 2 tanesi “ORTALAMA”, 4 tanesi de “BASARISIZ” olarak tahmin edilmiştir.

Bu durumda 8 (5+1+2) adet veri Çok Katmanlı Algılayıcı Algoritması ile oluşturulan modele göre yanlış, 63 (30+29+4) adet veri doğru sınıflandırılmıştır. Buna göre modelin doğruluk sınıflandırma yüzdesi aşağıdaki gibi hesaplanmıştır;

$$\text{Doğruluk Yüzdesi} = (63/71) * 100 = \%88,73$$

Şekil 3.5.'de sınıflandırıcı çıkış ekranında görülen sınıflara göre 63 adet verinin detaylı olarak doğruluk tablosu Çizelge 3.9.'da verilmiştir.

**Çizelge 3.9.** Çok Katmanlı Algılayıcı İçin Detaylı Doğruluk Tablosu

Doğruluk Yüzdesi	Duyarlılık (DP Oranı)	Yanlış Hassaslık Oranı (YP Oranı)	Hassaslık (Precision)	Sınıf (Class)
88,73%	0,857	0,028	0,968	a=BASARILI
	0,967	0,171	0,806	b=ORTALAMA
	0,667	0	1	c=BASARISIZ

Çizelge 3.8.'e göre, Çizelge 3.9.'daki değerler aşağıdaki gibi hesaplanmıştır:

Duyarlılık (Doğru Pozitif Oranı)	a sınıfı için $30/35 = 0,857$
	b sınıfı için $29/30 = 0,967$
	c sınıfı için $4/6 = 0,667$
Yanlış Hassaslık Oranı (Yanlış Pozitif Oranı)	a sınıfı için $1/36 = 0,028$
	b sınıfı için $7/41 = 0,171$
	c sınıfı için $0/65 = 0$
Hassaslık (Precision)	a sınıfı için $30 / 31 = 0,968$
	b sınıfı için $29/ 36 = 0,806$
	c sınıfı için $4 / 4 = 1$

### 3.3.2. k- En Yakın Komşu Algoritması İle Oluşturulan Veri Modelleme

Çizelge 3.10.'da k- En Yakın Komşu Algoritması için düzensizlik matrisi verilmiştir.

**Çizelge 3.10.** k- En Yakın Komşu Algoritması İçin Düzensizlik Matrisi

a	b	C	
30	5	0	a=BASARILI
3	27	0	b=ORTALAMA
1	1	4	c=BASARISIZ

Çizelge 3.10.'a göre;

35 adet başarı durumu "BASARILI" değerli test verisinin 30 tanesi "BASARILI", 5 tanesi "ORTALAMA",

30 adet başarı durumu "ORTALAMA" değerli test verisinin 3 tanesi "BASARILI" 27 tanesi "ORTALAMA",

6 adet başarı durumu “BASARISIZ” test verisinin 1 tanesi “BASARILI” 1 tanesi “ORTALAMA”, 4 tanesi de “BASARISIZ” olarak tahmin edilmiştir.

Çizelge 3.10.’da verilen düzensizlik matrisine göre 71 adet verinin detaylandırılmış doğruluk tablosu aşağıda Çizelge 3.11.’de verilmiştir.

**Çizelge 3.11.** k- En Yakın Komşu İçin Detaylı Doğruluk Tablosu

Doğruluk Yüzdesi	Duyarlılık (DP Oranı)	Yanlış Hassaslık Oranı (YP Oranı)	Hassaslık (Precision)	Sınıf (Class)
85,91%	0,857	0,111	0,882	a=BASARILI
	0,9	0,146	0,818	b=ORTALAMA
	0,667	0	1	c=BASARISIZ

Çizelge 3.11.’e göre, k- En Yakın Komşu Algoritması için doğruluk yüzdesi, %85,91 olarak hesaplanmıştır. Duyarlılık, yanlış hassaslık oranı ve hassaslık değerleri, BASARILI, ORTALAMA ve BASARISIZ sınıfları için ayrı ayrı hesaplanmıştır.

### 3.3.3. J48 Algoritması İle Oluşturulan Veri Modelleme

Çizelge 3.12.’de J48 Algoritması için düzensizlik matrisi verilmiştir.

**Çizelge 3.12.** J48 Algoritması İçin Düzensizlik Matrisi

A	B	c	
30	5	0	a=BASARILI
2	26	2	b=ORTALAMA
0	2	4	c=BASARISIZ

Çizelge 3.12.'ye göre;

35 adet başarı durumu "BASARILI" değerli test verisinin 30 tanesi "BASARILI", 5 tanesi "ORTALAMA",

30 adet başarı durumu "ORTALAMA" değerli test verisinin 2 tanesi "BASARILI", 26 tanesi "ORTALAMA", 2 tanesi "BASARISIZ"

6 adet başarı durumu "BASARISIZ" test verisinin 2 tanesi "ORTALAMA", 4 tanesi de "BASARISIZ" olarak tahmin edilmiştir.

Çizelge 3.12.'de verilen düzensizlik matrisine göre 71 adet verinin detaylandırılmış doğruluk tablosu aşağıda Çizelge 3.13.'de verilmiştir.

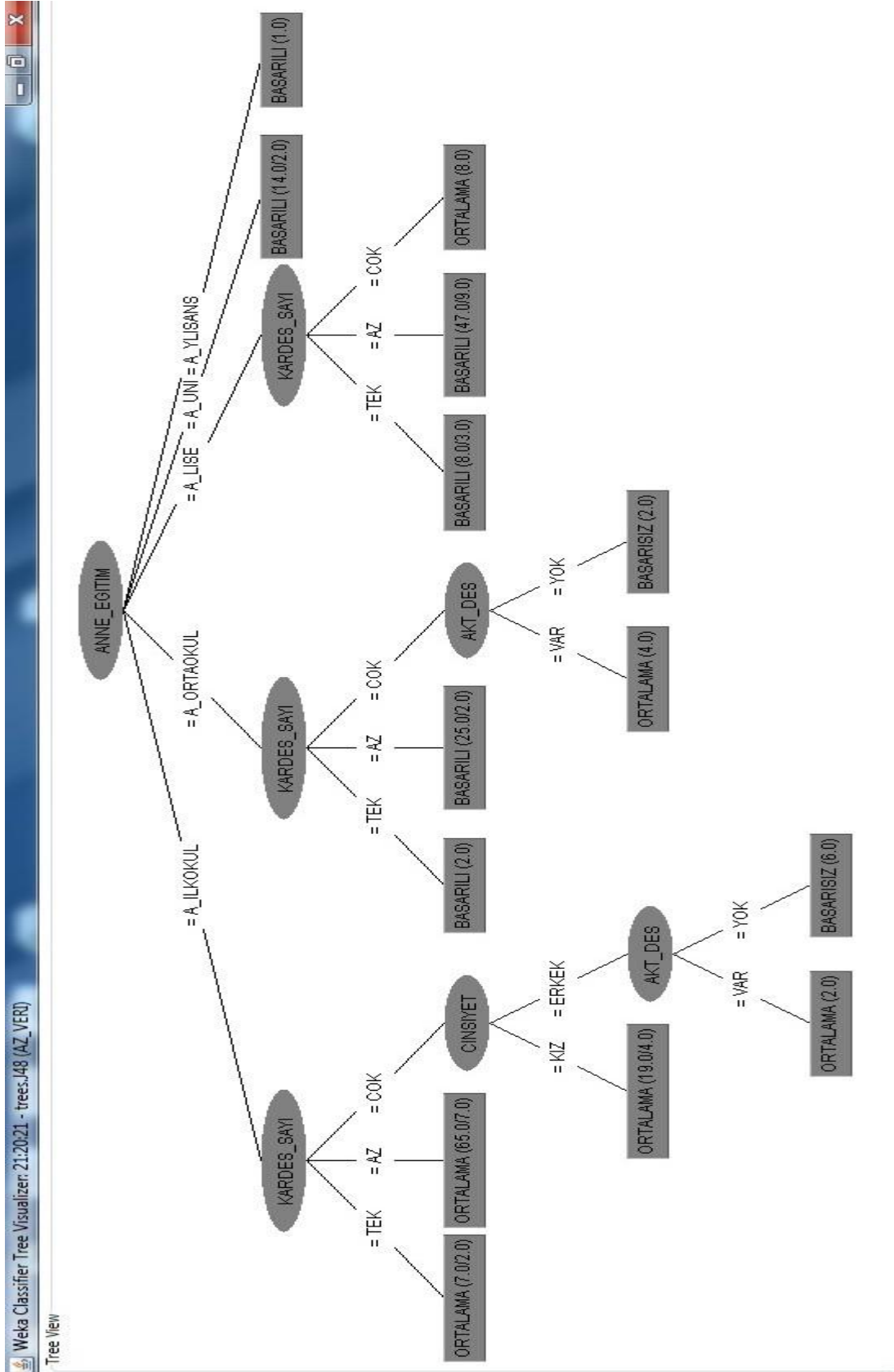
**Çizelge 3.13.** J48 İçin Detaylı Doğruluk Tablosu

<b>Doğruluk Yüzdesi</b>	<b>Duyarlılık (DP Oranı)</b>	<b>Yanlış Hassaslık Oranı (YP Oranı)</b>	<b>Hassaslık (Precision)</b>	<b>Sınıf (Class)</b>
84,50%	0,857	0,056	0,938	a=BASARILI
	0,867	0,171	0,788	b=ORTALAMA
	0,667	0,031	0,667	c=BASARISIZ

Çizelge 3.13.'e göre, J48 Algoritması için doğruluk yüzdesi, %84,50 olarak hesaplanmıştır. Duyarlılık, Yanlış Hassaslık Oranı ve Hassaslık değerleri, BASARILI, ORTALAMA VE BASARISIZ sınıfları için ayrı ayrı hesaplanmıştır.

Şekil 3.6.'da cinsiyet, kardeş sayısı, anne eğitim durumu, ders dışı aktivitelere destek ve başarı durumunun ele alındığı verilerle J48 Algoritması için karar ağacı sonuç ekranı görülmektedir.





Şekil 3.6. J48 Algoritması İçin Karar Ağacı Sonuç ekranı

Her satır, ağaçtaki bir düğümü, alt satırlar, ilk satırın çocuk düğümlerini; düğümlerde parantezin içindeki ilk sayı veri kümesindeki kaç durumun bu düğüm için doğru olarak sınıflandırıldığını; eğer varsa; parantezin içindeki ikinci sayı, düğüm tarafından yanlış olarak sınıflandırılan durumların sayısını göstermektedir.

Şekil 3.6.'da görüldüğü gibi kademeler en üstten aşağıya doğru anne eğitim durumu, kardeş sayısı, cinsiyet, ders dışı aktivitelere destek şeklindedir.

### 3.3.4. JRip Algoritması İle Oluşturulan Veri Modelleme

Çizelge 3.14.'de JRip Algoritması için düzensizlik matrisi verilmiştir.

**Çizelge 3.14.** JRip Algoritması İçin Düzensizlik Matrisi

A	b	C	
30	5	0	a=BASARILI
0	28	2	b=ORTALAMA
0	2	4	c=BASARISIZ

Çizelge 3.14.'e göre;

35 adet başarı durumu "BASARILI" değerli test verisinin 30 tanesi "BASARILI", 5 tanesi "ORTALAMA",

30 adet başarı durumu "ORTALAMA" değerli test verisinin 28 tanesi "ORTALAMA", 2 tanesi "BASARISIZ"

6 adet başarı durumu "BASARISIZ" test verisinin 2 tanesi "ORTALAMA", 4 tanesi de "BASARISIZ" olarak tahmin edilmiştir.

Çizelge 3.14.'de verilen düzensizlik matrisine göre 71 adet verinin detaylandırılmış doğruluk tablosu aşağıda Çizelge 3.15.'de verilmiştir.

**Çizelge 3.15.** JRip İçin Detaylı Doğruluk Tablosu

Doğruluk Yüzdesi	Duyarlılık (DP Oranı)	Yanlış Hassaslık Oranı (YP Oranı)	Hassaslık (Precision)	Sınıf (Class)
87,32%	0,857	0	1	a=BASARILI
	0,933	0,171	0,8	b=ORTALAMA
	0,667	0,031	0,667	c=BASARISIZ

Çizelge 3.15.'e göre, JRip Algoritması için doğruluk yüzdesi, %87,32 olarak hesaplanmıştır. Duyarlılık, Yanlış Hassaslık Oranı ve Hassaslık değerleri, BASARILI, ORTALAMA VE BASARISIZ sınıfları için ayrı ayrı hesaplanmıştır.

### 3.3.5. Saf Bayes Algoritması İle Oluşturulan Veri Modelleme

Çizelge 3.16.'da Saf Bayes Algoritması için düzensizlik matrisi verilmiştir.

**Çizelge 3.16.** Saf Bayes Algoritması İçin Düzensizlik Matrisi

A	b	c	
31	4	0	a=BASARILI
4	23	3	b=ORTALAMA
0	2	4	c=BASARISIZ

Çizelge 3.16.'e göre;

35 adet başarı durumu "BASARILI" değerli test verisinin 31 tanesi "BASARILI", 4 tanesi "ORTALAMA",

30 adet başarı durumu "ORTALAMA" değerli test verisinin 4 tanesi "BASARILI" 23 tanesi "ORTALAMA", 3 tanesi "BASARISIZ"

6 adet başarı durumu "BASARISIZ" test verisinin 2 tanesi "ORTALAMA", 4 tanesi de "BASARISIZ" olarak tahmin edilmiştir.

Çizelge 3.16.'da verilen düzensizlik matrisine göre 71 adet verinin detaylandırılmış doğruluk tablosu aşağıda Çizelge 3.17.'de verilmiştir.

**Çizelge 3.17.** Saf Bayes İçin Detaylı Doğruluk Tablosu

<b>Doğruluk Yüzdesi</b>	<b>Duyarlılık (DP Oranı)</b>	<b>Yanlış Hassaslık Oranı (YP Oranı)</b>	<b>Hassaslık (Precision)</b>	<b>Sınıf (Class)</b>
% 81,69	0,886	0,111	0,886	a=BASARILI
	0,767	0,146	0,793	b=ORTALAMA
	0,667	0,046	0,571	c=BASARISIZ

Çizelge 3.17.'ye göre, Saf Bayes Algoritması için doğruluk yüzdesi, %81,69 olarak hesaplanmıştır. Duyarlılık, yanlış hassaslık oranı ve hassaslık değerleri, BASARILI, ORTALAMA VE BASARISIZ sınıfları için ayrı ayrı hesaplanmıştır.

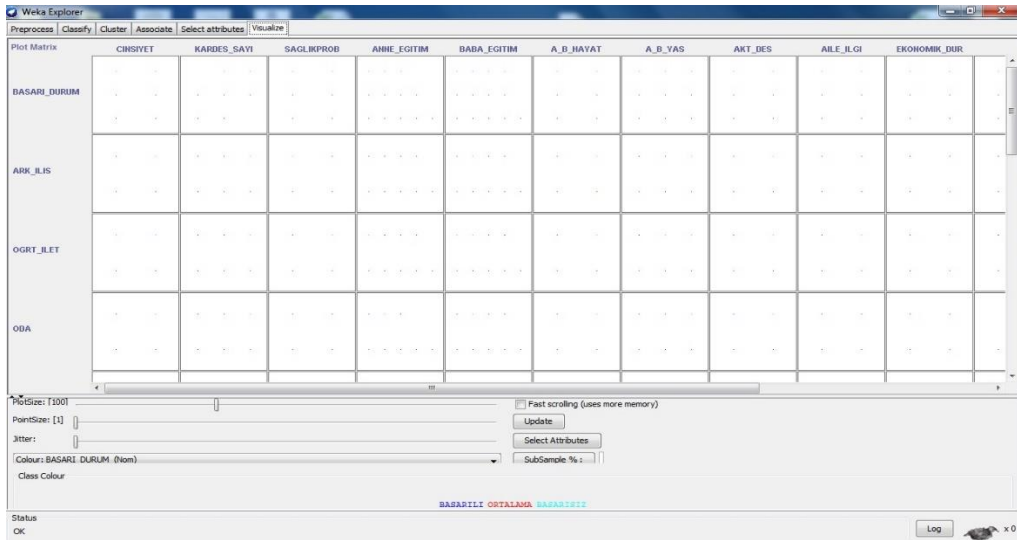
**Çizelge 3.18.** Seçilen Sınıflandırma Algoritmaları ve Doğruluk Yüzdeleri

Algoritma Adı	Doğru Sınıflandırılan Örnek Sayısı	Yanlış Sınıflandırılan Örnek Sayısı	Doğruluk Yüzdesi
<b>Multilayer Perceptron</b>	63	8	88,7324
<b>JRip</b>	62	9	87,3239
<b>IBk</b>	61	10	85,9155
<b>J48</b>	60	11	84,507
<b>Naive Bayes</b>	58	13	81,6901

### 3.4. Weka Programı İle Elde Edilen Görsel Sonuçlar

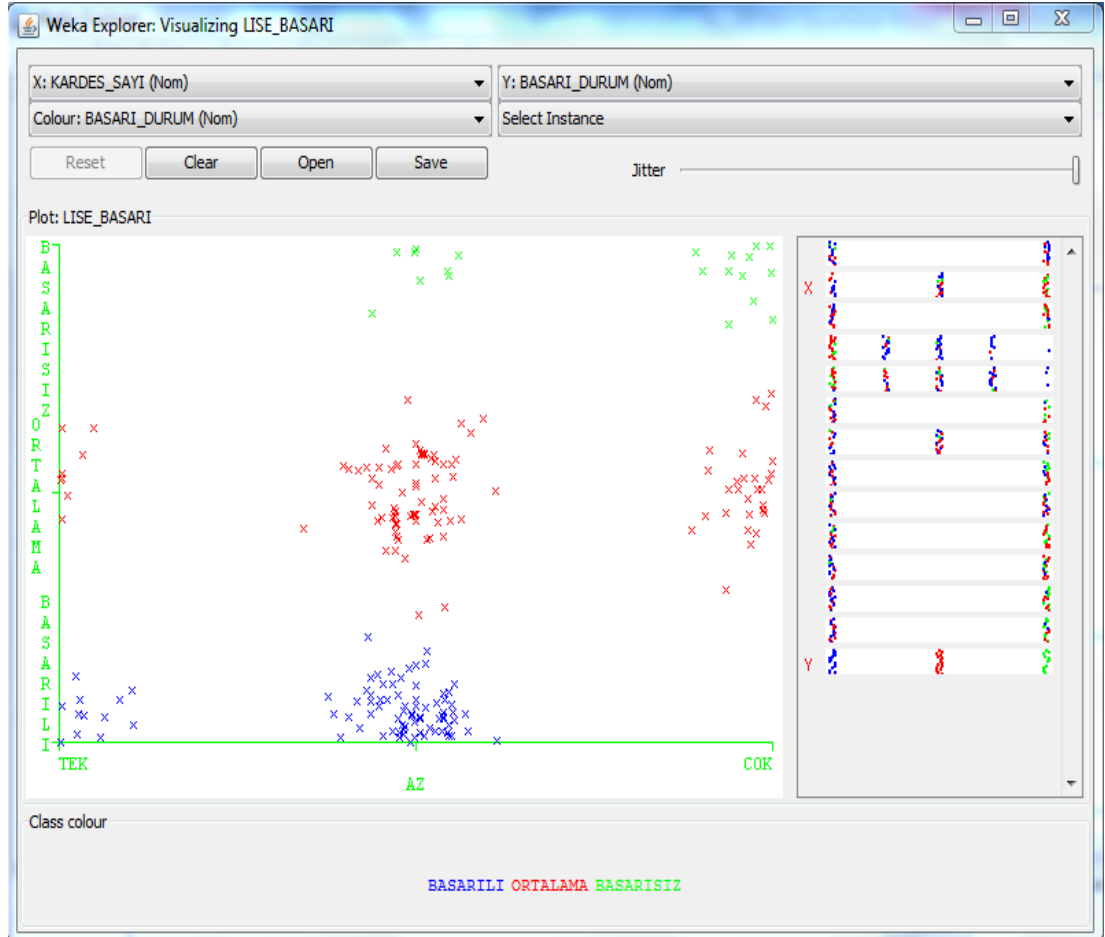
Weka görsel olarak da sonuçlar üretebilmektedir. Weka ile elde edilen görsel sonuçlar bu bölümde incelenecektir.

Şekil 3.7.'de görüldüğü gibi Weka görselleştirme sayfasında farklı veri alanları özelliklerine göre üç boyutlu olarak analiz edilmesine olanak sağlamaktadır.



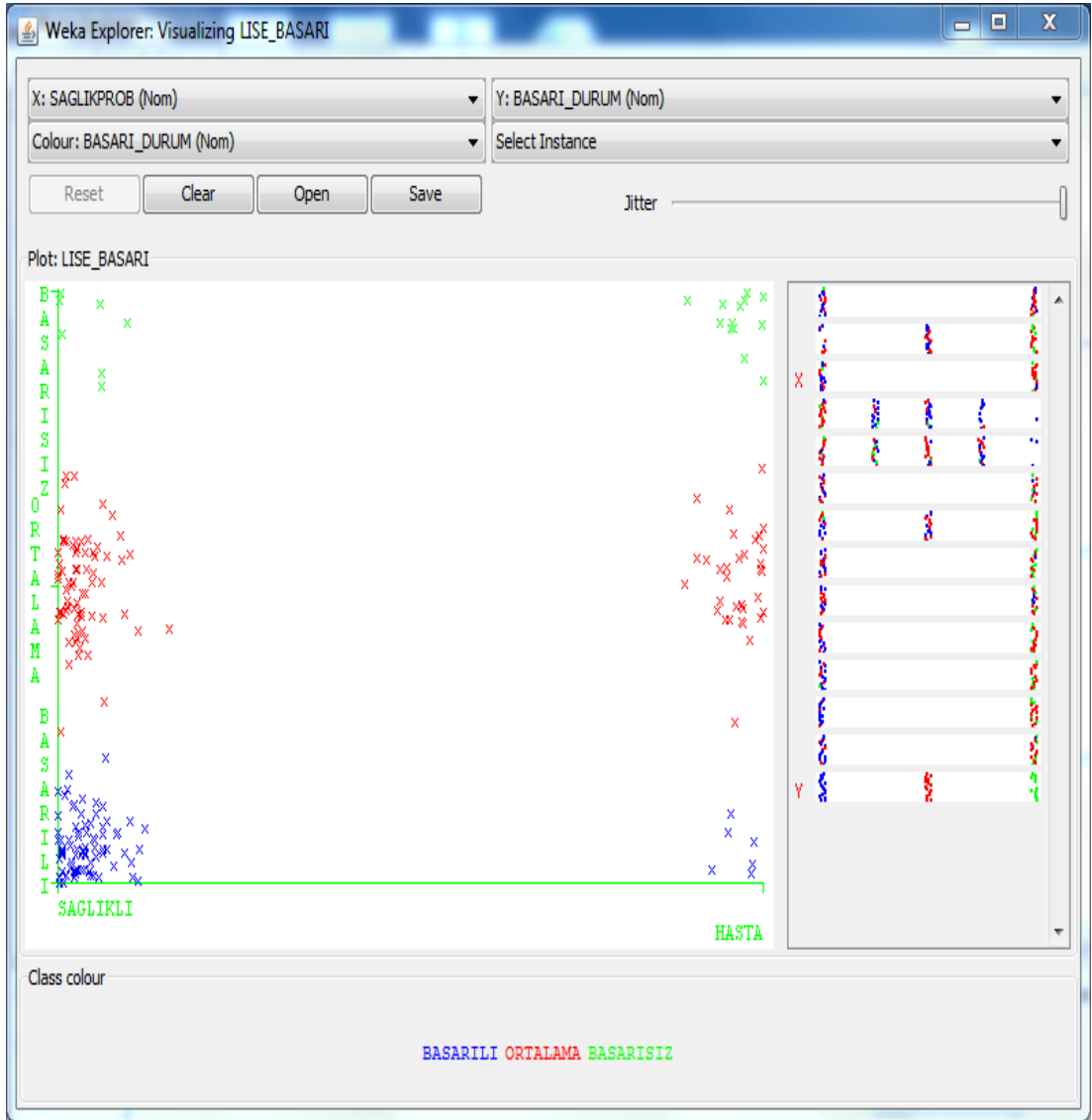
**Şekil 3.7.** Weka Görselleştirme (Visualize) Paneli

Şekil 3.8.'de öğrencilerin kardeş sayıları ile başarı durumları arasındaki ilişkiyel grafik raporlanmıştır. Bu rapora göre tek çocuk olanların başarı durumlarının başarılı ağırlıklı olduğu, kardeş sayısı çok olan öğrencilerin ise başarılı olamadıkları ortalama ve başarısız düzeyde oldukları görülmektedir.



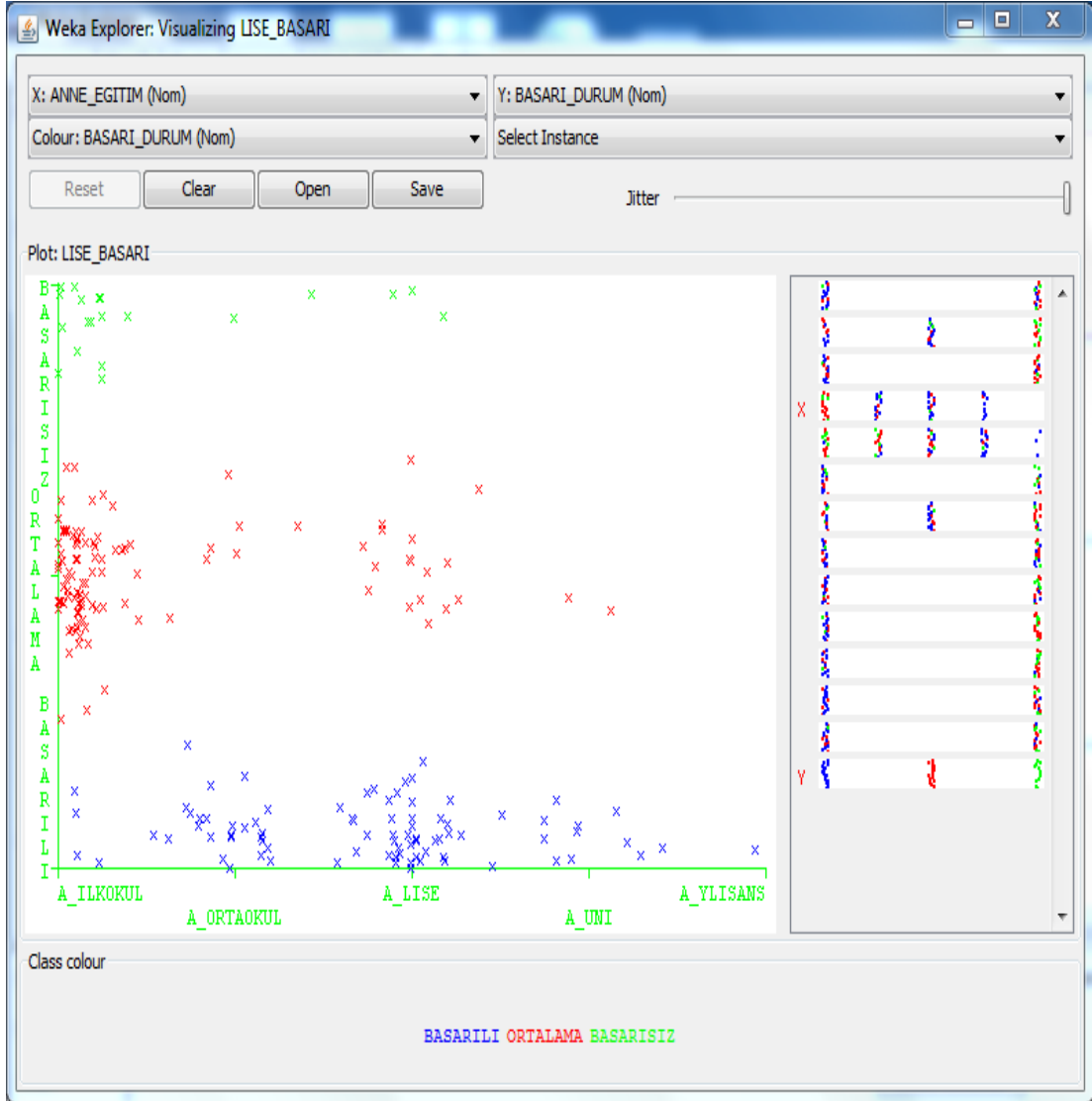
**Şekil 3.8.** Kardeş Sayısına Göre Başarı Durumu

Şekil 3.9.'da öğrencilerin sağlık problemleri ile başarı durumları arasındaki ilişkiyel grafik raporlanmıştır. Bu rapora göre sağlıklı öğrencilerin başarı durumlarının başarılı ve ortalama ağırlıklı olduğu, sağlık problemi olan öğrencilerin ise başarı durumlarının daha çok ortalama ve başarısız düzeyde olduğu görülmüştür.



**Şekil 3.9.** Sağlık Problemine Göre Başarı Durumu

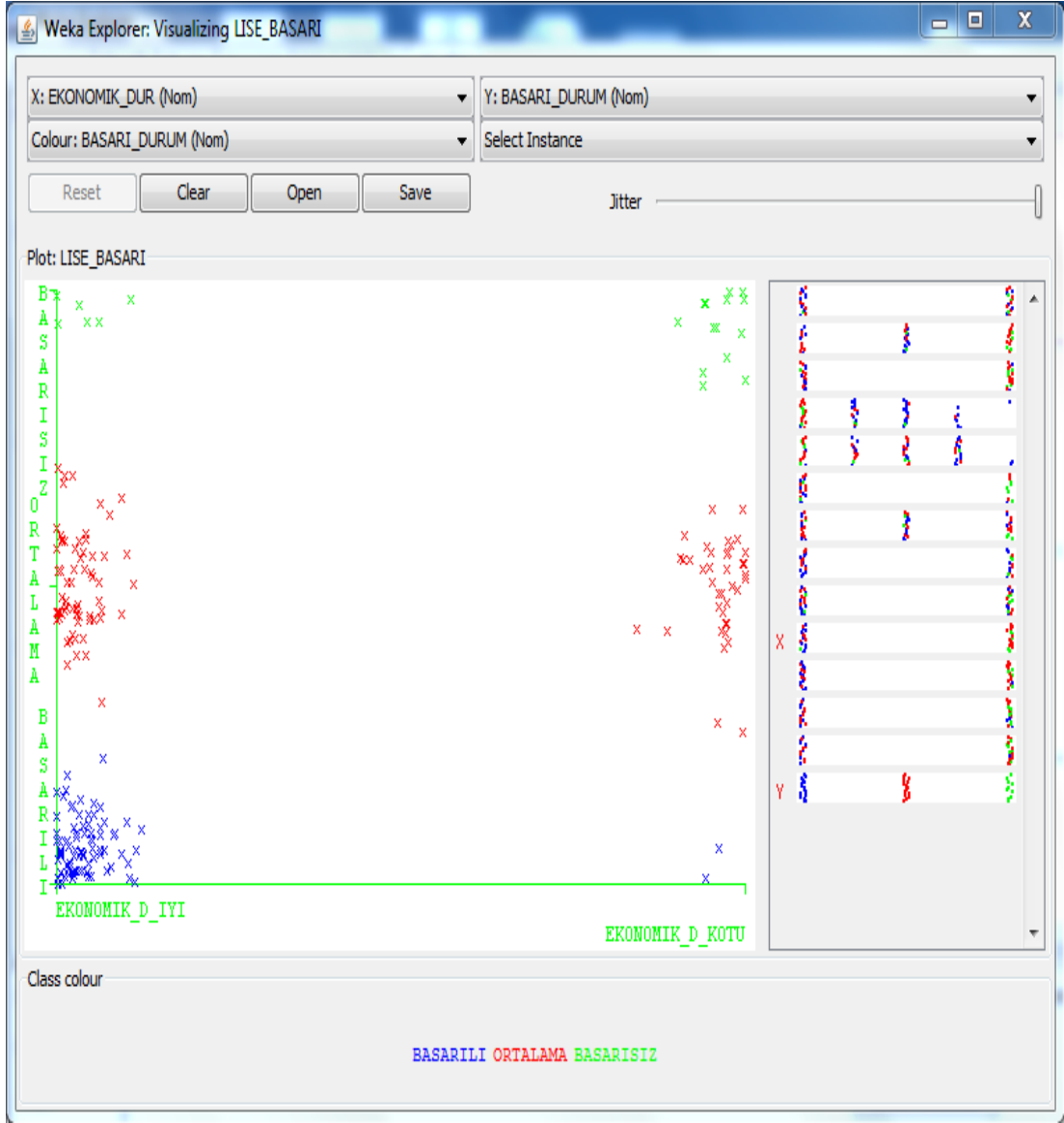
Şekil 3.10'da öğrencilerin anne eğitim durumları ile başarı durumları arasındaki ilişkiyel grafik raporlanmıştır. Bu rapora göre annesi ilkokul mezunu olan öğrencilerin başarılarının daha çok ortalama ve başarısız düzeyde olduğu, annesi üniversite ve yüksek lisans mezunu olan öğrencilerin ise başarılı olduğu görülmüştür. Bu rapordaki sonuca göre anne eğitim düzeyi arttıkça öğrenci başarı düzeyi artmaktadır.



**Şekil 3.10.** Anne Eğitim Düzeyine Göre Başarı Durumu

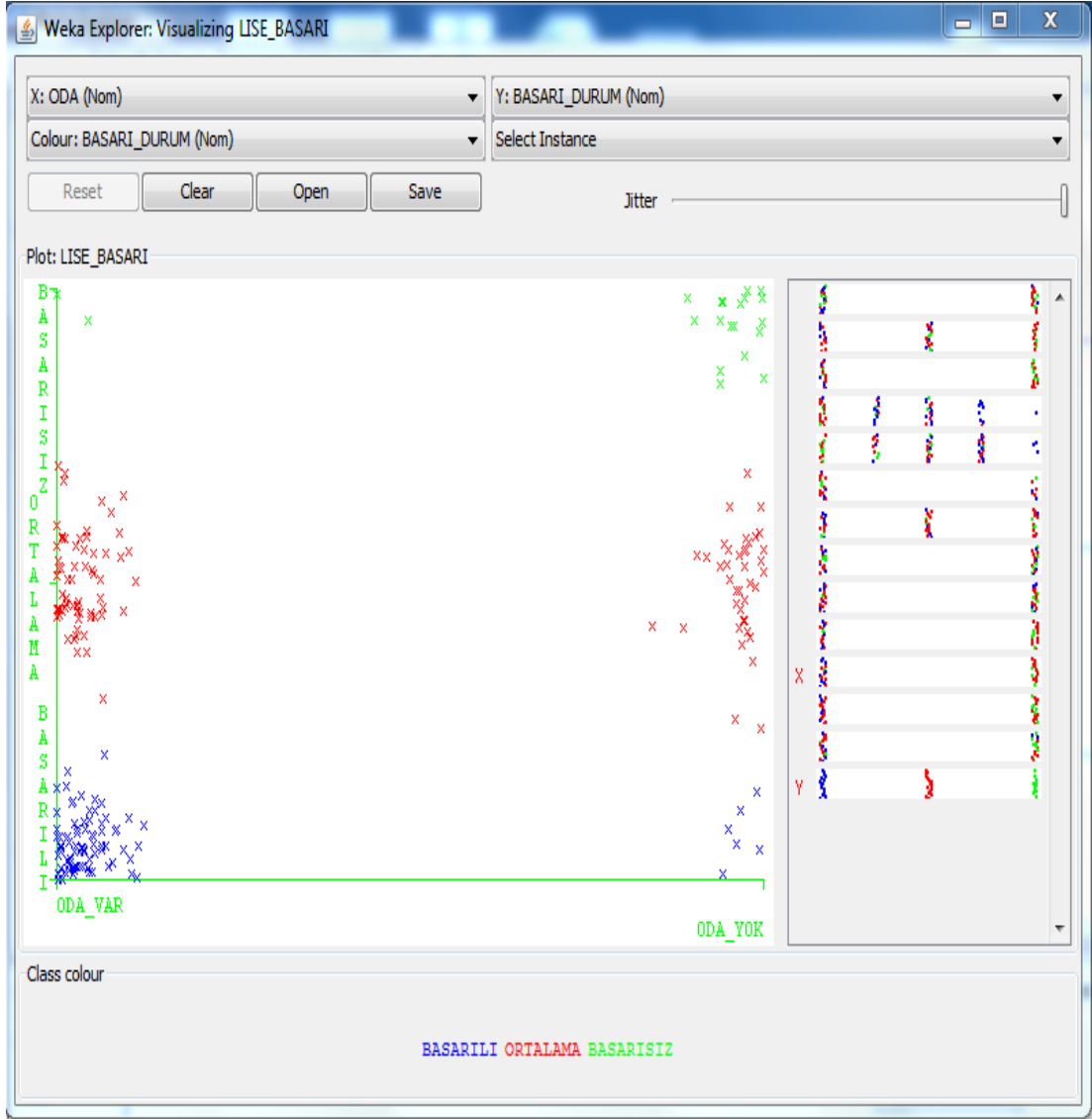
Şekil 3.11’de ekonomik durum ile başarı durumu arasındaki ilişki grafik raporlanmıştır. Bu rapora göre ekonomik durumu iyi olan öğrencilerin başarı durumlarının başarılı ve ortalama düzeyde olduğu, ekonomik durumu kötü olan öğrencilerin ise başarı durumlarının ortalama ve başarısız düzeyde olduğu görülmüştür.





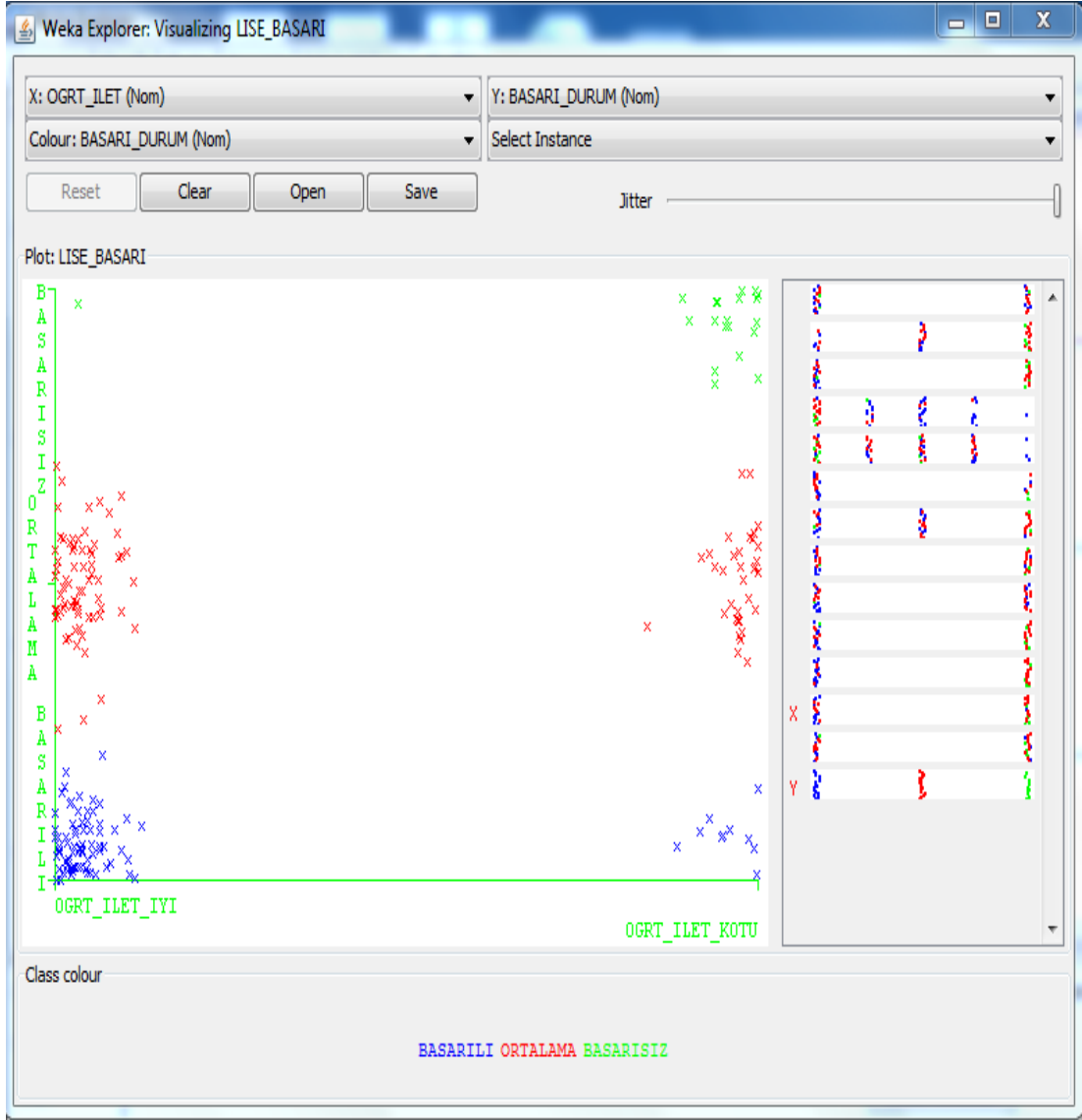
**Şekil 3.11.** Ekonomik Duruma Göre Başarı Durumu

Şekil 3.12’de öğrencinin kendisine ait odası olup olmaması ile başarı durumu arasındaki ilişkisel grafik raporlanmıştır. Bu rapora göre kendisine ait odası olan öğrencilerin başarı durumları başarılı ve ortalama düzeyde iken kendisine ait odası olmayan öğrencilerin başarısız ve ortalama düzeyde oldukları görülmüştür.



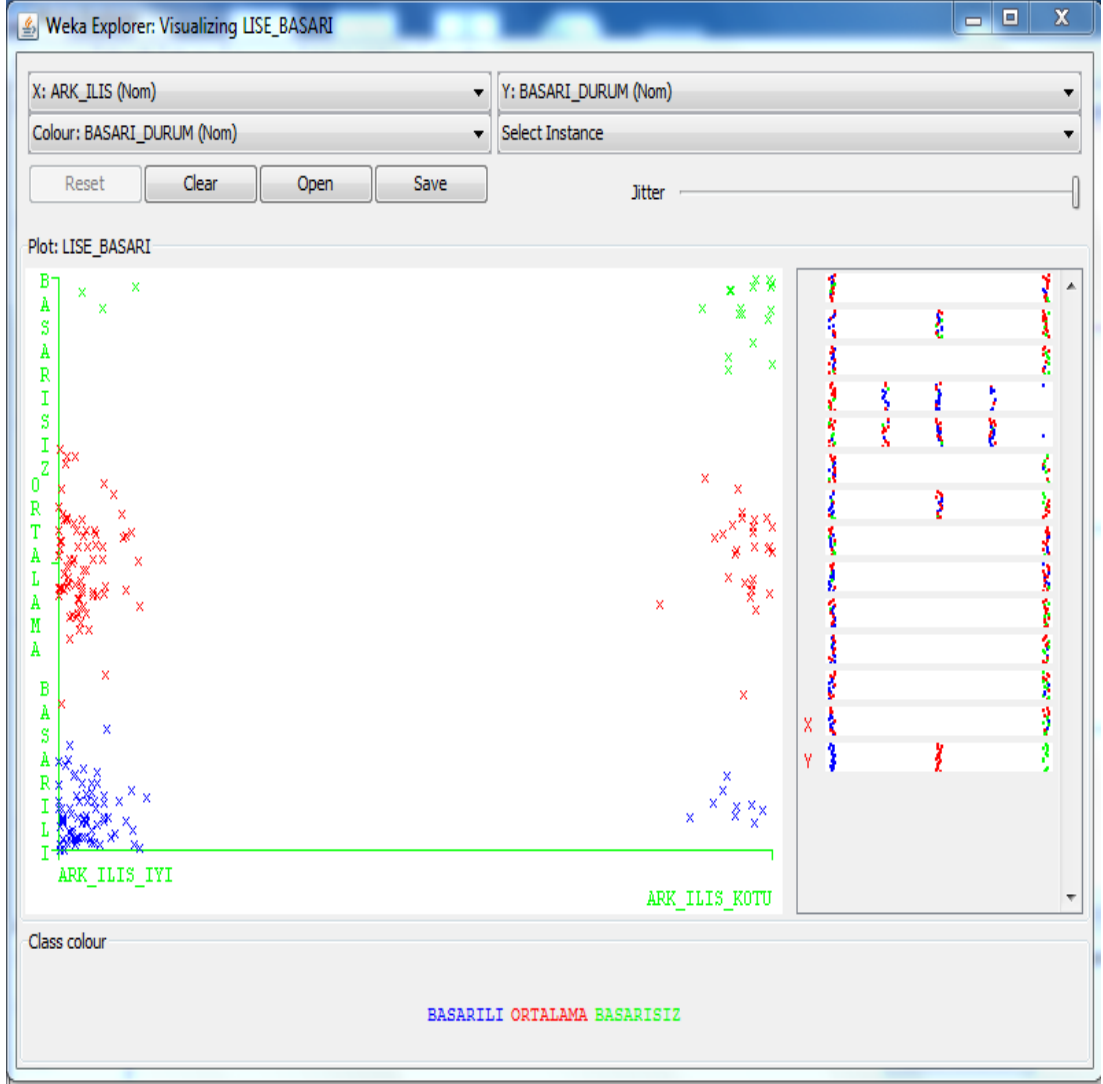
**Şekil 3.12.** Öğrencinin Kendi Odası Olup Olmamasına Göre Başarı Durumu

Şekil 3.13’de öğrencinin öğretmenleriyle iletişimi ile başarı durumu arasındaki ilişkiyel grafik raporlanmıştır. Bu rapora göre öğretmenleriyle iletişimi iyi olan öğrencilerin başarı durumları başarılı ve ortalama düzeyde iken öğretmenleriyle iletişimi kötü olan öğrencilerin başarı durumları daha çok ortalama ve başarısız düzeydedir.



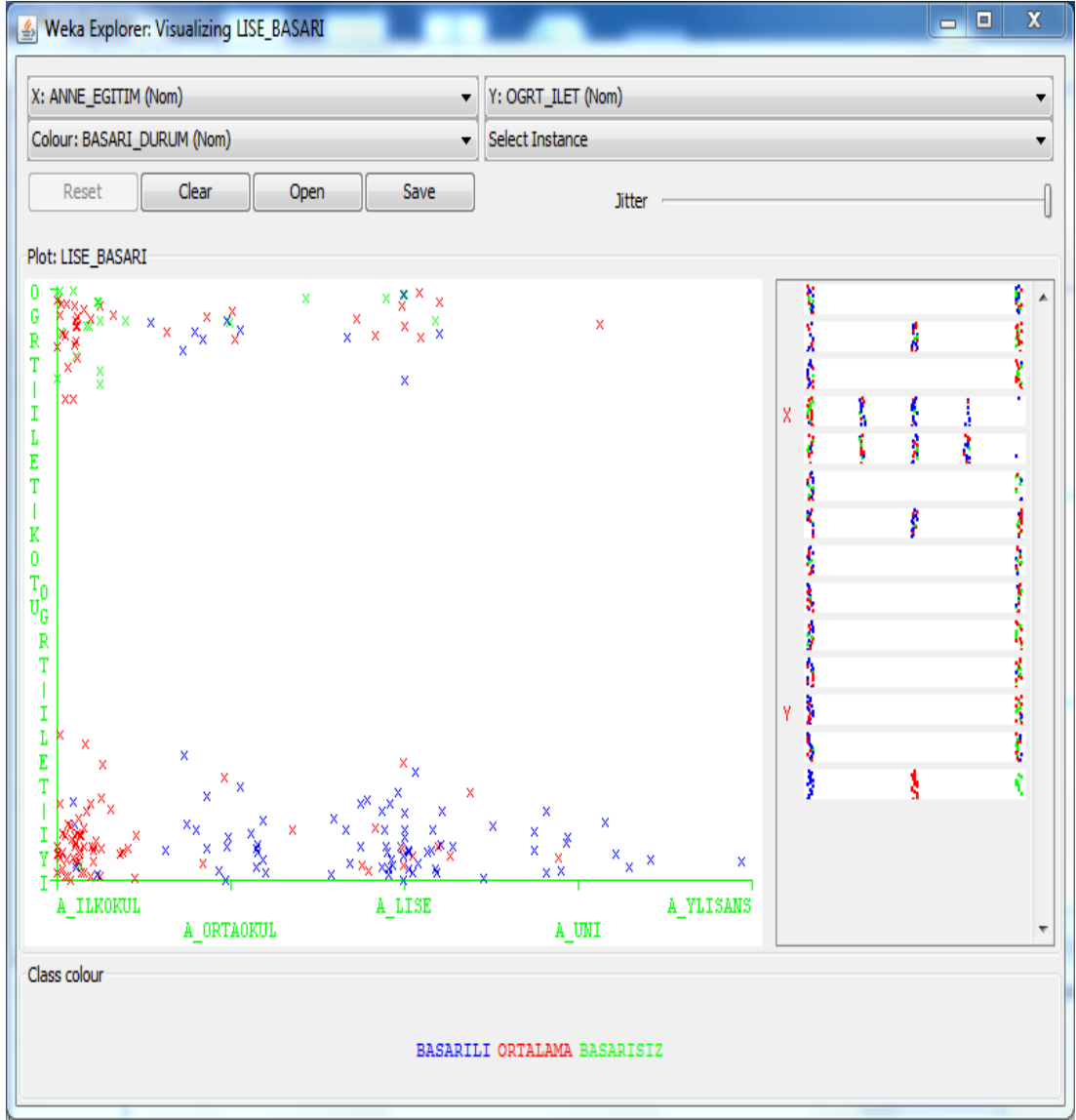
**Şekil 3.13.** Öğrencinin Öğretmenlerle Olan İletişimine Göre Başarı Durumu

Şekil 3.14'de öğrencinin arkadaşlarıyla olan iletişimi ile başarı durumu arasındaki ilişkiyel grafik raporlanmıştır. Bu rapora göre arkadaşlarıyla iletişimi iyi olan öğrencilerin başarı durumları başarılı ve ortalama düzeyde iken öğretmenleriyle iletişimi kötü olan öğrencilerin başarı durumları daha çok ortalama ve başarısız düzeydedir.



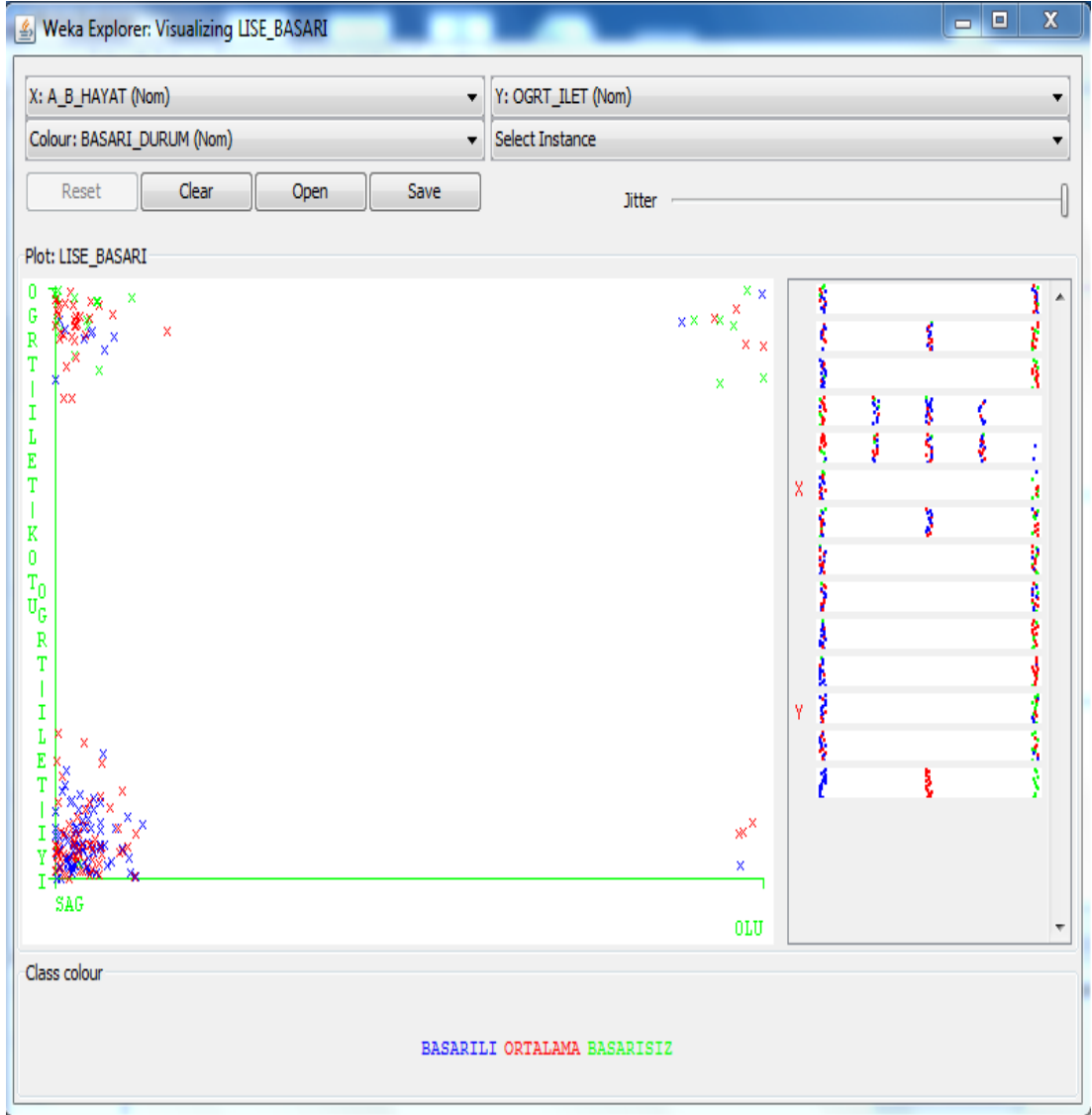
**Şekil 3.14.** Öğrencinin Arkadaşlarıyla Olan İletişimine Göre Başarı Durumu

Şekil 3.15’de öğrencinin annesinin eğitim durumu ile öğretmenleriyle olan iletişimi arasındaki ilişkiyel grafik raporlanmıştır. Bu rapora göre anne eğitim durumu lise, üniversite ve yüksek lisans düzeyinde olan öğrencilerin öğretmenleriyle olan iletişiminin iyi düzeyde olduğu görülmüştür. Bu rapora göre anne eğitim düzeyinin öğrencinin öğretmenleriyle olan iletişimde olumlu yönde etkiye sahip olduğu görülmektedir.



**Şekil 3.15.** Anne Eğitim Düzeyi İle Öğrencinin Öğretmenlerle Olan İletişimi

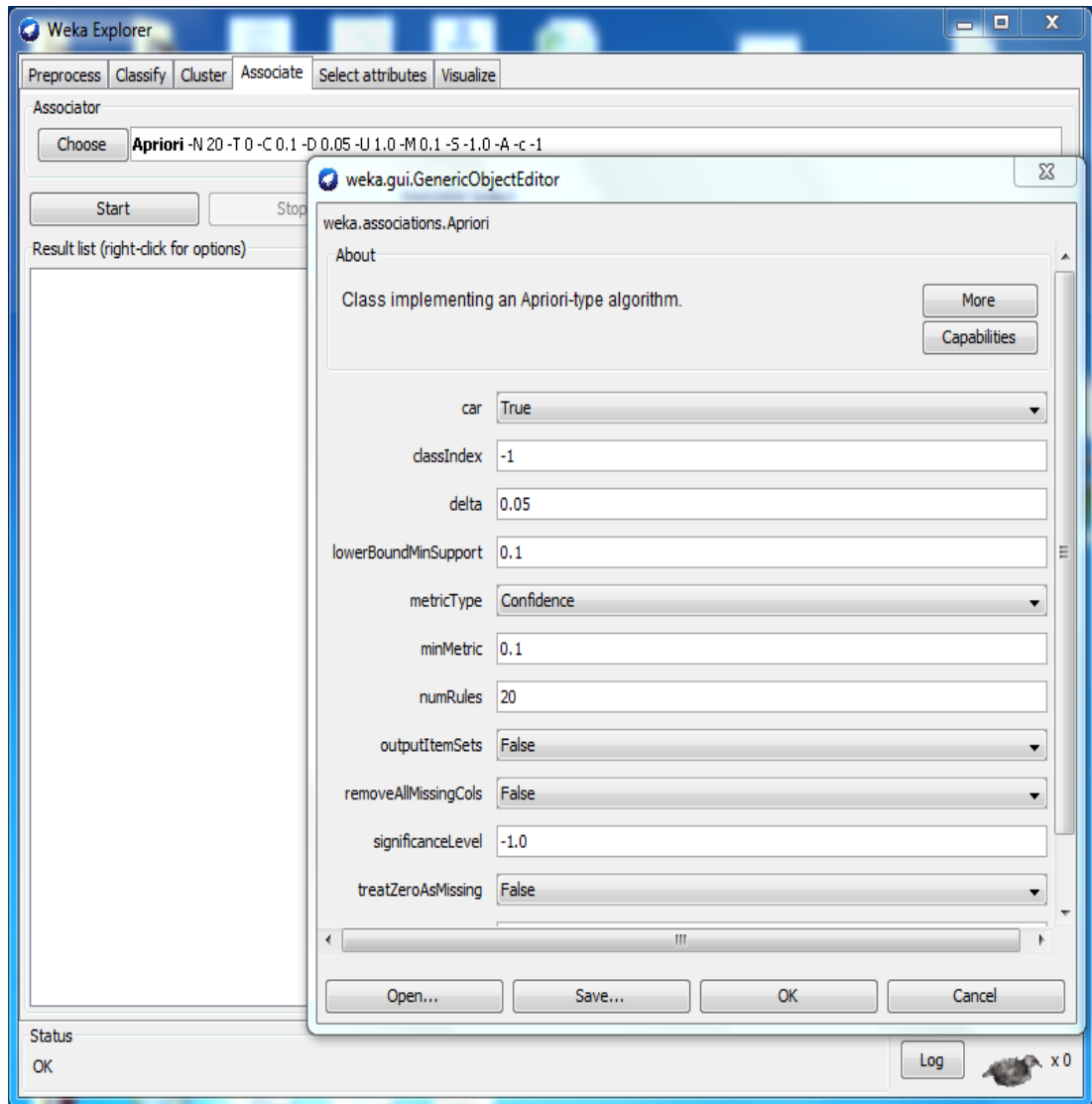
Şekil 3.16'de anne baba hayatta olma durumu ile öğrencinin öğretmenleriyle olan iletişimi arasındaki ilişkiyel grafik raporlanmıştır. Anne ve babası hayatta olan öğrencilerin öğretmenleriyle olan iletişiminin iyi olduğu, anne veya babası hayatta olmayan öğrencilerin ise öğretmenleriyle olan iletişimin kötü olduğu görülmüştür.



**Şekil 3.16.** Ailenin Hayatta Olma Durumu İle Öğrencinin Öğretmenleriyle İletişimi

### 3.5. Birliktelik Kuralları (Associate) ile Elde Edilen Kurallar

Şekil 3.17'deki gösterildiği gibi Birliktelik kuralları ile Apriori yöntemini tercih edip kural sayısını belirledikten sonra elde edilen sonuçlar şekil 3.18'de verilmiştir.



Şekil 3.17. Weka Birliktelik Kuralları Paneli

Best rules found:

1. ANNE\_EGITIM=A\_ILKOKUL A\_B\_HAYAT=SAG 87 ==> BASARI\_DURUM=ORTALAMA 74 conf:(0.85)
2. ANNE\_EGITIM=A\_ILKOKUL 99 ==> BASARI\_DURUM=ORTALAMA 80 conf:(0.81)
3. SAGLIKPROB=SAGLIKLI EKONOMIK\_DUR=EKONOMIK\_D\_IYI ODA=ODA\_VAR 115 ==> BASARI\_DURUM=BASARILI 74 conf:(0.64)
4. SAGLIKPROB=SAGLIKLI A\_B\_HAYAT=SAG EKONOMIK\_DUR=EKONOMIK\_D\_IYI 116 ==> BASARI\_DURUM=BASARILI 74 conf:(0.64)
5. SAGLIKPROB=SAGLIKLI EKONOMIK\_DUR=EKONOMIK\_D\_IYI 121 ==> BASARI\_DURUM=BASARILI 77 conf:(0.64)
6. SAGLIKPROB=SAGLIKLI ODA=ODA\_VAR 121 ==> BASARI\_DURUM=BASARILI 74 conf:(0.61)
7. EKONOMIK\_DUR=EKONOMIK\_D\_IYI ARK\_ILIS=ARK\_ILIS\_IYI 125 ==> BASARI\_DURUM=BASARILI 74 conf:(0.59)
8. A\_B\_HAYAT=SAG EKONOMIK\_DUR=EKONOMIK\_D\_IYI ODA=ODA\_VAR 135 ==> BASARI\_DURUM=BASARILI 76 conf:(0.56)
9. EKONOMIK\_DUR=EKONOMIK\_D\_IYI ODA=ODA\_VAR 141 ==> BASARI\_DURUM=BASARILI 79 conf:(0.56)
10. A\_B\_HAYAT=SAG ODA=ODA\_VAR 139 ==> BASARI\_DURUM=BASARILI 76 conf:(0.55)
11. A\_B\_HAYAT=SAG EKONOMIK\_DUR=EKONOMIK\_D\_IYI 148 ==> BASARI\_DURUM=BASARILI 80 conf:(0.54)
12. ODA=ODA\_VAR 147 ==> BASARI\_DURUM=BASARILI 79 conf:(0.54)
13. EKONOMIK\_DUR=EKONOMIK\_D\_IYI 155 ==> BASARI\_DURUM=BASARILI 83 conf:(0.54)
14. A\_B\_HAYAT=SAG OGRT\_ILET=OGRT\_ILET\_IYI 143 ==> BASARI\_DURUM=BASARILI 74 conf:(0.52)
15. SAGLIKPROB=SAGLIKLI A\_B\_HAYAT=SAG 148 ==> BASARI\_DURUM=BASARILI 76 conf:(0.51)
16. OGRT\_ILET=OGRT\_ILET\_IYI 147 ==> BASARI\_DURUM=BASARILI 75 conf:(0.51)
17. A\_B\_HAYAT=SAG 194 ==> BASARI\_DURUM=ORTALAMA 98 conf:(0.51)
18. SAGLIKPROB=SAGLIKLI 158 ==> BASARI\_DURUM=BASARILI 79 conf:(0.5)
19. A\_B\_HAYAT=SAG ARK\_ILIS=ARK\_ILIS\_IYI 148 ==> BASARI\_DURUM=BASARILI 74 conf:(0.5)
20. ARK\_ILIS=ARK\_ILIS\_IYI 156 ==> BASARI\_DURUM=ORTALAMA 77 conf:(0.49)

**Şekil 3.18.** Weka İle Elde Edilen Birliktelik Kuralları



#### 4. TARTIŞMA VE SONUÇ

Günümüzde inanılmaz oranda artan veri miktarından dolayı istenen bilgiye kolayca ulaşabilme ve veriden anlamlı sonuçlar elde edebilme zorlaşmıştır. Bunun daha kolay gerçekleştirilmesi için farklı yazılımlardan faydalanılmaktadır. Bu çalışmada, bu yazılımlardan biri olan WEKA yazılımı tanıtılmış ve Kırıkkale ili Anadolu Liselerinde okuyan 11. Sınıf öğrencilerinin performanslarının değerlendirilmesi ve anlamlı sonuçlar alınması hedeflenmiştir.

İlk olarak eğitim alanında yapılmış çalışmalarla ilgili bir literatür taraması yapılmış sonra Kırıkkale ili Anadolu Liselerinde okuyan öğrencilere uygulanmak üzere öğrenci başarısına etki edebilecek etkenlerin ve başarı notlarının belirlenebileceği bir anket hazırlanmıştır.

İkinci aşamada veri madenciliği üzerinde durulmuştur. Veri madenciliği, gizli, önemli, önceden bilinmeyen, yararlı bilgileri ortaya koyan bir veri analiz tekniğidir. Bu yöntemle, alışlagelmiş analiz tekniklerinden farklı olarak, yalnızca sayısal verilerle değil, sayısal olmayan verilerle de analizler yapılabilmekte ve gizli örüntüler ortaya çıkarılabilmektedir. Veri madenciliği analizinde kullanılan algoritmaları destekleyen pek çok program geliştirilmiştir. Bu alanda yaygın olarak kullanılan programlardan biri de WEKA'dır. WEKA hemen hemen bütün veri madenciliği yöntemlerini desteklemekte ve içerisinde bulundurduğu birçok algoritma sayesinde verilerden bilgi çıkarımı yapılmasına olanak sağlamaktadır. WEKA görsel olarak da yararlı sonuçlar üretebilmektedir. Diğer yazılımlara göre kolay elde edilebilir ve açık kaynak kodlu olması sebebiyle kullanıcı sayısı oldukça fazladır.

Tezin son bölümünde ise Kırıkkale ili Anadolu Lisesi öğrencilerine uygulanan anket sonucu elde edilen bilgiler çeşitli veri madenciliği yöntemleri ile incelenmiş ve karşılaştırma yapılmıştır. En iyi sonucu veren algoritmalar tespit edilmiştir.

Dünyadaki teknolojik gelişmelere paralel olarak ülkemizde de veri madenciliği uygulamaları gittikçe artmaktadır. Veri madenciliğinin eğitim alanındaki kullanımını

artırarak, etkili eğitim politikalarının geliştirilebilmesi, eğitim ve öğretim için kullanılacak materyallerin tespitinin yapılması, müfredat dışı öğrenme aktivitelerinin verimliliği sınanarak öğrencilerin bu aktivitelere teşvik edilmesi gibi çalışmalarla eğitim ve öğretim süreçlerine katkı sağlanabileceği ve verimliliğin artırılacağı öngörülmektedir.

Bu çalışmada öğrencinin başarısına etki eden faktörler hep birlikte ve ayrı ayrı ele alınıp kıyaslamalar yapılmıştır. Öğrencinin başarılı olmasına etki eden faktörlerin desteklenmesi, başarısızlığa neden olan faktörlerin ise çözümlenmesi hedeflenmiştir. Çalışma kapsamında Kırıkkale ilinde bulunan farklı Anadolu Liselerinde okuyan öğrencilere uygulanan anket sonucu elde edilen veriler kullanılmıştır. Uygulama WEKA 3.7. programı kullanılarak gerçekleştirilmiştir.

Genel olarak sonuçlar değerlendirilip öğrencilerin başarı durumları incelendiğinde;

Tek çocuk olan öğrencilerin,

Sağlık problemi olmayan öğrencilerin,

Anne eğitim durumu lise ve üzeri seviyede olan öğrencilerin,

Ekonomik durumu iyi olan öğrencilerin,

Kendisine ait odası olan öğrencilerin,

Öğretmenlerle iletişimi iyi olan öğrencilerin,

Arkadaşlarıyla iletişimi iyi olan öğrencilerin daha başarılı oldukları tespit edilmiştir.

Başarıya etki eden faktörlerin birbiri ile olan ilişkisi araştırıldığında etkileşim oranı en yüksek olan ifadeler şu şekildedir;

Öğrencinin öğretmenlerle ve arkadaşlarıyla olan iletişiminin iyi olması, kendisine ait çalışma odasının olması arasında başarıyı artıran bir ilişki bulunmaktadır.

Öğrencinin kardeş sayısının az olması, anne eğitim durumunun üniversite ve üstü olması, ailenin aktivitelere destek vermesi arasında başarıyı artıran bir ilişki bulunmaktadır.

Yapılan incelemede çok katmanlı algılayıcı algoritması ile yapılan sınıflama sonucunda %88,73, JRip algoritması ile yapılan sınıflama sonucunda % 87,32, k- En Yakın Komşu algoritması ile yapılan sınıflama sonucunda % 85,91, J48 algoritması ile yapılan sınıflama sonucunda % 84,50, Saf Bayes algoritması ile yapılan sınıflama sonucunda ise % 81,69 başarı sağlanmıştır.

Çalışma neticesinde elde edilen sonuçlara göre, ailenin eğitim seviyesi ve ekonomik durumu öğrenci başarısına yüksek oranda etki etmektedir. Ancak bu kriterlerin yanı sıra öğrenci başarısına etki eden pek çok etken vardır. Ailenin eğitim seviyesi ilkokul, ortaokul düzeyinde ve ekonomik durumu kötü olsa dahi öğrencinin ders dışı aktivitelere katılımı desteklendiğinde, öğrenciye uygun çalışma ortamı sağlandığında ve ailenin öğretmenlerle iletişimde olduğu durumlarda öğrenci başarısının arttığı gözlemlenmiştir. Başarısızlığa neden olan değiştirilemez faktörlerin etkisinin, başarıya katkı sağlayacak diğer faktörler üzerinde iyileştirilmeye gidildiğinde önemli oranda yok edilebileceği gözlemlenmiştir.

Bugüne kadar yapılan çalışmalara ek olarak incelenen faktörlerin başarıya etkisinin yanı sıra, veri madenciliği programı olan WEKA kullanılarak bu kriterlerin birbiri üzerindeki etkisi birden fazla boyutta incelenerek başarısızlığa neden olan faktörlerin çözümlenmesi için bilgiler paylaşılmıştır.

Yapılan çalışmada aynı ilde bulunan farklı Anadolu Liselerinde okuyan öğrencilere ait veriler kullanılmıştır. Öğrencilerin bu okullara yerleştirildikleri yılın taban ve tavan puanları incelendiğinde farklılıklar gözlemlenmiştir. Ancak öğrenciler aynı türdeki okullarda okuduklarından ortak dersler aldıkları düşünülürse yerleştirme puanlarında gözlemlenen farklılık, bulunulan dönemdeki başarı ortalamaları karşılaştırıldığında gözlemlenmemiştir. Bu çalışma geliştirilerek öğrencilerin üniversitelere yerleştirme sınavında aldıkları puanlar da çalışmaya dahil edilerek liselere yerleştirilme sınavında alınan puanın üniversiteye yerleştirme sınavında alınacak olan puanın tahmin edilmesinde ne derece etkili olacağı incelenebilir. Bu çalışmada bulunan sonuçlar da işe koşularak öğrencilerin üniversitelere yerleştirilme sınavında başarıyı yakalama oranı artırılabilir.

## KAYNAKLAR

- [1] Şimşek, A., Becit, G., Kılıçer, K., Özdamar, N., Akbulut, Y. ve Yıldırım, Y., Türkiye'deki Eğitim Teknolojisi Araştırmalarında Güncel Eğilimler, Anadolu Üniversitesi, Eskişehir, 2007.
- [2] Anonim, <http://www.slideshare.net/MuratAzimli/murat-azimli-veri-madenciligi-data-mining> (Erişim Tarihi:14.10.2014)
- [3] Kaya, H. ve Köymen, K., Veri Madenciliği Kavramı ve Uygulama Alanları, Doğu Anadolu Bölgesi Araştırmaları Dergisi, 2008.
- [4] Aydın, S., Veri Madenciliği ve Anadolu Üniversitesi Uzaktan Eğitim Sisteminde Bir Uygulama, Doktora Tezi, Anadolu Üniversitesi Sosyal Bilimleri Enstitüsü, Sayısal Yöntemler Anabilim Dalı, 2007.
- [5] İnan, O., Veri Madenciliği, Yüksek Lisans Tezi, Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, 2003.
- [6] Karabatak, M. ve İnce, M. C., Apriori Algoritması ile Öğrenci Başarısının Analizi, ELOCO International Conference, Bursa, 2004.
- [7] Erdoğan, Ş. Z. ve Timor, M., A Data Mining Application in a Student Database, Journale of Aeronautics and Space Technologies, 2(2) : 53-57, 2005.
- [8] Özçınar, H., KPSS Sonuçlarının Veri Madenciliği Yöntemleriyle Tahmin Edilmesi, Yüksek Lisans Tezi, Pamukkale Üniversitesi, Fen Bilimleri Enstitüsü, 2006.
- [9] Çiftci, S., Uzaktan Eğitimde Öğrencilerin Ders Çalışma Etkinliklerinin Log Verilerinin Analiz Edilerek İncelenmesi, Yüksek Lisans Tezi, Gazi Üniversitesi, Eğitim Bilimleri Enstitüsü, 2006.

- [10] Ayık, Y. Z., Özdemir, A. ve Yavuz, U., Lise Türü ve Lise Mezuniyet Başarısının, Kazanılan Fakülte İle İlişkisinin Veri Madenciliği Tekniği İle Analizi, Sosyal Bilimler Enstitüsü Dergisi, 10(2): 441-454, 2007.
- [11] Kayri, M., Elektronik Portfolyo Değerlendirmeleri İçin Veri Madenciliği Yaklaşımı, Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi, 5(1): 98-110, 2008.
- [12] Bozkır, A. S., Sezer, E. ve Gök, B., Öğrenci Seçme Sınavında (ÖSS) Öğrenci Başarımını Etkileyen Faktörlerin Veri Madenciliği Yöntemleriyle Tespiti, 5. Uluslararası İleri Teknolojiler Sempozyumu (IATS'09), 13-15 Mayıs, Karabük Üniversitesi, Karabük, 37-43, 2009.
- [13] Savaş, S., Arıcı, N., Web Tabanlı Uzaktan Eğitimde İki Farklı Öğretim Modelinin Öğrenci Başarısı Üzerindeki Etkilerinin İncelenmesi, 5. Uluslararası İleri Teknolojiler Sempozyumu (IATS'09), Karabük Üniversitesi, Karabük, 1229, 2009.
- [14] Kurt, Ç., Erdem, O. A., Öğrenci Başarısını Etkileyen Faktörlerin Veri Madenciliği Yöntemleriyle İncelenmesi, Politeknik Dergisi, 15(2): 111-116, 2012.
- [15] Barışçı, S. N., Öğrenci Performansının Veri Madenciliği ile Belirlenmesi, ISITES'2014, Karabük Üniversitesi, 1958-1965, 2014.
- [16] Yurdakul, S., Topal, T., Veri Madenciliği ile Lise Öğrenci Performanslarının Değerlendirilmesi, AB 2015 Akademik Bilişim Konferansı, Anadolu Üniversitesi, Şubat 2015.
- [17] Gürsoy, Ş., Veri Madenciliği ve Bilgi Keşfi, 2009.
- [18] Akpınar, H., Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği, İstanbul Üniversitesi İşletme Fakültesi Dergisi, Nisan 2000.

- [19] Özenç, Ç., SQL Server, Veritabanı Yedekleme Stratejileri, 2013.
- [20] Zafer, H. R., Veritabanı ve SQL, 2009.
- [21] Usgurlu, Ü. B., Veri Ambarı ve Veri Madenciliği, Başkent Üniversitesi, 2011.
- [22] Han, J., Kamber, M., Data Mining, Concepts and Techniques, Morgan Kaufmann, Multiscience Press, San Francisco, 2006.
- [23] Azimli, M., Özdemir, S., Tıpta Veri Madenciliği Uygulamaları, Yüksek Lisans Tezi, Gazi Üniversitesi, Bilişim Enstitüsü, 2014.
- [24] Akyokuş, S., Veri Madenciliği Yöntemlerine Genel Bakış, 2006.
- [25] Özkan, Y., Veri Madenciliği Yöntemleri, Papatya Yayınları, İstanbul, 2008.
- [26] Anonim, Kocabaş, K., Veri Madenciliği, <http://www.misjournal.com/?p=738> (Erişim Tarihi:17.10.2014)
- [27] Özmen, Ş., İş Hayatı Veri Madenciliği İle İstatistik Uygulamaları, V. Ulusal Ekonometri ve İstatistik Sempozyumu, Adana, 2001.
- [28] Larose, Daniel, T., Data Mining Methods and Models, A John Wiley & Sons, Inc, Publication, New Jersey, 2006.
- [29] Lior, R., Oded, M., Data Mining With Decision Trees, World Scientific, New Jersey, 2008.
- [30] Larose, Daniel, T., Discovering Knowledge in Data, A John Wiley & Sons, Inc, Publication, New Jersey, 2005.

- [31] Genesh, S., "Data Mining: Should it be included in the 'Statistics' Curriculum", The Sixt International Conference on Teaching Statistics, Cape Town, South Africa, 2002.
- [32] Oğuzlar, A., Veri Madenciliğine Giriş, Ekin Kitabevi, Bursa, 2004.
- [33] Anonim, "Data Mining and Statistics: What is the Connection?", <http://www.tdan.com/view-articles/5226/>, (Erişim Tarihi: 17.10.2014), 2004
- [34] Şentürk, A., Veri Madenciliği Kavram ve Teknikler, Ekin Kitabevi, Bursa, 2006.
- [35] Koyuncugil, A. S., Veri Madenciliği ve Sermaye Piyasalarına Uygulaması, Sermaye Piyasası Kurulu Araştırma Raporu, Araştırma Dergisi, 2007.
- [36] Yıldırım, P., Uludağ, M., Görür, A., Hastane Bilgi Sistemlerinde Veri Madenciliği, Akademik Bilişim Kongresi, Çanakkale Onsekiz Mart Üniversitesi, Çanakkale, 2007.
- [37] Sholom, M., Tong, Z., Performance Analysis And Evaluation, The Handbook of Data Mining, Edited by Nong Ye Arizona State University, Lawrence Erlbaum Associates, Mahwah, 2003.
- [38] Savaş, S., Topaloğlu, N., Yılmaz, M., Veri Madenciliği ve Türkiye'deki Uygulama Örnekleri, İstanbul Ticaret Üniversitesi, Fen Bilimleri Dergisi, 11(21): 1-23, 2012.
- [39] Tekerek, A., Veri Madenciliği Süreçleri ve Açık Kaynak Kodlu Veri Madenciliği Araçları, Gazi Üniversitesi, Elektronik-Bilgisayar Eğitimi Bölümü, Ankara, 2011.

- [40] Küçüksille, E., Veri Madenciliği Süreci Kullanılarak Portföy Performansının Değerlendirilmesi ve İMKB Hisse Senetleri Piyasasında Bir Uygulama, Doktora Tezi, Süleyman Demirel Üniversitesi, Sosyal Bilimler Enstitüsü, İşletme Anabilim Dalı, Isparta, 2009.
- [41] Özmen, U., Verileri Anlamak, Anonim, <http://ugurozmen.com/crm/verileri-anlamak>, (Erişim Tarihi: 21.10.2014)
- [42] Shearer, C., The Crisp-DM Model: The New Blueprint for Data Mining, Journal of Data Warehousing, 5(4): 13-23, 2000.
- [43] Şimşek, U. T. G., Veri Madenciliği ve Bilgi Keşfi, Pegem Akademi, Ankara, 2009.
- [44] Piramuthu, S., Evaluating Feature Selection Methods For Learning in Data Mining Applications, Thirty-First Annual Hawaii International Conference on System Sciences, IEEE Computer Society, 6-9 January, Kohala Coast Hawaii USA, 294, 1998.
- [45] Methodologies for Knowledge Discovery and Data Mining: Third Pacific-Asia Conference, Pakdd-99, Beijing, China, April 26-28, 1999.
- [46] Zhong, N., Zhou, L., Proceedings, Springer Verlag, 1999.
- [47] Yılmaz, E., Kütahya İlinde Sosyal Sınıfların Belirlenmesi ve Veri Madenciliği İle Tüketici Profiline Çıkarılmasına Yönelik Bir Uygulama, Dumlupınar Üniversitesi, Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, 91, 2006.
- [48] Argüden, Y., Erşahin, B., Veri Madenciliği, Arge Danışmanlık, 2008.
- [49] Berry, M.J.A., Linoff, G.S., Mastering Data Mining: The Art and Science of Customer Relationship Management, John Wiley Sons, 1st Ed. 1999.



- [50] Quinlan, J. R., Improved Use of Continuous Attributes in C4.5, Basser Department of Computer Science University of Sydney, Sydney Australia, 2006.
- [51] Yaralıođlu, K., Veri Madenciliđi, Anonim, [http://www.deu.edu.tr/userweb/k.yaralioglu/dosyalar/ver\\_mad.doc](http://www.deu.edu.tr/userweb/k.yaralioglu/dosyalar/ver_mad.doc) (Eriřim Tarihi: 26.10.2014)
- [52] Erdem, O. A., Uzun, E., Yapay Sinir Ađları ile Trke Times New Roman, Arial ve Elyazısı Karakterleri Tanıma, Gazi niversitesi Mhendislik Mimarlık Fakltesi Dergisi, 20(1):13, 2005.
- [53] Engin, O., Fıđlalı, A., Akıř tipi izelgeleme Problemlerinin Genetik Algoritma Yardımı ile özmnde Uygun Operatrnn Belirlenmesi, Dođus niveritesi, 28, 2006.
- [54] Harypis, G., Han, E., Kumar, V., Hierarchical Clustering Using Dynamic Modeling, IEEE Computer, 69-75, 1999.
- [55] Ramkumar, G. D., Swami, A., Clustering Data Without Distance Functions, IEEE Bulletin of the Technical Committee on Data Engineering, 21(1):9-14, 1998.
- [56] Seidman, C., Data Mining with Microsoft SQL Server 2000, Microsoft Press, 1 st Ed.; Washington, USA, 2001.
- [57] zekes, S., Veri Madenciliđi Modelleri ve Uygulama Alanları, İstanbul Ticaret niversitesi Dergisi, 76, 2003.
- [58] Akal, Z., İřletmelerde Performans lm ve Denetimi, ok ynl performans Gstergeleri, MPM Yayınları, Ankara, 2005.
- [59] Aslan, G., T., Cavit, Yeřilyurt, H., Sađlık Sektrnde Veri Zarflama Analiz Yntemi ile Greceli Verimlilik Analizi, Sekin Yayıncılık, Ankara, 2004.

- [60] Baysal, M. E., Veri Zarflama Analizi ile Orta Öğretimde Performans Ölçümü, Yayınlanmamış Yüksek Lisans Tezi, Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Ankara, 1999.
- [61] Erden, M., Öğretmenlik Mesleğine Giriş, Alkim Yayınları, İstanbul, 1996.
- [62] Silah, M., Üniversite Öğrencilerinin Akademik Başarılarını Etkileyen Çeşitli Nedenler Arasından Süreksiz Durumluk Kaygısının Yeri ve Önemi, Eğitim Araştırmaları Dergisi, 10: 102-115, 2003.
- [63] Kadı, Z., Adana Dil Merkezi İlköğretim Okulu Öğrencilerinin Sürekli Devamsızlık Nedenleri, Yayınlanmamış Yüksek Lisans Tezi, İnönü Üniversitesi, Sosyal Bilimler Enstitüsü, Malatya, 2000.
- [64] Balcı, A., Etkili Okul ve Okul Geliştirme, PegemA Yayınları, 2002.
- [65] Ataman, A., Sınıf İçinde Karşılaşılan Davranış Problemleri ve Bunlara Karşı Geliştirilen Önlemler, Nobel Yayınları, Ankara, 2001.
- [66] Arkonaç, S., Sosyal Psikoloji, Alfa Yayınları, İstanbul, 2001.
- [67] Dener, M., Açık kaynak kodlu veri madenciliği programları: WEKA’da örnek uygulama, 787-796. Akademik Bilişim’09 - XI. Akademik Bilişim Konferansı Bildirileri, Harran Üniversitesi, Şanlıurfa, Şubat 11-13, 2009.
- [68] Alfred, R., Knowledge Discovery: Enhancing Data Mining and Decision Support Integration, The University of York, United Kingdom, 2005.
- [69] Göktepe, A.B., Agar, E., Lav, A.H., Comparison of Multilayer Perceptron and Adaptive Neuro-Fuzzy System on Backcalculating the Mechanical Properties of Flexible Pavements. ARI The Bulletin of the Istanbul Technical University, 54(3), 65-77, 2004.

- [70] Haykin, S., "Neural Networks and Learning Machines", PHI Learning Private Limited, 2010.
- [71] Anonim, <http://kent.dl.sourceforge.net/sourceforge/weka/weka-3-4-5jre.exe>, (Erişim Tarihi: 16.11.2014)
- [72] Aha, D.W., Kibler, D., Albert, M.K., Instance-based learning algorithms. *Machine Learning*, 6, 37-66, 1991.
- [73] Tuncer, T. ve Tatar, Y., Karar Ağacı Kullanılarak Saldırı Tespit Sistemlerinin Performans Değerlendirmesi, 4. İletişim Teknolojileri Ulusal Sempozyumu, Adana, 2009.
- [74] Çöllüoğlu, Ö. ve Özdemir, S., Analysis of Gifted Students' Interest Areas Using Data Mining Techniques, *Journal of Gifted Education Research*, 1 (3): 213-226, 2013.
- [75] Sasaki, M. ve Kıtı, K., Rule-Based Text Categorization Using Hierarchical Categories, 1998 IEEE International Conference on Systems, Man, and Cybernetics, San Diego, CA, 1998.
- [76] Wang, J., (Editor), "Encyclopedia of Data Warehousing and Mining", Information Science Reference, 49, 140, 2006.

## EKLER

### EK 1. ANKET UYGULAMASI İÇİN VALİLİK ONAYI

T.C.  
KIRIKKALE VALİLİĞİ  
Milli Eğitim Müdürlüğü

Sayı : 12774561-044- 3619  
Konu : Anket Çalışması

28 Şubat 2013

#### VALİLİK MAKAMINA

İlgi : a) 28.02.2007 tarih ve B.08.0.EGD.0.33.05.311-311/1084 sayılı Makam Onayı ile uygulamaya konulan "Milli Eğitim Bakanlığına Bağlı Okul ve Kurumlarda Yapılacak Araştırma ve Araştırma Desteğine Yönelik İzin ve Uygulama Yönergesi."  
b) Kırıkkale Üniversitesi'nin 07/02/2013 tarih ve 122 sayılı yazısı.

İlgi (a) yönerge kapsamında; araştırma bir ili kapsıyorsa izin işlemlerinin ilgili İl Millî Eğitim Müdürlüğünce, birden çok ili kapsıyorsa Bakanlığımız Eğitimi Araştırma ve Geliştirme Dairesi Başkanlığınca sonuçlandırılması hükme bağlanmıştır.

İlgi (b) yazı ile; Kırıkkale Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı Yüksek Lisans Programı öğrencilerinden Semra YURDAKUL'un; "Veri Madenciliği ile Lise Öğrenci Performanslarının Değerlendirilmesi" konulu tez çalışması gereği; İlimiz Merkezinde bulunan Anadolu Liselerinin 11. Sınıf öğrencilerine yönelik "Öğrenci Tanıma Anketi" izin başvurusunda belirttiği sorularla ilgili 25/02/2013-15/03/2013 tarihleri arasında anket çalışmasını yapabilmesi talep edilmektedir.

İlgi (a) yönerge doğrultusunda oluşturulan Araştırma Değerlendirme Komisyonu tarafından düzenlenen Araştırma Değerlendirme Formunda adı geçen Kırıkkale Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı Yüksek Lisans Programı öğrencilerinden Semra YURDAKUL'un yukarıda belirtilen çalışmasının Müdürlüğümüze bağlı İlimiz Merkezinde bulunan Anadolu Liselerinin 11. Sınıf öğrencilerine yönelik, veli izin belgeleri alınarak tüm sorumluluğun okul müdürlerine ait olması kaydıyla ve eğitim-öğretimi aksatmadan, gönüllülük esasına göre yapılması Müdürlüğümüze uygun görülmektedir.

Makamlarınızca da uygun görüldüğü takdirde olurlarınıza arz ve teklif ederim.

  
H. Ali OKUR  
Milli Eğitim Müdürü V.

OLUR  
19/02/2013  
  
Yusuf Ziya KARACA EV  
Vali'nin  
Vali Yardımcısı

	Cumhuriyet Meydanı 71100 KIRIKKALE Tel : (0318) 224 61 03-04-07-08 Faks : (0318) 224 25 59	Web: <a href="http://kirikkale.meb.gov.tr">http://kirikkale.meb.gov.tr</a> e-posta: <a href="mailto:kirikkalemem@meb.gov.tr">kirikkalemem@meb.gov.tr</a>	<b>EGITIME</b> <b>%100</b> <b>DESTEK</b>	<b>Daha aydınlık</b> <b>gelecek!</b>
---	---	---	--	---

## EK 2. ANKET ÖRNEĞİ

### ÖĞRENCİ TANIMA ANKETİ

Doğum Yılı:

Cinsiyeti:

Kardeş sayısı:

- 1) Hangi mesleğe yöneleceğiniz konusunda karar verdiniz mi?  
Evet  
Hayır  
Kararsızım
- 2) Başarılı/Başarısız olmanızdaki etken hangisidir?  
Kişisel etkenler  
Çevresel etkenler
- 3) Sizin için hangisi daha önemlidir?  
Kısa süreli başarılar  
Gelecekle ilgili planlar
- 4) Sorumluluk almaktan kaçınır mısınız?  
Evet  
Hayır
- 5) Kendi ilgi ve yeteneklerinizden haberdar mısınız?  
Evet  
Hayır
- 6) 5.soruda cevabınız evetse...  
İlginizin olduğu alanları yazar mısınız?  
Yetenekli olduğunuz alanları yazar mısınız?
- 7) Çalışma şekliniz aileniz tarafından onaylanıyor mu?  
Evet  
Hayır
- 8) Ailenizin sizin için belirlediği başarı seviyesi size göre ne seviyededir?  
Düşük  
Normal  
Yüksek
- 9) Yazılılardan alacağınız puanları önceden tahmin etmedeki düzeyiniz nedir?  
Beklediğimden düşük alırım  
Beklediğim düzeyde notlar alırım  
Beklediğimin üzerinde notlar alırım
- 10) Bir işi yaparken yönlendirmelere ihtiyaç duyar mısınız?  
Evet, ara ara  
Eve, sık sık  
Hayır

- 11) Başarısız olduğunuz durumlarda sizce başarısızlığın nedeni genelde neye bağlıdır?  
Dışsal faktörlere (örn: Sınavın zor, kötü olmasına)  
İçsel faktörlere (örn: yeteri kadar çalışmamış olmanıza)
- 12) Sizce başarılı olmanızdaki öncelikli etken hangisidir?  
Yetenekleriniz  
Harcadığınız emek ve çabalarınız
- 13) Aileniz çalışmalarınızı sürekli kontrol eden bir yapıda mıdır?  
Evet  
Hayır
- 14) 13.soruya cevabınız evetse, bu kontrol edicilik şekilleri nasıldır?  
Hayatınızda sürekli sınırlılıklar oluşturacak şekilde  
Anlaşılı davranıp doğruları bulmanızı sağlayacak şekilde
- 15) Aileniz sizi ders dışı aktivitelere de katılmanız konusunda destekler mi?  
Evet  
Hayır
- 16) Başarısız olduğunuz durumlarda aileniz tarafından olumsuz eleştiriler alır mısınız?  
Evet  
Hayır
- 17) Derslerdeki başarınız ne olursa olsun aileniz tarafından gerekli desteği alabiliyor musunuz?  
Evet  
Hayır
- 18) Başarısızlıklar karşısında ümitsizliğe kapılmak yerine başarılı olma konusunda kendinizi motive edebiliyor musunuz?  
Evet  
Hayır
- 19) Aşağıda başarıya etki edebilecek unsurlar sıralanmıştır. Sizi ilgilendiren maddelerin yanına ✓ işareti koyunuz.

- Yazılı sınavlarda başarılı olamıyorum
- Sınavlarda heyecanlanmam nedeni ile başarısızım
- Bazı öğretmenlerin anlattıklarını anlayamadığımdan başarısızım
- Ders sayısının fazla olması nedeninden dolayı başarısızım
- Sınıfın kalabalık olmasından dolayı başarısızım
- Bazı derslere karşı ilgimin olmayışından dolayı başarısızım
- Öğretmenlerime kendimi yeterince iyi ifade edemediğimden dolayı başarısızım
- Evde ders çalışacak ortamım olmayışından başarısızım
- Ders çalışmaya yeterince zaman ayıramadığımdan başarısızım
- Öğrenimim sonucunda bir meslek edineceğim düşüncesinde olmadığım için başarısızım
- Ailemin ekonomik problemleri nedeniyle başarısızım
- Ailemin eğitime yeterli değeri vermeyişinden başarısızım

- Ailemin sađlık problemleri nedeniyle başarısızım
- Kardeřim/kardeřlerimle aramda ayırım yapıldığı için başarısızım
- Kardeřimden/kardeřlerimden ayrı olduğumdan başarısızım
- Ailemde kız-erkek ayırımını yapıldığı için başarısızım
- Bazı dersleri başaramayacağım kaygısından dolayı başarısızım
- Sözlü anlatımda kendime güvenim olmadığından dolayı başarısızım
- Verimli ders çalışma yollarını bilmediğimden başarısızım
- Okuduklarımı anlamakta güçlük çektiğimden başarısızım
- İlgi duymadığım derslerden başarısızım
- Dikkatim dađınık olduğu için başarısızım
- Arkadař grubumdan dolayı başarısızım
- Not tutma tekniđini bilmeyiřimden başarısızım
- Arkadařlarımdan kendimi soyutladıđımdan başarısızım
- Öğretmenlerimin önerilerine uymadığım için başarısızım
- Öğrenme isteđim olmadığından başarısızım

20) Herhangi bir sađlık sorunuz var mı?

Evet

Hayır

21) 21.soruda cevabınız evetse, rahatsızlığınızın ne olduğunu yazar mısınız?

22)

	Anne	Baba
Sađ/Ölü		
Yaşı		
Eđitim Durumu		
Mesleđi		

23) Birinci dönem ders notlarınızı derslerin adlarıyla birlikte yazar mısınız?