



**AMASYA ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ**

**UZAKTAN EĞİTİM ÖĞRENCİLERİNİN MEZUNİYET DURUMLARININ VERİ  
MADENCİLİĞİ YÖNTEMLERİ İLE TAHMİNİ:AMASYA ÜNİVERSİTESİ  
ÖRNEĞİ**

**YÜKSEK LİSANS TEZİ**

**OSMAN KAYHAN**

**HAZİRAN**

**OSMAN KAYHAN**

**TEKNOLOJİ VE İNOVASYON  
YÖNETİMİ**

**HAZİRAN 2019**

**UZAKTAN EĐİTİM ÖĐRENCİLERİNİN MEZUNİYET DURUMLARININ VERİ  
MADENCİLİĐİ YÖNTEMLERİ İLE TAHMİNİ : AMASYA ÜNİVERSİTESİ  
ÖRNEĐİ**

**Osman KAYHAN**

**YÜKSEK LİSANS TEZİ  
TEKNOLOĐİ VE İNOVASYON YÖNETİMİ ANABİLİM DALI**

**Danışman**

**Dr. Öğr. Üyesi Yavuz ÜNAL**

**AMASYA ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ**

**HAZİRAN 2019**

Osman KAYHAN tarafından hazırlanan “Uzaktan Eğitim Öğrencilerinin Mezuniyet Durumlarının Veri Madenciliği Yöntemleri ile Tahmini: Amasya Üniversitesi Örneği” adlı tez çalışması aşağıdaki jüri tarafından OY BİRLİĞİ/OY ÇOKLUĞU ile Amasya Üniversitesi Teknoloji ve İnovasyon Yönetimi Anabilim Dalında YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

**Danışman:**Dr. Öğr. Üyesi Yavuz ÜNAL

Anabilim Dalı, Üniversite Adı

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum/onaylamıyorum .....

**Başkan :** Unvanı Adı SOYADI

Anabilim Dalı, Üniversite Adı

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum/onaylamıyorum .....

**Üye :** Unvanı Adı SOYADI

Anabilim Dalı, Üniversite Adı

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum/onaylamıyorum .....

Tez Savunma Tarihi: .../.../...

Jüri tarafından kabul edilen bu tezin Yüksek Lisans Tezi olması için gerekli şartları yerine getirdiğini onaylıyorum.

.....

Doç. Dr. Meryem EVECEN

Fen Bilimleri Enstitüsü Müdürü

## ETİK BEYAN

Amasya Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada:

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Osman KAYHAN

23.05.2019

# UZAKTAN EĞİTİM ÖĞRENCİLERİNİN MEZUNİYET DURUMLARININ VERİ MADENCİLİĞİ YÖNTEMLERİ İLE TAHMİNİ: AMASYA ÜNİVERSİTESİ ÖRNEĞİ

(Yüksek Lisans Tezi)

Osman KAYHAN

AMASYA ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ

Haziran 2019

## ÖZET

Eğitim-öğretim süresince öğrencilerin doğru yönlendirilmesi, ileride karşılaşılabilecek başarısızlıkların önlenmesi açısından oldukça önemlidir. Zamanında mezun olamayan öğrenciler, genç iş gücünün azalmasının yanı sıra hem aile hem de ülke ekonomisinin olumsuz yönde etkilenmesine neden olabilmektedir. Bu durum zamanında mezun olamayan öğrenciler ile ilgili çalışmaları gerekli kılmaktadır. Öğrenciler ile ilgili eğitsel verilerin analiz edilmesi bu amaca yönelik çalışmalar içerisinde yer almaktadır. Özellikle yükseköğretim kurumlarında her yıl çok sayıda eğitsel verinin biriktiği değerlendirildiğinde, bu verilerin çeşitli yöntemler ile analiz edilmesi daha önemli hale gelmektedir. Bu çalışma kapsamında, Amasya Üniversitesi Uzaktan Eğitim Çocuk Gelişimi, Tıbbi Dokümantasyon ve Sekreterlik, İnternet ve Ağ Teknolojileri, Mekatronik ve Elektrik programlarına 2016-2017 güz döneminde kayıt yaptıran ön lisans öğrencilerinin zamanında mezun olabilme durumlarına ilişkin çıkarımlarda bulunulmaya çalışılmıştır. Tahmin için Karar ağacı, Naive bayes, Destek vektör makinesi, Random forest ve Yapay sinir ağları algoritmaları kullanılmıştır. Algoritmalar ile yapılan analiz sonucunda oluşan sınıflandırma performans ölçütlerine (accuracy, kappa, recall, precision, f-measure) karşılaştırılmalı olarak çalışmada yer verilmiştir.

Sayfa Adedi : 92

Anahtar Kelimeler : Veri madenciliği, uzaktan eğitim, mezuniyet tahmini,

Danışman : Dr. Öğr. Üyesi Yavuz ÜNAL

PREDICTING THE STATUS OF THE GRADUATING STUDENTS IN DISTANCE  
LEARNING WITH THE HELP OF DATA MINING METHODS: AMASYA  
UNIVERSITY SAMPLE  
(M.SC. THESIS)

Osman KAYHAN

AMASYA UNIVERSITY  
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES  
June 2019

ABSTRACT

Correct orientation of students during education is very crucial for preventing future failures. Students who cannot graduate at the time may cause negative impacts on the family and the country's economy as well as the decrease of the young labor force. This situation requires the studies to be held concerning the students who cannot graduate in time. Analyzing the educational data related to the students is included in this kind studies. Considering the accumulation of large number of educational data in higher education institutions, it becomes more important to analyze these data with various methods. In data mining, which is one of the methods used to analyze the data, estimation, classification and clustering methods are benefited. In this study, inferences were tried to be made about whether the students who enrolled to Amasya University Distance Education Child Development, Medical Documentation and Secretariat Associated Degree Programs in 2016-2017 will be able to graduate on time or not. Decision tree, Naive Bayes, Support vector machine, Random forest and Artificial neural networks algorithms are used for estimation. The classification performance criteria (accuracy, kappa, recall, precision, f-measure) which occurs as the result of the analysis with algorithms are included comparatively in the study.

PageNumber : 92  
KeyWords : Data Mining, Distance Education, Graduation Forecast  
Supervisor : Assist Prof. Üyesi Yavuz ÜNAL

## ÖN SÖZ ve TEŞEKKÜR

Akademik gelişimime büyük katkı sağlayan, nezaketleri ve sonsuz sabırları ile beni her zaman destekleyen, bu araştırmanın gerçekleştirilmesinde çok büyük emekleri olan, birlikte çalışmaktan mutlu olduğum çok değerli hocam tez danışmanım Dr. Öğr. Üyesi Yavuz ÜNAL' a en içten teşekkürlerimi sunarım.

Tezimin oluşum aşamasında benden desteklerini esirgemeyen sevgili arkadaşlarım Öğr. Gör. Fatih ÇEMÇEM ve Emre KERMEN'e, her zaman kendisiyle gurur duyduğum kardeşim Öğr. Gör. Selçuk KAYHAN'a çok teşekkür ederim.

Tüm süreci benimle yaşayan, her durumda bana inanan ve destekleyen sevgili eşim Hatice KAYHAN' a sabrı, desteği ve sevgisi için; hayatımdaki en değerli varlıklarım biricik kızlarım Azra KAYHAN ve Asya KAYHAN'a varlıkları ve yaşattıkları her an için minnettirim. İyi ki varsınız.

## İÇİNDEKİLER

	<b>Sayfa</b>
ÖZET .....	iv
ABSTRACT.....	v
ÖN SÖZ VE TEŞEKKÜR.....	vi
İÇİNDEKİLER .....	vii
ÇİZELGELER DİZİNİ .....	xi
ŞEKİLLER DİZİNİ .....	xiii
<b>1. GİRİŞ .....</b>	<b>14</b>
<b>2. UZAKTAN EĞİTİM .....</b>	<b>17</b>
2.1.Uzaktan Eğitim Nedir? .....	17
2.2.Uzaktan Eğitimin Özellikleri .....	17
2.3.Uzaktan Eğitimin Tarihi ve Modern Uzaktan Eğitim.....	18
2.4.Uzaktan Eğitimin Avantajları .....	20
2.5.Uzaktan Eğitim Dezavantajları .....	21
<b>3. VERİ MADENCİLİĞİ.....</b>	<b>22</b>
3.1.Verit Madenciligi Nedir? .....	22
3.2.Verit Madenciligi Uygulama Alanları .....	23
3.3.Verit Madenciligi Süreci.....	25
3.3.1. Verit temizleme:.....	26
3.3.2. Verit bütünleştirme .....	26
3.3.3. Verit indirgeme .....	27
3.3.4. Verit dönüştürme: .....	27
3.3.5. Verit madenciligi yöntemi: .....	27
3.3.6. Sunum ve değerlendirme .....	27
3.4.Verit Madenciligi Modelleri .....	28

	<b>Sayfa</b>
3.4.1. Sınıflandırma.....	28
3.4.2. Kümeleme .....	30
3.4.3. Birliktelik Kuralı.....	32
3.5. Veri Madenciliğinde Kullanılan Yöntemler .....	32
3.5.1. Karar ağaçları.....	33
3.5.2. Yapay sinir ağları.....	36
3.5.3. Naive bayes .....	40
3.5.4. Destek vektör makinesi.....	42
3.5.5. Random forest.....	43
3.6. Veri Madenciliği Yazılımları .....	45
3.6.1. Spss clementine.....	45
3.6.2. Rapidminer.....	45
3.6.3. Knime.....	46
3.6.4. Weka .....	47
3.6.5. Sas .....	47
3.6.6. Enterprise miner.....	48
3.6.7. Kxen.....	49
3.6.8. Insightful miner.....	49
3.6.9. Affinium model.....	49
3.6.10. Statistica data miner.....	50
3.6.11. Inlen .....	50
3.6.12. Dbminer .....	50
3.6.13. Darwin.....	51
3.7. Eğitimde Veri Madenciliği Alanında Yapılan Çalışmalar .....	51
4. MATERYAL VE YÖNTEM .....	55

	<b>Sayfa</b>
4.1. Araştırmanın Amacı.....	55
4.2. Veri Toplama Süreci.....	55
4.3. Veri Temizleme .....	56
4.4. Veri Dönüştürme.....	60
4.5. Veri Analizinde Kullanılan Program .....	60
4.6. Araştırma Modelin Oluşturulması .....	61
4.6.1. Araştırma modelinde kullanılan sınıflandırma kriterleri .....	62
<b>5. BULGULAR .....</b>	<b>64</b>
5.1. Çocuk Gelişimi Bölüm Öğrencilerinin Zamanında Mezun Olma Durumlarına İlişkin Bulgular .....	64
5.1.1. Karar ağacı modelinin başarıml ölçütü .....	64
5.1.2. Naive bayes modelinin başarıml ölçütü .....	64
5.1.3. Random forest modelinin başarıml ölçütü .....	65
5.1.4. Destek vektör makinesi modelinin başarıml ölçütü .....	66
5.1.5. Yapay sinir ağları modelinin başarıml ölçütü .....	66
5.1.6. Çocuk gelişimi bölümü için oluşturulan modellerin karşılaştırılması .....	67
5.2. Tıbbi Dokümantasyon ve Sekreterlik Bölüm Öğrencilerinin Başarıml Ölçütlerinin İncelenmesi .....	68
5.2.1. Karar ağacı modelinin başarıml ölçütü .....	68
5.2.2. Naive bayes modelinin başarıml ölçütü .....	69
5.2.3. Random forest modelinin başarıml ölçütü .....	69
5.2.4. Destek vektör makinesi modelinin başarıml ölçütü .....	70
5.2.5. Yapay sinir ağları modelinin başarıml ölçütü .....	71
5.2.6. Tıbbi dokümantasyon ve sekreterlik bölümü için oluşturulan modellerin karşılaştırılması.....	71
5.3. İnternet ve Ağ Teknolojileri Bölüm Öğrencilerinin Başarıml Ölçütlerinin İncelenmesi .....	73

	<b>Sayfa</b>
5.3.1. Karar ağacı modelinin başarıml ölçütü .....	73
5.3.2. Naive bayes modelinin başarıml ölçütü .....	73
5.3.3. Random forest modelinin başarıml ölçütü .....	74
5.3.4. Destek vektör makinesi modelinin başarıml ölçütü .....	75
5.3.5. Yapay sinir ağları modelinin başarıml ölçütü .....	75
5.3.6. İnternet ve ağ teknolojileri bölüm öğrencileri için oluşturulan modellerin karşılaştırılması .....	76
5.4. Elektrik Bölüm Öğrencilerinin Başarıml Ölçütlerinin İncelenmesi .....	77
5.4.1. Karar ağacı modelinin başarıml ölçütü .....	77
5.4.2. Naive bayes modelinin başarıml ölçütü .....	78
5.4.3. Random forest modelinin başarıml ölçütü .....	79
5.4.4. Destek vektör makinesi modelinin başarıml ölçütü .....	79
5.4.5. Yapay sinir ağları modelinin başarıml ölçütü .....	80
5.4.6. Elektrik bölüm öğrencileri için oluşturulan modellerin karşılaştırılması .....	80
5.5. Mekatronik Bölüm Öğrencilerinin Başarıml Ölçütlerinin İncelenmesi .....	81
5.5.1. Karar ağacı modelinin başarıml ölçütü .....	81
5.5.2. Naive bayes modelinin başarıml ölçütü .....	82
5.5.3. Random forest modelinin başarıml ölçütü .....	83
5.5.4. Destek vektör makinesi modelinin başarıml ölçütü .....	83
5.5.5. Yapay sinir ağları modelinin başarıml ölçütü .....	84
5.5.6. Mekatronik bölüm öğrencileri için oluşturulan modellerin karşılaştırılması .....	84
6. SONUÇ VE ÖNERİLER .....	86
KAYNAKLAR .....	88
ÖZ GEÇMİŞ .....	93

## ÇİZELGELER DİZİNİ

<b>Çizelge</b>	<b>Sayfa</b>
Çizelge 3.1 Veri madenciliği uygulama alanları.....	24
Çizelge 4.1 Çocuk gelişimi bölümü için çalışmada kullanılan nitelikler ve alabileceği değerler .....	57
Çizelge 4.2 Tıbbi dokümantasyon ve sekreterlik bölümü için çalışmada kullanılan nitelikler ve alabileceği değerler .....	58
Çizelge 4.3 Elektrik bölümü için çalışmada kullanılan nitelikler ve alabileceği değerler.....	59
Çizelge 4.4 İnternet ve ağ teknolojileri bölümü için çalışmada kullanılan nitelikler ve alabileceği değerler .....	59
Çizelge 4.5 Mekatronik bölümü için çalışmada kullanılan nitelikler ve alabileceği değerler.....	60
Çizelge 4.6 Hata Matrisi .....	63
Çizelge 5.1 Çocuk gelişim bölümü için karar ağacı algoritması sonuçları.....	64
Çizelge 5.2 Çocuk gelişim bölümü için naive bayes algoritması sonuçları .....	65
Çizelge 5.3 Çocuk gelişim bölümü için random forest algoritması sonuçları .....	65
Çizelge 5.4 Çocuk gelişim bölümü için destek vektör makinesi algoritması sonuçları	66
Çizelge 5.5 Çocuk gelişim bölümü için yapay sinir ağları modelinin algoritması sonuçları.....	66
Çizelge 5.6 Çocuk gelişim bölümü için oluşturulan modellerin karşılaştırılması .....	67
Çizelge 5.7 Tıbbi dokümantasyon ve sekreterlik bölümü için karar ağacı algoritması sonuçları.....	68
Çizelge 5.8 Tıbbi dokümantasyon ve sekreterlik bölümü için naive bayes algoritması sonuçları.....	69
Çizelge 5.9 Tıbbi dokümantasyon ve sekreterlik bölümü için random algoritması sonuçları .....	70
Çizelge 5.10 Tıbbi dokümantasyon ve sekreterlik bölümü için destek vektör makinesi algoritması sonuçları.....	70
Çizelge 5.11 Tıbbi dokümantasyon ve sekreterlik bölümü için yapay sinir ağları algoritması sonuçları.....	71

<b>Çizelge</b>	<b>Sayfa</b>
Çizelge 5.12 Tıbbi Dokümantasyon ve Sekreterlik bölümü için oluşturulan modellerin karşılaştırılması.....	72
Çizelge 5.13 İnternet ve ağ teknolojileri bölümü için karar ağacı algoritması sonuçları.....	73
Çizelge 5.14 İnternet ve ağ teknolojileri bölümü için karar ağacı algoritması sonuçları.....	74
Çizelge 5.15 İnternet ve ağ teknolojileri bölümü için random forest algoritması sonuçları.....	74
Çizelge 5.16 İnternet ve ağ teknolojileri bölümü için destek vektör makinesi algoritması sonuçları.....	75
Çizelge 5.17 İnternet ve ağ teknolojileri bölümü için yapay sinir ağları algoritması sonuçları.....	76
Çizelge 5.18 İnternet ve ağ teknolojileri bölümü için oluşturulan modellerin karşılaştırılması.....	77
Çizelge 5.19 Elektrik bölümü için karar ağaçları algoritması sonuçları.....	78
Çizelge 5.20 Elektrik bölümü için naive bayes algoritması sonuçları.....	78
Çizelge 5.21 Elektrik bölümü için random forest algoritması sonuçları .....	79
Çizelge 5.22 Elektrik bölümü için destek vektör makinesi algoritması sonuçları.....	79
Çizelge 5.23 Elektrik bölümü için yapay sinir ağları algoritması sonuçları.....	80
Çizelge 5.24 Elektrik bölümü için oluşturulan modellerin karşılaştırılması .....	81
Çizelge 5.25 Mekatronik bölümü için karar ağacı algoritması sonuçları .....	82
Çizelge 5.26 Mekatronik bölümü için naive bayes algoritması sonuçları .....	82
Çizelge 5.27 Mekatronik bölümü için random forest algoritması sonuçları .....	83
Çizelge 5.28 Mekatronik bölümü için destek vektör makinesi algoritması sonuçları ...	83
Çizelge 5.29 Mekatronik bölümü için yapay sinir ağları algoritması sonuçları.....	84
Çizelge 5.30 Mekatronik bölümü için oluşturulan modellerin karşılaştırılması.....	85

## ŞEKİLLER DİZİNİ

Şekil	Sayfa
Şekil 3.1 Veri madenciliği süreci.....	25
Şekil 3.2 Sınıflandırma modelinde kullanılan teknik ve algoritmalar .....	29
Şekil 3.3. Kümeleme modelinde kullanılan teknik ve algoritmalar.....	31
Şekil 3.4 Birliktelik modelinde kullanılan teknik ve algoritmalar.....	32
Şekil 3.5 Eğitim verilerine uygun karar ağacı .....	34
Şekil 3.6 Yapay sinir ağları.....	38
Şekil 3.7 Random forest modeli oluşturma algoritması.....	44
Şekil 4.1 RapidMiner programının açılış ekranı.....	61
Şekil 4.2 Analiz ekranı.....	62

## 1. GİRİŞ

Bilgi ve veri iletişim teknolojilerinin hızlı bir biçimde değişmesi ve gelişmesi tüm toplumları bilgi üretmeye doğru yöneltmiştir. Bilgiyi üretebilen ve faydalı bir şekilde kullanabilen toplumlar, teknolojinin gelişmesine katkıda bulunmaktadır. Üretilen teknolojinin kullanılması insanların hayatlarını kolaylaştırmakta ve her türlü olay karşısında hazır olmalarını sağlamaktadır. Teknolojik gelişimlere uyum sağlayabilen toplumlar, teknolojinin sağladığı yararları yaşamlarına entegre ederek bu gelişime uyum sağlayamayan toplumların her zaman bir adım önünde olmuşturlar [1].

Teknolojinin gelişmesiyle beraber tüm toplumlarda kullanılan bilginin dijital ortamlarda kolayca saklanabilmesiyle birlikte oluşan veri tabanları sayısını da gün geçtikçe arttırmaktadır. Veriye ulaşmanın ve veriyi saklamanın kolay olması yüksek kapasiteli işlem yapabilme gücüne katkı sağlamaktadır. Veri tabanlarında saklanan bu ham veriler devasa boyutlara gelmektedir. Bu veri yığınları herhangi bir amaç için sistemli bir biçimde işlenerek analiz edilmesi ve bu amaç doğrultusunda ham veriden çok değerli veri elde edilmediği sürece çok önem arz etmeyebilir [2]. Bu verilerin nasıl faydalı bir şekilde kullanılacağı ve bu ham verilerin nasıl anlamlı bir biçime getirileceği sorularını ortaya çıkarmıştır. Bu soruların cevabına imkân sağlayan veri madenciliği son zamanlarda yoğun olarak kullanılmaya başlanmıştır.

Veri madenciliği, ham verileri çeşitli yöntemler aracılığıyla analiz etme, o verilerden anlamlı ve kıymetli bilgiler meydana getirme süreci olarak ifade edilmektedir. Veri madenciliği, devasa işlenmemiş veriler içerisinden anlamlı ve gizli kalmış bilgiye ulaşılabilme imkanı sağlamaktadır. Veri madenciliği eğitim, finans, elektronik ticaret, sağlık, sosyal medya, savunma sistemleri, pazarlama gibi büyük veri yığınlarının oluşturulmasına olanak sağlayan tüm alanlarda sıklıkça kullanılmaktadır.

Eğitimde veri madenciliği, eğitim ortamlarından gelen işlenmemiş bilgiyi keşfederek veri madenciliği algoritmalarıyla geliştirilmesiyle alakalıdır. Günümüzde yaşanan teknolojik gelişmeler neticesinde diğer alanlarda olduğu gibi eğitim alanında da büyük veri tabanları bulunmaktadır. Eğitim alanında yer alan veri madenciliği çalışmaları, keşfedilmemiş bilgilerin var olduğu düşüncesinden yola çıkarak eğitim sistemlerinde bulunan veri tabanlarında öğrencilere, akademik sorumlulara ve eğitimcilere faydalı olabilecek bilgilere

ulaşmaya imkan sağlar. Eğitim alanında veri madenciliği uygulama alanlarından bazıları ölçme ve değerlendirme çalışmaları, öğrenci mezuniyet durumlarını tahmin etme, mesleki rehberlik etkinlikleri, sınav başarısı analizleri, öğrenci başarı ve başarısızlık nedenlerinin belirlenmesi, öğrenci başarılarının yükseltilmesi, eğitim-öğretim ortamlarındaki eksikliklerin belirlenmesi, daha etkileyici eğitim-öğretim ortamlarının oluşturulmasıdır.

Eğitim, uzmanlar tarafından kişinin davranışlarında kendi yaşantısı yoluyla istendik davranışları meydana getirme süreci olarak tanımlanmaktadır [3]. Uzmanlar tarafından yapılan bu tanımda vurgulanmak istenen belli bir program ve plana göre öğrencilere istendik davranışları kazandırmak ve öğrencilerin bu davranışları yaşantılarında sergilemesini beklemektir. Öğrencilerdeki tüm bu değişimin meydana gelmesi belli bir plana göre icra edilmesi gerekir. Bu plana eğitim ve öğretim programı denir. Eğitim türleri içerisinde yer alan formal eğitim örgün ve yaygın eğitim olmak üzere ikiye ayrılır. Örgün eğitim içerisinde okul öncesi, ilköğretim, ortaöğretim, yükseköğretim, açık öğretim ve uzaktan eğitim yer almaktadır.

Uzaktan Eğitim zaman ve mekan kavramlarını tamamen göz ardı ederek öğrencinin ve eğitmenin eğitim alanına fiziki olarak gelme mecburiyeti olmaksızın hâlihazırdaki var olan bilgisayar ve bilişim imkanları aracılığı ile tamamen web ortamında canlı, görüntülü, sesli olarak derslerin anlatıldığı yada dinlendiği, öğrencilerin istedikleri an bu ders içeriklerini tekrar görüntüleyebileceği ve izleyebileceği, eğitim ve öğretimin günümüz şartlarında hızlı bir şekilde dijital ortama aktarıldığı akılcı, modern, inovatif bir eğitim sistemidir.

Bu tez kapsamında veri madenciliğinin eğitim alanında kullanımı incelenmiş ve eğitim alanında bir uygulama yapılması amaçlanmıştır. Çalışma kapsamında 2016/2017 eğitim öğretim yılında Amasya Üniversitesi uzaktan eğitim çocuk gelişimi, tıbbi dokümantasyon ve sekreterlik, internet ve ağ teknolojileri, elektrik ve mekatronik ön lisans programlarına kayıt yaptıran öğrencilere ait veriler, veri madenciliği makine öğrenme algoritmaları ile sınıflandırılmaya çalışılacak ve doğruluk oranları kıyaslanacaktır.

Bu araştırmanın amacı Amasya Üniversitesi uzaktan eğitim ön lisans öğrencilerinin zamanında mezun olabilme durumlarını veri madenciliği uygulamalarını kullanarak önceden tahmin etmektir. Böylece, zamanında mezun olma potansiyelini barındırmayan öğrenciler tespit edilebilecektir. Bu durumda bulunan öğrencilere gerekli danışmanlık ve

rehberlik faaliyetleri uygulanarak öğrencilerin zamanında mezun olmaları sağlanabilecektir. Bu doğrultuda Amasya Üniversitesi uzaktan eğitim çocuk gelişimi, tıbbi dokümantasyon ve sekreterlik, internet ve ağ teknolojileri, elektrik ve mekatronik bölümlerinde öğrenim gören öğrencilere ait bilgiler, veri madenciliği yöntemlerinden karar ağaçları, naive bayes, random forest, destek vektör makinesi ve yapay sinir ağıları algoritmaları kullanılarak öğrencilerin zamanında mezun olup olamayacakları tahmin edilmeye çalışılmıştır.

Bu çalışma; uzaktan eğitim, veri madenciliği, materyal ve yöntem, bulgular, sonuç ve öneriler bölümlerinden oluşmaktadır. Tezin ikinci bölümünde uzaktan eğitimin tanımı ve uzaktan eğitimle ilgili temel kavramlar hakkında genel bilgiler verilmiştir. Tezin üçüncü bölümünde veri madenciliğinin tanımı, uygulama alanları, süreci, modeli, teknikleri ve veri madenciliği alanında kullanılan programlara ilişkin bilgiler verilmiştir. Materyal ve yöntem bölümünde araştırmanın temel amacından, veri toplama, veriyi temizleme, veriyi dönüştürme işlemlerinin nasıl yapıldığı, analiz için kullanılan programın özelliklerinden ve araştırma modelinin nasıl oluşturulacağı anlatılmıştır. Bulgular ve Tartışma bölümünde yapılan analizler sonucunda elde edilen bulgular karşılaştırılarak modellerin sınıflandırma kriter değerleri verilmiş, birbirlerine üstünlükleri ortaya konmaya çalışılmıştır. Araştırmanın son bölümünde elde edilen bulgular, sonuç ve öneriler başlığı altında değerlendirilerek yorumlanmakta ve gelecekte yapılacak çalışmalarla ilgili önerilerde bulunulmaktadır.

## 2. UZAKTAN EĞİTİM

### 2.1. Uzaktan Eğitim Nedir?

Uzaktan eğitim zaman ve mekan kavramlarını tamamen göz ardı ederek öğrencinin ve öğretmenin eğitim alanına fiziki olarak gelme mecburiyeti olmaksızın hâlihazırdaki var olan bilgisayar ve bilişim imkanları aracılığı ile tamamen web ortamında canlı, görüntülü, sesli olarak derslerin anlatıldığı yada dinlendiği, öğrencilerin istedikleri an bu ders içeriklerini tekrar görüntüleyebileceği ve izleyebileceği, eğitim ve öğretimin günümüz şartlarında hızlı bir şekilde dijital ortama aktarıldığı akılcı, modern, inovatif bir eğitim sistemidir.

Uzaktan eğitim yalnız açık öğretim sistemi ya da örgün öğretimin yanında başvuru alan ikinci sınıf bir öğretim sistemi olarak görülmemelidir. Bunun tersine uzaktan eğitim sistemi uzun yıllardır eğitimde lider olan ülkelerde uygulanan ve modern ölçme ve değerlendirme yöntemleri kullanan, en yenilikçi işleve sahip eğitim sistemidir. Uzaktan eğitim, yazılı veya sanal yazışmalar kullanımı ile oluşan öğrenme uygulaması olarak da tanımlanır. Bu uygulamayla herhangi bir şehirdeki bir öğretmen, dünyanın başka yerlerindeki öğrencilere ders verebilir. Bu, öğrencilerin coğrafi olarak bir sınıfa katılamayacak öğretmenlere erişmelerine yardımcı olur. Ayrıca çalışma saatleri veya diğer sorumlulukları nedeniyle geleneksel saatlerde ders alamayan öğrencilere yardımcı olur [4].

### 2.2. Uzaktan Eğitimin Özellikleri

Uzaktan eğitim olgusunu tanımlamak için çeşitli terimler kullanılmıştır. Açıkçası, uzaktan eğitim (öğrencinin etkinliği) ve uzaktan öğretim (öğretmenin etkinliği) birlikte uzaktan eğitimi oluşturur. Yaygın değişimler arasında, internet ortamındayken kullanılan e-öğrenme veya çevrimiçi öğrenme; genellikle ilkökul veya ortaokul öğrencileri tarafından (ve genellikle interneti kullanarak) sınıf dışında alınan dersleri ifade eden sanal öğrenme; yazışma eğitimi, bireysel öğretimin posta yoluyla yapıldığı uzun süredir devam eden yöntem ve “açık” üniversite yoluyla öğrenme için Avrupa'da yaygın olan sistem açık öğrenme.

Dört özellik uzaktan öğrenmeyi ayırt eder. İlk olarak, uzaktan öğrenme, kurumlar aracılığıyla gerçekleştirilen tanım gereğidir; bireysel çalışma veya akademik olmayan bir

öğrenme ortamı değildir. Kurumlar geleneksel sınıf temelli öğretimi de sunabilir veya sunmayabilirler; ancak geleneksel yöntemler kullananlarla aynı kurumlar tarafından akreditasyona hak kazanırlar.

İkincisi, coğrafi ayrılık uzaktan eğitimde doğaldır ve zaman öğrencileri ve öğretmenleri de ayırabilir. Erişilebilirlik ve rahatlık bu eğitim tarzının önemli avantajlarıdır. İyi tasarlanmış programlar, öğrenciler arasındaki entelektüel, kültürel ve sosyal farklılıkları da köprüleyebilir.

Üçüncüsü, etkileşimli telekomünikasyon, bir öğrenme grubundaki ve öğretmenle bireyleri birbirine bağlar. Çoğu zaman e-posta gibi elektronik iletişimler kullanılır, ancak posta sistemi gibi geleneksel iletişim biçimleri de rol oynayabilir. Ortam ne olursa olsun, etkileşim herhangi bir eğitimde olduğu gibi uzaktan eğitimde de önemlidir. İletişim sistemleri daha sofistike ve yaygın bir şekilde kullanılabilir hale geldikçe öğrencilerin, öğretmenlerin ve öğretim kaynaklarının bağlantıları fiziksel yakınlığa daha az bağımlı hale gelir. Sonuç olarak, İnternet, cep telefonları ve e-posta, uzaktan öğrenmedeki hızlı büyümeye katkıda bulunmuştur [5].

Son olarak, uzaktan eğitim, herhangi bir eğitim gibi, bazen öğrencilerden, öğretmenlerden ve öğretim kaynaklarından oluşan bir öğrenme topluluğu olarak adlandırılan, yani öğrencinin yapmasına izin veren kitaplar, ses, video ve grafik gösterimlerden oluşan bir öğrenme grubu oluşturur. İnternetteki sosyal ağlar topluluk oluşturma fikrini desteklemektedir. Facebook ve Youtube gibi sitelerde, kullanıcılar profiller kurar, bir bağlantı paylaştığı üyeleri tanımlar ve düşünen kişilerin yeni topluluklarını oluşturur. Uzaktan eğitim ortamında, bu tür bir ağ bağlantısı öğrencilerin birbirleriyle bağlantılarını sağlayabilir ve böylece izolasyon duygularını azaltabilir.

### **2.3. Uzaktan Eğitimin Tarihi ve Modern Uzaktan Eğitim**

21. yüzyılın başlarında, Amerika Birleşik Devletleri'nde iki yıllık ve dört yıllık yüksek lisans derecesi veren yükseköğretim kurumlarının yarısından fazlası, esas olarak İnternet aracılığıyla uzaktan eğitim kursları vermiştir. Seçilebilecek 100.000'den fazla farklı çevrimiçi kursla Amerikalı öğrencilerin yaklaşık dörtte biri her dönem en az bir tane ders aldı. Uzaktan eğitim için ortak hedef popülasyonlar, yeniden sertifikalandırma isteyen

profesyonelleri, istihdam becerilerini güncelleyen çalışanları, engelli bireyleri ve aktif askeri personeli içerir.

1990'lı yıllarda başlayan teorik eğilim, video, ses ve diğer multimedya daha güçlü bir şekilde güveniyor gibi görünse de pratikte çoğu başarılı program ağırlıklı olarak elektronik metinlerden ve basit metin tabanlı iletişimden yararlanmışır. Bunun nedenleri kısmen pratiktir - bireysel eğitmenler çoğu zaman kendi multimedyaalarını üretme yükünü yükler - ancak uzaktan öğrenmenin merkezi yararlarına dair gelişen bir anlayışı yansıtır. Şimdi geleneksel sınıflarda veya eğitmenlerin çalışma saatlerinde bilgi paylaşımıyla ilgili zaman kısıtlamalarını kaldırarak, eğitmenler ve öğrenciler arasında iletişimi kolaylaştırmanın bir yöntemi olarak görülmektedir. Benzer şekilde, kendi kendine yeten yazılım eğitim sistemleri, yine de belirli dar eğitim türleri için kullanılmasına rağmen genellikle resmi eğitim ortamlarında diğer insanlarla etkileşimi talep eden bireysel öğrencilere cevap verme ve uyum sağlama konusunda sınırlı esnekliğe sahiptir. Modern uzaktan eğitim kursları, dijital okuma materyallerini, podcast'leri (öğrencinin eğlencesinde elektronik dinleme veya görüntüleme için kaydedilmiş oturumlar), e-posta, dışli (bağlantılı) tartışma forumlarını, sohbet odalarını ve test içeren Web tabanlı kurs yönetim sistemlerini kullanır. Sanal (bilgisayarla simüle edilmiş) sınıflarda işlevsellik kazanmak. Hem tescilli hem de açık kaynaklı sistemler yaygındır. Çoğu sistem genel olarak eş zamansız olmasına rağmen öğrencilerin istedikleri zaman çoğu özelliğe erişmesine izin vermek, zaman zaman canlı video, ses ve elektronik belgelere paylaşılan erişim içeren eşzamanlı teknolojiler de kullanılır. Bloglar, wikiler (tüm sınıf katılımcıları tarafından değiştirilebilen web siteleri) ve işbirliğine dayalı olarak düzenlenmiş belgeler şeklindeki paylaşılan sosyal alanlar, eğitim ortamlarında da, ancak sosyalleşme için internetteki benzer alanlardan daha az bir ölçüde kullanılır.

Modern kurumsal uzaktan öğrenmedeki büyümenin yanı sıra, e-egitim, e-mentorluk ve araştırma yardımı dahil olmak üzere Web tabanlı veya kolaylaştırılmış kişisel eğitim hizmetleri geldi. Ayrıca, ebeveynlerin çocukları için yerel öğretmenlerle seçim yapmasına ve onlarla iletişim kurmasına yardımcı olan birçok yardım şirketi vardır. Uzaktan eğitim programlarının ve özel ders hizmetlerinin kullanımı, özellikle çocuklarına ev sahipliği yapan ebeveynler arasında artmıştır. Birçok üniversitede okuma, yazma ve temel matematik konularında yardımcı yardım için bazı çevrimiçi özel ders hizmetleri, hatta bazılarında doktora adaylarına tez sürecinde yardımcı olacak çevrimiçi rehberlik

programları vardır. Son olarak, web tabanlı kişisel asistanlık yapan birçok şirket, sürekli eğitim veya mesleki gelişim isteyen yetişkinler için bir dizi hizmet sunmaktadır.

Uzaktan eğitim alanı, son on yılda önemli ölçüde değişmiştir. Uzaktan eğitim, öğrencinin ve öğretmenin yere göre ayrıldığı ve bazen de zaman zaman yerel ve uluslararası eğitimin en hızlı büyüyen şekli olan yapılandırılmış öğrenme. Bir zamanlar geleneksel olmayan dağıtım sistemlerini kullanarak özel bir eğitim şekli olarak kabul edilen şey, şimdi ana eğitimde önemli bir kavram haline geliyor [6].

#### 2.4. Uzaktan Eğitimin Avantajları

Uzaktan eğitimin avantajları şu şekilde sıralanmaktadır [7]:

- Çok esneklik. Uzaktan eğitim kurslarıyla, öğrenciler bilgisayar ve internet bağlantısı olması koşuluyla ders çalışmalarını hemen hemen her yerden yapabilirler. Bu, öğrencilerin zamanlanmış derslerde zaten yoğun bir yaşama sıkılmak zorunda kalmadan kendileri için ne zaman ve nerede uygun çalışmalarına olanak sağlar.
- Yolculuk yok. Çevrimiçi bir kursa katılmak, pahalı gazı veya toplu taşımayı azaltmanın bir yolu olabilir. Öğrenciler, sınıf ödevlerini tamamlamak için sık sık evden çalışabileceklerinden, hem derse hem de sınıfa yapılan gezilerin kesilmesinde hem zaman hem de para tasarrufu sağlar.
- Okullar için çok sayıda seçenek. Kolejlerin az olduğu veya hiç olmadığı bir toplulukta yaşıyor olsanız bile, uzaktan öğrenim, eğitiminizi tamamlamak için çok çeşitli okullardan seçim yapmanıza olanak sağlar. Kendi alanınızda uzmanlaşmış çevrimiçi okulları veya mükemmel bir genel eğitim sağlayabilen okulları bulabilirsiniz. Her iki durumda da eğitim seçenekleriniz büyük ölçüde genişletilecektir.
- Düşük maliyetler. Çevrimiçi kursların fiyatları genellikle kampüs içi meslektaşlarına göre daha ucuzdur ve kampüste işe gidip gelmek, taşımak veya yemek planları almaktan endişelenmenize gerek kalmaz, evden öğrenmenin bazı ek faydaları vardır.
- Çalışırken öğrenin. Uzaktan eğitim genellikle kendi programınıza göre tamamlanabildiği için, uzaktan eğitim kurslarını tamamlamak, geleneksel eğitim programlarına göre çalışmaktan çok daha kolaydır. Mesleğinizi sürdürmek size daha fazla gelir, deneyim ve istikrar kazandırırken derecenizi tamamlar, endişelenmenize daha az zaman verir ve çalışmalarınıza odaklanmanız için daha fazla zaman verir.

## 2.5. Uzaktan Eğitim Dezavantajları

Uzaktan eğitimin dezavantajları şu şekilde sıralanmaktadır [7]:

- Sosyal etkileşim eksikliği, sınıf ortamı, öğrenmeyi en çok sevdiğiniz şeyse, bir adım geriye gitmek ve uzaktan öğrenmeyi yeniden düşünmek isteyebilirsiniz. Muhtemelen sohbet odaları, tartışma panoları ve e-posta yoluyla etkileşimde bulunursunuz, ancak deneyim geleneksel kurslardan oldukça farklı olacaktır.
- Format, tüm öğrenciler için ideal değildir. Herkes çevrimiçi öğrenme için ideal bir aday değildir. Motivasyon, erteleme ve bir eğitmen tarafından kişisel ilgi gerektiren birçok sorun olduğuna biliyorsanız, bir çevrimiçi öğrenme programına kaydolmadan önce uzun ve zor düşünmek isteyebilirsiniz.
- Bazı işverenler çevrimiçi dereceleri kabul etmemektedir. İşverenlerin çoğunluğu varken, hala uzaktan öğrenmeye bağlı bir leke görüyorlar. Çevrimiçi derecenizin bazı iş alanları veya ileride öğrenim için ideal bir araç olmayabileceğini anlayın.
- Yeni teknolojilere adaptasyon gerektirir. Teknolojiyle çalışmayı hiç sevmeyen bir kişi olmadıysanız, çevrimiçi bir dersten daha fazla teknoloji meraklısı meslektaşlarınızdan çok daha azını alırsınız. Bir sınıfa kaydolmadan önce bilgisayarlarla ve çevrimiçi programlarla çalışmaktan rahat hissettiğinizden emin olun.
- Dereceyi tamamlamak için gereken tüm dersler çevrimiçi olarak sunulamaz. Hemşirelik gibi daha pratik ana dalların tamamen çevrimiçi olarak sunulmadığı anlaşılıyor, sonuçta derecenin bir kısmı doğrudan hastalarla çalışmayı öğreniyor. Çevrimdışı olarak nelerin tamamlanması gerektiğini görmek için derecenizin tüm gereksinimlerini öğrenin.

### 3. VERİ MADENCİLİĞİ

#### 3.1. Veri Madenciliği Nedir?

Günümüzde bilişim ve veri iletişim teknolojilerinde yaşanan gelişmelerin çok hızlı bir şekilde ilerlediği gözlemlenmektedir. Bu teknolojilerde her gün bir başka yenilikle karşılaşmaktadır. Bu yeniliklerle birlikte teknolojiye dayalı ürünlerin fiyatlarındaki düşüşleri meydana gelmesi üzerinde durulması gerekli olan bir diğer kritik konudur. Bu gelişmeler, kullanıcıların daha hızlı, daha yetenekli ve daha kullanışlı bilişim teknolojisi ürünlerine daha kolay sahip olabilmelerine neden olabilmektedir.

Bilişim teknolojilerinin sağladığı kolaylıklar ve yaşantımızda daha fazla yer edinmesiyle birlikte gerçekleştirilen tüm uygulama dijital koşulda belge olarak kaydedilmektedir. Örneğin bir markette satın alınan her bir ürün, geri verilen ürün, herhangi bir ürünle birlikte aldığımız diğer ürünler veri tabanlarında saklanmaya başlanmıştır. Müşteri sayısı fazla olan işletmeler çokça veri üretmek ve bu veriler de dijital ortamda çok büyük veriler oluşturmaktadır.

Bilişim teknolojisi günümüz dünyasında üretilen bu devasa verileri saklamak için yeterli olabilir. Fakat bu yığınsal veriler ne işe yarayacaktır? Bu verilerden firmalar kazanç sağlayabilecek midir? Biriken bu veriler işimize yarar gerçek anlamda işe yarar bilgiye dönüştürülebilir midir? Bu devasa verilerin bilgisayar yardımı ile çeşitli işlemlere tabi tutularak ve değerlendirilerek bu tür sorulara olumlu yanıt vermek mümkündür. Bu verilerden yararlanarak kurum içi destek sistemleri oluşturulabilir. Çeşitli çözümler yapılarak özellikle stratejik seviyedeki kararlara destek sağlanabilir.

Veri kayıtlarının çok büyük oranlarda artması, otomatik veri toplama merkezlerinin geliştirilmesi, iş zekâsı uygulamalarının yaygınlaşması, verilerin niteliksel ve niceliksel olarak sınıflandırılması, gen teknolojilerindeki gelişmelerin artması, büyük verinin bulut teknolojisi ile yaygınlık kazanması, ticari, sosyal ve ekonomik alanlarda verilerin artması, süper bilgisayarlardaki ucuzlama ve müşteri memnuniyeti ve tatmini kavramlarının öneminin artması veri madenciliği kullanımını zorunlu kılan unsurlardır [8].

Veri madenciliği 1990'lı yılların başından itibaren veri saklama araçları, RFID ve barkod teknolojileri ile birlikte gelişmekte ve kullanım alanı artmakta olan bir konu olduğundan kullanılan yer ve zamana göre çeşitli tanımları yapılmıştır. Çünkü zaman ilerledikçe şurada verilen betimleme bir sonraki gün eksik olabilmektedir. Genel olarak kullanılan betimlemelerden birine göre: “Veri madenciliği evvelinde malumat sahibi olunmayan, muteber ve pratik bilginin büyük veri tabanlarından toplanması ve bunların işletme sürecinde pratiğe dönüştürülmesidir.” İlgili tanımlamada dikkat edilmesi gereken husus, toplanan verilerin evvelinde bilinmemesidir; buradan anlaşılması gereken detay ise sonucun önceden bilinmemesi durumudur.

Başka bir tanımlamaya göre veri madenciliği matematiksel ve istatistiksel metotlarla beraber örüntü oluşturma teknolojilerini işe koşarak, depolama ortamlarında korunmuş olan bilgi topluluklarının elimine edilmesi sonucunda manalı yeni korelasyon, örüntü ve eğilimlere ulaşılması işlemleridir [9]. Veri madenciliğini sadece bir tanımlama ile ifade etmeye çalışmak sahip olduğu önemi tam olarak ortaya koymada sorunlara neden olabilir. Veri madenciliği bir yordama aracı olarak düşünülebilir veya gösterişsiz bir bilgisayar uygulaması olarak ele alınabilir. Bununla birlikte sadece bir programı anlamak ve o programın ara yüzünü kullanmak bu alana dair yeterli donanıma sahip olunduğu anlamına gelmez. Bu konunun sadece bu özelliğiyle ele alınması veri madenciliği alanına yapılan en büyük haksızlık olacaktır. Bu sebeple öncelikle veri madenciliği kullanım alanlarını ve amaçlarını kavrayıp daha sonra bu amaçla geliştirilmiş teknik algoritma ve yazılımları kullanmakta fayda vardır.

### **3.2. Veri Madenciliği Uygulama Alanları**

Veri Madenciliği sektör farkı gözetmeden, büyük veri yığınlarının oluşturulmasına olanak sağlayan tüm alanlarda sıklıkça kullanılmaktadır. Bu alanlar Çizelge 3.1’ de gösterilmektedir [10].

Çizelge 3.1 Veri madenciliği uygulama alanları

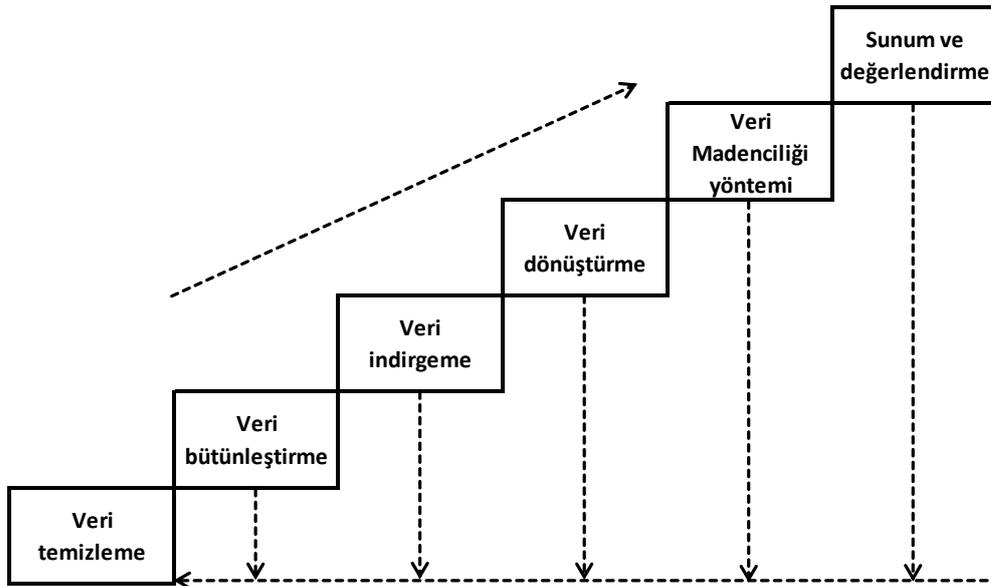
<b>Pazarlama</b>	<ul style="list-style-type: none"> <li>• Müşterilerin satın alma eğilimlerinin belirlenmesi</li> <li>• Müşteri kayıp analizleri</li> <li>• Pazar sepeti analizleri</li> <li>• Satışlardaki anormal gelişmelerin saptanması</li> <li>• Müşteri segmentasyonu</li> <li>• Müşteri memnuniyet araştırılması</li> </ul>
<b>Finans</b>	<ul style="list-style-type: none"> <li>• Sigorta dolandırıcılıklarının tespiti</li> <li>• Kredi kartı dolandırıcılıkların belirlenmesi</li> <li>• Harcamalara göre müşteri gruplarının belirlenmesi</li> <li>• Kredi taleplerinin değerlendirilmesi</li> <li>• Personel kayıp analizleri</li> <li>• Finansal göstergeler arasındaki gizli ilişkilerin belirlenmesi</li> <li>• Tehlikeli müşteri topluluklarının tespit edilmesi</li> </ul>
<b>Elektronik Ticaret</b>	<ul style="list-style-type: none"> <li>• Elektronik mağazalar için tehditlerin tespiti</li> <li>• Web sayfalarına ilişkin kullanımların analiz edilmesi</li> <li>• Web sayfalarına yapılan saldırıların belirlenmesi</li> <li>• Müşteri görüşlerin çözümlenmesi</li> <li>• Müşteri memnuniyetin araştırması</li> <li>• Okuyucu görüş analizleri</li> </ul>
<b>Sağlık</b>	<ul style="list-style-type: none"> <li>• DNA mikro dizi analizleri</li> <li>• Hastalıkların tanılanması</li> <li>• İlaç çözümlenmeleri</li> <li>• İlaç yan etki analizleri</li> <li>• Sağlıkta dolandırıcılık analizleri</li> </ul>
<b>Sosyal Medya</b>	<ul style="list-style-type: none"> <li>• Duygu çözümlenmeleri</li> <li>• Kutuplaşma çözümlenmeleri</li> <li>• Politik gelişmelerin analizleri</li> <li>• Oylama kampanyalarının analizleri</li> <li>• Seçim neticelerine yönelik tahmin araştırması</li> <li>• Politikacıların popülerlik çalışması</li> <li>• Sanatçıların popülerlik araştırması</li> <li>• TV yayınları popülerlik araştırması</li> <li>• İnternet gazetelerdeki okuyucu görüşleri</li> <li>• Sermaye piyasa analizleri</li> </ul>
<b>Savunma Sistemleri</b>	<ul style="list-style-type: none"> <li>• Düşman eylemleri ve terör eylemlerinin modellenmesi ve öngörülmesi</li> <li>• Uçak kazalarında hataların belirlenmesi ve tedbirlerin alınması</li> </ul>
<b>Eğitim</b>	<ul style="list-style-type: none"> <li>• Ölçme ve değerlendirme çalışmaları</li> <li>• Mesleki rehberlik faaliyetleri</li> <li>• Sınav başarısı analizleri</li> <li>• Öğrenci başarı ve başarısızlık nedenlerinin belirlenmesi,</li> <li>• Öğrenci başarılarının yükseltilmesi,</li> <li>• Eğitim-öğretim ortamlarındaki eksikliklerin belirlenmesi,</li> <li>• Daha etkileyici eğitim-öğretim ortamlarının oluşturulması</li> <li>• Öğrenci mezuniyet durumlarının tahmin edilmesi</li> </ul>

### 3.3. Veri Madenciliği Süreci

Son yıllarda büyük veri tabanlarına kolayca ulaşılabilir olması veri madenciliğinin ortaya çıkmasına ve veri madenciliği yöntemlerinin sıkça kullanılıp birçok konu üzerinde araştırmalar yapılmasına neden olmaktadır. Tüm sektörlerle ilgili tüm bilgiler bilgisayar hafızalarında saklanmaktadır. Fakat bu verilerin tamamının doğru olduğunu kimse söyleyemez ve bu verilerin %100 doğru olduğunu ifade edemez; ayrıca bu bilgilerin hâlihazırdaki durumu yapılacak olan araştırmalara hizmet edeceği de kesin olarak söylenmez. Bu sebepten elimizde ki ham verilerin bazı işlemlerden geçirilmesi gerekmektedir [11].

Veri madenciliği belirli bir adım değil, bir süreçtir. Veri madenciliği uygulaması birtakım basamaklardan meydana gelmektedir. Şekil 3.1’ de gösterilen basamaklar aşağıda sıralanmaktadır [12]:

- Veri temizleme
- Veri bütünleştirme
- Veri indirmene
- Veri dönüştürme
- Veri madenciliği yöntemi
- Sunum ve değerlendirme



Şekil 3.1 Veri madenciliği süreci

### 3.3.1. Veri temizleme:

Veri madenciliği uygulamalarında, üzerinde çalışma gerçekleştirilecek verilerin beklenen niteliklerde ve yanlış olduğu anlaşılabilir. Bu ham veriler hatalı veya eksik verilerin oluşturduğu tutarsızlıklara karşılaşılabılır. Veri tabanlarında bulunan bu tür tutarsız ve hatalı veriler gürültü veriler olarak adlandırılmaktadır. Doğru ya da tam olmayan bu gürültü veriler yapılan uygulamalarda doğru sonuçlara ulaşmamızı engeller. Analizlerin olumlu sonuçlar verebilmesi için bu tür gürültü verilerin temizlenmesi gerekir. Bu ham verilerin düzeltilmesinde kullanılabilir teknikler aşağıda sıralanmıştır [13].

- Eksik değer barındıran belge veya belgeler çıkarılabilir. Bu uygulamaya daha çok sınıf etiketi yeterli olmayan koşullarda başvurulabilir; ancak bu uygulama satır birden çok nitelik eksik veriye sahip değilse etkili olmaz.
- Eksik değer manüel şekilde betimlenebilir. Bu uygulama vakit zaman gerektirir ve büyük belge yapılarında pratik değildir.
- Eksik değer bütün bir stabil ile doldurulur. Bütün eksik değerler “Bilinmiyor”, “∞” gibi aynı stabil ile ifade edilir. Bu uygulamada Veri Madenciliği yazılımı verilerin tamamının ortak “Bilinmiyor” verisine sahip olduğu sonucuna varılabilir.
- Aynı sınıfa bağlı bütün örneklem adına parametrelerin ortalaması işe koşulabilir. Sözelimi farksız kredi tehlike sınıflamasına dâhil olan bireyler için ortalama kazanım değeri eksik değerler yerine kullanılabilir.
- Eksik veri ortalama veriyle doldurulur. Meselâ bir kişinin kazanımı eksik ise bütün kişilerin kazanımlarının ortalaması eksik veriye kaydedilir.
- Bulunan değerlere bağlı olarak en ideal veri uygulanabilir. Burada bahsedilen en ideal verinin tespit edilmesi amacıyla regresyon ya da karar ağacı gibi yöntemlere başvurulabilir. Mesela yaş  $x$ , eğitim seviyesi  $y$  olan bir birey için ücret konusu, var olan değerlerden yukarıdaki yöntemlerden bir tanesinin işe koşulmasıyla bilinebilir.

### 3.3.2. Veri bütünleştirme:

Veri bütünleştirme işlemi, çeşitli değer tabanlarından toplanan değerlerin beraber değerlendirilmesi amacıyla çeşitli cinsteki değerlerin bir cinse dönüştürülmesidir. Mesela, bir değer kaynağında cinsiyet “E” ve “K” kodlarıyla, başka bir değer kaynağında ise bu alan (bir) veya 2 (iki) değerleriyle, farklı bir veri tabanında ise cinsiyet alanı “Bay” ve “Bayan” sözcükleri ile ifade edilmiş olabilir. Farklı veri tabanlarında aynı veri alanı için farklı

semboller kullanılmış olması bu tür veriler üzerinden analiz yapılmasını imkansız hale getirebilir. Bu nedenden farklı tiplerde bulunan verilerin analiz yapılmadan önce aynı türe dönüştürülmesi yani veri bütünleştirilmesi gerekmektedir [14].

### **3.3.3. Veri indirgeme:**

Veri madenciliği yöntemleri uygulanırken çoğunlukla büyük veri setleri kullanılmak istenir. Fakat bu husus çoğu zaman büyük problemleri beraberinde getirmektedir. Eğer analizden çıkarılacak sonucun farklılık göstermeyeceği düşünülüyorsa, uygulayıcının değişken sayılarında ve veri sayılarında yönetilebilir rakamlara indirgemesi gerekmektedir[15].

### **3.3.4. Veri dönüştürme:**

Veri madenciliği algoritmaları her türdeki verilerle çalışmamaktadır. Parametrelerin averaj ve varyansları birbirinden kayda değer şekilde başka olması durumunda büyük averaj ve varyansa sahip parametrelerin diğer değişkenler üzerinde baskısı daha çok olur ve diğerlerinin görevlerini ciddi anlamda düşürür. Ayrıca parametrelerde var olan çok büyük ve küçük verilerde analizlerin doğru bir şekilde gerçekleşmesini sekteye uğratar. Bu sebepten model için uygun olan seçim yapıldıktan sonra, verilerin istatistiksel analizleri yapılarak çıkan sonuçlara göre gerekiyorsa veriler dönüştürülmelidir [16].

### **3.3.5. Veri madenciliği yöntemi:**

Veri madenciliği yöntemlerini sağlıklı bir şekilde uygulayabilmek için değer temizleme, değer bütünleştirme, değer indirgeme ve değer dönüştürme işlemlerinin tamamı veya bir kısmı veriyi hazır hale getirmek için uygulanır. Bu aşama, değerler düzenlenmesinin devamında veri madenciliği prosedürleri kullanılarak analizin yapıldığı aşamadır. Bu algoritmalar birliktelik kuralları, kümeleme ve sınıflandırmadır [16].

### **3.3.6. Sunum ve değerlendirme:**

Sonuç alma ve sonuç değerlendirme veri madenciliği sürecinin en önemli aşamalarından biridir. Veri madenciliği teknikleri ile yapılan analiz sonucunda elde edilen sonuçlar

değerlendirilerek oluşturulan veri madenciliği modelin kullanılıp kullanılmayacağına karar verilir. Analiz sonucunda elde edilen bilginin sürecin başında ortaya konan hedeflere uygun olmadığı tespit edilirse problem tanımlama aşamasına dönülebilir [15].

### 3.4. Veri Madenciliği Modelleri

Veri madenciliğinin iki temel amacı bulunmaktadır: Bu amaçlardan ilki hali hazırda bulunan veri tabanından ham verileri analiz ederek tahminlerde bulunmak buna tahmin edici model denmektedir. Diğer amacı ise veriler arasında bulunan ilişkilerden çeşitli davranışlar tanımlamak buna da tanımlayıcı model denmektedir [17].

Birçoğu istatistiksel tabanlı olmak üzere veri madenciliğinde çok sayıda algoritmalar ve yöntemler geliştirilmiştir. Geleneksel olarak veri madenciliği modellerini işlevlerine göre 3 ana başlıkta gruplandırabiliriz [18].

- Sınıflandırma
- Kümeleme
- Birliktelik Kuralı

#### 3.4.1. Sınıflandırma

Veri madenciliği modelleri içerisinde en çok bilenen ve en sık kullanılan sınıflandırma modeli, veri tabanlarındaki saklı kalmış örüntüleri ortaya çıkarmak için kullanılır. Sınıflandırma, elimizde bulunan ham verileri daha önceden belirlenen bir özelliğini dikkate alarak kategorilere ayırmak ve yeni eklenecek verilerin hangi kategoriye ekleneceğini belirleme işlemidir. Başka bir deyişle, yeni bir verinin hangi kategoriye dahil edileceğini belirleme olayıdır [19]. Sınıflandırma örüntü tanıma, kalite kontrol çalışmaları, resim, hastalık tanıları, pazarlama ve dolandırıcılık tespiti gibi alanlarda sıkça kullanılmaktadır.

Matematiksel olarak sınıflandırma şu şekilde tanımlanabilir:

$X = \{a_1, a_2, \dots, a_n\}$  bir veri tabanı olsun ve her bir  $a_i$  bir kaydı temsil etsin.

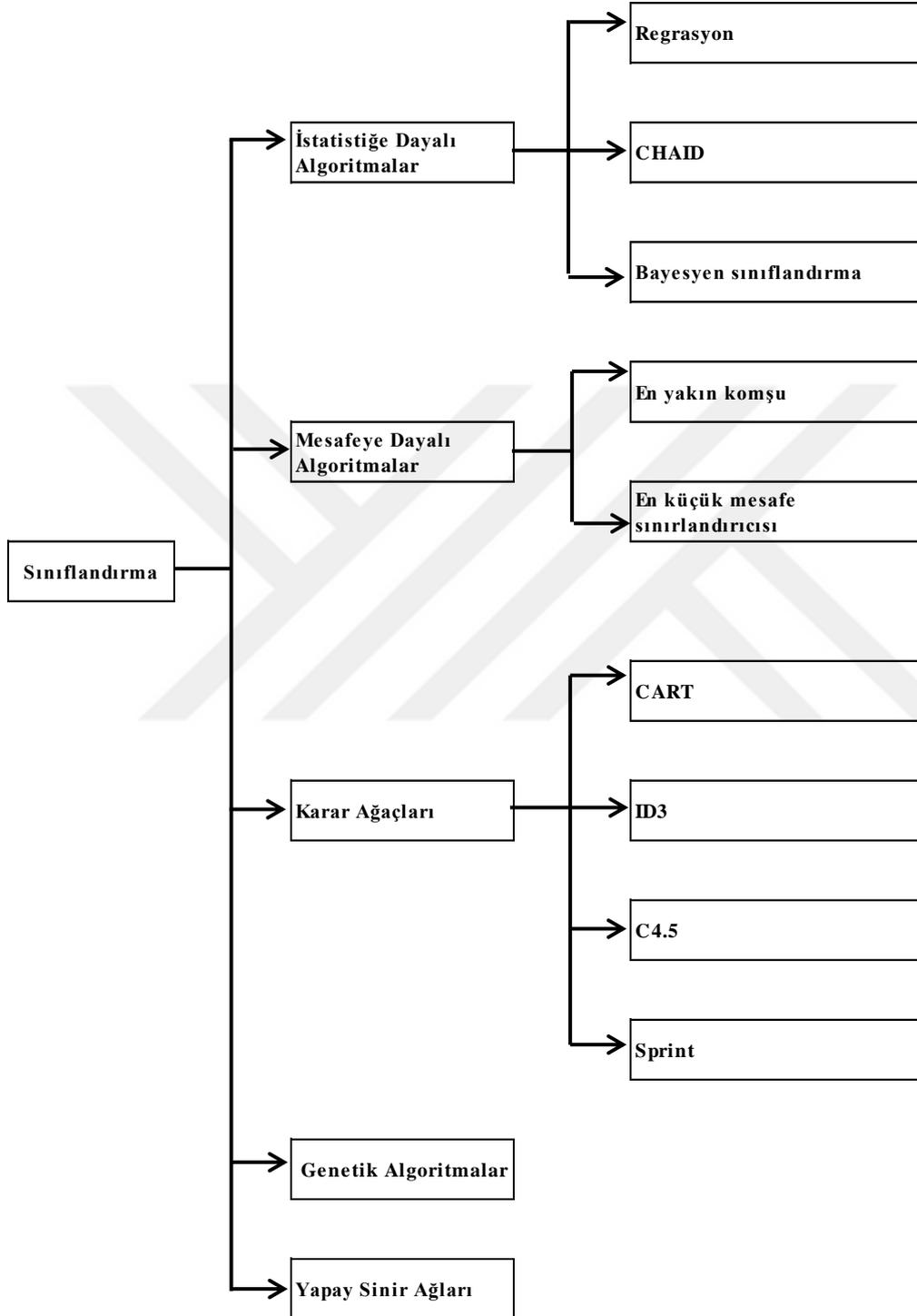
$Y = \{y_1, y_2, \dots, y_m\}$  ise  $m$  adet sınıftan oluşan sınıflar kümesini temsil etsin.

$f: X \rightarrow Y$  ve her bir  $a_i$  bir sınıfa dahil olmalıdır.

Her bir  $Y_j$  ayrı bir sınıftır ve her bir sınıf kendisine ait kayıt içerir.

Yani,  $Y_j = \{ a_i \mid f(a_i) = Y_j, 1 \leq i \leq n, \text{ ve } a_i \in X \}$  [20].

Sınıflandırma modelinde kullanılan teknik ve algoritmaları Şekil 3.2’ de gösterilmiştir [11].



Şekil 3.2 Sınıflandırma modelinde kullanılan teknik ve algoritmalar

### 3.4.2. Kümeleme

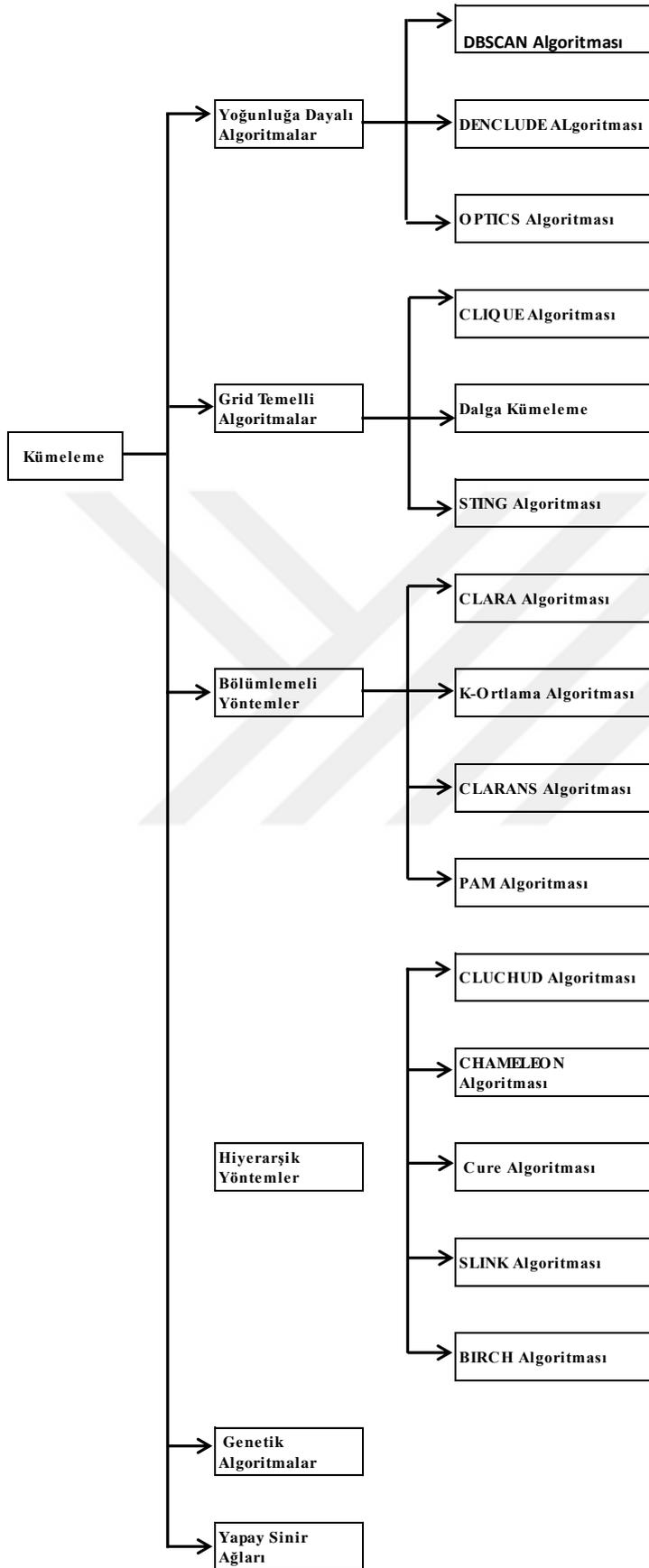
Kümeleme, veri dizisinde bulunan benzer nesnelerin aynı gruplar içerisinde yer alacak şekilde ayrıştırılmasıdır. Sınıflandırma modeline göre sınıflar önceden belli iken, kümeleme modelinde bu sınıflar belli değildir. Yani kümeleme modeli sınıflandırma modelinden farklı olarak kümeleme analizinde kaç tane grup oluşacağı tespit edilememektedir [21].

Bu modelin temel amacı, veri yığınlarında doğal olarak oluşabilecek alt sınıfları bulmaktır. Genellikle büyük veri yığınlarında ki gizli örüntülerin keşfedildiği uygulamalarda kullanılan kümeleme, denetimsiz öğrenme olarak ta tanımlanmaktadır. Denetimsiz öğrenme veri kümesinin sınıflar haline dönüştürülmesi ve bölünmesi sürecidir. Her kümede bulunan üyeler ortak özellikleri paylaşabilmektedir. Bilimsel bilgi keşfinde, pazar araştırmalarında, desen tanımlamada, enformasyon erişiminde, resim işlemede, sayısal biyolojide, web kayıtları analizinde, uzaysal harita, suç analizlerinde ve bunun gibi pek çok alanda kümeleme analizi uygun bir yöntem olarak kullanılmaktadır [22].

Kümelemenin matematiksel tanımı ise şu şekilde yapılabilir:

Elimizde  $D = \{X_1, X_2, X_3, \dots, X_n\}$ ,  $n = 1, 2, \dots, m$  veri tabanı olsun, her bir  $X_n$  bir kaydı temsil etsin  $X = \{x_1, x_2, \dots, x_i\}$ ,  $i = 1, 2, \dots, m$ , her bir  $x_i$ , ad, soyad, yaş ve gelir gibi özellikler olsun. Kümelerdeki amaç  $D$  veri tabanını,  $j$  adet  $K$  kümesine bölmek ve  $K_j \subseteq D$  koşulunun sağlanmasıdır [20].

Kümeleme modelinde kullanılan teknik ve algoritmalar Şekil 3.3' de gösterilmiştir [11].



Şekil 3.3 Kümeleme modelinde kullanılan teknik ve algoritmalar

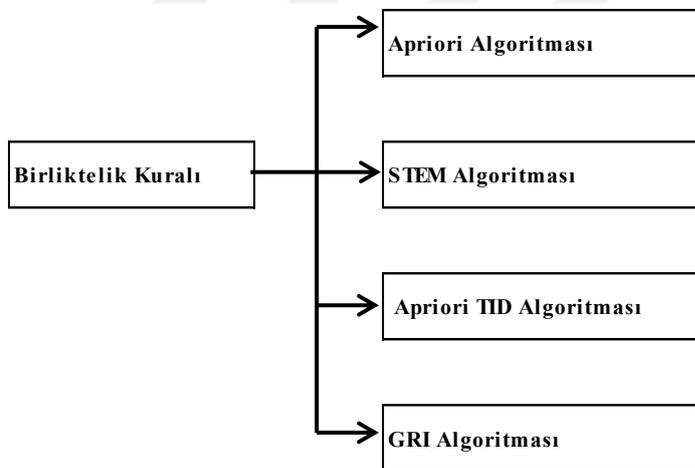
### 3.4.3. Birliktelik Kuralı

Birliktelik kuralı, bir veri tabanında ki kayıtların aralarında ki ilişkileri inceleyerek, bu kayıtlar arasındaki bağlantıları denetimsiz biçimde arayan veri madenciliği modelidir. Genellikle pazarlama ve market sektöründe kullanılan bu yöntem ‘‘Pazar sepeti analizi’’ olarak da isimlendirilir [23].

Birliktelik kuralları kayıtları arasında fark edilmesi zor olan bağlantıları kurmaktadır. Bu bağlantılar firmalar için birçok fayda sağlamakta, stratejik kararlar almaya ve müşteri memnuniyetine kadar farklı aşamalarda kullanılmaktadır.

Bebek maması alan bir müşterinin çocuk bezi de alacağı rahatlıkla tahmin edilebilir fakat birliktelik kuralı çocuk bezi ile yumuşatıcı arasındaki ilişkiyi bulmamıza olanak sağlamaktadır [24].

Birliktelik modelinde kullanılan teknik ve algoritmalar Şekil 3.4’ de gösterilmiştir [11].



Şekil 3.4 Birliktelik modelinde kullanılan teknik ve algoritmalar

### 3.5. Veri Madenciliğinde Kullanılan Yöntemler

Veri madenciliğinde kullanılan çok sayıda yöntem ve algoritmalar bulunmaktadır. Bu çalışma da karar ağaçları, naive bayes, destek vektör makinesi, random forest ve yapay

sinir ağırları algoritmalarıyla analizler yapılmıştır. Bu kapsamda bu yöntemler aşağıda açıklanmıştır:

### 3.5.1. Karar ağaçları

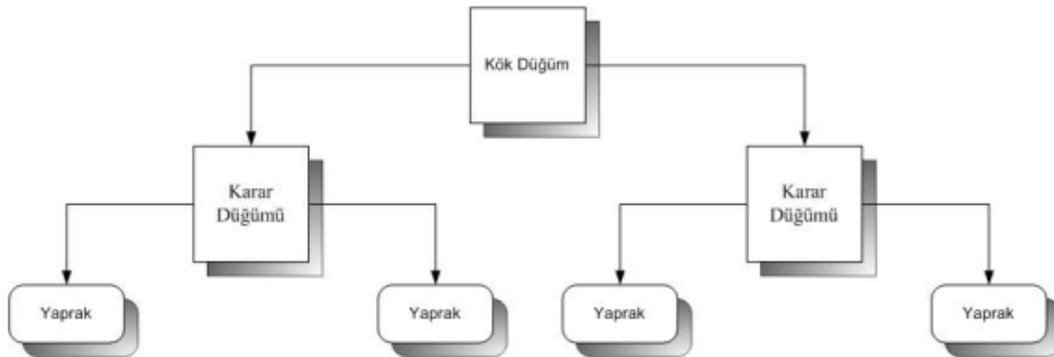
Karar ağaçları, Bierman ve Friedman tarafından 1973 yılında ileri sürülmüştür. Karar ağaçlarında bağımlı değişken üzerindeki farklılıkların maksimize edilmesi amacıyla değişkenler parçalanmakta ve bir ağaç elde edilmektedir. Karar ağaçları, veri madenciliğinde oluşumlarının pahalı olmaması, açıklamalarının basit olması, değer kaynağı sistemleri ile basitçe birleştirilmeleri ve kredibilitelerinin üstün olması nedeniyle sınıflama modelleri arasında çoklukla tercih edilen bir yöntemdir [25]. Karar ağacı, söyleminden de görüldüğü gibi bir ağaç şeklinde, kestirimde bulunmak için başvurulacak bir yöntemdir. Ağaç yapısı ile, kolayca anlaşılabilen kurallar oluşturan, bilişim teknolojileri ile uyumlu olabilen en sık kullanılan sınıflama tekniğidir. Karar ağaçları, çoğunlukla kategorize etme, kümeleme, kestirim modellerinde ve problem bazlı araştırma alanını alt bölümlere dağıtmak amacıyla sıkça tercih edilen bir tekniktir [26].

Karar ağacı; karar düğümleri, dal ve yapraklar olmak üzere 3 kısımdan oluşmaktadır. Karar düğümü, başvurulacak uygulamayı ortaya koyar. Bunun neticesinde sonucu veri kaybı olmaksızın ağacın dallara bölünmesi söz konusudur. Her düğümden test ve bölümlere dağılma uygulaması birbirini takip eder ve bu dağılım üst düzeydeki ayrışmalara tâbidir. Her dal kategorize işlemini bitirmek için çabalar; ancak bir dalın ucunda kategorize etme durumu söz konusu değilse, ilgili dalın devamında bir karar ilmeği meydana gelir. Ancak dalın nihayetinde bir sınıf meydana geliyorsa, o dalın neticesinde yaprak mevcuttur. Bu yaprak, değer üzerinde meydana gelmesi beklenen kategorize etme yöntemlerindedir. Karar ağacı uygulaması kök ilmeğinden çıkar ve yukarıdan aşağıya olacak şekilde yaprağa varana kadar meydana gelir [27].

Karar ağacı tekniğine başvurularak verinin kategorize edilmesi iki aşama olarak gerçekleşmektedir. Birinci basamakta öğrenme işlemi gerçekleşir. Öğrenme aşamasında evvelinde malumat sahibi olunan bir eğitim değeri, model meydana getirmek için kategorize etme prosedürü ile çözümlenir. Öğrenilen model, sınıflama normları ya da karar ağacı şeklinde ifade edilir. Bunun devamında ise kategorize edilme işlemi gerçekleşir. Kategorize etme aşamasında test verisi, sınıflama normlarının ya da karar ağacının

doğruluğunu göstermek için işe koşulur. Eğer doğruluk makul ölçülerdeyse ilgili normlar diğer değerlerin kategorize edilmesi için yeniden işe koşulur [2].

Karar ağaçları geçmişteki bilgileri esas alarak yeni girdilerin hangi sınıfa ait olduğuna, belirli kurallar oluşturarak karar verme sürecidir. Karar ağacı, sorular ve bu sorulara verilen yanıtlar doğrultusunda hareket ederek, elde edilen yanıtları bir araya getirerek bazı kurallar oluşturur. Ortaya çıkan ağaç “eğer-ise”(if-then)’den oluşan kurallar bütünü olarak adlandırılabilir. Kök düğümünü oluşturmak amacıyla soru sormaya hangi değişkenden başlanılacağına karar verilmesi gerekir. Diğer bir ifadeyle bu kararın sonucunda kök düğüm oluşturulmuş olur. Cevabı veri tabanında bulunan sorular kök düğümden başlayarak sorulur ve alınan cevaplara göre yeni düğümler oluşturulur. Her düğüm kendisinden sonra en az iki dala ayrılmaktadır. Eğer oluşan düğümden sonra yeni sorular gelmiyorsa dallanma sona ermiştir ve bir sınıfı temsil eden yaprağa ulaşılmıştır [28]. Şekil 3.5’ de örnek bir karar ağacı yapısı gösterilmiştir.



Şekil 3.5 Eğitim verilerine uygun karar ağacı

Karar ağaçları herhangi bir sınıfa ait olan elemanların belirlenmesi, çeşitli risk durumlarına göre yüksek, orta, düşük şeklinde farklı sınıflamalara bölünmesi, gelecekte gerçekleşebilecek durumların öngöründe bulunulabilmesi amacıyla çeşitli kurallar oluşturulması, kategorilerin birleştirilmesi gibi alanlarda sıkça kullanılmaktadır. Bunların dışında karar ağacı tekniğinin sıkça kullanıldığı bazı uygulama alanları aşağıda sıralanmıştır [29]:

- Hangi insani oluşumların mektup vasıtasıyla gerçekleştirilen pazarlama uygulamalarında belirgin şekilde yüksek yanıtlanma oranına sahip olduğunun ortaya çıkarılması,
- Kredi geçmişlerinin incelenmesi neticesinde kişilere dair kredi kararlarının alınması,

- Kurumsal fayda dikkate alarak kuruma katkısı yüksek kişilerin özelliklerinden faydalanarak yeni personellerin işe kabul süreçlerinin ortaya çıkarılması,
- Tıp alanına dair inceleme bilgilerinden yola çıkarak en faydalı kararların alınması,
- Satış sürecini etkileyen varyantların yapmış olduğu etkinin belirlenmesi,
- Ürün hatalarına sebep olan varyantların tespit edilmesi.

Karar ağaçlarının sayısal ve sayısal olmayan verilerle işlem yapabilme, kendini açıklayabilecek özellikte olma, eksik değerleri içeren veri setlerini analiz edebilme ve bu analizlerin kolayca yorumlama gibi birçok avantajı bulunmaktadır. Bu avantajlarının yanı sıra, karar ağaçlarının bağımlı değişkenler için kesikli değerler gerektirmesi, sınıflandırma sonucunda kök düğümlerinde daha az bilgi kalacağı bulunacağından karar vermek, karar ağacı yaklaşımı için ciddi bir sorun olabilmesi gibi birçok dezavantajı da bulunmaktadır [26].

Karar ağaçlarının avantajları ve dezavantajları şu şekilde açıklanmaktadır [29, 30]: Karar ağaçlarının avantajları:

- Karar ağacından elde edilen veriler hakkında fikir belirtmek ve analiz yapmak son derece kolaydır.
- Karar ağaçları vasıtasıyla makul yönergeler elde etmek kolaydır.
- Karar ağaçlarında varyantların özelliği pek değer taşımaz yani ayrık ve bütüncül değerlerin kullanılması manasında son derece kullanışlıdır.
- Karar ağaçlarının genişliği yüksek oranda olsa da doğal süreçte kolaydır.
- Karar ağaçları ihtiyaç duyulan değişimleri sağlama açısından elverişlidir.
- Karar ağaçlarının ortaya çıkardığı yaprak adedi olması gereken miktarda ise fazla bir çaba göstermeksizin kavranabilir özelliktedir.
- Karar ağaçları eksik veri setlerini tamamlama açısından başarılıdır.
- Sezgisel düşünce metodu ile karar ağacını kavramak mümkün olabilmektedir.
- Çok varyantlı tekniklerde istatistik varsayımların oluşturulması elzem olsa da bu durum karar ağaçları için gerekli değildir.
- Karar ağacı neticesinde sağlanan bütün kurallar uygulanmalıdır.
- Karar ağaçları algoritmalarında bağımlı ve bağımsız değişkenler arasındaki bağların durumunun görselleştirilebilir olması.

Zayıf yönleri ise,

- Devamlı varyant değerlerini öngörmeye fazla fonksiyonel olmaması,
- Kategori sayısı çok ve öğrenme kümesi örnekleri sayısı düşük olduğunda model meydana getirmesinin yeterli olmaması,
- Vakit ve konum karmaşayı kavrama kümesi örnekleri adetine, varyant miktarına ve ortaya çıkan ağacın çatısına tabiidir,
- Ağaç meydana getirme kompleksliği ve ağaç budama kompleksliği fazladır.

Ağaç tabanlı tekniklerin tabanının meydana getiren karar ağaçları formlarının birincil işletimleri AID (Automatic Interaction Detector) prosedürüyle gerçekleştirilmiş ve çeşitli algoritmalar ile sürdürülmüştür. Geliştirilen bu algoritmalar içerisinde, Exhaustive CHAID, CART (Classification and Regression Trees), ID3 (Iterative Dichotomiser 3), C4.5, CHAID (Chi-Squared Automatic Interaction Detector), C5.0, MARS (Multivariate Adaptive Regression Splines), SLIQ (Supervised Learning in Quest), QUEST (Quick, Unbiased, Efficient Statistical Tree), SPRINT (Scalable Parallelizable Induction of Decision Trees) başlıcalarıdır [31].

### 3.5.2. Yapay sinir ağları

İnsan beyninin temel niteliklerinde bulunan; öğrenme yöntemiyle modern bilgiler yaratabilme, yeni bilgiler meydana getirebilme ve bulmak gibi mezziyetleri bir yardım almaksızın kendiliğinden meydana gelmesi için oluşturulan bilgisayar sistemi yapay sinir ağları olarak adlandırılmaktadır [32].

Yapay sinir ağlarının basit anlamda tanımlamak gerekirse beyinde bulunan sinirlerin çalışma şekline yola çıkarak sistemlere öğrenme, malûmat kazanma, bilgiler arası irtibat ve bağ yaratma, hatırlamak gibi mezziyetleri edinmeyi güden yan bilgi prosedürleridir.

Yapay sinir ağlarının kullanılan ağ modeline göre bazı farklı özellikleri sahip olsa da her durumda geçerli olan temel özellikleri şu şekilde sıralanabilir [33]:

- Yapay sinir ağları makine öğrenmesi gerçekleştirirler.
- Programlama yöntemleri klasik programlama yöntemleri ile benzerlik göstermez.
- Bilgileri saklarlar.

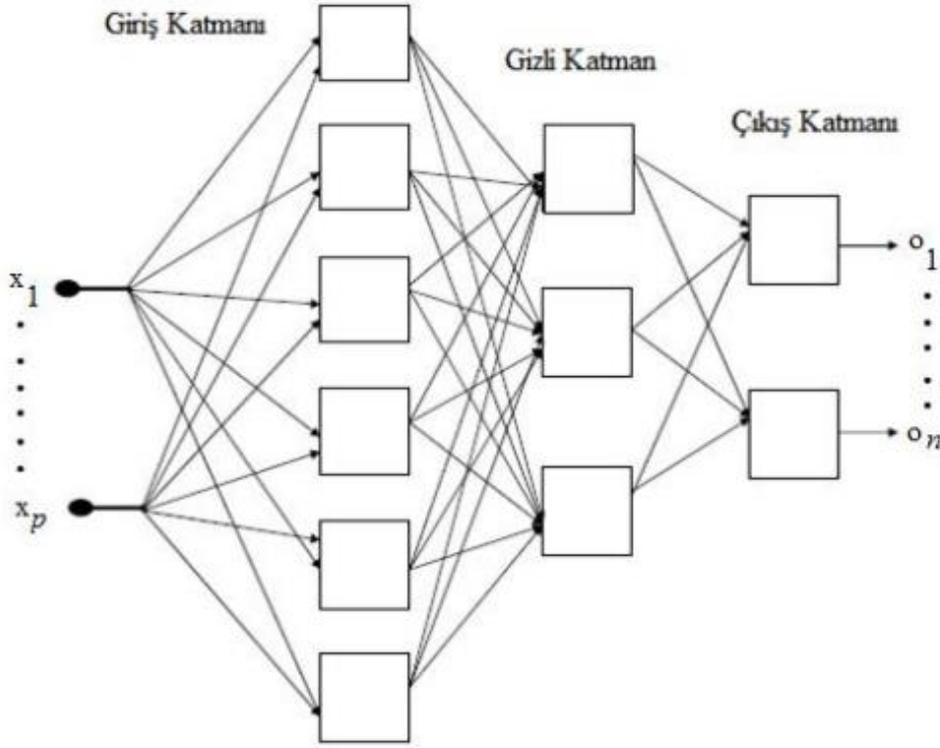
- Örnekleri kullanırlar ve bu örnekler sayesinde öğrenirler.
- Güvenle çalıştırılabilmeleri için ağın eğitimi ve performansı önemlidir.
- Bilinmeyen örnekler hakkında bilgi üretirler. Bu yönleriyle genelleme özelliğine sahiptir.
- Kavramaya yönelik hâdiselerde işe koşulabilirler.
- Şekil ilişkilendirebilme ve sınıflandırma yapabilirler.
- Kendi kendine öğrenme meziyetleri söz konusudur.
- Noksan veriyle çalışabilme özelliği vardır.
- Bozukluk ve hengâmeye karşı hassaslığa ve hoşgörüyeye sahiptir.
- Tam olmayan müphem olan verileri de işlemeleri mümkündür.
- Dereceli bozulma gösterirler.
- Sadece nümerik bilgiler ile çalışabilmektedirler.
- Lineer ve lineer olmayan problemleri çözümünde kullanılmaktadır.

Yapay sinir bağlarının bir olayla ilgili hammadde ve ürünler arasındaki irtibatı, doğrusal olsun ya da doğrusal olmasın, var olan numunelerden bilinerek daha evvel hiç rastlanmamış hâdiseleri, evvelki olaylardan ya da numunelerden esinlenerek alâkalı hadiseye çareler bulabilmesi, prosedürel çözüme ulaşamamış sorunların üstesinden gelinmesinde işe koşulmasını popüler hale getirmiştir.

Yapay sinir ağları, yapay sinir hücrelerinin birleşmesinden oluşmaktadır. Sinir hücrelerinin bütün oluşturması tesadüfü değildir. Çoğunlukla hücreler 3 tabaka şeklinde ve her tabaka içinde özdeş şekilde birleşerek ağı meydana getirirler. Bu tabakalar [34]:

- Hammadde tabakası: Bu tabakadaki işleyiş öğeleri harici verileri edinerek ara tabakalara aktarmakla görevlidir. Bazı ağlarda hammadde tabakasında herhangi bir bilgi işleme söz konusu değildir.
- Gizli tabakalar: Hammadde tabakasından aktarılan veriler üzerinde çalışılarak ürün tabakasına transfer edilirler. Bu verilerin çalışılması ara tabakalarda meydana gelir. Bir ağ için birden çok ara tabaka bulunabilir.
- Ürün tabakası: Bu tabakadaki işleyiş öğeleri ara tabakadan aktarılan verileri çalışarak ağın hammadde tabakasından getirilen girdi seti (örnek) için oluşturması beklenen ürünü oluştururlar. Oluşturulan bu ürün dış dünyaya aktarılır.

Örnek bir yapay sinir ağları modeli Şekil 3.6' da gösterilmiştir.



Şekil 3.6 Yapay sinir ağları

Yapay sinir ağları, çeşitli alanlarda sorunları çözmek için kullanılabilecek esnek ve çok güçlü sayısal araçlardan biridir. Yapay sinir ağlarını kullanmanın amaçları arasında örüntü tanıma ve eşleştirme, sınıflandırma, optimizasyon ve kontrol örüntü tamamlama, sayılabilir [35].

Matematiksel olarak modellenmesi çok zor olan soruları kolayca çözümlenebilir olanağına sahiptirler. Yapay sinir ağlarının avantajları aşağıdaki sıralanmıştır [36]:

- Kural tabanı kullanımı gerektirmezler.
- Öğrenme konusunda yeteneklidir ve çeşitli öğrenme algoritmalarıyla kolayca öğrenebilirler.
- Matematiksel olarak modellenmesi kolay olmayan problemler rahatça çözülebilir.
- Herhangi bir sorun için ağ modelleneninde herhangi bir bilgiye gereksinim duymaz sadece konu ile ilgili örnek olması yeterlidir. Örnek bulmaksa bilgi bulmaktan kolaydır.
- Gerçek hayatta yaşanan olaylar ve bu olayların arkasındaki bağlantıları bulmak kolay değildir. Ancak bu olaylar arasında bulunan bağlantı yapay sinir ağları ile otomatik olarak yapılır.

- Yapay sinir ağı uygulamaları pratik olmasının yanı sıra çokta ekonomiktir. Sadece sorun ile ilgili örneklerin toplanması yeterlidir. Basit bir program aracılığıyla problem çözülebilir.
- Olaylar ve olaylar arasındaki bağlantılar doğrusal olmadığı için bu tür problemleri modellemek ve çözmek zordur. Çözüm için bazı varsayımlarda bulunmak gerekir. Bu da modellenen sistem ile gerçek sistem arasında çeşitli farklılıklar göstermesine sebep olur. Hâlbuki yapay sinir ağı doğrusal olmayan bağlantılar için geleneksel yöntemlerden daha iyi ve gerçekçi çözüm önerileri üretir.
- Zaman bakımından Yapay sinir ağını diğer sistemlere karşılaştırdığımızda yapay sinir ağlarının daha verimli olduğu görülmektedir. Örneklerin bulunması, problemlere uygun biçimde ağı oluşturulması, ağı öğrenmesi, diğer örnekler için kullanıma geçilmesi çok kısa sürede gerçekleştirilebilmektedir.
- Yapay sinir ağı aynı olay için yeni bilgilerin olması ve bazı değişiklikler söz konusu olduğunda yeniden eğitilip ortama uyum sağlayabilirler.
- Yapay sinir ağı, aynı olay için yeni bilgiler ve bazı değişiklikler olması durumunda yeniden eğitilerek çevreye yeniden uyarlanabilir
- Paralel olarak çalışan yapay sinir ağı gerçek-zamanlı uygulamalar için rahatça kullanılabilir.

Yapay Sinir Ağlarının dezavantajları aşağıdaki gibidir [36]:

- Sistemin içerisinde ne olduğu bilinemez.
- Birkaç ağ dışında kararlılık analizi yapamaz.
- Farklı sistemlere uyarlanması kolay olmayabilir.
- Yapay sinir ağı ile bir uygulama geliştirirken, model seçerken ve ağların topolojisini belirlerken özel bir kural yoktur. Doğru seçimleri yapmak tamamen kullanıcının deneyimine bağlıdır.
- Ağı davranışlarının açıklanması mümkün olmadığı için ağa güven azalmaktadır.
- Bazen eğitim gerçekleşmesi çok uzun zaman diliminde gerçekleştirilebilir.
- Bulunan çözümün en uygun çözüm olduğunu söylemek mümkün değildir. En iyi çözümlerden sadece bir tanesi olduğu söylenebilir. Bununla birlikte geleneksel yöntemler en uygun sonuçlar üretebilirler.
- Örnekleri bulmanın zor olduğu durumlarda veya doğru örneklere karar verilmediğinde sağlıklı bir çözüm üretilemez.

### 3.5.3. Naive bayes

18. yüzyılda Thomas Bayes tarafından ortaya konulan yaklaşımlar, Naive bayes algoritmasının mantıksal temellerini oluşturmaktadır. Naive bayes yöntemi genellikle olayların ihtimallerinin değerlendirilmesi için kullanılmaktadır. Ayrıca her yeni ek bilgiyi yeniden incelemek için kullanılan yöntemdir [37].

Naive bayes, metinlerin kategorize edilmesi konusunda bilinen ve çok sık kullanılan bir metottur. Hedef fonksiyona ulaşmak amacıyla eğitim veri kümesi belirlenmekte, nitelik değerleri tarafından tanımlanan yeni örneklemeler sunulmakta ve öğrenci sınıf hedef değeri veya sınıfı tahmin etmektedir [38].

Bir olayın gerçekleşme ihtimali, başka bir olayın gerçekleşme ihtimaline etkide bulunabiliyorsa burada koşullu olasılıktan bahsetmek mümkündür. İki olay arasında bu biçimde bir bağlantı bulunması, bir olayın gerçekleşmesi diğer bir olayın gerçekleşmesini engellemekle yani bağdaşan iki olay olmakla beraber olayın yalnız başına gerçekleşme olasılığı ile başka bir olayın meydana gelme bilgisinin var olduğu durumda gerçekleşme ihtimalinin değişmesi, şarta bağlı bir olasılık hesabını gerektirmektedir. Bağdaşan X ve Y olayları ele alınırsa; Y olayının meydana gelme olasılığı X olayının meydana gelmesine bağlıdır. Bu koşullu olasılık şu şekilde belirtilir:

$$P(Y|X) = \frac{P(Y \cap X)}{P(X)} \quad (3.1)$$

Bu eşitlikten yararlanılarak olayların kesişim olasılığı elde edilir.

$$P(X \cap Y) = P(Y|X).P(X) \quad (3.2)$$

Aynı mantıkla hareket ederek X olayına ilişkin olasılık ve sonucunda kesişim olasılığı aşağıdaki şekilde elde edilir.

$$P(X|Y) = \frac{P(X \cap Y)}{P(Y)} \quad (3.3)$$

$$P(X \cap Y) = P(X|Y).P(Y) \quad (3.4)$$

Böylece Y olayının koşullu olasılığı şu şekilde bulunur:

$$P(Y|X) = \frac{P(X|Y) \cdot P(Y)}{P(X)} \quad (3.5)$$

Bayes teoremi istatistik ve olasılıkta büyük bir öneme sahiptir. Bu teorem koşullu olasılıkların hesaplanmasında kullanılmaktadır. Teoremden faydalanılarak sınıflandırma işlemi yapılabilen, koşullu olasılıklar aşağıdaki gibi hesaplanabilmektedir:

P(X): X olayının gerçekleşme olasılığı

P(Y): Y olayının gerçekleşme olasılığı

P(X|Y): Y olayından sonra X olayının gerçekleşme olasılığı

P(Y|X): X olayından sonra Y olayının gerçekleşme olasılığı

X olayının önsel olasılığı, tüm ayrık parçalarının birleşimidir [39].

$$P(X) = P(X \cap Y) + P(X \cap \bar{Y}) \quad (3.6)$$

$$P(Y|X) = \frac{P(X \cap Y)}{P(X)} \quad (3.7)$$

P(Y|X) eşitliği, koşullu olasılık eşitliğini kullanarak açıldıktan sonra bayes formülü elde edilir. Elde edilen bayes formülü, iki ve daha fazla olay için aşağıdaki gibi genelleştirilebilir:

$$P(X|Y) = \frac{P(X \cap Y)}{P(X \cap Y) + P(X \cap \bar{Y})} \quad (3.8)$$

$$P(Y|X) = \frac{P(X|Y) \cdot P(Y)}{P(X|Y) \cdot P(Y) + P(X|\bar{Y}) \cdot P(\bar{Y})} \quad (3.9)$$

$$P(Y_i|X) = \frac{P(X|Y_i) \cdot P(Y_i)}{\sum_{j=1}^n P(X|Y_j) \cdot P(Y_j)} \quad (3.10)$$

Naive bayes algoritmasının sıkça kullanılmasını sağlayan bazı avantajları bulunmaktadır.

Bu avantajlar:

- Anlaşılması kolaydır
- Uygulanması kolay bir metottur.
- Eğitimi çok hızlıdır.
- İki ya da çoklu sınıflamalar için kullanılabilir.
- Sınıflandırma performansını artırmak için ilişkili olmayan özellikleri ortadan kaldırır.
- Olasılık tahmini durumunda, kayıp değerle mücadele eder.

- Hesaplama süresi kısa olduğundan dolayı hızlı çalışır.

Bu metodun avantajların yanı sıra bazı dezavantajları da bulunmaktadır. Bu dezavantajlar:

- Büyük veri yığınlarına ihtiyaç duymaktadır.
- Tüm eğitim datalarını sakladıkları için tembeldir.

#### 3.5.4. Destek vektör makinesi

İlk kez 1992 yılında COLT (Computational Learning Theory) konferansında Vladimir Vapnik ve çalışma arkadaşları tarafından önerilen destek vektör makinesi, günümüzde kullanım biçimini, Corinna Cortes ve Vladimir Vapnik tarafından 1995 yılında almıştır. 1997 yılında ise, destek vektör makinesi metodu yine Vapnik ve arkadaşları tarafından regresyon uygulamalarını kapsayacak bir biçimde genişletilmiş ve bu tarihten sonra gelişen bilgisayar teknolojisi ile birlikte kullanımı artmaya başlanmıştır [40].

Destek vektör makinesi algoritması son zamanlarda sınıflandırma problemlerinin çözümü için kullanılan yaygın olarak kullanılan algoritmalarından birisidir. Sınıflandırma problemleri çözüldükçe, Destek vektör makinesi algoritmasının başarı oranının çok yüksek çıkması bu algoritmanın yaygın bir şekilde kullanılmasını sağlamıştır. Destek vektör makinesi yöntemini diğer yöntemlerden ayıran özelliklerinden birisi de, algoritma yapısının karışık olmaması ve öğrenme sırasında gerçekleşen işlem sayısının az olmasıdır [41]. Bu nedenlerden dolayı büyük veri setlerinin sınıflandırması için son derece kullanışlı bir metottur.

Başlangıçta destek vektör makinesi sadece sınıflandırma problemlerinin çözümünde kullanılmasına rağmen daha sonra yapılan çalışmalarda özellikle zaman serilerinin tahmin edilmesi ve regresyon uygulamalarında çok iyi sonuçlar verdiği görülmüştür.

Destek vektör makinesi algoritmasından karşılaşılabilecek iki durum, verilerin doğrusal olarak ayrılabilen bir yapıda olması veya doğrusal olarak ayrılamayan yapıda olmasıdır.

Bununla birlikte gerçek yaşamda karşılaştığımız problemlerin çoğu, farklı bileşenlerden meydana gelip doğrusal olarak ayrılmış bir biçimde karşımıza çıkmazlar. Doğrusal olarak ayrılan veriler arasında doğrudan maksimum sınırının bulunması oldukça basittir, ancak

doğrusal olarak ayrılmayan veriler öncelikle doğrusal olarak ayrılacakları farklı bir uzaya aktarılmalıdırlar [42].

### 3.5.5. Random forest

1996 yılında Breiman tarafından önerilen Bagging tekniği ile 1998 yılında Ho tarafından önerilen Random Subspace tekniğinin birleşmesiyle Random forest yöntemi meydana gelmiştir. Bagging metodu ile karar ağaçları oluşturulurken örnekler data setinden bootstrap tekniği ile seçilmektedir. Bu sayede karar ağaçları bağımsız bir şekilde oluşturulmaktadır. Random Subspace metodunda ise karar ağacı her bir düğümündeki en iyi dallanma değişkeninin, tüm değişkenler arasında rastgele seçilen birkaç değişkenden seçilmesiyle oluşur [28].

Random forest metodunda sonradan eklenen verinin tahminlenmesiyle birlikte, değişkenlerin önem derecesi de hesaplanmaktadır. Veri kümesinde çokça değişken bulunuyorsa bu değişkenlerin önem derecesinin hesaplanması model indirgemesi açısından oldukça kullanışlıdır. Örneğin yüzlerce değişkenin bulunduğu veri kümesinde, Random forest metoduyla bulunan önem derecesine göre, kurulacak yeni modelde önem derecesi yüksek olan değişkenler kullanılarak daha doğru tahminlerin yapılması mümkün olabilir.

Random forest yöntemi 2 parametre üzerine inşa edilir. Bu parametreler, oluşturulacak ağaç sayısı (B) ve her düğümden rastgele seçilecek olan tahmin sayısıdır (m). Karar ağaçları oluşturulurken, orijinal veri kümesinden gözlem sayısı (n) ile aynı sayıda olacak şekilde bootstrap metodu ile örneklem oluşturulur. Ağacı oluşturulurken örneklemelerin 2/3'ü eğitim veri kümesi (InBag), 1/3'ü ise modelin iç hata oranını hesaplayabilmek için test veri kümesi (out of bag veya OOB) olarak kullanılır.

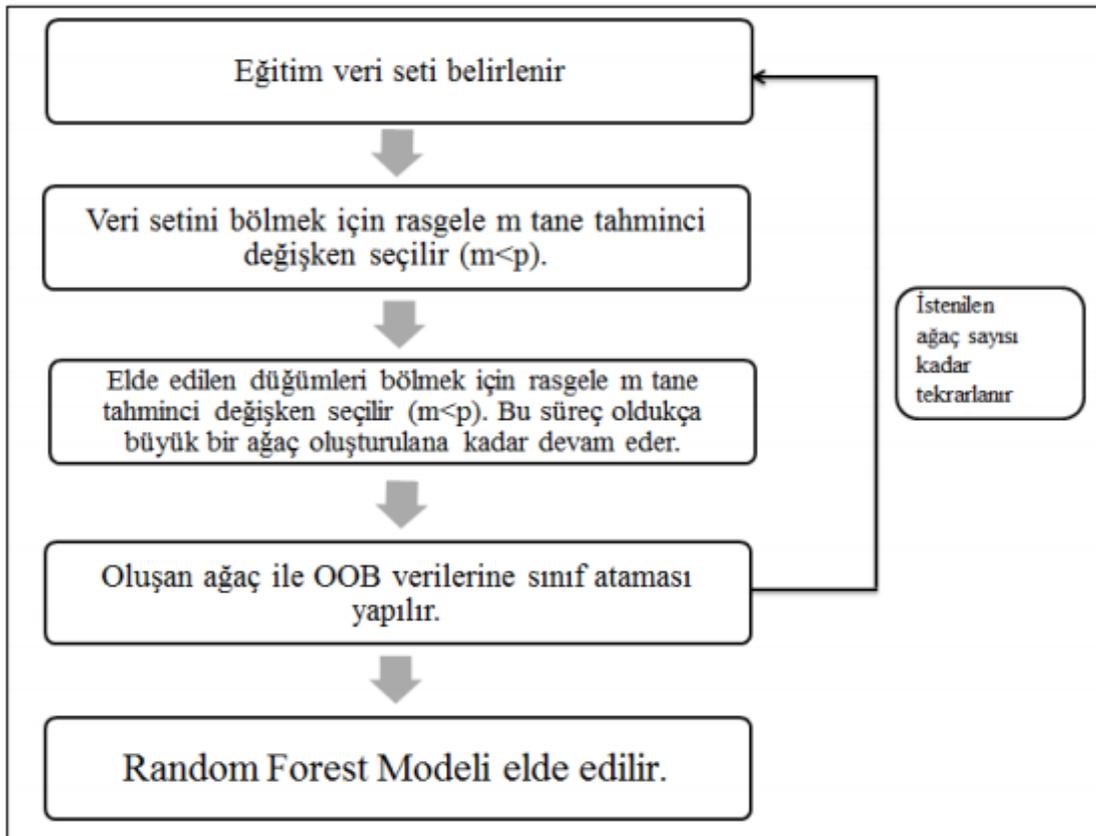
Random forest modeli şu şekilde kurulur:

- Öncelikle bootstrap metodu veri seti oluşturulur. Bu veri seti, InBag (eğitim veri seti) ve OOB (test veri seti) olarak iki kısma ayrılır.
- InBag (eğitim veri seti) aracılığıyla olabildiğince büyük bir karar ağacı (CART) oluşturulur. Oluşturulan bu karar ağacı budanmaz. Bu ağaç oluşturulurken her düğüm ayırımında toplam p tane tahmin edici değişkenden m tanesi random bir biçimde seçilir. Burada  $m < p$  koşulu mutlaka sağlanmalıdır. Çünkü ağacın çok fazla büyümesi ve fazla

uyum gözlenmesi isten bir durum değildir. Seçilen bu m tane tahmin ediciden bilgi kazancı en yüksekten başlanarak dallara ayrılma gerçekleşir. Belirlenen bu değişkenin hangi değerine göre ayrılacağına Gini indeksi ile karar verilir. Bu işlem her düğüm için yeni meydana gelebilecek dal olmayana kadar tekrarlanır.

- Oluşturulan her yaprak düğüme bir sınıf ataması yapılır. Daha sonra OOB(test veri seti) ağacın en tepesinden bırakılır ve bu veri setinde bulunan her gözleme atanan sınıf kaydedilir.
- İlk 3 adıma kadar tüm aşamalar B sayısı kadar tekrarlanır.
- Ağaç oluşturulurken kullanılmayan test verileri (OOB) ile bir değerlendirme yapılır. İncelenen bir test verisinin hangi kategorilerde kaç defa sınıflandırıldığı sayılır.
- Her test verisi, ağaç setleri üzerinden belirlenen bir oy çoğunluğu ile sınıf ataması yapılır. Örneğin 2 kategoriye ayrılan bir sınıflandırma modelinde, bir test verisi tüm ağaçlar üzerinden oyların çoğunu aldığı sınıfın etiketini taşır ve bu sınıf onun tahmin edilmiş sınıf değeri olur.

Random forest modeli oluşturma algoritması Şekil 3.7' de gösterilmiştir:



Şekil 3.7 Random forest modeli oluşturma algoritması

### **3.6. Veri Madenciliği Yazılımları**

Veri madenciliği algoritmaları ile analiz yapmak için çok sayıda bilgisayar yazılımı geliştirilmiştir. Bunlardan kısmı ticari amaç için yazılırken, bir kısım yazılımlar ise açık kaynak kodlu olarak sunulmuştur.

#### **3.6.1. Spss clementine**

SPSS Clementie ya da IBM SPSS Modeler, IBM tarafından geliştirilmiş bir veri madenciliği ve metin analitiği yazılımı uygulamasıdır. Tahmine dayalı modeller oluşturmak ve diğer analitik görevleri yapmak için kullanılır. Kullanıcılara programlama ve istatistik işlemleri olmadan veri madenciliği algoritmalarından faydalanmalarını sağlayan görsel bir ara yüz sağlar. Başlıca amaçlarından biri, veri dönüşümlerinde gereksiz karmaşıklıktan kurtulmak ve karmaşık öngörü modelleri kullanımını daha kolay hale getirmektir.

SPSS, programlama yapmadan hızlı ve sezgisel olarak doğru tahmin modelleri oluşturmaya yardımcı olan, güçlü, çok yönlü bir veri ve metin analitiği programıdır. Program IBM SPSS Modeler adlı sezgisel grafik ara-yüzünü kullanarak veri madenciliği sürecinde görselleştirme sağlamaktadır. Bu ara-yüzden, hem yapılandırılmış (sayılar ve tarihler) hem de yapılandırılmamış (metin) işlemsel veri-tabanları, anket verileri, dosyalar gibi çeşitli kaynaklardan kolayca erişim ve geliştirmek için modelleme araçları sağlar. Müşteri ilişkileri yönetimi, pazarlama, kaynak planlaması, dolandırıcılık ve risk azaltma, tıbbi araştırma, kanun uygulama ve ulusal güvenlik gibi çeşitli alanlarda yaygın olarak kullanılmaktadır [43].

#### **3.6.2. Rapidminer**

RapidMiner, aynı adı taşıyan şirket tarafından geliştirilen ve veri hazırlama, makine öğrenmesi, derin öğrenme, metin madenciliği ve tahmine dayalı analitik için entegre bir ortam sağlayan bir veri bilimi yazılım platformudur. Araştırma, eğitim, öğretim, hızlı prototipleme ve uygulama geliştirme ile birlikte ticari ve ticari olmayan uygulamalar için kullanılır. Programda veri hazırlama, sonuç görselleştirme, model doğrulama ve

optimizasyon dahil olmak üzere makine öğrenme sürecinin tüm adımlarını destekleyici araçlar bulunmaktadır [44].

RapidMiner, veri madenciliği ve makine öğrenmesi prosedürlerini sağlar: Veri yükleme ve dönüştürme (Çıkar, Dönüştür, Yükle, 'Extract, Transform, Load – ETL'), veri ön işleme ve görselleştirme, tahmine dayalı analitik ve istatistiksel modelleme, değerlendirme ve dağıtım. RapidMiner, Java programlama dilinde yazılmıştır. RapidMiner, analitik iş akışlarını tasarlamak ve yürütmek için bir GUI sağlar. Bu iş akışlarına RapidMiner'da "İşlemler" denir ve birden fazla "Operatör" den oluşur. Her operatör işlem içinde tek bir görevi yerine getirir ve her operatörün çıktısı bir sonrakinin girişini oluşturur. Alternatif olarak, temel fonksiyonları başka programlardan çağrılabilir veya bir API olarak kullanılabilir. Bireysel fonksiyonlar komut satırından çağrılabilir. RapidMiner öğrenme planları, modelleri ve algoritmaları sağlar ve R ve Python scriptleri kullanılarak genişletilebilir.

### 3.6.3. Knime

KNIME Analytics Platform, veri bilimi uygulamaları ve hizmetleri oluşturmak için açık kaynaklı bir yazılımdır. Sezgisel, açık ve sürekli olarak yeni gelişmeleri bütünleştiren KNIME, verileri anlamayı ve veri bilimi iş akışlarını ve yeniden kullanılabilir bileşenleri tasarlamayı herkes için erişilebilir hale getirmektedir.

KNIME, kullanıcıların veri akışlarını görsel olarak oluşturmasını, analiz adımlarının bir kısmını veya tamamını seçerek yürütmesini ve daha sonra sonuçları, modelleri ve etkileşimli görünümleri incelemesini sağlar. KNIME, Java dilinde yazılmıştır ve Eclipse'i temel alır ve ek işlevsellik sağlayan eklentiler eklemek için uzantı mekanizmasından yararlanır. Veri dönüşümü, filtre, dönüştürücü, ayırıcı, birleştirici yanı sıra istatistik, veri madenciliği, analiz ve metin analitiği için kullanılan yöntemleri destekler. KNIME iş akışları, doc, ppt, xls, pdf ve diğerleri gibi belge biçimlerine verilebilecek rapor şablonları oluşturmak için veri kümeleri olarak kullanılabilir [45].

### 3.6.4. Weka

Weka, veri analizi, makine öğrenmesi gibi fonksiyonlara kolay erişim için grafiksel kullanıcı ara-yüzleri ile birlikte öngörü modelleme için görselleştirme araçları ve algoritmaları koleksiyonu içermektedir. Weka' nın orijinal Java olmayan sürümü, diğer programlama dillerinde uygulanan (çoğunlukla üçüncü parti) modelleme algoritmalarının yanı sıra, C' de veri ön işleme programları ve makine öğrenimi deneylerini yürütme işleminde Makefile tabanlı bir sistem için bir platform sağlamaktadır [46].

Weka, daha spesifik olarak veri ön işleme, kümeleme, sınıflandırma, regresyon, görselleştirme ve özellik seçimi gibi birçok standart veri madenciliği görevlerini desteklemektedir. Tüm Weka teknikleri, verinin tek bir düz dosya veya ilişki olarak mevcut olduğu varsayımına dayanır; burada her veri noktası sabit sayıda nitelikle tanımlanır (normalde, sayısal veya nominal özellikler, ancak diğer bazı nitelik türleri de desteklenir). Weka, Java Database Connectivity kullanarak SQL veri-tabanlarına erişim sağlar ve bir veri-tabanı sorgusu tarafından döndürülen sonucu işleyebilir. Weka, Deeplearning4j ile derin öğrenmeye de erişim sağlar. Yoğun ilişkisel veri madenciliğine sahip değildir, ancak bağlantılı veri-tabanı tablolarının bir koleksiyonunu Weka kullanarak işlemek için uygun olan tek bir tabloya dönüştürmek için yazılım modülleri vardır. Weka dağılımında yer alan algoritmaların kapsamadığı bir diğer önemli alan dizi modellemesidir.

### 3.6.5. Sas

SAS ("İstatistiksel Analiz Sistemi"), SAS Institute tarafından geliştirilmiş analitik, çok değişkenli analiz, iş zekası, veri yönetimi ve tahmine dayalı analitik için geliştirilen bir yazılım paketidir. SAS, çeşitli kaynaklardan veri toplayabilen, değiştirebilen, yöneten ve alan ve üzerinde istatistiksel analiz yapabilen bir yazılım paketidir. SAS, teknik olmayan kullanıcılar ve SAS dili aracılığıyla daha gelişmiş seçenekler için grafiksel bir kullanıcı arabirimi sağlar. SAS programları, verileri alan ve işleyen DATA/VERİ adımlarına ve verileri analiz eden PROC adımlarına sahiptir. Her adım bir dizi ifadeden oluşur.

VERİ basamağı, yazılımın bir işlem gerçekleştirmesine neden olan çalıştırılabilir ifadeler ve bir veri kümesini okumak veya verilerin görünümünü değiştirmek için talimatlar sağlayan bildirici ifadelerle sahiptir. VERİ basamağının iki aşaması vardır: derleme ve

yürütme. Derleme aşamasında, bildirici ifadeler işlenir ve sözdizimi hataları tanımlanır. Daha sonra, yürütme aşaması her çalıştırılabilir ifadeyi sırayla işler. Veri kümeleri, "gözlem" olarak adlandırılan satırlar ve "değişken" olarak adlandırılan sütunlar içeren tablolara ayrılmıştır. Ek olarak, her veri parçası bir tanımlayıcıya ve bir değere sahiptir.

PROC adımı, adlandırılmış prosedürleri çağıran PROC ifadelerinden oluşur. Prosedürler, istatistikler, analizler ve grafikler üretmek için veri kümeleri üzerinde analiz ve raporlama yapar. 300'den fazla prosedür var ve her biri önemli miktarda programlama ve istatistik çalışması içermektedir. PROC ifadeleri ayrıca sonuçları görüntüleyebilir, verileri sıralayabilir veya diğer işlemleri gerçekleştirebilir [47].

### 3.6.6. Enterprise miner

Enterprise miner, farklı kaynaklar arasında büyük miktarda veri üzerinde doğru kestirici ve açıklayıcı modeller oluşturmak için bir çözüm ortamı sağlar. SAS; zengin, etkileşimli görselleştirme ve veri keşfi ile modeller oluşturmak için esnek bir veri hazırlama ve yönetim özellikleri sunar. SAS, daha hızlı yanıt ve sonuçlar için bellek içi, veri-tabanı içi ve grid/ızgara yeteneklerini kullanır. Kurumların karmaşık verileri analiz etmelerini, yararlı bilgiler bulmalarını ve güvenle hareket etmelerini sağlayan yüksek teknoloji tahmin analizleri ve veri madenciliği yeteneklerine sahiptir. Enterprise Miner'ın grafik ara-yüzü, kullanıcıların beş aşamalı SAS SEMMA (örnekleme, keşif, değiştirme, modelleme ve değerlendirme) yaklaşımını kullanarak veri madenciliği sürecinde mantıksal olarak hareket etmelerini sağlar. Bu, veri madencileri ve iş analistlerinin doğru tahminde ve tanımlayıcı modeller oluşturmak için daha hızlı çalışabilmesi ve iş süreçlerinde paylaşılacak ve birleştirilebilecek sonuçlar üretebilmesi için optimize edilmiş bir mimariyi sunar. Kullanımı kolay, sürükle ve bırak ara-yüzü, analitik profesyonellere hitap etmek için tasarlanmıştır. Gelişmiş analitik algoritmalar, başarılı bir veri madenciliği girişiminde gerçekleştirilen temel görevler altında düzenlenir. R dili kodu, bir SAS Kurumsal Madenci süreç akış şemasının içine entegre edilebilir. Bu, veri dönüşümü ve keşif işlemlerinin yanı sıra R'de denetlenen ve denetimsiz modelleri eğitme ve puanlama gerçekleştirilmesini de sağlar [48].

### 3.6.7. Kxen

InfiniteInsight, KXEN tarafından geliştirilen ve analitik profesyonellere yardımcı olan öngörücü bir modelleme paketi ve verilerden bilgi almak için bir platform sağlar. Diğer işlevler arasında, Platform, Java Veri Madenciliği ara-yüzü tarafından tanımlanan ve ifade edilen değişken önem, sınıflandırma, regresyon, segmentasyon, zaman serileri, ürün önerileri ve sosyal ağ analizi için kullanılır. Program bir davranışın veya değer tahminini, bir zaman serisinin tahminini veya benzer davranışa sahip bir grup davranış örüntüsünün anlaşılmasını sağlar. Gelişmiş işlevleri arasında davranışsal modelleme, model kodunu farklı hedef ortamlara verme veya SAS veya SPSS veri dosyalarının üstüne öngörücü modeller oluşturma yer alır [49].

### 3.6.8. Insightful miner

Insightful Miner, karar vericilere iş performansını artırmak ve finansal riskleri azaltmak için gereken yanıtları veren analitik uygulamalar oluşturmak için ölçeklenebilir bir veri analizi platformudur. Insightful Miner, tüm veri analizi sürecini destekleyerek, geniş bir kuruluş yelpazesinin faaliyetlerinde maliyetleri düşürmek ve üst düzey istikrar ve büyümeyi teşvik etmek için tahmine dayalı analitik olarak dağıtılmalarını sağlar.

Insightful Miner, modern ve sezgisel görsel çalışma ara-yüzü ile sofistike veri analizi uygulamaları oluşturma sürecini hızlandırır ve basitleştirir. Sağladığı kolaylıklar ile yeni veri analistleri bile geleneksel yöntemlerle gereken sürenin bir bölümünde titiz analitik uygulamalar oluşturabilir. Özel programlama uzmanlığına ihtiyaç duyulmaz, böylece mevcut kaynaklarınızı kullanarak derhal sonuç alabilmektedir [50].

### 3.6.9. Affinium model

Affinium Model, otomatik tahmin modelleme çözümleri koleksiyonudur. Genel olarak, Affinium Model, tüm veri-tabanı pazarlama modelleme ve profil oluşturma ihtiyaçları için eksiksiz bir çözüm sunar. İhtiyaç olan belirli bir modülü, ihtiyaç duyulduğunda seçebilir veya modülleri birlikte kullanılabilir.

Affinium Model, iki sunucu gerektiren web tabanlı bir uygulamadır: Arka uç sunucu ve web sunucusu. Sunucuları kurduktan sonra, Affinium Modelini makinenizde en iyi şekilde çalışacak şekilde yapılandırmak gerekebilir. Ardından Affinium Model'i, tarayıcı ve ağ sunucusuna ağ erişimi olan herhangi bir makineden çalıştırabilirsiniz [51].

### **3.6.10. Statistica data miner**

Statistica, orijinal olarak StatSoft tarafından geliştirilen ve Mart 2014'te Dell tarafından satın alınan bir analitik yazılım ürünleri ve çözümleri paketidir. Yazılım, çeşitli tahmine dayalı modelleme, kümeleme, sınıflandırma ve keşif teknikleri yanında, bir dizi veri analizi, veri yönetimi, veri görselleştirme ve veri madenciliği prosedürlerini içerir. Ücretsiz, açık kaynaklı R programlama ortamı ile entegrasyon yoluyla ek teknikler kullanımına olanak sağlayabilir [52].

### **3.6.11. Inlen**

Bu platform, veri tabanlarından bilgi elde etmek için INLEN adında geniş çaplı, çok tipli bir akıl yürütme sisteminin geliştirilmesi ile ilgilidir. Sistem, genel veri-tabanları veya eğilimleri, anlamlı ilişkileri, kavramsal veya sayısal düzenleri veya büyük veri-tabanlarındaki anomalileri keşfetmesinde kullanıcıya yardımcı olur. Veri-tabanındaki bilgi hacmi genellikle bir veri analistinin bu tür kalıpları veya düzenlilikleri tespit edebilmesi için çok büyüktür. INLEN, sembolik öğrenme ve istatistik tekniklerini veri-tabanı ve bilgi tabanı teknolojilerine entegre eder. Bir veri kümesini karakterize eden kuralları keşfetmek, anlamlı kavramsal sınıflandırmalar oluşturmak, benzerlikleri tespit etmek ve kurallar için açıklamaları oluşturmak, verileri karakterize etmek, yeni ilgili değişkenleri seçmek ve/veya üretmek için bir kullanıcıya “bilgi oluşturma operatörleri” (KGO'lar) sunar [53].

### **3.6.12. Dbminer**

Büyük ilişkisel veri tabanlarında çok seviyeli bilginin etkileşimli madenciliği için bir veri madenciliği sistemi olan DBMiner, yıllarca süren araştırmalarımıza dayanarak geliştirilmiştir. Sistem, genelleme, karakterizasyon, ayrımcılık, ilişkilendirme, sınıflandırma ve tahmin dâhil olmak üzere geniş bir yelpazede veri madenciliği işlevleri sağlar. Öznitelik odaklı indüksiyon, madenciliğin çok seviyeli kurallar için aşamalı

derinleşmesi ve meta-kural güdümlü bilgi madenciliği dâhil olmak üzere birçok ilginç veri madenciliği tekniğinin eklenmesiyle, sistem, iyi bir performans ile kullanıcı dostu, etkileşimli veri madenciliği ortamı sağlamaktadır [54].

### 3.6.13. Darwin

Darwin, evrimsel farklılıklar temelinde çeşitlilik ve filo genetik analiz için geliştirilmiş bir yazılım paketidir. Ağaç ve faktöring temsili için standart yöntemler önerilmiş, özellikle veri doğruluğuna duyarlılık sorununu ele alan özgün ve özel yaklaşımlarla geliştirilmiştir. Farklı veriler için çeşitli farklılıklar ve uzaklık tahminleri önerir: nicel, nitel, ikili, DNA dizisi gibi. Farklılıkların özellikleri büyük ölçüde araştırılmış ve uygun özellikleri geri kazanmak için dönüşümler önerilmiştir. Temel Koordinat analizi, birimler arasındaki mesafeyi en iyi şekilde koruyan Öklid planlarında grafiksel gösterimler üretir [55].

Ağaç yapım yöntemleri, çeşitli toplama kriterleri (ağırlıklı veya ağırlıksız), Komşu Birleştirme ağacı (ağırlıklı veya ağırlıksız), Puanlar yöntemini içeren hiyerarşik ağaçları içerir. NJTree ve Skorların sıralı uzantıları, veri hatasına duyarlılığı azaltmaya çalışır. Topolojik kısıtlamalar altındaki NJTree, bazı veri alt kümelerinin önceden bilinen bir ağaç yapısının zorlanmasına izin verir. NJTree'deki önyükleme ve etkili birimleri tespit etmek için orijinal bir yöntem, ağacın veriler tarafından nasıl desteklendiğini tahmin etmek için kullanılabilir.

## 3.7. Eğitimde Veri Madenciliği Alanında Yapılan Çalışmalar

Eğitimde veri madenciliği alanında yapılan çalışmalar öğrencilerin eğitim sürecinde kaliteyi artırmanın yanı sıra bu bireylerin eğitimden aldıkları katkıyı en üst seviyeye çıkarmak amacıyla yapılmaktadır. Son zamanlarda sıklıkla yapıldığını gördüğümüz bu alanda gerçekleştirilmiş olan bazı çalışmalar kaynak ve içerik manasında incelenmiştir.

Bresfelean ve arkadaşları (2008) tarafından yapılan çalışmada veri madenciliği modellerinden olan veri kümeleme ve sınıflandırma metotlarından faydalanılarak öğrenci durumları belirlenmeye çalışılmış ve ders başarısızlığı sebeplerinin ortaya çıkarılarak öğrenci başarısızlıklarının önlenmesi ve giderilmesi amaçlanmıştır [56].

Şengür ve Tekin (2013) tarafından yürütülen çalışmada ise Fırat Üniversitesi Bilgisayar ve Öğretim Teknolojileri Öğretmenliği programını 2011 yılında bitiren 127 öğrencinin lisans eğitimi boyunca sorumlu oldukları 49 kültür ve mesleki derse ait dönem sonu notları VM yöntemlerinden ikisi kullanılarak mezuniyet notlarının tahmin edilmesi ile gerçekleştirilmiştir [57].

Bozkır ve arkadaşları (2009) tarafınca gerçekleştirilen çalışmada, 2008 yılında Öğrenci Seçme Sınavına katılmış olan 10.000 öğrencinin sosyal statüleri, eğitim süreçleri ve anne-baba eğitim seviyesi gibi bilgiler öğrenilmiş ve VM metotları ile incelenmiştir. ÖSS’de başarı durumunu gösteren eşit ağırlık, sözel ve sayısal puanları ve bu puan türlerinin Türkiye geneli başarı sırasıyla OÖBP ve AOBP gibi sonuçlarına doğrudan etki eden en önemli unsurlar belirlenmiş ve araştırma neticesinde dikkat çekici bağıntılar elde edilmiştir [58].

Bray (2007), özel ders ve olası sonuçlarının incelemek üzere gerçekleştirdiği çalışmasında, Hindistan’da özel ders imkanına sahip öğrencilerin başarı seviyelerinin Singapur, Çin, Malezya, Sri Lanka ve Japonya’daki öğrencilere nazaran göreceli şekilde yüksek olduğu sonucuna ulaşmıştır. Ayrıca özel dersin sıklığına bağlı olarak ders başarısında yükselme olduğu gözlemlenmiş ve bu özel ders sıklığının değişkenleri; sosyoekonomik imkanlar olarak adlandırılan toplu unsura dayandırılmıştır. [59].

Kristiansen v.d (1991 akt. Ramaswami ve Bhaskaran, 2010), ergenlerde, öz saygı, vücut kitle endeksi (BMI), sağlık tutumları ve akademik başarı arasındaki bağlantıyı ortaya koyabilme amacıyla bir çalışma yürütülmüştür. Bu çalışmada, İzlanda’da 6346 ergen birey ile araştırma yapılmış ve düşük vücut kitle endeksi, fiziksel hareket ve iyi beslenme davranışlarının eğitim alanında başarılı olma açısından oldukça bağlantılı olduğu sonucuna ulaşılmıştır [60].

Can (2017), tarafından yürütülen çalışmada bir kamu eğitim kurumunun öğrencilerine dönem sonu ders değerlendirme anketi uygulamışlardır. Anket 5820 adet üniversite öğrencisine uygulanmıştır. SPSS Clementine ve WEKA programları kullanılarak veriler üzerinde lojistik regresyon ve karar ağacı algoritmaları oluşturulmuş ve elde edilen sonuçlar yorumlanmıştır. Yapılan analiz sonuçlarına göre, derslerin başarısının öğrencinin

mesleki gelişimlerini artırması ve dünya görüşünü geliştirmesi ile bağlantılı olduğu görülmüştür [21].

Kurt ve Erdem (2012), tarafından yapılmış olan bu çalışmada öğrenci başarılarını etkileyebilecek unsurlar değişik VM algoritma ve modelleriyle ele alınmıştır. Ekonomik, sosyal, kişisel, çevresel varyantlar üzerindeki yapılan uygulamada mezuniyet sonrası mezun olunan bölümle ilgili bir işte çalışıp çalışmama olasılığının başarıyı yüksek oranda etkilediği, araştırmacı bir karaktere sahip olmanın başarıya ulaşma manasında dikkate değer bir etkiye sahip olduğu, başarısız bireylerin büyük çoğunluğunun hedeflemediği programda eğitim aldığı, öğrencilerin orta öğretim transkript ortalamaları ile lisans ortalamaları arasında bağlantı olduğu ve cinsiyetin başarıyı belirlemede dikkate değer herhangi bir etkisi olmadığı saptanmıştır [61].

Aydın (2007), tarafından yapılan bu çalışmada Anadolu Üniversitesi Uzaktan Eğitim Sisteminde okuyan öğrencilere ilişkin çeşitli kaynaklarda yer alan veriler birleştirilerek bir uygulama meydana getirilmiştir. Uzaktan Eğitim Sisteminin planlama aşamasına katkıda bulunabilecek öğrenci başarılarını tahmin etmek amacıyla bir veri madenciliği model geliştirilmiş ve mezun olan öğrencilerin profillerini belirlemek için veri madenciliği modellerinden olan kümeleme çalışması yapılmıştır. Öğrenci başarısını tahmin etme de model olarak C5.0 karar ağacı algoritması kullanılmıştır. Önerilen bu modelin karar kuralları sisteme eklenerek öğrenci başarılarını tahmin etmek amacıyla kullanılacağı öngörülmektedir. Mezun olan öğrencilere dair yapılan uygulamada “K-means” tekniği kullanılarak beş küme meydana gelmiştir. Yapılan kümeleme analizi sonucunda öğrenci başarısı ve bilgisayar kullanımı arasında sonuçları doğrular nitelikte bir ilişki olduğu görülmüştür [62].

Can (2017), tarafından yapılan bu çalışmada temel eğitim öğrenim gören öğrencilerin her ders için ulaşması beklenen genel amaçlara yönelik belirlenen kazanımların gerçekleşip gerçekleşmediği, eksik olan kazanımlar ve bu kazanımlar arasındaki anlamlı ilişkiler araştırılarak öğrencilerin ortaöğretime geçişinden önce eksik kalan kazanımlarının tespit edilmesi amaçlanmıştır. Araştırmada 2015-2016 eğitim-öğretim yılı II. dönem Temel Eğitimden Ortaöğretime Geçiş sınavında yer alan Din Kültürü ve Ahlak Bilgisi, Matematik, Türkçe, T.C. İnkılap Tarihi ve Atatürkçülük, Fen ve Teknoloji derslerinden sınava giren öğrencilerin bilgileri kullanılmıştır. Bu sınavda sorulan sorular incelenmiştir.

Talim Terbiye Kurulu Başkanlığınca 2015/2016 eğitim öğretim yılı başlangıcında yayınlanan kazanım çizelgelerinde bulunan sorularla ilgili kazanımlar belirlenmiştir. Belirlenen kazanımlar RapidMiner ve Weka veri madenciliği programları aracılığıyla Fp-Growth ve Apriori birliktelik analizi algoritmaları kullanarak incelenmiştir. Öğrencilerin en çok doğru cevap verdiği dersin Din Kültürü ve Ahlak Bilgisi en az doğru cevap verdiği dersin ise Matematik dersi olduğu görülmüştür [20].

Ayık, Özdemir ve Yavuz (2007) tarafından yapılan bu çalışmada Atatürk Üniversitesi örgün eğitim öğrencilerinin mezun oldukları lise türleri ve diploma notları ile kazandıkları fakülteler arasındaki bağlantı, veri madenciliği tekniklerinden yararlanarak incelenmiştir. Lise türü diploma notlarının, kazanılan fakülte üzerindeki öneminin tespit edilebilmesi amaçlanmıştır. Bu çalışmaya göre, lise türünün istenilen bir üniversite kazanılmasında etkisinin çok büyük olduğu, yine diploma notunun da aynı derecede önemli olduğu tespit edilmiştir [63].

## 4. MATERYAL VE YÖNTEM

### 4.1. Araştırmanın Amacı

Bu araştırmanın amacı Amasya Üniversitesi uzaktan eğitim ön lisans öğrencilerinin zamanında mezun olabilme durumlarını veri madenciliği uygulamalarını kullanarak önceden tahmin etmektir. Böylece, zamanın da mezun olma potansiyelini barındırmayan öğrenciler tespit edilebilecektir. Bu durumda bulunan öğrencilere gerekli danışmanlık ve rehberlik faaliyetleri uygulanarak öğrencilerin zamanında mezun olmaları sağlanabilecektir. Bu doğrultuda veri madenciliği algoritmaları ile öğrencilerin zamanında mezun olma durumları tahmin edilerek, bu algoritmalara ait sonuçlar karşılaştırılmıştır.

Bu kapsamda, çalışmada 2016/2017 eğitim öğretim yılında Amasya Üniversitesi uzaktan eğitim ön lisans bölümlerine kayıt yaptıran öğrencilerin bilgileri, Amasya Üniversitesi bilgi işlem merkezinden alınarak bu uygulamada kullanılmıştır. Bilgi işlem merkezinden alınan bu ham veriler; Amasya Üniversitesi uzaktan eğitim ön lisans bölümlerinden kayıt yaptıran öğrencilerin 2016/2017 güz dönemi ders notları, 2016/2017 bahar dönemi ders notları, yaş, cinsiyet ve mezuniyet durum bilgilerini içermektedir.

Öğrencilerin mezuniyet durumlarını tahmin etmek amacıyla her bölüm öğrencilerine ait veriler ayrı ayrı analiz edilmiştir. Bu süreçte öğrencilerin I. ve II. dönem ders notları, yaş ve cinsiyet bilgileri kullanarak IV. dönem sonunda mezuniyet durumlarını tahmin etmek için 5 farklı algoritma ile analiz yapılmıştır. Analizlerde Rapidminer 9.2 sürümü içerisinde yer alan algoritmalarından karar ağaçları, naive bayes, random forest, destek vektör makinesi ve yapay sinir ağları algoritmaları ile tahminde bulunmaya çalışılmıştır. Ayrıca her bölüme uygulanan algoritmaların sonuçları karşılaştırılarak en iyi performansı gösteren algoritmalar da bulunmaya çalışılmıştır.

### 4.2. Veri Toplama Süreci

Bu araştırmanın evrenini 2016 yılında Amasya Üniversitesi uzaktan eğitim ön lisans öğrencilerine ait bilgiler oluşturmaktadır. Belirtilen yılda Amasya Üniversitesi uzaktan eğitim ön lisans programına 5 farklı bölümde öğrenci alımı gerçekleştirilmiştir. Bu bölümler aşağıda belirtilmiştir:

- Tıbbi Dokümantasyon ve Sekreterlik
- Mekatronik
- İnternet ve Ağ Teknolojileri
- Çocuk Gelişimi
- Elektrik

Bu bölümlere kayıt yaptıran tüm öğrencilere ait ders yılsonu notları, mezuniyet durumları, yaş ve cinsiyet bilgileri üniversitenin bilgi işlem birimine başvuru yapılarak talep edilmiştir. Amasya Üniversitesi bilgi işlem merkezi tarafından tüm bilgiler bir excel dosyası formatında verilmiştir.

### **4.3. Veri Temizleme**

Amasya Üniversitesi bilgi işlem merkezi tarafından excel formatında sunulan dosya incelendiğinde öğrencilere ait olan bazı özelliklerin sütunlarda yer aldığı, bazı özelliklerin satırlarda yer aldığı görülmüştür. Veriler üzerinde yapılacak işlemlerde niteliklerin tamamının aynı düzende olması gerektiğinden tüm nitelikler sütunlar haline dönüştürülmüştür. Bu işlemi gerçekleştirmek için visual basic script'ler kullanılmıştır. Excel dosyası içine yazılan visual basic script'ler ile satırlar halinde olan nitelikler sütunlara dönüştürülmüştür.

İlk olarak veri setinde yer alan verilerden hatalı olan veya null değerli veriler excel programında bu alanlara filtre uygulanarak tespit edilmiş ve silinmiştir. Ayrıca tüm üniversitelerde ortak olarak bulunan derslerden ( Türk Dili I, Türk Dili II, Atatürk İlkeleri ve İnkılap Tarihi I, Atatürk İlkeleri ve İnkılap Tarihi II, Yabancı Dil I, Yabancı Dil II) bazı öğrencilerin muaf olması, bazı öğrencilerin de notlarının eksik olması nedeniyle bu dersler de silinerek veri setinden çıkarılmıştır. Yapılacak çalışmaya katkı sağlayacak nitelikler ve alabileceği değerler çocuk gelişimi bölümü için Çizelge 4.1' de, tıbbi dokümantasyon ve sekreterlik bölümü için Çizelge 4.2' de, elektrik bölümü için Çizelge 4.3' de, internet ve ağ teknolojileri bölümü için Çizelge 4.4' de, mekatronik bölümü için Çizelge 4.5' de gösterilmiştir.

Çizelge 4.1 Çocuk gelişimi bölümü için çalışmada kullanılan nitelikler ve alabileceği değerler

Cinsiyet	1,2
Yas	17-40
Aile Planlaması ve Cinsel Eğitim(Seçmeli II)	0-100
Çocuk Gelişimi I	0-100
Çocuk Gelişimi II	0-100
Çocuk Hakları Ve Koruma (Seçmeli I)	0-100
Çocuk Sağlığı Ve Hastalıkları	0-100
Çocuk ve Drama	0-100
Çocuk ve Oyun	0-100
İletişim	0-100
İlk Yardım	0-100
Meslek Etiği (Seçmeli I)	0-100
Mesleki Uygulamalar	0-100
Özel Eğitim I	0-100
Özel Eğitim II	0-100
Stres Ve Kriz Yönetimi (Seçmeli II)	0-100
Mezun_Durumu	0, 1

Çizelge 4.1’ de kullanılan cinsiyet alanında bayanlar için “1” erkekler için “2” rakamı kullanılmıştır. Yaş alanında, öğrencilerin yaş aralığı belirtilmiştir. Her dersin notu “0-100” arasında değer alırken, mezun\_durumu zamanında mezun olamayanlar için “0”, zamanında mezun olanlar için ise “1” rakamı ile ifade edilmiştir. Benzer durum Çizelge 4.2, Çizelge 4.3, Çizelge 4.4 ve Çizelge 4.5 için de söz konusudur.

Çizelge 4.2 Tıbbi dokümantasyon ve sekreterlik bölümü için çalışmada kullanılan nitelikler ve alabileceği değerler

Cinsiyet	1, 2
Yas	17-45
Aile Planlaması ve Cinsel Eğitim (Seçmeli II)	0-100
Anatomi	0-100
Biyoistatistik	0-100
Davranış Bilimleri (Seçmeli I)	0-100
Halkla İlişkiler	0-100
Hastalıklar Bilgisi	0-100
Hastalıkların Uluslar Arası Sınıflandırılması I	0-100
İlk Yardım (Seçmeli I)	0-100
Klavye Kullanım Teknikleri I	0-100
Klavye Kullanım Teknikleri II	0-100
Stres Ve Kriz Yönetimi (Seçmeli II)	0-100
Temel Bilgi Teknolojisi Kullanımı	0-100
Tıbbi Dokümantasyon I	0-100
Tıbbi Terminoloji	0-100
Tıbbi Dokümantasyon Iı	0-100
Yazışma Teknikleri	0-100
Mezun_Durumu	0, 1

Çizelge 4.3 Elektrik bölümü için çalışmada kullanılan nitelikler ve alabileceği değerler

Cinsiyet	1, 2
Yas	18-42
Alternatif Akım Devreleri	0-100
Bilgisayar Destekli Tasarım	0-100
Doğru Akım Devreleri	0-100
Ev Cihazları - I	0-100
İş Güvenliği	0-100
Ölçme Tekniği	0-100
Sayısal Elektronik	0-100
Temel Elektronik	0-100
Tesisata Giriş	0-100
Trafo Ve Doğru Akım Makinaları	0-100
Mezun_Durumu	0, 1

Çizelge 4.4 İnternet ve ağ teknolojileri bölümü için çalışmada kullanılan nitelikler ve alabileceği değerler

Cinsiyet	1, 2
Yas	18-35
Ağ Teknolojilerine Giriş	0-100
Bilgisayar Donanımı	0-100
E-Ticaret	0-100
Grafik Ve Animasyon	0-100
İşletim Sistemleri	0-100
Matematik	0-100
Nesne Tabanlı Programlama	0-100
Ofis Yazılımları	0-100
Programlama Temelleri	0-100
Web Tasarımının Temelleri	0-100
Mezun_Durumu	0, 1

Çizelge 4.5 Mekatronik bölümü için çalışmada kullanılan nitelikler ve alabileceği değerler

Cinsiyet	1, 2
Yas	18-38
Alternatif Akım Devre Analizi	0-100
Analog Elektronik	0-100
Bilgisayar Destekli Devre Tasarımı	0-100
Bilgisayarda Programlama	0-100
Doğru Akım Devre Analizi	0-100
Elektrik Elektronik Ölçme	0-100
Matematik	0-100
Mekanizmalar	0-100
Mekatroniğin Temelleri	0-100
Teknik Meslek Resmi	0-100
Mezun Durumu	0, 1

#### 4.4. Veri Dönüştürme

Ham verileri yapılacak olan çalışmaya uygun hale getirmek amacıyla bazı veriler üzerinde verilerin niteliğini değiştiren dönüşümler yapılmıştır. Veri kümesinde zamanında mezun olan öğrencilerin mezun durumu alanı 1 olarak dönüştürülmüşken, zamanında mezun olamayan ve aktif olarak görünen öğrencilerin mezun durumu alanı 0 olarak dönüştürülmüştür. Ayrıca verilerde yer alan doğum tarihi bilgisi veri kümesinde kullanılan yaş bilgisine gerekli hesaplamalar yapılarak dönüştürülmüş ve uygulama da yaş bilgisi olarak kullanılmıştır. Verilerde bulunan cinsiyet alanı kadınlar için 1 rakamına erkekler için 2 rakamına dönüştürülmüştür. Bu dönüşüm işlemi excel dosyasında yer alan verilere bazı basit formüllerin uygulanması ile gerçekleştirilmiştir.

#### 4.5. Veri Analizinde Kullanılan Program

Bu çalışma kapsamında analiz yapılırken makine öğrenmesi, derin öğrenme, metin madenciliği ve tahmine dayalı analitik için entegre bir ortam sağlayan bir veri bilimi yazılım platformu olan RapidMiner programının kullanılmıştır. Halen yeni sürümleri geliştirilmeye devam eden RapidMiner programının en güncel sürümü

<http://www.rapidminer.com> adresinden indirilmektedir. RapidMiner yazılımı öğrencilere, akademisyenlere ve araştırmacılara akademik kullanım için ücretsiz RapidMiner ürün lisansları sağlamaktadır. Analiz yapılırken RapidMiner programının en güncel versiyonu olan 9.2 sürümü kullanılmıştır.



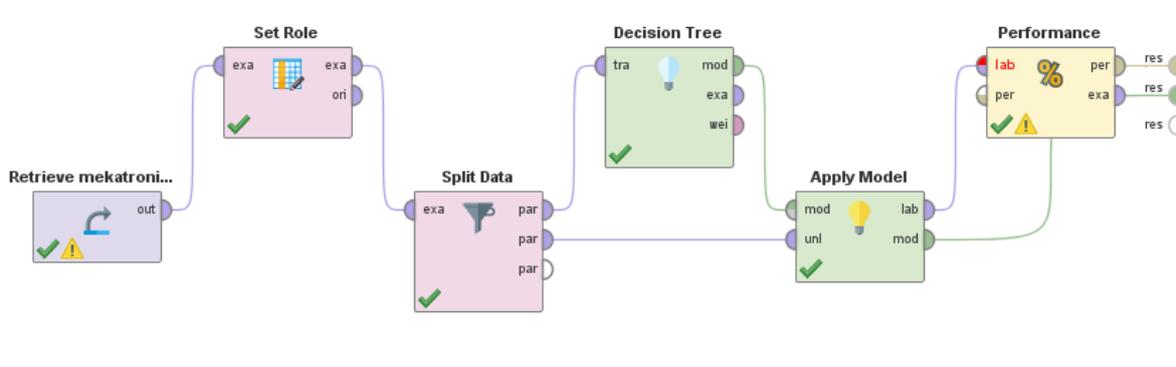
Şekil 4.1. RapidMiner programının açılış ekranı

#### 4.6. Araştırma Modelin Oluşturulması

Amasya Üniversitesi uzaktan eğitim ön lisans öğrencilerinin zamanında mezun olabilme durumlarını veri madenciliği uygulamalarını kullanarak önceden tahmin etmektir. Bu amaç doğrultusunda bu çalışmada veri madenciliği algoritmaları kullanılmıştır. Ön işlemler sonucunda 5 bölüm öğrencilerine ait kayıtlar üzerinde karar ağaçları, naive bayes, random forest, yapay sinir ağları, destek vektör makinesi algoritmaları uygulanmıştır. Oluşturulan veri kümesinde bulunan kayıtlar Rapidminer programı içerisinde bulunan Splitdata özelliği ile verilerin %70'i eğitim, %30'u test için kullanılacak tabloda bölünmüştür. Bu bölme işlemi yapılırken her gruptan örneklem sayısı mezuniyet durumlarına göre eşit bir biçimde ayarlanmıştır.

Bu çalışmada öğrencilerin 2016/2017 güz dönemi ve 2016/2017 bahar dönemi ders notları, cinsiyet ve yaşları dikkate alınarak öğrencilerin mezuniyet durum tahmin edilmeye çalışılmıştır. Her bölüm için karar ağaçları, naive bayes, random forest, destek vektör makinesi ve yapay sinir ağları algoritmaları kullanılmıştır. Analiz aşamasında her bölüm için oluşturulan veriler ayrı ayrı RapidMiner programına import edilmiştir. Daha sonra

RapidMiner programı içerisinde yer alan Set Role ile label alanı yani tahmin edilecek alan seçilmiştir. Yapılan analizde Mezun\_durumu alanı label alan olarak belirlenmiştir. Daha sonra Split Data özelliği ile verilerin %70'i eğitim verisi, %30'u test verisi olacak şekilde ayarlanmıştır. Bir sonra ki aşama da kullanılan algoritma seçilip, algoritma sonucunda oluşan performans değerlerini hesaplamak için program içerisinde yer alan Performance(Binominal Classification) kullanılmıştır. Mekatronik bölümü için karar ağaçları algoritmasıyla yapılan analizde kullanılan model örnek olarak Şekil 4.2' de gösterilmiştir. Diğer bölüm ve algoritmalar için veriler ve uygulanan algoritmalar seçilerek modeller oluşturulmuştur.



Şekil 4.2 Analiz ekranı

#### 4.6.1. Araştırma modelinde kullanılan sınıflandırma kriterleri

Model başarımını değerlendirirken kullanılan temel kavramlar doğruluk oranı(Accuracy), kesinlik (Precision), duyarlılık (Recall), F-ölçütü (F-Measure) ve kappa istatistiğidir (Kappa Statistics).

Yapılan analiz sonucunda sınıflandırma modelinin başarısı hata matrisindeki doğru sınıflandırılan değişkenler ile yanlış sınıflandırılan değişkenlerin sayısı ile hesaplanmaktadır. Hata matrisinde satırlar test kümesindeki örneklere ait gerçek sayıları, kolonlar ise modelin tahminlemesini ifade etmektedir. Çizelge 4.6 da iki sınıflı bir veri kümesinde oluşturulan bir modelin hata matrisi verilmiştir.  $n \times n$  boyutlarındaki bir hata matrisinde ana köşegen doğru tahminlenmiş örnek sayılarını; ana köşegen dışında kalan matris elemanları ise hatalı sonuçları vermektedir. TP (True Pozitif) ve TN(True Negatif) değerleri doğru sınıflandırılmış örnek sayısıdır. False Pozitif (FP), aslında 0 (negatif) sınıftayken 1 (pozitif) olarak tahminlenmiş örneklerin sayısıdır. False Negative (FN) ise

1 (pozitif) sınıfındayken 0 (negatif) olarak tahminlenmiş örneklerin sayısını ifade etmektedir [64]. Hata matrisi Çizelge 4.6’ da gösterilmiştir.

Çizelge 4.6 Hata Matrisi

		Tahmin Edilen Sınıf	
		Sınıf=0	Sınıf=1
Gerçek Sınıf	Sınıf=0	TN	FP
	Sınıf=1	FN	TP

**Doğruluk Oranı(Accuracy):** Sınıfı doğru bulunmuş örnek sayısının (TP +TN), toplam örnek sayısına (TP+TN+FP+FN) oranıdır [64].

$$Accuracy = \frac{TN + TP}{FN + TN + TP + FP} \quad (4.1)$$

**Kesinlik(Precision):** Doğru sınıflandırılmış pozitif örneklem sayısının sınıfı pozitif olarak öngörölmüş toplam örneklem sayısına oranıdır [65].

$$Precision = \frac{TP}{TP + FN} \quad (4.2)$$

**Duyarlılık (Recall):** Doğru sınıflandırılmış pozitif örneklem sayısının gerçek sınıfı pozitif olan tüm örneklem sayısına oranıdır. Duyarlılık, gerçek pozitiflik oranı (True Positive Rate) olarak da adlandırılır [65].

$$Recall = \frac{TP}{TP + FP} \quad (4.3)$$

**F-Ölçütü (F-Measure):** Çoğunlukla kesinlik ve duyarlılık arasında zıt bir ilişki vardır ki birinin değerini arttırmak diğerinin değerini düşürebilir. Bu nedenle daha kesin ve duyarlı sonuçlar elde etmek için her iki ölçütün harmonik ortalaması olan F-ölçütü kullanılır [64].

$$F - Measure = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (4.4)$$

**Kappa İstatistiği:** Kappa istatistiği, yapılan tahminin doğruluk ölçüsüdür. Kappa sonuçları [0,1] aralığında değişir. Kappa değeri 0,4 ile 0,6 arasında ise orta seviyede bir uyum vardır. 0,6 ile 0,8 arasında ise iyi seviyede bir uyum vardır. 0,8 ile 1 arasında ise çok iyi seviyede bir uyum vardır.

## 5. BULGULAR

### 5.1. Çocuk Gelişimi Bölüm Öğrencilerinin Zamanında Mezun Olma Durumlarına İlişkin Bulgular

Bu başlıkta uzaktan eğitim çocuk gelişimi bölüm öğrencilerin zamanında mezun olma durumlarına ilişkin yapılan analizlere yer verilmektedir.

#### 5.1.1. Karar ağacı modelinin başarımlı ölçütü

Amasya Üniversitesi uzaktan eğitim çocuk gelişimi bölümünde öğrenim gören öğrencilerin karar ağaçları algoritmasıyla yapılan analiz sonuçları Çizelge 5.1’ de gösterilmiştir.

Çizelge 5.1 Çocuk gelişim bölümü için karar ağacı algoritması sonuçları

PerformanceVector:	
<b>accuracy: 84,85%</b> ConfusionMatrix: True:     0         1 0:         6         3 1:         2         22	<b>kappa: 0,604</b> ConfusionMatrix: True:     0         1 0:         6         3 1:         2         22
<b>precision: 91,67% (positive class: 1)</b> ConfusionMatrix: True:     0         1 0:         6         3 1:         2         22	<b>recall: 88,00% (positive class: 1)</b> ConfusionMatrix: True:     0         1 0:         6         3 1:         2         22
<b>f_measure: 89,80% (positive class: 1)</b> ConfusionMatrix: True:     0         1 0:         6         3 1:         2         22	

Elde edilen analiz sonuçları incelendiğinde, karar ağaçları algoritması, çocuk gelişimi bölüm öğrencilerinin mezuniyet durumunu %84,85 ile doğru sınıflandırmıştır.

#### 5.1.2. Naive bayes modelinin başarımlı ölçütü

Amasya Üniversitesi uzaktan eğitim çocuk gelişimi bölümünde öğrenim öğrencilerin naive bayes algoritmasıyla yapılan analiz sonuçları Çizelge 5.2’ de gösterilmiştir.

Çizelge 5.2 Çocuk gelişim bölümü için naive bayes algoritması sonuçları

PerformanceVector:	
<b>accuracy: 78,79%</b> ConfusionMatrix: True: 0 1 0: 6 5 1: 2 20	<b>kappa: 0,488</b> ConfusionMatrix: True: 0 1 0: 6 5 1: 2 20
<b>precision: 90,91% (positive class: 1)</b> ConfusionMatrix: True: 0 1 0: 6 5 1: 2 20	<b>recall: 80,00% (positive class: 1)</b> ConfusionMatrix: True: 0 1 0: 6 5 1: 2 20
<b>f_measure: 85,11% (positive class: 1)</b> ConfusionMatrix: True: 0 1 0: 6 5 1: 2 20	

Elde edilen analiz sonuçları incelendiğinde, naive bayes algoritması, çocuk gelişimi bölüm öğrencilerinin mezuniyet durumunu %78,79 ile doğru sınıflandırmıştır.

### 5.1.3. Random forest modelinin başarımlı ölçütü

Amasya Üniversitesi uzaktan eğitim çocuk gelişimi bölümünde okuyan öğrencilerin random forest algoritmasıyla yapılan analiz sonuçları ölçütleri Çizelge 5.3' de gösterilmiştir.

Çizelge 5.3 Çocuk gelişim bölümü için random forest algoritması sonuçları

PerformanceVector:	
<b>accuracy: 81,82%</b> ConfusionMatrix: True: 0 1 0: 5 3 1: 3 22	<b>kappa: 0,505</b> ConfusionMatrix: True: 0 1 0: 5 3 1: 3 22
<b>precision: 88,00% (positive class: 1)</b> ConfusionMatrix: True: 0 1 0: 5 3 1: 3 22	<b>recall: 88,00% (positive class: 1)</b> ConfusionMatrix: True: 0 1 0: 5 3 1: 3 22
<b>f_measure: 88,00% (positive class: 1)</b> ConfusionMatrix: True: 0 1 0: 5 3 1: 3 22	

Elde edilen analiz sonuçları incelendiğinde, random forest algoritması, çocuk gelişimi bölüm öğrencilerinin mezuniyet durumunu %81,82 ile doğru sınıflandırmıştır.

### 5.1.4. Destek vektör makinesi modelinin başarıml ölçütü

Amasya Üniversitesi uzaktan eğitim çocuk gelişimi bölümünde okuyan öğrencilerin destek vektör makinesi algoritmasıyla yapılan analiz sonuçları Çizelge 5.4' de gösterilmiştir.

Çizelge 5.4 Çocuk gelişim bölümü için destek vektör makinesi algoritması sonuçları

PerformanceVector:	
<b>accuracy: 90,91%</b> ConfusionMatrix: True: 0 1 0: 5 0 1: 3 25	<b>kappa: 0,716</b> ConfusionMatrix: True: 0 1 0: 5 0 1: 3 25
<b>precision: 89,29% (positive class: 1)</b> ConfusionMatrix: True: 0 1 0: 5 0 1: 3 25	<b>recall: 100,00% (positive class: 1)</b> ConfusionMatrix: True: 0 1 0: 5 0 1: 3 25
<b>f_measure: 94,34% (positive class: 1)</b> ConfusionMatrix: True: 0 1 0: 5 0 1: 3 25	

Elde edilen analiz sonuçları incelendiğinde, destek vektör makinesi algoritması, çocuk gelişimi bölüm öğrencilerinin mezuniyet durumunu %90,91 ile doğru sınıflandırmıştır.

### 5.1.5. Yapay sinir ağırları modelinin başarıml ölçütü

Amasya Üniversitesi uzaktan eğitim çocuk gelişimi bölümünde okuyan öğrencilerin yapay sinir ağırları algoritmasıyla yapılan analiz sonuçları Çizelge 5.5' de gösterilmiştir.

Çizelge 5.5 Çocuk gelişim bölümü için yapay sinir ağırları modelinin algoritması sonuçları

PerformanceVector:	
<b>accuracy: 84,85%</b> ConfusionMatrix: True: 0 1 0: 5 2 1: 3 23	<b>kappa: 0,569</b> ConfusionMatrix: True: 0 1 0: 5 2 1: 3 23
<b>precision: 88,46% (positive class: 1)</b> ConfusionMatrix: True: 0 1 0: 5 2 1: 3 23	<b>recall: 92,00% (positive class: 1)</b> ConfusionMatrix: True: 0 1 0: 5 2 1: 3 23
<b>f_measure: 90,20% (positive class: 1)</b> ConfusionMatrix: True: 0 1 0: 5 2 1: 3 23	

Elde edilen analiz sonuçları incelendiğinde, yapay sinir ağları algoritması, çocuk gelişimi bölüm öğrencilerinin mezuniyet durumunu %84,85 ile doğru sınıflandırmıştır.

### 5.1.6. Çocuk gelişimi bölümü için oluşturulan modellerin karşılaştırılması

Çocuk gelişimi bölümünde öğrenim gören 111 öğrencinin mezun olma durumlarını tahmin etmede şu on altı özellik kullanılmıştır: “Cinsiyet” , ”Yas”, ”Aile Planlaması Ve Cinsel Eğitim (Seçmeli II)”, ”Çocuk Gelişimi I”, “Çocuk Gelişimi II”, “Çocuk Hakları Ve Koruma (Seçmeli I)”, “Çocuk Sağlığı Ve Hastalıklar”, “Çocuk Ve Drama”, “Çocuk Ve Oyun”, “İletişim”, “İlk Yardım“, “Meslek Etiği (Seçmeli I)”, “Mesleki Uygulamalar”, “Özel Eğitim I”, “Özel Eğitim II”, “Stres Ve Kriz Yönetimi (Seçmeli II)”. Bu özellikler üzerinde Karar ağaçları, Random Forest, Naive Bayes, Yapay Sinir Ağları ve Destek Vektör Makinesi algoritmaları ile analiz edilerek her algoritma için oluşturulan modele ait performans göstergeleri daha önceki bölümlerde verilmiştir. Algoritmalar arasında karşılaştırma yapabilmek için her modele ait Doğruluk Oranı(Accuracy), Kesinlik(Precision), Duyarlılık (Recall), F-Ölçütü (F-Measure), Kappa İstatistiği değerleri alınmıştır. Bu değerler aşağıdaki çizelgede gösterilmiştir.

Çizelge 5.6 Çocuk gelişim bölümü için oluşturulan modellerin karşılaştırılması

Algoritmalar	Doğruluk	Kappa	Kesinlik	Duyarlılık	F-Ölçütü
Karar Ağaçları	%84,85	0,604	%91,67	%88,00	%89,80
Naive Bayes	%78,78	0,488	%90,91	%80,00	%85,11
Random Forest	%81,82	0,505	%88,00	%88,00	%88,00
Destek Vektör Makinesi	%90,91	0,716	%89,29	%100,00	%94,34
Yapay Sinir Ağları	%84,85	0,569	%88,46	%92,00	%90,20

Algoritmaların doğru sınıflandırma yüzdelerine bakıldığında en yüksek doğruluk oranının destek vektör makinesi algoritmasına ait olduğu görülmektedir. Tahmin doğruluğunu ölçen Kappa istatistiğine göre en iyi sonucu destek vektör makinesi algoritması vermektedir. Kesinlik(Precision) değerine bakıldığında karar ağaçları diğer algoritmalarından daha iyi performans göstermektedir. Algoritmaların duyarlılıklarına bakıldığında destek vektör makinesinin diğer algoritmalarından daha iyi olduğu görülmektedir. F-ölçütü duyarlılık ile kesinlik değerlerinin harmonik bir ortalamasından

oluşmaktadır. Bu değere bakıldığında yine en iyi sonucu destek vektör makinesi algoritmasının verdiği anlaşılmaktadır. Doğru sınıflandırma yüzdesi ve Kappa istatistiğine bakıldığında en kötü sonucu veren algoritmanın naive bayes algoritması olduğu görülmektedir.

## 5.2. Tıbbi Dokümantasyon ve Sekreterlik Bölüm Öğrencilerinin Başarım Ölçütlerinin İncelenmesi

Bu başlıkta uzaktan eğitim tıbbi dokümantasyon ve sekreterlik bölüm öğrencilerin zamanında mezun olma durumlarına ilişkin yapılan analizlere yer verilmektedir.

### 5.2.1. Karar ağacı modelinin başarım ölçütü

Amasya Üniversitesi uzaktan eğitim tıbbi dokümantasyon ve sekreterlik bölümünde okuyan öğrencilerin karar ağaçları algoritmasıyla yapılan analiz sonuçları Çizelge 5.7’ de gösterilmiştir.

Çizelge 5.7 Tıbbi dokümantasyon ve sekreterlik bölümü için karar ağacı algoritması sonuçları

PerformanceVector:	
<b>accuracy: 93,10%</b> ConfusionMatrix: True: 1 0 1: 11 2 0: 0 16	<b>kappa: 0,859</b> ConfusionMatrix: True: 1 0 1: 11 2 0: 0 16
<b>precision: 100,00% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 11 2 0: 0 16	<b>recall: 88,89% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 11 2 0: 0 16
<b>f_measure: 94,12% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 11 2 0: 0 16	

Elde edilen analiz sonuçları incelendiğinde, karar ağaçları algoritması, tıbbi dokümantasyon ve sekreterlik bölüm öğrencilerinin mezuniyet durumunu %93,10 ile doğru sınıflandırmıştır.

### 5.2.2. Naive bayes modelinin başarıml ölçütü

Amasya Üniversitesi uzaktan eğitim tıbbi dokümantasyon ve sekreterlik bölümünde okuyan öğrencilerin naive bayes algoritmasıyla yapılan analiz sonuçları Çizelge 5.8’ de gösterilmiştir.

Çizelge 5.8 Tıbbi dokümantasyon ve sekreterlik bölümü için naive bayes algoritması sonuçları

PerformanceVector:	
<b>accuracy: 75,86%</b> ConfusionMatrix: True: 1 0 1: 11 7 0: 0 11	<b>kappa: 0,544</b> ConfusionMatrix: True: 1 0 1: 11 7 0: 0 11
<b>precision: 100,00% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 11 7 0: 0 11	<b>recall: 61,11% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 11 7 0: 0 11
<b>f_measure: 75,86% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 11 7 0: 0 11	

Elde edilen analiz sonuçları incelendiğinde, naive bayes algoritması, tıbbi dokümantasyon ve sekreterlik bölüm öğrencilerinin mezuniyet durumunu %75,86 ile doğru sınıflandırmıştır.

### 5.2.3. Random forest modelinin başarıml ölçütü

Amasya Üniversitesi uzaktan eğitim tıbbi dokümantasyon ve sekreterlik bölümünde okuyan öğrencilerin random forest algoritmasıyla yapılan analiz sonuçları Çizelge 5.9’ da gösterilmiştir.

Çizelge 5.9 Tıbbi dokümantasyon ve sekreterlik bölümü için random algoritması sonuçları

PerformanceVector:	
<b>accuracy: 93,10%</b> ConfusionMatrix: True: 1 0 1: 11 2 0: 0 16	<b>kappa: 0,859</b> ConfusionMatrix: True: 1 0 1: 11 2 0: 0 16
<b>precision: 100,00% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 11 2 0: 0 16	<b>recall: 88,89% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 11 2 0: 0 16
<b>f_measure: 94,12% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 11 2 0: 0 16	

Elde edilen analiz sonuçları incelendiğinde, random forest algoritması, tıbbi dokümantasyon ve sekreterlik bölüm öğrencilerinin mezuniyet durumunu %93,10 ile doğru sınıflandırmıştır.

#### 5.2.4. Destek vektör makinesi modelinin başarımlı ölçütü

Amasya Üniversitesi uzaktan eğitim tıbbi dokümantasyon ve sekreterlik bölümünde okuyan öğrencilerin destek vektör makinesi algoritmasıyla yapılan analiz sonuçları Çizelge 5.10' da gösterilmiştir.

Çizelge 5.10 Tıbbi dokümantasyon ve sekreterlik bölümü için destek vektör makinesi algoritması sonuçları

PerformanceVector:	
<b>accuracy: 86,21%</b> ConfusionMatrix: True: 1 0 1: 10 3 0: 1 15	<b>kappa: 0,717</b> ConfusionMatrix: True: 1 0 1: 10 3 0: 1 15
<b>precision: 93,75% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 10 3 0: 1 15	<b>recall: 83,33% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 10 3 0: 1 15
<b>f_measure: 88,24% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 10 3 0: 1 15	

Elde edilen analiz sonuçları incelendiğinde, destek vektör makinesi algoritması, tıbbi dokümantasyon ve sekreterlik bölüm öğrencilerinin mezuniyet durumunu %86,21 ile doğru sınıflandırmıştır.

### 5.2.5. Yapay sinir ağları modelinin başarımlı ölçütü

Amasya Üniversitesi uzaktan eğitim tıbbi dokümantasyon ve sekreterlik bölümünde okuyan öğrencilerin yapay sinir ağları algoritmasıyla yapılan analiz sonuçları Çizelge 5.11’de gösterilmiştir.

Çizelge 5.11 Tıbbi dokümantasyon ve sekreterlik bölümü için yapay sinir ağları algoritması sonuçları

PerformanceVector:	
<b>accuracy: 89,66%</b> ConfusionMatrix: True:     1         0 1:         11        3 0:         0         15	<b>kappa: 0,791</b> ConfusionMatrix: True:     1         0 1:         11        3 0:         0         15
<b>precision: 100,00% (positive class: 0)</b> ConfusionMatrix: True:     1         0 1:         11        3 0:         0         15	<b>recall: 83,33% (positive class: 0)</b> ConfusionMatrix: True:     1         0 1:         11        3 0:         0         15
<b>f_measure: 90,91% (positive class: 0)</b> ConfusionMatrix: True:     1         0 1:         11        3 0:         0         15	

Elde edilen analiz sonuçları incelendiğinde, yapay sinir ağları algoritması, tıbbi dokümantasyon ve sekreterlik bölüm öğrencilerinin mezuniyet durumunu %89,66 ile doğru sınıflandırmıştır.

### 5.2.6. Tıbbi dokümantasyon ve sekreterlik bölümü için oluşturulan modellerin karşılaştırılması

Tıbbi Dokümantasyon ve Sekreterlik bölümünde öğrenim gören 97 öğrencinin mezun olma durumlarını tahmin etmede şu on sekiz özellik kullanılmıştır: “Cinsiyet”, “Yas”, “Aile Planlaması ve Cinsel Eğitim (Seçmeli II)”, “Anatomi”, “Biyostatistik”, “Davranış Bilimleri (Seçmeli I)”, “Halkla İlişkiler”, “Hastalıklar Bilgisi”, “Hastalıkların Uluslar Arası Sınıflandırılması I”, “İlk Yardım (Seçmeli I)”, “Klavye Kullanım Teknikleri I”, “Klavye

Kullanım Teknikleri II”, “Stres ve Kriz Yönetimi (Seçmeli II)”, Temel Bilgi Teknolojisi Kullanımı”, “Tıbbi Dokümantasyon I”, “Tıbbi Terminoloji”, “Tıbbi Dokümantasyon II”, “Yazışma Teknikleri”. Bu özellikler üzerinde Karar ağaçları, Random Forest, Naive Bayes, Yapay Sinir Ağları ve Destek Vektör Makinesi algoritmaları ile analiz edilerek her algoritma için oluşturulan modele ait performans göstergeleri daha önceki bölümlerde verilmiştir. Algoritmalar arasında karşılaştırma yapabilmek için her modele ait Doğruluk Oranı(Accuracy), Kesinlik(Precision), Duyarlılık (Recall), F-Ölçütü (F-Measure), Kappa İstatistiği değerleri alınmıştır. Bu değerler Çizelge 5.12’ de gösterilmiştir.

Çizelge 5.12 Tıbbi Dokümantasyon ve Sekreterlik bölümü için oluşturulan modellerin karşılaştırılması

Algoritmalar	Doğruluk	Kappa	Kesinlik	Duyarlılık	F-Ölçütü
Karar Ağaçları	%93,10	0,859	%100,00	%88,89	%94,12
Naive Bayes	%75,86	0,544	%100,00	%61,11	%75,86
Random Forest	%93.10	0.859	%100.00	%88.89	%94,12
Destek Vektör Makinesi	%86,21	0,717	%93,75	%83,33	%88,24
Yapay Sinir Ağları	%89,66	0,791	%100,00	%83,33	%90,91

Algoritmaların doğru sınıflandırma yüzdelerine baktığımızda en yüksek doğruluk oranını aynı oranla karar ağaçları ve random forest algoritması vermektedir. Tahmin doğruluğunu ölçen Kappa istatistiğine göre en iyi sonucu karar ağacı ve random forest algoritması vermektedir. Kesinlik(Precision) değerine baktığımızda destek vektör makinesi algoritması diğer algoritmalara göre daha düşük bir performans göstermektedir. Algoritmaların duyarlılıklarına baktığımızda random forest ve karar ağaçları algoritmalarının ilk sırada yer aldığı görülmektedir. Duyarlılık ile kesinlik değerlerinin harmonik bir ortalamasından oluşan F-ölçütüne baktığımızda karar ağaçları ile random forest ve destek vektör makinesi ile yapay sinir ağlarının aynı sonucu verdiği görülmüştür. Doğru sınıflandırma yüzdesi ve Kappa istatistiğine baktığımızda en düşük performansı gösteren algoritmanın naive bayes algoritması olduğu görülmektedir.

### 5.3. İnternet ve Ağ Teknolojileri Bölüm Öğrencilerinin Başarım Ölçütlerinin İncelenmesi

Bu başlıkta uzaktan eğitim internet ve ağ teknolojileri bölüm öğrencilerin zamanında mezun olma durumlarına ilişkin yapılan analizlere yer verilmektedir.

#### 5.3.1. Karar ağacı modelinin başarım ölçütü

Amasya Üniversitesi uzaktan eğitim internet ve ağ teknolojileri bölümünde okuyan öğrencilerin karar ağaçları algoritmasıyla yapılan analiz sonuçları Çizelge 5.13' de gösterilmiştir.

Çizelge 5.13 İnternet ve ağ teknolojileri bölümü için karar ağacı algoritması sonuçları

PerformanceVector:	
<b>accuracy: 83,33%</b> ConfusionMatrix: True: 1 0 1: 4 0 0: 1 1	<b>kappa: 0,571</b> ConfusionMatrix: True: 1 0 1: 4 0 0: 1 1
<b>precision: 50,00% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 4 0 0: 1 1	<b>recall: 100,00% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 4 0 0: 1 1
<b>f_measure: 66,67% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 4 0 0: 1 1	

Elde edilen analiz sonuçları incelendiğinde, karar ağaçları algoritması, internet ve ağ teknolojileri bölüm öğrencilerinin mezuniyet durumunu %83,33 ile doğru sınıflandırmıştır.

#### 5.3.2. Naive bayes modelinin başarım ölçütü

Amasya Üniversitesi uzaktan eğitim internet ve ağ teknolojileri bölümünde okuyan öğrencilerin naive bayes algoritmasıyla yapılan analiz sonuçları Çizelge 5.14' de gösterilmiştir.

Çizelge 5.14 İnternet ve ağ teknolojileri bölümü için karar ağacı algoritması sonuçları

PerformanceVector:	
<b>accuracy: 80,00%</b> ConfusionMatrix: True: 1 0 1: 3 0 0: 1 1	<b>kappa: 0,545</b> ConfusionMatrix: True: 1 0 1: 3 0 0: 1 1
<b>precision: 50,00% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 3 0 0: 1 1	<b>recall: 100,00% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 3 0 0: 1 1
<b>f_measure: 66,67% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 3 0 0: 1 1	

Elde edilen analiz sonuçları incelendiğinde, naive bayes algoritması, internet ve ağ teknolojileri bölüm öğrencilerinin mezuniyet durumunu %80,00 ile doğru sınıflandırmıştır.

### 5.3.3. Random forest modelinin başarıml ölçütü

Amasya Üniversitesi uzaktan eğitim internet ve ağ teknolojileri bölümünde okuyan öğrencilerin random forest algoritmasıyla yapılan analiz sonuçları Çizelge 5.15' de gösterilmiştir.

Çizelge 5.15 İnternet ve ağ teknolojileri bölümü için random forest algoritması sonuçları

PerformanceVector:	
<b>accuracy: 80,00%</b> ConfusionMatrix: True: 1 0 1: 3 0 0: 1 1	<b>kappa: 0,545</b> ConfusionMatrix: True: 1 0 1: 3 0 0: 1 1
<b>precision: 50,00% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 3 0 0: 1 1	<b>recall: 100,00% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 3 0 0: 1 1
<b>f_measure: 66,67% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 3 0 0: 1 1	

Elde edilen analiz sonuçlarına göre, random forest algoritması, internet ve ağ teknolojileri bölüm öğrencilerinin mezuniyet durumunu %80,00 ile doğru sınıflandırmıştır.

### 5.3.4. Destek vektör makinesi modelinin başarıml ölçütü

Amasya Üniversitesi uzaktan eğitim internet ve ağ teknolojileri bölümünde okuyan öğrencilerin destek vektör makinesi algoritmasıyla yapılan analiz sonuçları Çizelge 5.16' da gösterilmiştir.

Çizelge 5.16 İnternet ve ağ teknolojileri bölümü için destek vektör makinesi algoritması sonuçları

PerformanceVector:	
<b>accuracy: 60,00%</b> ConfusionMatrix: True: 1 0 1: 3 1 0: 1 0	<b>kappa: -0,250</b> ConfusionMatrix: True: 1 0 1: 3 1 0: 1 0
<b>precision: 0,00% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 3 1 0: 1 0	<b>recall: 0,00% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 3 1 0: 1 0
<b>f_measure: unknown (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 3 1 0: 1 0	

Elde edilen analiz sonuçları incelendiğinde, destek vektör makinesi algoritması, internet ve ağ teknolojileri bölüm öğrencilerinin mezuniyet durumunu %60,00 ile doğru sınıflandırmıştır.

### 5.3.5. Yapay sinir ağları modelinin başarıml ölçütü

Amasya Üniversitesi uzaktan eğitim internet ve ağ teknolojileri bölümünde okuyan öğrencilerin yapay sinir ağları algoritmasıyla yapılan analiz sonuçları Çizelge 5.17' de gösterilmiştir.

Çizelge 5.17 İnternet ve ağ teknolojileri bölümü için yapay sinir ağları algoritması sonuçları

PerformanceVector:	
<b>accuracy: 40,00%</b> ConfusionMatrix: True: 1 0 1: 2 1 0: 2 0	<b>kappa: -0,364</b> ConfusionMatrix: True: 1 0 1: 2 1 0: 2 0
<b>precision: 0,00% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 2 1 0: 2 0	<b>recall: 0,00% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 2 1 0: 2 0
<b>f_measure: unknown (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 2 1 0: 2 0	

Elde edilen analiz sonuçları incelendiğinde, yapay sinir ağları algoritması, internet ve ağ teknolojileri bölüm öğrencilerinin mezuniyet durumunu %40,00 ile doğru sınıflandırmıştır.

### 5.3.6. İnternet ve ağ teknolojileri bölüm öğrencileri için oluşturulan modellerin karşılaştırılması

İnternet ve Ağ Teknolojileri bölümünde öğrenim gören 21 öğrencinin mezun olma durumlarını tahmin etmede şu on iki özellik kullanılmıştır: “Cinsiyet”, “Yas”, “Ağ Teknolojilerine Giriş”, “Bilgisayar Donanımı”, “E-Ticaret”, “Grafik Ve Animasyon”, “İşletim Sistemleri”, “Matematik”, “Nesne Tabanlı Programlama”, “Ofis Yazılımları”, “Programlama Temelleri”, “Web Tasarımının Temelleri”. Bu özellikler üzerinde Karar ağaçları, Random Forest, Naive Bayes, Yapay Sinir Ağları ve Destek Vektör Makinesi algoritmaları ile analiz edilerek her algoritma için oluşturulan modele ait performans göstergeleri daha önceki bölümlerde verilmiştir. Algoritmalar arasında karşılaştırma yapabilmek için her modele ait Doğruluk Oranı(Accuracy), Kesinlik(Precision), Duyarlılık (Recall), F-Ölçütü (F-Measure), Kappa İstatistiği değerleri alınmıştır. Bu değerler Çizelge 5.18’ de gösterilmiştir.

Çizelge 5.18 İnternet ve ağ teknolojileri bölümü için oluşturulan modellerin karşılaştırılması

Algoritmalar	Doğruluk	Kappa	Kesinlik	Duyarlılık	F-Ölçütü
Karar Ağaçları	%83,33	0,571	%50,00	%100,00	%66,67
Naive Bayes	%80,00	0,545	%50,00	%100,00	%66,67
Random Forest	%80,00	0,545	%50,00	%100,00	%66,67
Destek Vektör Makinesi	%60,00	-0,250	%0,00	%0,00	-
Yapay Sinir Ağları	%40,00	-0,364	%0,00	%0,00	-

Algoritmaların doğru sınıflandırma yüzdelerine baktığımızda en yüksek doğruluk oranını %83.33 oran ile karar ağaçları algoritması vermektedir. Kappa istatistiği incelendiğinde karar ağaçları, naive bayes ve random forest pozitif sonuç verirken destek vektör makinesi ve yapay sinir ağları algoritması negatif sonuç verdiği görülmektedir. Kesinlik(Precision), duyarlılık(recall) ve f-ölçütü değerlerine baktığımızda karar ağaçları, naive bayes ve random forest algoritmaları aynı sonucu verdiği görülmüştür. Diğer iki algoritmanın kesinlik, duyarlık ve f- ölçütü değerleri diğer algoritmalara göre düşük çıkmıştır. Doğru sınıflandırma yüzdesine baktığımızda en düşük performansı gösteren algoritmanın yapay sinir ağları algoritması olduğu görülmektedir.

#### 5.4. Elektrik Bölüm Öğrencilerinin Başarım Ölçütlerinin İncelenmesi

Bu başlıkta uzaktan eğitim elektrik bölüm öğrencilerin zamanında mezun olma durumlarına ilişkin yapılan analizlere yer verilmektedir.

##### 5.4.1. Karar ağacı modelinin başarım ölçütü

Amasya Üniversitesi uzaktan eğitim elektrik bölümünde okuyan öğrencilerin karar ağaçları algoritmasıyla yapılan analiz sonuçları Çizelge 5.19' da gösterilmiştir.

Çizelge 5.19 Elektrik bölümü için karar ağaçları algoritması sonuçları

PerformanceVector:	
<b>accuracy: 85,71%</b> ConfusionMatrix: True: 0 1 0: 4 0 1: 1 2	<b>kappa: 0,696</b> ConfusionMatrix: True: 0 1 0: 4 0 1: 1 2
<b>precision: 66,67% (positive class: 1)</b> ConfusionMatrix: True: 0 1 0: 4 0 1: 1 2	<b>recall: 100,00% (positive class: 1)</b> ConfusionMatrix: True: 0 1 0: 4 0 1: 1 2
<b>f_measure: 80,00% (positive class: 1)</b> ConfusionMatrix: True: 0 1 0: 4 0 1: 1 2	

Elde edilen analiz sonuçları incelendiğinde, karar ağaçları algoritması elektrik bölüm öğrencilerinin mezuniyet durumunu %85,71 ile doğru sınıflandırmıştır.

#### 5.4.2. Naive bayes modelinin başarımlı ölçütü

Amasya Üniversitesi uzaktan eğitim elektrik bölümünde okuyan öğrencilerin naive bayes algoritmasıyla yapılan analiz sonuçları Çizelge 5.20' de gösterilmiştir.

Çizelge 5.20 Elektrik bölümü için naive bayes algoritması sonuçları

PerformanceVector:	
<b>accuracy: 85,71%</b> ConfusionMatrix: True: 0 1 0: 4 0 1: 1 2	<b>kappa: 0,696</b> ConfusionMatrix: True: 0 1 0: 4 0 1: 1 2
<b>precision: 66,67% (positive class: 1)</b> ConfusionMatrix: True: 0 1 0: 4 0 1: 1 2	<b>recall: 100,00% (positive class: 1)</b> ConfusionMatrix: True: 0 1 0: 4 0 1: 1 2
<b>f_measure: 80,00% (positive class: 1)</b> ConfusionMatrix: True: 0 1 0: 4 0 1: 1 2	

Elde edilen analiz sonuçları incelendiğinde, naive bayes algoritması, elektrik bölüm öğrencilerinin mezuniyet durumunu %85,71 ile doğru sınıflandırmıştır.

### 5.4.3. Random forest modelinin başarıml ölçütü

Amasya Üniversitesi uzaktan eğitim elektrik bölümünde okuyan öğrencilerin random forest modeliyle yapılan analiz sonucunda analiz sonuçları Çizelge 5.21' de gösterilmiştir.

Çizelge 5.21 Elektrik bölümü için random forest algoritması sonuçları

PerformanceVector:	
<b>accuracy: 85,71%</b> ConfusionMatrix: True:     0         1 0:         4         1 1:         0         2	<b>kappa: 0,696</b> ConfusionMatrix: True:     0         1 0:         4         1 1:         0         2
<b>precision: 100,00% (positive class: 1)</b> ConfusionMatrix: True:     0         1 0:         4         1 1:         0         2	<b>recall: 66,67% (positive class: 1)</b> ConfusionMatrix: True:     0         1 0:         4         1 1:         0         2
<b>f_measure: 80,00% (positive class: 1)</b> ConfusionMatrix: True:     0         1 0:         4         1 1:         0         2	

Elde edilen analiz sonuçları incelendiğinde, random forest algoritması, elektrik bölüm öğrencilerinin mezuniyet durumunu %85,71 ile doğru sınıflandırmıştır.

### 5.4.4. Destek vektör makinesi modelinin başarıml ölçütü

Amasya Üniversitesi uzaktan eğitim elektrik bölümünde okuyan öğrencilerin destek vektör makinesi algoritmasıyla yapılan analiz sonuçları Çizelge 5.22' de gösterilmiştir.

Çizelge 5.22 Elektrik bölümü için destek vektör makinesi algoritması sonuçları

PerformanceVector:	
<b>accuracy: 71,43%</b> ConfusionMatrix: True:     0         1 0:         3         1 1:         1         2	<b>kappa: 0,417</b> ConfusionMatrix: True:     0         1 0:         3         1 1:         1         2
<b>precision: 66,67% (positive class: 1)</b> ConfusionMatrix: True:     0         1 0:         3         1 1:         1         2	<b>recall: 66,67% (positive class: 1)</b> ConfusionMatrix: True:     0         1 0:         3         1 1:         1         2
<b>f_measure: 66,67% (positive class: 1)</b> ConfusionMatrix: True:     0         1 0:         3         1 1:         1         2	

Elde edilen analiz sonuçları incelendiğinde, Destek Vektör Makinesi algoritması, Elektrik bölüm öğrencilerinin mezuniyet durumunu %71,43 ile doğru sınıflandırmıştır.

#### 5.4.5. Yapay sinir ağları modelinin başarımlı ölçütü

Amasya Üniversitesi uzaktan eğitim elektrik bölümünde okuyan öğrencilerin yapay sinir ağları algoritmasıyla yapılan analiz sonuçları Çizelge 5.23’ de gösterilmiştir.

Çizelge 5.23 Elektrik bölümü için yapay sinir ağları algoritması sonuçları

PerformanceVector:	
<b>accuracy: 71,43%</b> ConfusionMatrix: True: 0 1 0: 4 1 1: 1 1	<b>kappa: 0,300</b> ConfusionMatrix: True: 0 1 0: 4 1 1: 1 1
<b>precision: 50,00% (positive class: 1)</b> ConfusionMatrix: True: 0 1 0: 4 1 1: 1 1	<b>recall: 50,00% (positive class: 1)</b> ConfusionMatrix: True: 0 1 0: 4 1 1: 1 1
<b>f_measure: 50,00% (positive class: 1)</b> ConfusionMatrix: True: 0 1 0: 4 1 1: 1 1	

Elde edilen analiz sonuçları incelendiğinde, yapay sinir ağları algoritması, elektrik bölüm öğrencilerinin mezuniyet durumunu %71,43 ile doğru sınıflandırmıştır.

#### 5.4.6. Elektrik bölüm öğrencileri için oluşturulan modellerin karşılaştırılması

Elektrik bölümünde öğrenim gören 25 öğrencinin mezun olma durumlarını tahmin etmede şu on iki özellik kullanılmıştır: “Cinsiyet”, “Yas”, Alternatif Akım Devreleri”, “Bilgisayar Destekli Tasarım”, “Doğru Akım Devreleri” “Ev Cihazları-I”, “İş Güvenliği”, “Ölçme Tekniği”, “Sayısal Elektronik”, “Temel Elektronik”, “Tesisata Giriş”, “Trafo ve Doğru Akım Makinaları”. Bu özellikler üzerinde Karar ağaçları, Random Forest, Naive Bayes, Yapay Sinir Ağları ve Destek Vektör Makinesi algoritmaları ile analiz edilerek her algoritma için oluşturulan modele ait performans göstergeleri daha önceki bölümlerde verilmiştir. Algoritmalar arasında karşılaştırma yapabilmek için her modele ait Doğruluk

Oranı(Accuracy), Kesinlik(Precision), Duyarlılık (Recall), F-Ölçütü (F-Measure), Kappa İstatistiği değerleri alınmıştır. Bu değerler Çizelge 5.24' de gösterilmiştir.

Çizelge 5.24 Elektrik bölümü için oluşturulan modellerin karşılaştırılması

Algoritmalar	Doğruluk	Kappa	Kesinlik	Duyarlılık	F-Ölçütü
Karar Ağaçları	%85,71	0,696	%66,67	%100,00	%80,00
Naive Bayes	%85,71	0,696	%66,67	%100,00	%8000
Random Forest	%85,71	0,696	%66,67	%100,00	%80,00
Destek Vektör Makinesi	%71,43	0,417	%66,67	%66,67	%66,67
Yapay Sinir Ağları	%71,43	0,300	%50,00	%50,00	%50,00

Algoritmaların doğru sınıflandırma yüzdelerine baktığımızda karar ağaçları, naive bayes random forest algoritmaların en iyi sonucu verdiği görülmektedir. Kappa istatistiği incelendiğinde karar ağaçları, naive bayes ve random forest 0,696 değeriyle en iyi sonucu vermektedir. Kesinlik(Precision) değeri incelendiğinde yapay sinir ağları diğer algoritmalara göre daha düşük performans ölçütüne sahip olduğu, duyarlılık(recall) ve f-ölçütü değerlerine baktığımızda Karar Ağaçları, Naive Bayes ve Random Forest algoritmalarının aynı sonucu verdiği görülmüştür Doğru sınıflandırma yüzdesine ve Kappa istatistiğine göre en düşük performansı gösteren algoritmanın Yapay Sinir Ağları algoritması olduğu görülmektedir.

## 5.5. Mekatronik Bölüm Öğrencilerinin Başarım Ölçütlerinin İncelenmesi

Bu başlıkta uzaktan eğitim mekatronik bölüm öğrencilerin zamanında mezun olma durumlarına ilişkin yapılan analizlere yer verilmektedir.

### 5.5.1. Karar ağacı modelinin başarım ölçütü

Amasya üniversitesi uzaktan eğitim mekatronik bölümünde okuyan öğrencilerin karar ağaçları algoritmasıyla yapılan analiz sonuçları Çizelge 5.25' de gösterilmiştir.

Çizelge 5.25 Mekatronik bölümü için karar ağacı algoritması sonuçları

PerformanceVector:	
<b>accuracy: 93,33%</b> ConfusionMatrix: True: 1 0 1: 8 1 0: 0 6	<b>kappa: 0,865</b> ConfusionMatrix: True: 1 0 1: 8 1 0: 0 6
<b>precision: 100,00% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 8 1 0: 0 6	<b>recall: 85,71% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 8 1 0: 0 6
<b>f_measure: 92,31% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 8 1 0: 0 6	

Elde edilen analiz sonuçları incelendiğinde, karar ağaçları algoritması, mekatronik bölüm öğrencilerinin mezuniyet durumunu %93,33 ile doğru sınıflandırmıştır.

### 5.5.2. Naive bayes modelinin başarımlı ölçütü

Amasya Üniversitesi uzaktan eğitim mekatronik bölümünde okuyan öğrencilerin naive bayes algoritmasıyla yapılan analiz sonuçları Çizelge 5.26' da gösterilmiştir.

Çizelge 5.26 Mekatronik bölümü için naive bayes algoritması sonuçları

PerformanceVector:	
<b>accuracy: 80,00%</b> ConfusionMatrix: True: 1 0 1: 3 0 0: 3 9	<b>kappa: 0,545</b> ConfusionMatrix: True: 1 0 1: 3 0 0: 3 9
<b>precision: 75,00% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 3 0 0: 3 9	<b>recall: 100,00% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 3 0 0: 3 9
<b>f_measure: 85,71% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 3 0 0: 3 9	

Elde edilen analiz sonuçları incelendiğinde, naive bayes algoritması, mekatronik bölüm öğrencilerinin mezuniyet durumunu %80,00 ile doğru sınıflandırmıştır.

### 5.5.3. Random forest modelinin başarıml ölçütü

Amasya Üniversitesi uzaktan eğitim mekatronik bölümünde okuyan öğrencilerin random forest algoritmasıyla yapılan analiz sonuçları Çizelge 5.27’ de gösterilmiştir.

Çizelge 5.27 Mekatronik bölümü için random forest algoritması sonuçları

PerformanceVector:	
<b>accuracy: 86,67%</b> ConfusionMatrix: True: 1 0 1: 5 1 0: 1 8	<b>kappa: 0,722</b> ConfusionMatrix: True: 1 0 1: 5 1 0: 1 8
<b>precision: 88,89% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 5 1 0: 1 8	<b>recall: 88,89% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 5 1 0: 1 8
<b>f_measure: 88,89% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 5 1 0: 1 8	

Elde edilen analiz sonuçları incelendiğinde, random forest algoritması, mekatronik bölüm öğrencilerinin mezuniyet durumunu %86,67 ile doğru sınıflandırmıştır.

### 5.5.4. Destek vektör makinesi modelinin başarıml ölçütü

Amasya Üniversitesi uzaktan eğitim mekatronik bölümünde okuyan öğrencilerin destek vektör makinesi algoritmasıyla yapılan analiz sonuçları Çizelge 5.28’ de gösterilmiştir.

Çizelge 5.28 Mekatronik bölümü için destek vektör makinesi algoritması sonuçları

PerformanceVector:	
<b>accuracy: 80,00%</b> ConfusionMatrix: True: 1 0 1: 6 1 0: 2 6	<b>kappa: 0,602</b> ConfusionMatrix: True: 1 0 1: 6 1 0: 2 6
<b>precision: 75,00% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 6 1 0: 2 6	<b>recall: 85,71% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 6 1 0: 2 6
<b>f_measure: 80,00% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 6 1 0: 2 6	

Elde edilen analiz sonuçları incelendiğinde, destek vektör makinesi algoritması, mekatronik bölüm öğrencilerinin mezuniyet durumunu %80,00 ile doğru sınıflandırmıştır.

### 5.5.5. Yapay sinir ağları modelinin başarımlı ölçütü

Amasya Üniversitesi uzaktan eğitim mekatronik bölümünde okuyan öğrencilerin yapay sinir ağları algoritmasıyla yapılan analiz sonuçları Çizelge 5.29’ da gösterilmiştir.

Çizelge 5.29 Mekatronik bölümü için yapay sinir ağları algoritması sonuçları

<b>PerformanceVector:</b> <b>accuracy: 86,67%</b> ConfusionMatrix: True: 1 0 1: 7 1 0: 1 6	<b>kappa: 0,732</b> ConfusionMatrix: True: 1 0 1: 7 1 0: 1 6
<b>precision: 85,71% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 7 1 0: 1 6	<b>recall: 85,71% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 7 1 0: 1 6
<b>f_measure: 85,71% (positive class: 0)</b> ConfusionMatrix: True: 1 0 1: 7 1 0: 1 6	

Elde edilen analiz sonuçları incelendiğinde, yapay sinir ağları algoritması, mekatronik bölüm öğrencilerinin mezuniyet durumunu %86,67 ile doğru sınıflandırmıştır.

### 5.5.6. Mekatronik bölüm öğrencileri için oluşturulan modellerin karşılaştırılması

Mekatronik bölümünde öğrenim gören 50 öğrencinin mezun olma durumlarını tahmin etmede şu on iki özellik kullanılmıştır: “Cinsiyet”, “Yas”, “Alternatif Akım Devre Analizi”, “Analog Elektronik”, “Bilgisayar Destekli Devre Tasarımı”, “Bilgisayarda Programlama”, “Doğru Akım Devre Analizi”, “Elektrik Elektronik Ölçme”, “Matematik”, “Mekanizmalar”, “Mekatroniğin Temelleri”, “Teknik Meslek Resmi”. Bu özellikler üzerinde Karar ağaçları, Random Forest, Naive Bayes, Yapay Sinir Ağları ve Destek Vektör Makinesi algoritmaları ile analiz edilerek her algoritma için oluşturulan modele ait performans göstergeleri daha önceki bölümlerde verilmiştir. Algoritmalar arasında karşılaştırma yapabilmek için her modele ait Doğruluk Oranı(Accuracy), Kesinlik(Precision), Duyarlılık (Recall), F-Ölçütü (F-Measure), Kappa İstatistiği değerleri alınmıştır. Bu değerler Çizelge 5.30’ da gösterilmiştir.

Çizelge 5.30 Mekatronik bölümü için oluşturulan modellerin karşılaştırılması

Algoritmalar	Doğruluk	Kappa	Kesinlik	Duyarlılık	F-Ölçütü
Karar Ağaçları	%93,33	0,865	%100,00	%85,71	%92,31
Naive Bayes	%80,00	0,545	%75,00	%100,00	%85,71
Random Forest	%86,67	0,722	%88,99	%88,99	%88,99
Destek Vektör Makinesi	%80,00	0,602	%75,00	%85,71	%80,00
Yapay Sinir Ağları	%86,67	0,732	%85,71	%85,71	%85,71

Algoritmaların doğru sınıflandırma yüzdelere bakıldığında en yüksek doğruluk oranını karar ağaçları algoritmasına ait olduğu görülmektedir. Tahmin doğruluğunu ölçen Kappa istatistiğine göre en iyi sonucu karar ağaçları algoritması vermektedir. Kesinlik(Precision) değerine bakıldığında karar ağaçları diğer algoritmalarından daha iyi performans göstermektedir. Algoritmaların duyarlılıklarına bakıldığında naive bayes algoritmasının diğer algoritmalara göre daha performans ölçütüne sahip olduğu görülmektedir. F-ölçütü duyarlılık ile kesinlik değerlerinin harmonik bir ortalamasından oluşmaktadır. Bu değere bakıldığında yine en iyi sonucu karar ağaçları algoritmasının verdiği anlaşılmaktadır. Doğru sınıflandırma yüzdesi ve Kappa istatistiğine bakıldığında en kötü sonucu veren algoritmanın naive bayes algoritması olduğu görülmektedir.

## 6. SONUÇ VE ÖNERİLER

Öğrencilerin eğitim süreci içerisinde akademik başarılarını ve mezuniyet durumlarını tahmin etmek oldukça zorlu bir süreçtir. Başarı soyut bir kavram olmakla birlikte başarıyı etkileyen pek çok faktör bulunmaktadır. Öğrencilerin akademik başarısını etkileyen faktörleri bulmak ve bu faktörlerden öğrencinin başarısını ve mezuniyet durumunu tahmin etmeye çalışmak çokça bilginin bir arada kullanılmasını gerekli kılmaktadır. Büyük veri yığınlarından anlamlı bir örüntü ortaya çıkarmak amacıyla kullanılan yöntemlerin en uygunu veri madenciliğidir. Bu çalışmada Amasya Üniversitesi Uzaktan Eğitim Merkezi ön lisans öğrencilerinin zamanında mezun olabilme durumlarını tahmin etmek için veri madenciliği modelleri kullanıldı. Bunun için Amasya Üniversitesi Uzaktan Eğitim Merkezi ön lisans bölümlerine 2016/2017 akademik yılında kayıt olan öğrencilerin 2016/2017 güz dönemi ders notları, 2016/2017 bahar dönemi ders notları, yaş ve cinsiyet verileri kullanıldı. Bu bilgiler ışığında öğrencilerin zamanında mezun olup olamayacağı tahmin edilmeye çalışıldı. Bu amaçla veri madenciliği algoritmalarından sınıflandırma teknikleri kullanıldı ve bu tekniklerden öğrencilerin mezuniyet durumlarını en iyi tahmin eden algoritma belirlendi.

Bu çalışmada Çocuk Gelişimi, Tıbbi Dokümantasyon ve Sekreterlik, İnternet ve Ağ Teknolojileri, Elektrik ve Mekatronik bölümleri için ayrı ayrı analiz yapıldı. Çocuk Gelişim bölümünde öğrenim gören 111 öğrenci, Tıbbi Dokümantasyon ve Sekreterlik bölümünde öğrenim gören 97 öğrenci, İnternet ve Ağ Teknolojileri bölümünde öğrenim gören 21 öğrenci, Elektrik bölümünde öğrenim gören 25 öğrenci ve Mekatronik bölümünde öğrenim gören 50 öğrencinin bilgileri kullanıldı. Her bölüm için ayrı ayrı Karar Ağaçları, Naive Bayes, Random Forest, Destek Vektör Makinesi ve Yapay Sinir Ağları modelleri uygulandı. Tüm modellerde veri kümesinin %70'i eğitim, %30'u tahmin için kullanıldı.

Çocuk gelişimi bölümünde en iyi tahmini Destek Vektör Makinesi modeli yaparken, en düşük tahmin sonucunu Naive Bayes modeli verdi. Diğer bölümlerde en iyi tahmin Karar Ağaçları algoritmasıyla gerçekleşti. En düşük tahmin oranını ise Tıbbi Dokümantasyon ve Sekreterlik Bölümü ve Mekatronik bölümü için Naive Bayes algoritması, İnternet ve Ağ Teknolojileri bölümü ve Elektrik bölümü için en düşük tahmin doğruluk oranını Yapay Sinir Ağları algoritmaları verdi.

Zamanında mezun olamayan öğrencilerin büyük bir kısmı okulu bırakmaktadır. Ayrıca okulun zamanında bitirilmemesi kişilerin iş hayatına geç atılmasına sebep olmaktadır. Böylece kişi ve devlet maddi ve manevi açıdan zarara uğramaktadır. Bu çalışma ile zamanında mezun olamayacak öğrenciler için bir uyarı mekanizması oluşturularak öğrencilerin bu olumsuz olaylarla karşılaşmaması açısından önem arz etmektedir. Öğrencilere gerekli uyarılar yapılarak onların derslerine daha çok çalışmalarını sağlanarak, maddi manevi oluşabilecek zararların önüne geçilmektedir.

Bu alanda yapılacak çalışmalarda zamanında mezun olma durumları tahmin edilmesi için uygulanacak veri madenciliği sürecine genetik algoritmalar ve doğrusal olmayan regresyon yöntemleri dâhil edilerek tahmin doğrulukları artırılabilir.

Ayrıca yapılacak benzer çalışmalarda, uzaktan eğitim sisteminde bulunan öğrencinin sisteme kaç kere girdiği, uzaktan eğitim sisteminde ne kadar zaman geçirdiği, öğrencilerin lise diploma notları gibi bilgilerin tahmin edici değişken olarak kullanılması, benzer araştırmalar için önerilebilir.

## KAYNAKLAR

1. Gündüz, Ş. ve Odabaşı, F. (2004). Bilgi Çağında Öğretmen Adaylarının Eğitiminde Öğretim Teknolojileri ve Materyal Geliştirme Dersinin Önemi. *The Turkish Online Journal of Educational Technology*, 3(1), 43-48.
2. Özekes, S. (2003). Veri Madenciliği Modelleri ve Uygulama Alanları. *İstanbul Ticaret University Journal*, 3, 65-82.
3. Şimşek, A. S. (2012). *Bilişsel ve Duyuşsal Özelliklerin Yükseköğretimdeki Akademik Başarıyı Yordama Gücü*, Yüksek Lisans Tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
4. Elveren, Ş. (2018). *Dünyada ve Türkiye'de Adli Bilişim Eğitim Uygulamalarının İncelenmesi ve Türkiye için Uygulamaya Yönelik Öneriler*, Yüksek Lisans Tezi, Gazi Üniversitesi Bilişim Enstitüsü, Ankara.
5. Odabaş, H. (2003). İnternet Tabanlı Uzaktan Eğitim ve Bilgi ve Belge Yönetimi. *Türk Kütüphaneciliği*, 17(1), 22-36.
6. Gunawardena, C. N. and McIsaac, M. S. (2013). Distance education. In *Handbook of research on educational communications and technology* (pp. 361-401). Routledge.
7. Beldarrain, Y. (2006). Distance Education Trends: Integrating New Technologies to Foster Student Interaction and Collaboration. *Distance Education*, 27(2), 139-153.
8. Deitch, J.W. (2001). *A Nation of Inventors*. Carlise:Discovery Enterprise
9. Akpınar, H. (2000). Veri Tabanlarında Bilgi Keşfi ve Veri madenciliği. *İÜ İşletme Fakültesi Dergisi*, 29(1), 1-22.
10. Özkan, Y. (2016). *Veri Madenciliği Yöntemleri* (Üçüncü Baskı), İstanbul, Papatya Bilim, 25-26
11. Silahtaroglu, G. (2016). *Veri Madenciliği Kavram ve Algoritmaları* (Üçüncü Baskı), İstanbul, Papatya Yayıncılık 22, 63, 64
12. Han, J. and Kamber, M. (2006), *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers.
13. Taşdemir, M. (2012). *Veri Madenciliği (Öğrenci Başarısına Etki Eden Faktörlerin Regresyon Analizi ile Tespiti)*, Yüksek Lisans Tezi, Dicle Üniversitesi Sosyal Bilimler Enstitüsü, Diyarbakır.
14. Aydemir, B. (2017). *Veri Madenciliği Yöntemleri Kullanarak Meslek Yüksek Okulu Öğrencilerinin Akademik Başarı Tahmini*, Yüksek Lisans Tezi, Pamukkale Üniversitesi Fen Bilimleri Enstitüsü, Denizli.

15. Fakı, B. M. (2015). *Veri Madenciliği Yöntemlerini Kullanarak Anemi Sınıflandırılmasına Yönelik Bir Uygulama*, Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü.
16. Ünsal, Ö. (2011). *Mesleki Alan Seçimlerinin Makine Öğrenmesi Algoritması Kullanılarak Belirlenmesi*, Yüksek Lisans Tezi, Gazi Üniversitesi Bilişim Enstitüsü, Ankara.
17. Akın, Y.K. (2008). *Veri Madenciliğinde Kümeleme Algoritmaları ve Kümeleme Analizi*, Doktora Tezi, Marmara Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
18. Akpınar, H. (2017). *Veri Madenciliği Veri Analizi (İkinci Baskı)*, İstanbul, Papatya Yayıncılık, 74.
19. Keklik, Ç. (2017). *Eksik Değerleri En Olası Değer ile Doldurmanın Sınıflandırma Algoritmaları Üzerinden Karşılaştırılması*, Yüksek Lisans Tezi, Beykent Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
20. Can, E. (2017). *Temel Eğitimden Ortaöğretime Geçiş Sınavı Kazanımlarının Veri Madenciliği Yöntemleri ile Değerlendirilmesi*, Yüksek Lisans Tezi, Kocatepe Üniversitesi Fen Bilimleri Enstitüsü, Afyon.
21. Can, Ş. (2017). *Veri Madenciliği ve Eğitim Sektöründe Bir Uygulama*, Yüksek Lisans Tezi, Celal Bayar Üniversitesi Sosyal Bilimler Enstitüsü, Manisa.
22. Kiremitçi, B. (2005). *Veri Ambarlarında Veri Madenciliği ve Ulaştırma-Lojistik Sektöründe Bir Uygulama*, Yüksek Lisans Tezi, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
23. Gülce, A. C. (2010). *Veri madenciliğinde Apriori Algoritması ve Apriori Algoritmasının Farklı Veri Kümelerinde Uygulanması*, Yüksek Lisans Tezi, Trakya Üniversitesi Fen Bilimleri Enstitüsü, Edirne.
24. Argüden, Y. (2008). *Veri Madenciliği: Veriden Bilgiye, Masraftan Değere*. İstanbul, ARGE Danışmanlık Yayınları.
25. Safavian, S. R. and Landgrebe, D. (1991). A Survey of Decision Tree Classifier Methodology. *Ieee Transactions on Systems, Man, and Cybernetics*, 21(3), 660-674.
26. Quinlan, J. R. (1993). *Program for Machine Learning. C4.5* (Birinci Baskı) California, Morgan Kaufmann Publishers
27. Gülçe, G. (2010). *Veri Ambarı ve Veri Madenciliği Teknikleri Kullanılarak Öğrenci Karar Destek Sistemi Oluşturma*, Yüksek Lisans Tezi, Pamukkale Üniversitesi Fen Bilimleri Enstitüsü, Denizli.
28. Akman, M. (2010). *Veri Madenciliğine Genel Bakış ve Random Forest Yönteminin İncelenmesi: Sağlık Alanında Bir Uygulama*, Yüksek Lisans Tezi, Ankara Üniversitesi Sağlık Bilimleri Enstitüsü, Ankara.

29. Köktürk, F. (2012). *K-En Yakın Komşuluk, Yapay Sinir Ağları Ve Karar Ağaçları Yöntemlerinin Sınıflandırma Başarılarının Karşılaştırılması*, Doktora Tezi, Bülent Ecevit Üniversitesi Sağlık Bilimleri, Zonguldak.
30. Güler, Ç. (2017). *Ortaöğretim Başarısını Etkileyen Faktörlerin Karar Ağacı ile Sınıflandırılması*, Yüksek Lisans Tezi, Kahramanmaraş Sütçü İmam Üniversitesi Fen Bilimleri Enstitüsü, Kahramanmaraş.
31. Pehlivan, G. (2006). *CHAID Analizi ve Bir Uygulama*, Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
32. Çırak, G. ve Çokluk, Ö. (2012). Yüksek Öğretimde Öğrenci Başarılarının Sınıflandırılmasında Yapay Sinir Ağları ve Lojistik Regresyon Yöntemlerinin Kullanılması. *Akdeniz İnsani Bilimler Dergisi*, 3(2), 71-79.
33. Sağıroğlu Ş., Beşdok E., ve Erler M.,(2003). *Mühendislikte Yapay Zeka Uygulamaları1 Yapay Sinir Ağları*, Kayseri, Ufuk Kitap Kırtasiye-Yayıncılık.
34. Öztemel, E. (2003). *Yapay Sinir Ağları* (Üçüncü Baskı), İstanbul, Papatya Bilim, 32,33.
35. Irmak, S., Köksal, C. D., ve Asilkan, Ö. (2012). Hastanelerin Gelecekteki Hasta Yoğunluklarının Veri Madenciliği Yöntemleri İle Tahmin Edilmesi. *Journal of Alanya Faculty of Business/Alanya İstetme Fakültesi Dergisi*, 4(1).
36. Elmas, Ç. (2003). *Yapay Sinir Ağları* (Birinci Baskı), Ankara, Seçkin Yayıncılık.
37. Lantz, B. (2013). *Machine Learning with R*. (Birinci Baskı), Birmingham, Packt Publishing Ltd., 90.
38. Bidgoli, A. M. and Boraghi, M. (2010, December). A Language Independent Text Segmentation Technique Based on Naive Bayes Classifier. *In 2010 International Conference on Signal and Image Processing* (pp. 11-16). IEEE.
39. Akkoç, B. (2012). *Sosyal Ağ Analizi için Bayes Ağlarının Kullanımı*, Yayınlanmamış Yüksek Lisans Tezi, Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Konya.
40. Bilişik, M. T. (2011). Destek Vektör Makinesi, Çoklu Regresyon Ve Doğrusal Olmayan Programlama İle Perakendecilik Sektöründe Gelir Yönetimi İçin Dinamik Fiyatlandırma. *XI. Üretim Araştırmaları Sempozyumu*, 23-24 Haziran 2011.
41. Osowski, S., Siwekand, K., and Markiewicz, T. (2004). MLP and SVM Networks. *In: Proceedings of the 6th Nordic Signal Processing Symposium*, 37-40.
42. Demirci, D. A. (2007). *Destek Vektör Makineleri ile Karakter Tanıma*, Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
43. IBM SPSS Modeler, (2018), Accelerate Time to Value With Visual Data Science and Machine Learning, *IBM Analytics*, 1-8.

44. Ristoski, P., Bizer, C., and Paulheim, H. (2015). Mining The Web of Linked Data With Rapidminer. *Web Semantics: Science, Services and Agents on the World Wide Web*, 35, 142-151.
45. Berthold, M. R., Cebron, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., and Wiswedel, B. (2009). KNIME-the Konstanz Information Miner: version 2.0 and beyond. *AcM SIGKDD explorations Newsletter*, 11(1), 26-31.
46. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The WEKA Data Mining Software: an update. *ACM SIGKDD explorations newsletter*, 11(1), 10-18.
47. İnternet:  
URL:[http://www.webcitation.org/query?url=https%3A%2F%2Fwww.sas.com%2Ftr\\_tr%2Fhome.html&date=2019-05-23](http://www.webcitation.org/query?url=https%3A%2F%2Fwww.sas.com%2Ftr_tr%2Fhome.html&date=2019-05-23), Son Erişim Tarihi: 23.05.2019.
48. Abraham, C., and Poston, M. (2015). Assessing Real World Applications of Data Mining With SAS Enterprise Miner (EM): A Technical Report for Teaching the Big Data Generation. *Journal of Emerging Trends in Computing and Information Sciences*, 6(9).
49. İnternet:  
URL:<http://www.webcitation.org/query?url=https%3A%2F%2Finsideanalysis.com%2Fmarkets%2Foperational-intelligence%2Fkxen%2F&date=2019-05-23>, Son Erişim Tarihi: 23.05.2019.
50. İnternet:  
URL:<http://www.webcitation.org/query?url=http%3A%2F%2F5Bhttps%3A%2F%2Fwww.rdmag.com%2Fproduct-release%2F2006%2F01%2Finsightful-miner-7&date=2019-05-23>, Son Erişim Tarihi: 23.05.2019.
51. Affinium Model User's Guide 7.0, Unica Corporation 1996 - 2006. Waltham, MA 02451-1379.
52. İnternet:  
URL:<http://www.webcitation.org/query?url=http%3A%2F%2Fcedric.cnam.fr%2Ffichiers%2FRFC184.pdf&date=2019-05-23>, Son Erişim Tarihi: 23.05.2019.
53. Michalski, R. S., Kerschberg, L., Kaufman, K. A., and Ribeiro, J. S. (1992). Mining for Knowledge in Databases: The INLEN architecture, initial implementation and first results. *Journal of Intelligent Information Systems*, 1(1), 85-113.
54. Han, J., Fu, Y., Wang, W., Chiang, J., Zaiane, O. R., and Koperski, K. (1996). DBMiner: Interactive Mining of Multiple-Level Knowledge in Relational Databases. In *ACM SIGMOD Record* (Vol. 25, No. 2, p. 550). ACM.
55. İnternet:  
URL:<http://www.webcitation.org/query?url=http%3A%2F%2Fdarwin.cirad.fr%2F&date=2019-05-23>, Son Erişim Tarihi:23.05.2019.

56. Bresfelean, V. P., Bresfelean, M., Ghisoiu, N., and Comes, C. A. (2008, June). Determining students' Academic Failure Profile Founded on Data Mining Methods. *In ITI 2008-30th International Conference on Information Technology Interfaces* (pp. 317-322). IEEE.
57. Şengür, D. ve Tekin, A. (2013). Öğrencilerin Mezuniyet Notlarının Veri Madenciliği Metotları ile Tahmini. *Bilişim Teknolojileri Dergisi*, 6(3), 7-16.
58. Bozkır, A. S., Sezer, E. ve Bilge, G. (2009). Öğrenci Seçme Sınavında (ÖSS) Öğrenci Başarımını Etkileyen Faktörlerin Veri Madenciliği Yöntemleriyle Tespiti. 5. *Uluslararası İleri Teknolojiler Sempozyumu (IATS'09)*, 37-43.
59. Bray, T. M. (1999). *The Shadow Education System: Private Tutoring and Its Implications for Planners*. UNESCO International Institute for Educational Planning.
60. Ramaswami, M., & Bhaskaran, R. (2010). A CHAID based Performance Prediction Model İn Educational Data Mining. *International Journal of Computer Science*, 7(1), 10.
61. Kurt, Ç. ve Erdem, O. A. (2012). Öğrenci başarısını Etkileyen Faktörlerin Veri Madenciliği Yöntemleriyle İncelenmesi, *Politeknik Dergisi*, 15(2), 111-116.
62. Aydın, S. (2007). *Veri Madenciliği ve Anadolu Üniversitesi Uzaktan Eğitim Sisteminde Bir Uygulama*, Yayınlanmamış Doktora Tezi, Anadolu Üniversitesi Sosyal Bilimler Enstitüsü, Eskişehir.
63. Ayık, Y. Z., Özdemir, A., ve Yavuz, U. (2007). Lise Türü Ve Lise Mezuniyet Başarısının, Kazanılan Fakülte İle İlişisinin Veri Madenciliği Tekniği İle Analizi. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 10(2), 441-454.
64. Şık, M. Ş. (2014). *Veri Madenciliği ve Kansere Erken Teşhisinde Kullanımı*, Yayınlanmamış Yüksek Lisans Tezi, İnönü Üniversitesi Sosyal Bilimler Enstitüsü, Malatya.
65. İnternet:  
[URL:http://www.webcitation.org/query?url=http%3A%2F%2Fwww.rapidminer.com&date=2019-05-23](http://www.webcitation.org/query?url=http%3A%2F%2Fwww.rapidminer.com&date=2019-05-23), Son Erişim Tarihi:23.05.2019.

## ÖZ GEÇMİŞ

### Kişisel Bilgiler

Adı Soyadı : Osman KAYHAN

Doğum Yeri : Malatya

Doğum Tarihi : 01.01.1985

### Eğitim Derecesi

Lise : Malatya Hüseyin Köçük Anadolu Ticaret ve Meslek Lisesi (1998-2002)

Lisans : Süleyman Demirel Üniversitesi, Teknik Eğitim Fakültesi, Bilgisayar Sistemleri  
Öğretmenliği (2003 -2007)

### Çalıştığı Kurumlar

Amasya Ticaret Meslek Lisesi Merkez/AMASYA(2007-2014)

Amasya Anadolu Lisesi Merkez/AMASYA (2014-)

**Yabancı Dili:** İngilizce

### İletişim Bilgileri

E-posta: osmankayhan@msn.com

Adres: Ellibeş evler mah. Yeğenzade Sok. Lale Apt. No:6 / AMASYA

### Yayımlar

Kayhan, O. , Ünal, Y. ve Sağlam, A. (2019). Student Graduation Prediction with Data Mining: Amasya University Distance Education Application and Research Center Sample, *3.rd Innovate Approaches In Scientific Studies(Section:Engineering and Natural Sciens)*, 130-133.