





# **KARİYER PLANLAMA İÇİN KARAR DESTEK SİSTEMİ**

## **A DECISION SUPPORT SYSTEM FOR CAREER RECOMMENDATION**

**MUHAMMET AKGÜN**

**PROF. DR ÖZLEM MÜGE TESTİK**

**Tez Danışmanı**

**DOÇ. DR OUMOUT CHOUSEINOLOU**

**Tez Eş Danışmanı**

Hacettepe Üniversitesi

Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin

Endüstri Mühendisliği Anabilim Dalı için Öngördüğü

YÜKSEK LİSANS TEZİ olarak hazırlanmıştır.



**MUHAMMET AKGÜN**'ün hazırladığı "KARİYER PLANLAMA İÇİN KARAR DESTEK SİSTEMİ" adlı bu çalışma aşağıdaki jüri tarafından **ENDÜSTRİ MÜHENDİSLİĞİ ANABİLİM DALI**'nda **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Doç. Dr. Özgür Tolga PUSATLI

Başkan



Prof. Dr. Özlem Müge TESTİK

Danışman



Dr. Öğr. Üyesi Ceren TUNCER ŞAKAR

Üye



Dr. Öğr. Üyesi Barbaros YET

Üye



Dr. Öğr. Üyesi Diclehan TEZCANER ÖZTÜRK

Üye



Bu tez Hacettepe Üniversitesi Fen Bilimleri Enstitüsü tarafından **YÜKSEK LİSANS TEZİ** olarak ..... / ..... /..... tarihinde onaylanmıştır.

Prof. Dr. Menemşe GÜMÜŞDERELİOĞLU

Fen Bilimleri Enstitüsü Müdürü



## ETİK

Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmada,

- tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- başkalarının eserlerinden yararlanılması durumunda ilgili eslere bilimsel normlara uygun olarak atıfta bulunduğumu,
- atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,
- kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- ve bu tezin herhangi bir bölümünü bu üniversite veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

10 /09 / 2019

MUHAMMET AKGÜN







## YAYINLANMA FİKRİ MÜLKİYET HAKLARI BEYANI

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kağıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanması zorunlu metinlerin yazılı izin alarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan “*Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge*” kapsamında tezim aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H. Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- Enstitü / Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren 2 yıl ertelenmiştir.
- Enstitü / Fakülte yönetim kurulu gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren .... ay ertelenmiştir.
- Tezim ile ilgili gizlilik kararı verilmiştir.

10 / 09 /2019

MUHAMMET AKGÜN



# ÖZET

## KARİYER PLANLAMA İÇİN KARAR DESTEK SİSTEMİ

**Muhammet AKGÜN**

**Yüksek Lisans, Endüstri Mühendisliği Bölümü**

**Tez Danışmanı: Prof. Dr. Özlem Müge TESTİK**

**Eş Danışman: Doç. Dr. Oumout CHOUSEINOGLU**

**Eylül 2019, 56 sayfa**

Son yıllarda giderek artan yapay zekâ uygulamaları hayatın her alanına girmeye devam etmektedir. Bu çalışma, makine öğrenmesi uygulamalarının kariyer planlaması alanında kullanılmaya başlamasının sonucu olarak, üniversite mezunlarına çalışma sektörü önerisinde bulunacak bir tavsiye sistemi geliştirmek amacıyla yapılmıştır. Çalışmada makine öğrenmesi algoritmalarından faydalanılarak, mezunların ders notları bilgilerini ve ayrıca özlük bilgileri haricindeki eğitim ve iş bilgilerini kullanarak, yeni mezun olacak olan öğrencilere çalışabilecekleri sektör bazında kariyer planlamaları hakkında tavsiyede bulunan bir sistem tasarlanmıştır. Veri madenciliği süreçlerinde en çok kullanılan yöntemlerden biri olan Cross Industry Standard Process for Data Mining (Çapraz Endüstri Veri Madenciliği Standart Süreci – CRISP-DM) yöntemi çalışmanın özelliklerine uygun olarak seçilmiştir. CRISP-DM yöntemi içerisindeki işin anlaşılması, verinin anlaşılması, veri ön işleme, modelleme, değerlendirme ve konuşlandırma adımları uygulama kısmında çalışmanın akışını belirlemiştir. Modelleme kısmında makine öğrenmesi algoritmalarından (En yakın komşuluk, Rassal orman, Naive bayes, Karar vektör sınıflandırıcıları, Karar ağacı) faydalanılmıştır. Bu kapsamda tanımlanmış olan iki araştırma sorusunun cevaplandırılması için bir vaka çalışması olarak Hacettepe

Üniversitesi Endüstri Mühendisliği Bölümü'nden ve Hacettepe Üniversitesi Öğrenci İşleri Daire Başkanlığından (ÖİDB) alınmış olan veriler kullanılmıştır. Çalışma sonucunda gözetimli makine öğrenmesi algoritmalarının doğruluk oranları karmaşıklık matrisi yardımı ile incelenmiş ve karşılaştırılmış, en iyi sonucu Rassal Orman (%67,46 doğruluk oranı) vermiştir.

**Anahtar Kelimeler:** Öneri Sistemleri, Veri Analizi, Makine Öğrenmesi, Yapay Zekâ, Kariyer Tavsiye Sistemi.

# **ABSTRACT**

## **A DECISION SUPPORT SYSTEM FOR CAREER RECOMMENDATION**

**Muhammet AKGÜN**

**Master of Science, Department of Industrial Engineering**

**Supervisor: Prof. Dr. Özlem Müge TESTİK**

**Co- Supervisor: Assoc. Prof. Dr. Oumout CHOUSEINOGLU**

**September 2019, 56 pages**

In the last years, the increasing number of artificial intelligence applications have been continuously invading our daily lives. This thesis, as a result of the introduction of machine learning approaches to the career planning domain, has been undertaken in order to develop a recommender system that counsels and proposes a work industry to university graduates. A system based on machine learning algorithms that recommends to new graduates an industry to work at, based on the education history, grades and personal information of previous graduates is designed in this study. The Cross Industry Standard Process for Data Mining (CRISP-DM), which is one of the most common data mining processes, is employed after reviewing the characteristics of the problem at hand. The six steps of CRISP-DM, namely understanding the business, understanding the data, preparing the data, modelling, evaluation and setting out, have guided the research methodology. In the modeling phase KNN, Random Forest, Naive Bayes, Support Vector Machines and Decision Tree machine learning algorithms have been utilized. In order to

answer the research questions set by this thesis, a case study based on the data collected by Hacettepe University Department of Industrial Engineering and Hacettepe University Student Affairs Office (ÖİDB) has been designed and executed. At the end of the research, the accuracy of supervised machine learning algorithms has been examined with the use of a confusion matrix, and the best compared result has been obtained from Random Forest (with a 67,46% accuracy).

**Keywords:** Recommendation Systems, Data Analytics, Machine Learning, Artificial Intelligence, Career Recommendation Systems.

## TEŐEKKÜR

Bu alıŐma ve yksek lisans hayatım boyunca, her konuda desteklerini hi bir zaman esirgemeyen, bin minnet duyduėum, haklarını asla deyemeyeceėim, ok deėerli danıŐmanlarım Prof. Dr. zlem Mge Testik ve Do. Dr. Oumout Chouseinoglou'ya, lisans hayatımdan bu yana fikirlerini can kulaėıyla dinleyip uygulamaya alıŐtıėım ok deėerli Prof. Dr. Murat Caner Testik hocama, blmde zerimde emeėi olan tm hocalarıma saygılarımı sunar, ok teŐekkr ederim.

Ayrıca yksek lisansım boyunca hep yanımda olan arkadaŐlarıma, bnyelerinde alıŐmıŐ olduėum BiliŐim A.Ő. ve Icterra'daki yneticilerime ve alıŐma arkadaŐlarıma, hayatım boyunca hep arkamda olan aileme sonsuz teŐekkrlere.

# İÇİNDEKİLER

ABSTRACT .....	iii
TEŞEKKÜR .....	v
İÇİNDEKİLER.....	vi
ŞEKİLLER DİZİNİ.....	viii
ÇİZELGELER DİZİNİ .....	ix
SİMGELER VE KISALTMALAR.....	x
1. GİRİŞ .....	1
2. LİTERATÜR TARAMASI.....	4
2.1. Yapay Zekâ .....	4
2.2. Makine Öğrenmesi .....	5
2.3. Tavsiye Sistemleri .....	7
3.METODOLOJİ.....	10
3.1. Yapay Zekâ .....	10
3.2. Makine Öğrenmesi .....	12
3.2.1.Gözetimli Öğrenme .....	12
3.2.2.Gözetimsiz Öğrenme.....	13
3.2.3.Takviyeli Öğrenme.....	14
3.3. Tavsiye Sistemleri .....	17
3.3.1. İçerik Tabanlı Tavsiye Metotları .....	17
3.3.2. İşbirlikçi Tavsiye Metotları .....	18
3.3.3. Bilgiye Dayalı Tavsiye Metodu .....	19
3.3.4. Hibrit Tavsiye Metodu .....	19
3.3.5. Tavsiye Sistemlerinde Veri Analizi Süreci .....	19
3.3.5.1. Veri Ön İşleme .....	20
3.3.5.2. Veri Analizi .....	22
3.3.5.3. Sonuç Yorumlama.....	26
4. ÖNERİLEN KARİYER DESTEK SİSTEMİ .....	27
4.1. Veri Toplama.....	27



4.2. Veriyi Anlama.....	28
4.3. Veri Çerçevesi Hazırlama .....	37
4.4. Model Oluşturma .....	42
4.5. Değerlendirme .....	45
4.6. Konuşlandırma.....	45
5. SONUÇ VE TARTIŞMA .....	47
6. Kaynaklar.....	52
EKLER.....	55
EK 1 – Etik Kurul İzin Belgesi.....	55
ÖZGEÇMİŞ .....	56

## ŞEKİLLER DİZİNİ

Şekil 3.1 Gözetimli Öğrenme Süreci.....	13
Şekil 3.2 En Yakın Komşuluk Örneği [27] .....	15
Şekil 3.3 CRISP-DM Metodolojisi Süreç Akışı [30].....	16
Şekil 3.4 Veri Analizi Süreçleri .....	20
Şekil 4.1 Savunma Sanayinde Çalışan Mezunların Not Ortalamalarının Dağılım Grafiği .....	30
Şekil 4.2 Otomotiv Sektöründe Çalışan Mezunların Not Ortalamalarının Dağılım Grafiği .....	31
Şekil 4.3 Elektronik Sektöründe Çalışan Mezunların Not Ortalamalarının Dağılım Grafiği .....	32
Şekil 4.4 Bilişim Sektöründe Çalışan Mezunların Not Ortalamalarının Dağılım Grafiği .....	33
Şekil 4.5 Hava Yolları Sektöründe Çalışan Mezunların Not Ortalamaları Dağılım Grafiği .....	34
Şekil 4.6 Danışmanlık Sektöründe Çalışan Mezunların Not Ortalamalarının Dağılım Grafiği .....	35
Şekil 4.7 Diğer Kategorisinde Değerlendirilen Firmaların Not Ortalamalarının Dağılım Grafiği .....	36
Şekil 4.8 Cinsiyet Dağılımı .....	39
Şekil 4.9 Erasmus Programına Katılım Durumu Dağılımı .....	40

## ÇİZELGELER DİZİNİ

Çizelge 4.1 Savunma Sanayinde Çalışan Mezun Verilerinin Temel Özellikleri.....	30
Çizelge 4.2 Otomotiv Sektöründe Çalışan Mezunların Verilerinin Analizi.....	31
Çizelge 4.3 Elektronik Sektöründe Çalışan/Çalışmış Mezunların Verilerinin Analizi ..	32
Çizelge 4.4 Bilişim Sektöründe Çalışan Mezun Verilerinin Analizi.....	33
Çizelge 4.5 Hava Yolları Sektöründe Çalışan Mezunların Verilerinin Analizi .....	34
Çizelge 4.6 Danışmanlık Sektöründe Çalışan Mezunların Verilerinin Analizi.....	35
Çizelge 4.7 Diğer Kategorisinde Yer Alan Mezunların Verilerinin Analizi .....	36
Çizelge 4.8 Ders Notlarının Veri Dönüşümü.....	38
Çizelge 4.9 Bağımsız Değişkenlerin Veri Setine Benzer Örneği .....	42
Çizelge 5.1 Doğruluk Oranlarının Farklı Tekrarlarda Karşılaştırılması.....	47
Çizelge 5.2 Doğruluk Kriterlerinin Karşılaştırılması (1000 Çalıştırma) .....	48
Çizelge 5.3 Doğruluk Kriterlerinin Karşılaştırılması (500 Çalıştırma) .....	49
Çizelge 5.4 Doğruluk Kriterlerinin Karşılaştırılması (200 çalıştırma) .....	49
Çizelge 5.5 Doğruluk Kriterlerinin Karşılaştırılması (100 Çalıştırma) .....	50

## SİMGELER VE KISALTMALAR

### Kısaltmalar

CRISP-DM	Cross Industry Standard Process for Data Mining
KDD	Knowledge and Data Discovery
K-NN	$k$ En yakın komşu
MyEMU	Endüstri Mühendisliği Bölümü Öğrenci ve Mezun Portalı
OİDB	Öğrenci İşleri Daire Başkanlığı
SVM	Support Vector Machine
SEMMA	Sample, Explore, Modify, Model, Evaluate

# 1. GİRİŞ

Günümüzde üniversitelerin lisans bölümlerinden mezun olan öğrenciler kariyer rotalarını çizerken genel anlamda subjektif verilerine bakarak çalışacakları veya başvuracakları şirketleri belirlemekte, kariyerlerini ona göre şekillendirmektedirler. Fakat bu durum ilerleyen zamanlarda öğrencinin kariyer rotasında hatalı davrandığını ortaya çıkarabilmekte ve kişinin belki de bir kaç yılını boşuna harcaması gibi çeşitli kayıplar yaşatabilmektedir. Üstelik bu kayıplara öğrencinin sadece çalıştığı süre zarfı değil aynı zamanda iş arama süreci kaybı ve ileri süreçte kariyer rotasını değiştirip yeni kariyerinde başarılı olmak için harcayacağı zaman kaybı gibi kayıplar da eklenmektedir. Aynı zamanda yanlış başlanılan kariyer rotaları iş dünyasına personel sirkülasyonu olarak geri dönmekte, maddi ve manevi olarak iş dünyasına zarar vermektedir.

Yeni mezun durumunda olan öğrenciler kariyer rotalarını çizerken hangi pozisyonlarda başarılı olacaklarını ön göremeyebilirler. Örneğin, öğrenci sadece iş ihtiyacı olması sebebiyle birçok pozisyona başvurabilir, süreçleri geçebilir ve işe başladıktan sonra pozisyonun kendi benliğine uygun olmadığını veya bu pozisyonda başarılı olmayacağını anlayabilir. Bu durumun sonucunda kişi işini severek yapmayacak ve belki farklı pozisyonlarda çok daha başarılı olabilecek iken mevcut işinde istediği başarıyı yakalayamayacaktır.

Yapılacak olan çalışma ile öğrencilerin kendi bölüm mezunlarına ait verileri incelenecek, bu verilerin yorumlanmasından sonra mezun durumundaki öğrencilerin verileri analiz edilip benzerliklere göre öğrenciye çalışabileceği bir sektör tavsiye edilecektir. Böylelikle öğrenci çok fazla zaman kaybetmeyip tavsiye edilen sektör doğrultusunda iş arama sürecine odaklanabilecektir. Bunun başarılabilmesi için yapay zekâ ve makine öğrenmesi yöntemleri ile desteklenmiş bir tavsiye sisteminin geliştirilmesi planlanmıştır.

Tavsiye sistemleri artık hayatın her alanına girmiş durumdadır. Müzik dinlerken, film izlerken, bir ürün satın alırken, iş ararken, yemek seçerken vs. her yerde karşılaşılmaktadır. Geçmiş tarihi her ne kadar yeni gibi dursa da 10 yılı aşkın bir süredir

hayatın içindedir. Makine öğrenmesi sayesinde gelişen tavsiye sistemleri günümüzde popüler çalışma alanlarından biri haline gelmiştir. Büyük veri sayesinde gelişen algoritmalar öneri sistemlerinin üzerinde çalışılması gereken bir konu olduğunu belirtmektedir.

Öğrencilerin ve mezunların verileri ile veri analizi yapılarak mezun durumunda olan öğrencilere başvurabilecekleri ve çalışabilecekleri sektörlerin önerilmesini sağlayacak bir tavsiye sisteminin geliştirilmesi bu tezin temel amacını oluşturmaktadır. Bu motivasyon doğrultusunda aşağıdaki iki başlıca araştırma sorusu oluşturulmuştur:

- Araştırma Sorusu 1: Mezunların eğitim ve öğrenim bilgileri (ders notları ve ortalama), öğrenci topluluk aktiviteleri, Erasmus katılımları, dil ve beceri sınavları gibi bilgiler kullanılarak ve makine öğrenmesi teknikleri ile yeni mezun öğrencilere çalışabilecekleri ya da başvurabilecekleri sektör önerilebilir mi?
- Araştırma Sorusu 2: Araştırma Sorusu 1’de tanımlanmış olan tavsiye sisteminin oluşturulmasında kullanılmış olan alternatif algoritmaların doğruluk oranlarının karşılaştırılması nedir? Hangi algoritma ne kadar başarılı olmuştur?

Yukarıda belirtilen araştırma sorularının cevabının elde edilmesi için, gözetimli makine öğrenmesi tekniklerinden sınıflandırma yöntemine dayanan bir tavsiye sistemi geliştirilmiştir. Bu tavsiye sistemi için gerekli olan model, Cross Industry Standard Process for Data Mining (Çapraz Endüstri Veri Madenciliği Standart Süreci – CRISP-DM) veri madenciliği yöntemi kullanılarak oluşturulmuştur. Geliştirilen tavsiye sistemi yeni mezunlara, eski mezunların özellikleri kullanılarak bir sektör önerdiği için işbirlikçi tavsiye sistemi olarak çalışmaktadır.

Birinci ve ikinci araştırma sorularının cevabının elde edilmesi için Hacettepe Üniversitesi Endüstri Mühendisliği Bölümü mezunları üstünde bir vaka çalışması gerçekleştirilmiştir. Bu vaka çalışmasındaki mezunların eğitim öğretim ve iş verileri kullanılarak, ilk aşamada

tasarlanmış olan tavsiye sisteminin çalışıp çalışmadığı, yani bir kişiye eğitim öğretim bilgileri doğrultusunda çalışabileceği bir sektör önerip önermediği incelenmiştir (Araştırma Sorusu 1). Devamında, tavsiye sisteminin çalışabileceği farklı algoritmaların başarıları, literatürde bunların başarılarının karşılaştırılması için önerilen karmaşıklık matrisi (confusion matrix) kullanılarak değerlendirilmiştir (Araştırma Sorusu 2).

Tezin devamı şu bölümlerden oluşmaktadır: Bölüm 2’de tezde kullanılmış olan başlıca kavramlar olan yapay zekâ, makine öğrenmesi ve tavsiye sistemleri ile ilgili yapılmış olan literatür taraması verilmektedir. Aynı alt başlıklar ile ilgili temel tanımlar, yaklaşımlar ve yöntemler üçüncü bölümde sunulmaktadır. Önerilen kariyer destek sistemi ve vaka çalışmasının detayları, elde edilen veri çerçevesi ve oluşturulan model Bölüm 4’te verilmektedir. Bölüm 5 çalışmadan elde edilen bulgu ve sonuçların tartışmasını ve planlanmış olan gelecek çalışmaları sunmaktadır.

## 2. LİTERATÜR TARAMASI

Bu bölümde, tezde yer alan temel başlıklara ilişkin yapılmış olan literatür taraması yer almaktadır. Bu kapsamda, yapay zekâ, makine öğrenmesi ve tavsiye sistemleri üzerine literatürde yer alan başlıca yayınlar ilgili alt başlıklar altında incelenmiştir.

### 2.1. Yapay Zekâ

Yapay zekâ kavramı, son yıllarda daha önceki dönemlere oranla daha az hatalı çalışmalar yapmak, yeni teorilerde bulunmak ve modelleri deneysel kanıtlar yerine mevcut olanın üstüne inşa etmek amacıyla sıklıkla kullanılmaktadır. Metodoloji açısından, yapay zekâ bilimsel literatüre girmiştir ancak, bilimsel yöntem olarak kabul edilebilmesi için, hipotezler titiz deneylere tabi tutulmalı ve sonuçlar istatistiksel olarak analiz edilmelidir [1].

Yapay zekânın gelişim süresince yapılmış olan ilk çalışmalardan biri McCulloch ve Walter [2] tarafından gerçekleştirilmiştir. Bu çalışmada her birinin on/off olarak nitelenebildiği yapay sinirlerin, “on” durumuna geçmeleri modellenmiş olup, yapay sinir ağlarının başlangıç çalışması olarak kabul edilmektedir. Öte yandan, “yapay zekâ” terimi ilk kez 1956 yılında aralarında McCarthy ve Minsky’nin de bulunduğu alanında uzman on kişinin katıldığı bir konferansta ortaya atılmıştır [3]. McCarthy ve arkadaşları iki aylık bir çalışma organize etmişler ve makinelerin dil kullanımı, özet çıkarma ve insana ait problem çözme gibi özellikleri nasıl kullanabileceklerini incelemişlerdir.

Bu öncü çalışmalardan sonra, yapay zekâ alanında gerçekleştirilmiş birkaç önemli çalışma, yapay zekânın hem araştırma alanlarına nasıl etki ettiğini, hem de günlük hayatımızı nasıl şekillendirdiğini göstermektedir. Russel ve Norvig [4], yapay zekâyâ çağdaş bir yaklaşım sundukları çalışmalarında yapay zekâ tarihçesinden bahsederlerken Arthur Samuel’in kendi kendine dama oynamayı öğrenen ve geliştiricisinden daha iyi oynayan bir oyun programladığını ve bu sayede bilgisayarların sadece ne yapılacağını söyleyen aygıtlar olmaktan çıktığını belirtmişlerdir. Newell ve arkadaşları [5] “insancıl düşünme” yaklaşımının yer aldığı genel problem çözücü bir program geliştirmişlerdir. Programın alt hedefleri ve olası eylemleri ele alma sürecinin, insanların problem çözme



sürecine benzer olduğu ortaya çıkmıştır. McCarthy ve arkadaşları [6] bir sonraki 30 yıl için baskın yapay zekâ programlama dili olan üst seviye Lisp'i ve ilk eksiksiz yapay zekâ sistemi olarak görülebilen farazi bir program olan Advice Taker'ı tanımlamıştır . Bilgi ağırlıklı ilk başarılı sistem, DENDRAL programıdır [7]. Bu program geniş muhakeme adımlarına izin vermiş ve ortaya çıkan sorunlar için uzmanlık alanlarından elde edilen, etkili bilgileri kullanmıştır. Yapay zekânın tıbbi teşhis alanında da uygulamaları olmuştur. Fagan ve arkadaşları [8], kan enfeksiyonlarının teşhisi için MYCIN programını geliştirmişlerdir. Bu programın DENDRAL'den iki büyük önemli farkı bulunmaktadır. Birincisi, DENDRAL kurallarının aksine, MYCIN kuralları genel teorik modellerden çıkmamakta, kurallar uzmanlar görüşleri ve deneyimlere bağlı olarak tanımlanmıştır. İkinci fark ise, tanımlanan kurallara tıptaki belirsizliğin dahil edilmiş olmasıdır. Bir yapay zekâ uygulaması olan R1, ilk başarılı ticari uzman sistemidir ve Digital Equipment Corporation'da uygulanmıştır [9]. Program, yeni bilgisayar sistemleri için siparişleri yapılandırmış ve şirket yılda yaklaşık 40 milyon dolar tasarruf sağlamıştır. Yarowsky [10] kelime-anlam belirsizliği üzerine yaptığı çalışmada “*plant*” kelimesinin cümle içinde hangi anlamda kullanıldığını (“*bitki*” yada “*fabrika*”) sadece iki kelime (*works* ve *industrial plant*) kullanarak %96’lık doğrulukta tahmin edebilmiştir. Hays ve Efron [11] bir fotoğraftaki boşluğu kapatmak için bir algoritma önermişlerdir.

## 2.2. Makine Öğrenmesi

Samuel [12], ilk kez 1959 yılında makine öğrenmesi kavramını ortaya atmıştır. Samuel, geçmişte çok eskilere dayanan bilinen bir algoritması olmayan (satranç ve dama gibi) oyunları oynayabilen bir bilgisayar programı yaratmayı amaçlamıştır. Bilgisayarların ortalama bir insandan daha iyi performans göstereceği öğrenme ve karar verme algoritmalarının mümkün olduğunu kanıtlamış ve bu öğrenmenin gelecek yaşamda gerçekleşeceğini ön görmüştür.

Das ve arkadaşları [13], makine öğrenmesinin uygulamalarını incelemişlerdir. Makine öğrenmesinin temel amacının aslında insan beyni gibi düşünüp, insan gibi davranan bir bilgisayar olduğunu söylemişlerdir. Makine öğrenmesini, gözetimli öğrenme, gözetimsiz öğrenme, takviyeli öğrenme ve öneri sistemleri olarak dört başlık altında incelemişlerdir. Gözetimli öğrenme altında; spam postalarından kurtulma, yüz, konuşma ve el yazısı

tanıma, doğal dil işleme, bilgisayarları görü çalışmaları, metin filtreleme, verileri sınıflandırma, müzik gibi çalışmaların yürütüldüğünü; gözetimsiz öğrenme altında; gen tanıma, sosyal ağ analizi, pazar analizi, medikal kayıtlar, konuşma yakalama gibi çalışmaların olduğunu; takviyeli öğrenme altında; trafik tahmini, bilgisayar oyunları, market stok analizleri gibi çalışmaların yürütüldüğünü; öneri sistemleri altında ise; sayısal reklamlar, fikir analizi, kişisel kullanılan programlar gibi çalışmaların yürütüldüğünü açıklamışlardır.

Chen ve arkadaşları [14], Bitcoin için fiyat tahmini yapılmasını makine öğrenmesi teknikleri ile günlük ve beş dakikalık verileri kullanarak incelemişlerdir. Tahminde bulunmak için hem istatistiksel modelleri kullanırken (Lojistik Regresyon, Doğrusal Diskriminat Analizi), hem de makine öğrenmesi modellerini kullanmışlardır. Bitcoin alım satım sitelerinden almış oldukları verileri kullanmışlar, bu verilerin %75'i ile modelleri eğitirken, %25'ini test için kullanmışlardır. Günlük ve beş dakikalık modellerin farkını oluşturmak için iki farklı veri seti ile çalışmışlardır. Çalışmanın sonunda günlük veri seti ile oluşturdukları istatistiksel modeller için lojistik regresyon %66 başarıya ulaşırken, makine öğrenmesi modellerinden destek vektör makinesi %65.3 başarıya ulaşmıştır. Günlük veri ortalamasında istatistiksel modeller %65 ortalama doğruluk başarırken, makine öğrenmesi modelleri ortalama %55 doğruluk ile çalışmışlardır. Beş dakikalık veri setinde ise istatistiksel modeller ortalama %53 başarı sağlarken, makine öğrenmesi modelleri ortalama %62.2 başarı sağlamışlardır.

Chung ve arkadaşları [15], öğrencilerin okuldan ayrılmalarını önlemek amacıyla makine öğrenimi kullanarak tahmin edici bir model geliştirmişlerdir. Rassal orman tekniğini kullanarak ayrılma durumunda olan öğrencilerin riskini tahmin edip onlara yardım etmeyi planlamışlardır. 165.715 lise öğrencisinin verisi ile izinsiz gelmeme gibi verilerini alarak veri seti oluşturmuşlardır. Karmaşıklık matrisine göre %95 oranında doğruluk tespit etmişlerdir.

Tomasevic ve arkadaşları [16], gözetimli veri madenciliği ve makine öğrenmesi tekniklerini kullanarak öğrencilerin sınav performanslarını tahmin etmişlerdir. Yapmış oldukları çalışmada, farklı makine öğrenmesi modellerini hem sınıflandırma hem de

regresyon modellerinde incelemişlerdir. Sınıflandırma modellerinde en iyi performansı yapay sinir ağları göstermiştir. Ayrıca sınıflandırma modelleri için yapmış oldukları veri seti değişimleri ile demografik verilerin sınav tahmin sonucuna çok etki etmediğini gözlemlemişlerdir. Aynı şekilde regresyon modellerinde de benzer sonuçlar yakalamışlardır. Tam anlamıyla not tahmininden performans almak için yeterli miktarda veri olması gerektiğini savunmuşlardır.

### **2.3. Tavsiye Sistemleri**

Literatürdeki başlıca kariyer belirleme ve profesyonel anlamda tercih yapmada destek olan tavsiye sistemlerinin kısa bir taraması burada sunulmaktadır. Chien ve Chen [17], teknoloji şirketleri için en önemli sermayenin bilgi sahibi çalışanlar olduğunu ve bu çalışanların iyi seçilemediğini öne sürmüştür. Her ne kadar insan kaynakları birçok test yapıyor veya adayın yeterliliklerini ölçüyor olsa da özellikle yönetim pozisyonları için bulunan adayların bazı zamanlarda yeteri kadar performans alınmadığı görülmüştür. Çalışma, doğru yeteneği doğru pozisyon ihtiyacına göre almayı hedeflemiştir. Bu çalışmada doğru yeteneği işe alıp maksimum fayda sağlamak ve kişiyi elde tutmak için veri madenciliği kullanılmıştır. Kişilerin iş performanslarını, kişilik özelliklerini değerlendirmeye alıp çalışan önerisinde bulunacak bir otomasyon düzenlemiştir. Burada bir şirketin verilerinden yararlanılmış ve geliştirilen sistem ile daha iyi adaylar önerilmiş, alınan adaylardan daha iyi performans girdileri oluşmuştur. Demografik özelliklerden sadece yaş, cinsiyet, evlilik durumu, iş deneyimi, eğitim bilgileri gibi bilgiler kullanılan bu çalışmada diğer demografik özellikler de katıldığında doğruluğun muhtemelen artacağı gözlemlenmiştir.

Diğer yandan Vialardi ve arkadaşları [18], Lima Üniversitesi Sistem Mühendisliği Bölümünde son 7 yıl içerisinde kayıt olan öğrencilerin verilerinden veri madenciliği yaparak öğrencilere akademik güzergahlarının nasıl ilerlemesi gerektiğini ve hangi dersleri hangi öğretim üyelerinden görmesi gibi kararlar için, eski öğrencilerin verilerinden yararlanarak yeni öğrencilere tavsiyelerde bulunmuşlardır. Veri setlerini geçmiş öğrencilerin demografik özellikleri, aldığı dersler, aldığı puanlar, her dönem kaç ders aldığı, her yıl akademik ortalaması gibi metriklerden oluşturmuştur. Yaptıkları çalışmanın sonucunda yaklaşık olarak %80 başarılı tahminlerde bulunacağı ön

görülmüştür. Bu çalışmada eğitim alanında öneri sistemlerinin ve veri madenciliğinin önemini belirtilmiştir. Çalışma sonucunda %77.8 doğruluk oranına ulaşmıştır.

Paparrizos ve arkadaşları [19], yaptıkları çalışmada yeni iş arayan herkese bir iş öneren bir tavsiye sistemi geliştirmek istemiştir. Web’de bulunan, herkese açık kaynaklara dayanan veri setini oluşturmuştur. Geçmiş veri kümelerinden yararlanarak kişinin bir sonraki iş geçişini tahmin etmeye ve de o iş geçişi hakkında makine öğrenmesi sayesinde kişiye tavsiyelerde bulunmayı amaçlamışlardır. Yaptığı çalışmalar sonucunda iş geçişlerinin doğru bir şekilde tahmin edilebildiğini ve buna göre tavsiyelerde bulunabileceğini belirtmişlerdir. Veri seti olarak yaklaşık 5.000.000 kişinin verisini ve 1.000.000’den fazla şirketi incelemişlerdir. Gelecek çalışmalarda sadece objektif durumları değil aynı zamanda kişilerin sosyal yönlerini de çalışmaya dahil ederek incelemeyi ön görmüşlerdir.

Razak ve arkadaşları [20], bulanık mantık yaklaşımı ile öğrencilere kariyerlerini seçerken yönlendirme ve rehberlik sağlayan bir öneri sistemi oluşturmuştur. Bu öneriyi yaparken kariyer testlerini temel almışlardır. Bilgisayar mühendisliği öğrencilerine yapmış oldukları test sonuçlarını bulanık mantık ile analiz ederek çıkan sonuçlara göre kariyer tavsiyesinde bulunmuşlardır. 15 soruluk testi zayıf, orta ve iyi olarak değerlendirmişlerdir. Sonuçları bulanık mantık doğrultusunda değerlendirip bilgisayar mühendisliği öğrencilerine öneride bulunmuşlardır.

Zhang ve arkadaşları [21], işbirlikçi filtreleme ile iş öneri sistemi üzerine araştırma yapmışlardır. Daha önce başvuru ve beğenilen işlerin öğrenci özgeçmişi ile uyumunu inceleyerek işbirlikçi filtreleme ile kullanıcıya en iyi eşleşen işleri listelemişlerdir. Öğrencilerin başvurduğu işlerle kıyaslayınca oldukça başarılı bir sonuç elde etmişlerdir.

Almalis ve arkadaşları [22], iş arayanlar ile iş pozisyonların uygunluğu için makine öğrenmesi ile bir algoritma üzerinde çalışmışlardır. Algoritma, dört boyutlu öneri algoritması olarak isimlendirilmiştir. Bunun nedeni, algoritmasında bir iş arayanın iş pozisyonuna uygunluğunu, aday profili ile pozisyonun alt yapısı, iş tanımı ve adayın öz

geçmişini bulunmasıdır. Deneylerinin sonucunda algoritma içeriğinin faydalı olduğunu kanıtlamışlardır.

Deshpande ve Emmanuel [23], öneri sistemlerini genel bir bakış açısıyla incelemişlerdir. Özellikle de tahminler için benzerlik ölçütlerine sahip işbirlikçi filtreleme yöntemini ele almışlardır. Öneri sistemlerinin çeşitlerini işbirlikçi filtreleme, içerik tabanlı filtreleme ve hibrit filtreleme olarak belirtmişlerdir. Sistemlerde benzerlik ölçümleri ile tahminlerinde Öklit uzaklığı, Pearson korelasyon katsayısı ve oran tahmini olmak üzere üç farklı yaklaşım uygulanmıştır.

Patel ve arkadaşları [24], öğrencilerin iş arama sitelerinden yapmış oldukları aramalarda gereğinden fazla bilgiden dolayı çok zaman kaybettiklerini ve var olan iş öneri sistemlerinin çoğunun sadece kullanıcının ilgisine göre olduğunu, asıl önemli olanın kullanıcının yetkinliklerine göre önermesi gerektiğini savunmuşlardır. Bu durumdan yola çıkarak bir kariyer öneri sistemi geliştirmişlerdir. Bu sistemi geliştirirken metin madenciliği, doğal dil işleme ve makine öğrenmesi gibi alanlardan faydalanmışlardır. Amaçlarının öğrencilerin yetkinliklerine ve karakteristik özelliklerine göre iş önermek olduğunu belirtmişlerdir. San Jose State üniversitesinin kariyer merkezinde bulunan 10.000 iş ilanı ve 100 öğrencinin verisinden yararlanmışlardır.

### 3.METODOLOJİ

Tezin bu kısmında tezde kullanılan yöntemler olan yapay zekâ, makine öğrenmesi ve tavsiye sistemleri kavramları tanıtılmaya çalışılmıştır. Yapay zekâ kavramını tanıtırken yapay zekânın ne olduğu hakkında en genel haliyle bilgi verilmiş, hangi alanlarda hangi çalışmalar yapıldığı anlatılmıştır. Makine öğrenmesi başlığında; yapay zekânın alt dallarından biri olarak kabul edilen makine öğrenmesi kavramını, türlerini ve bu çalışmada nasıl kullanıldığı anlatılmıştır. Tavsiye sistemleri başlığında ise daha çok teknik detay verilmiş, bu çalışmada kullanılan yöntemler belirtilmiştir.

#### 3.1. Yapay Zekâ

Russell ve Norvig [4] yapay zekâ kavramını açıklarken dört ayrı yaklaşım kullanmışlardır. Bu yaklaşımları insan merkezli ve mantık odaklı olarak iki ayrı grupta açıklamışlardır. İnsan merkezli yaklaşımlar grubunun altında insanca düşünme ve insanca davranış varken, diğer yandan mantık odaklı yaklaşımlar grubunda mantıklı düşünme ve mantıklı davranış yaklaşımları bulunmaktadır. İnsan merkezli yaklaşım deneysel yöntemlerin insan davranışlarıyla etkileşimine dayanırken, mantıklı yaklaşım ise matematik ve mühendisliğe dayandırılmıştır.

İlk yaklaşım olarak tanımlanan ve insanca davranış olarak bilinen daha çok Alan Turing tarafından yapılmış olan Turing Testi ile yapay zekâyı tanımlamakta en ilham verici çalışma olmuştur. Yapay zekânın büyük bölümünü kapsayan doğal dil işleme, bilgi işlem, otomatikleştirilmiş muhakeme, makine öğrenmesi, bilgisayarlı görü ve robotik gibi disiplinler Turing Testi aracılığıyla ön görülmüştür. Birçok araştırmacı bu yaklaşımdan yola çıkmış ve yapay zekâya katkıda bulunmuştur. İnsanca düşünme yaklaşımı ise, uygulamalardaki girdi ve çıktıların insan davranışlarına uyumluluğunu kanıtladığı takdirde o uygulamanın bazı durumlarda insanlarda da çalışabileceğini ifade eder.

Mantıksal düşünme ise düşünce kanunları üzerinde duran bir yaklaşımdır. Bir kavramın girdileri doğru ise çıktıları da doğru olacaktır. Bu yaklaşım prensipte doğru olarak gözükse de pratikte çok doğru olmayacaktır. Ayrıca bilgi eksikliği ve bilgi yanlışlığı gibi durumlarda bilgisayarın hesaplama kaynakları tükenecektir. Mantıksal davranış ise

mantıksal karar verici üzerine odaklanır. Mantıksal karar vericiler birkaç olası durum çıktısı olduğunda en mantıklısını seçmeyi beklerler. Bu yaklaşımın diğer yaklaşımlara oranla avantajlı olduğu durumlar vardır. Örneğin düşünce kanunları gibi tek doğrusu yoktur birden fazla sonuç içerisinde doğru çıkarım yapmaya çalışır. Ayrıca insanca davranışlardan ziyade mantıksal rasyonellikleri iyi belirlenmiş sınırlar içerisinde daha doğru kararlar verebilecektir. Fakat bu yaklaşımda kesin bir doğru sonucu seçecek diye bir durumda kalmayacaktır. Tam net olarak kesin sonucu elde etmenin yeteri kadar zamanda olmayacağı kabul edilir. Yapay zekânın temellerinde farklı disiplinler yatar. Bu disiplinler kendi içinde bazı kavramlara ve sorulara cevap ararken aslında yapay zekâ kavramının temellerini atmışlardır. Felsefe, matematik, ekonomi, nörobilim, psikoloji, bilgisayar mühendisliği, kontrol teorisi ve dil bilimi gibi disiplinler yapay zekânın temellerini oluşturmuşlardır. Kısacası yapay zekâ bu dört yaklaşım altından ortak bir paydaya çıkartıldığı zaman makinelerin insan gibi karar alması ve aldığı kararları uygulaması denebilir. Yapay zekânın gelişimi süresince ortaya çıkan ve yapay zekânın alt alanları olarak kabul edilen önemli alanlar kısaca robotik araçlar (Mojave çölünün engebeli arazisini tek başına geçen sürücüsüz otomobil STANLEY), konuşma tanıma (United Airlines'ın uçuş rezervasyon sistemi), otonom planlama ve zamanlama (NASA'nın Uzak Ajan programı), oyunlar (ilk defa bir dünya şampiyonunu satrançta yenen IBM'in DeepBlue bilgisayarı), spamla mücadele (günde bir milyardan fazla mesajı spam olarak etiketleyip kullanıcıların zaman kaybetmesine engel olan spam öğrenme algoritmaları), lojistik planlama (1991'deki İran Körfezi krizi sırasında, ABD kuvvetlerinin kullandığı Dinamik Analiz ve Yeniden Planlama Aracı), robotik (ev kullanımı için iki milyondan fazla üretilmiş olan Roomba robot elektrik süpürgeler), çeviri (iki trilyondan fazla metin örneklerinden oluşturulmuş istatistiksel modeller kullanan çeviri programları) ve tavsiye sistemleri (Amazon, Netflix, Spotify gibi şirketlerin müşterilerine belli kurallar doğrultusunda yeni ürünler tavsiye etmeleri) olarak sıralanabilir. Bu örnekler şu anda yapay zekânın hangi alanda kullanıldığına dair kısa bilgilerdir. Bu çalışmada yeni mezun olacak öğrencilerin başvurabilecekleri sektörler tahmin edildiği için yapay zekâ biliminin alt dallarından biri olan makine öğrenmesi ile birlikte tavsiye sistemleri uygulamaları bir arada kullanılmıştır.

### **3.2. Makine Öğrenmesi**

Alpaydin [25] makine öğrenmesini; süreçlerin tam olarak belli olmadığı ya da tam anlamıyla algoritmalara dökülemediği, ancak iyi ve kullanışlı bir yaklaşımın ön sürülebildiği, eldeki verilerin bir kısmıyla tam bir sonuç gösterilemese dahi iyi bir varsayımın makineler yani bilgisayarlar tarafından önerebildiği yapay zekânın bir parçası olan bilim dalı olarak tanımlamıştır. Bir başka deyişler, makine öğrenmesi, bilgisayardan gerçek dünyaya veya gerçek dünyadan bilgisayara olan aradaki çizgi anlamına gelen yapay zekânın bir alt kolu olarak tanımlanabilir. Makine öğrenmesi, bilgisayarların olduğu ve verinin olduğu her yerde kullanılabilir. Dijital dünya geliştikçe üretilmekte, aktarılmakta ve saklanmakta olan veri artmakta ve insanların uğraşamayacağı kadar büyük veri yığınları ortaya çıkmaktadır. Bu artan veri ve sürekli daha yetenekli hale gelen donanım ve yazılımlar, makine öğrenmesinin son yıllarda çok daha etkin ve yaygın kullanılmasına olana vermiştir.

Bilgisayarlı görü makine öğrenmesinin ilk çalışmalarından biridir. Nesne tanıma, görüntü işleme veya bilgisayarlı görü derin öğrenme konusu altında işlenir. Sanal gerçeklik, co-op, arttırılmış gerçeklik, ses tanıma, robotik, pazarlama, tavsiye algoritmaları, sağlık sistemleri, coğrafi bilgi sistemleri, güvenlik sistemleri, arama motorları, Endüstri 4.0, ve nesnelerin interneti makine öğrenmesinin üstünde durduğu konuların başında gelir.

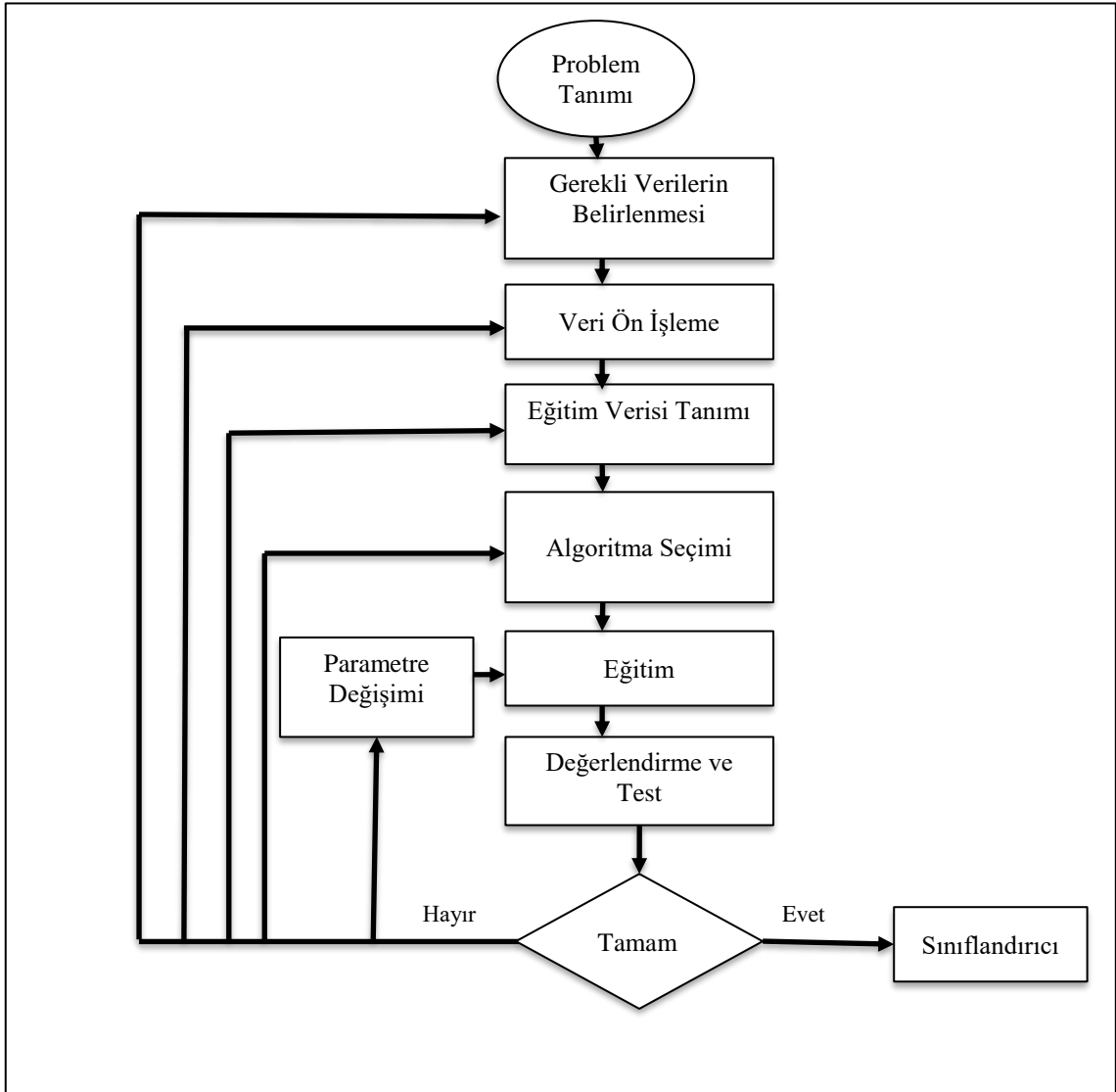
Makine öğrenmesi, gözetimli öğrenme, gözetimsiz öğrenme ve takviyeli öğrenme olmak üzere üç ana başlık altında incelenir.

#### **3.2.1. Gözetimli Öğrenme**

Gözetimli öğrenme, girdi verileri ile çıktı verisi tanımlanmış olup, girdi verilerinden çıktıya ulaşmayı sağlayan fonksiyonun ne olduğu bir algoritma yardımı ile keşfedilmeye çalışılmaktadır. Elde edilecek olan sonucun elde hazır olduğu öğrenme yöntemidir. Eğitim verisi ile makinenin öğrenmesi sağlanır, öğrenme işlemi tamamlandıktan sonra gelen veriye göre makine tahmin yapar. Gözetimli öğrenme içerisinde regresyon ve sınıflandırma olmak üzere iki yöntem bulunmaktadır. Regresyon yöntemi, üretilen çıktıların sayısal olduğu durumlarda kullanılır ve başta doğrusal regresyon, çoklu



regresyon, polinomial regresyon, karar ağacı ve rassal orman (random forest) olmak üzere farklı yöntemler kullanır. Sınıflandırma yöntemi ise, üretilen verinin kategorik olduğu durumlarda kullanılır ve lojistik regresyon, K-NN (en yakın komşu), destek vektör, Bayes, karar ağacı ve rassal orman gibi kategorik verilere uygun yöntemler kullanır. Gözetimli öğrenmenin adımları Şekil 3.1’de gösterilmektedir.



Şekil 3.1 Gözetimli Öğrenme Süreci

### 3.2.2. Gözetimsiz Öğrenme

Eğitim verisi hazırlanması çok mümkün olmayan algoritmalar için kullanılır. Gözetimli öğrenmeden farklı olarak sadece girdi verilerinin hazır bulunup çıktı verisinin olmadığı

durumlarda kullanılan öğrenme yöntemidir. Gözetimsiz öğrenme içinde kümeleme, birliktelik kuralı ve boyut azaltma olmak üzere üç yöntem bulunmaktadır.

- **Kümeleme:** Veri setini birbirine benzer gruplara ayırma anlamına gelir. Gruplar arasındaki benzerliğin olabildiğince az olması beklenir. *K-Means* algoritmasını kullanır.
- **Birliktelik kuralı:** Veri seti arasındaki ilişkileri inceleyen yöntemdir. Bir benzerlik yakalamaya çalışır. Veriler arasında bağlantı kurmaya çalışır. Kullandığı algoritmalar *Apriori* ve *Eclat*'dir.
- **Boyut azaltma:** Veri setini en iyi sonucu verecek olan algoritmalar için düzenlemektir. Yararsız olan öznelikler veri setinden çıkartılır. Temel bileşenler analizi en çok kullanılan algoritmadır.

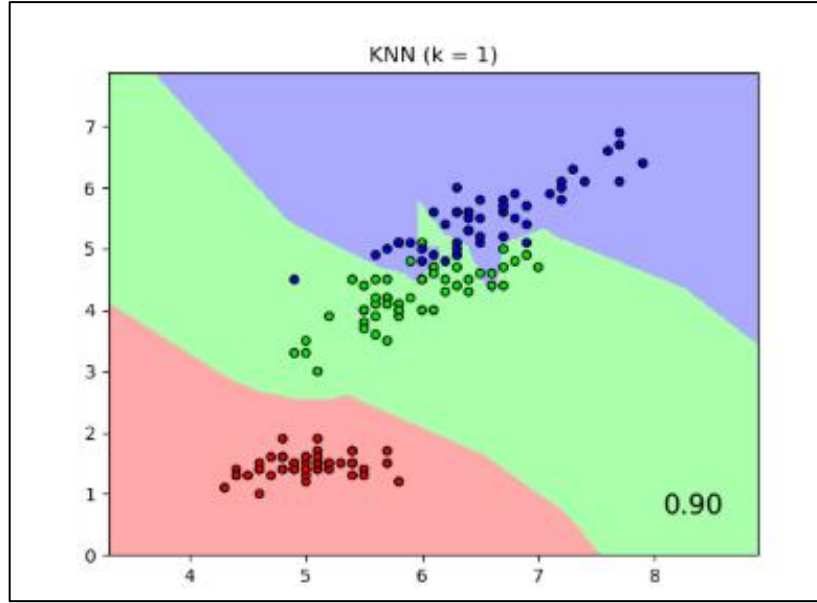
### 3.2.3. Takviyeli Öğrenme

Derin öğrenme algoritmaları takviyeli öğrenmenin bir sonucudur. İstenilen sonucu aldığı zaman devam eden alamadığı durumlarda hedefin o olmadığını saptayarak hep istenilen sonuca yönelmeye çalışan sürekli öğrenen bir yöntemdir.

Bu çalışmada çıktı olarak elde edilmesi hedeflenen bağımlı değişkenin kategorik ve nominal veri olması nedeniyle gözetimli öğrenme ve içerisindeki sınıflandırma yöntemi seçilmiştir. Sınıflandırma yönteminde farklı teknikler denenerek optimal sonucu veren algoritma kullanılmıştır.

Sınıflandırma yöntemindeki teknikleri Kotsiantis ve arkadaşları tarafından [26] aşağıdaki gibi açıklamıştır.

- *K-NN (En Yakın Komşu):* Hangi sınıfa ait olduğu bulunmak istenen verinin önceden sınıflandırılmış (LazyLearning) veri içerisindeki konumuna ve en yakın üç komşusuna (Öklid veya Minkowski uzaklığı) bakılarak hangi sınıfa ait olduğu belirlenir. Şekil 3.2'de örneği verilmiştir.



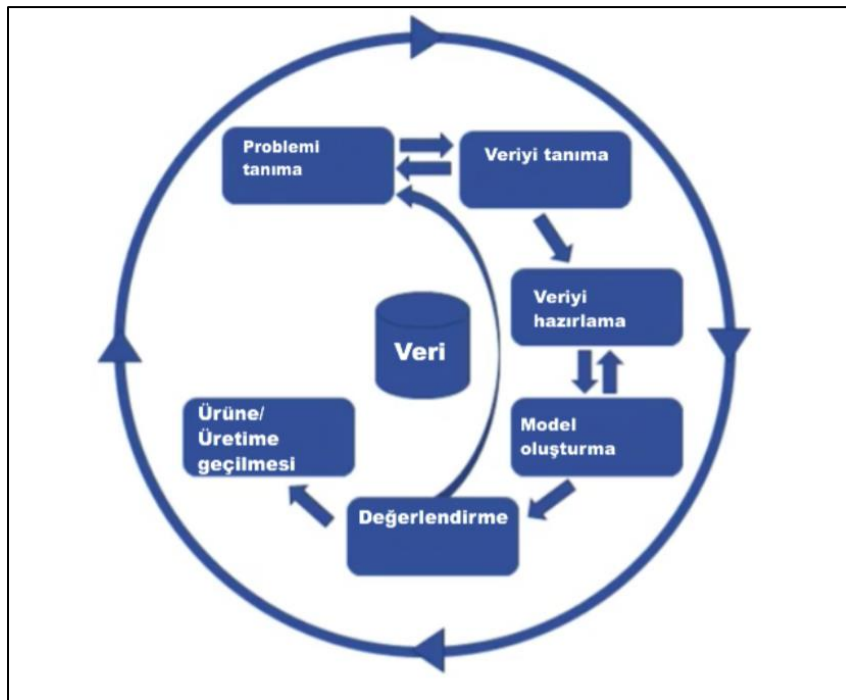
Şekil 3.2 En Yakın Komşuluk Örneği [27]

- *Naive Bayes Sınıflandırıcısı*: Temeli Bayes teoremine dayanır. Verideki her bir sonuç için diğer sonuçların olasılığını hesaplar ve en yüksek olasılık olanı döndürür.
- *Karar ağacı (Decision Tree)*: Veri kümesini gruplara bölerek ilerler. Bütün grubun etiketleri aynı olana kadar veya ağırlıkları farkı belirgin olana kadar devam eder.
- *Rassal Orman (Random Forest)*: Karar ağacı yönteminin bir problemi olan overfitting (aşırı öğrenme) problemini çözer. Veriyi defalarca bölüp alt setler şeklinde eğitir ve bunların hepsini karar ağacı yönteminde derler ve tahminler oluşturur. En sonunda sınıflandırma algoritması için en çok oy alan tahmini seçer.
- *Destek Vektör Makinesi (Support Vector Machine - SVM)*: Cortes ve Vapnik tarafından 1995 yılında önerilmiş olan yöntem, verileri sınıflandırırken aralarında bir doğru çizmeye çalışır [28].

Makine öğrenmesi uygulamalarında veri analizi yaparken CRISP-DM, “Sample, Explore, Modify, Model, Evaluate” (SEMMA ) ve Knowledge and Data Discovery (Bilgi ve Veri Keşfi – KDD) metodolojilerinden biri kullanılır.

- KDD süreci sırasıyla; veri seçimi, veri ön işleme, veri dönüşümü, veri madenciliği ve değerlendirme olarak ilerler.
- SEMMA süreci sırasıyla; örnekleme, keşfetme, dönüştürme, model oluşturma, değerlendirme aşamalarından oluşur. Sas firması tarafından ortaya çıkarılmıştır.
- En çok kullanılan CRISP-DM metodu ise, problemi tanıma ile başlar, ardından veriyi tanıma ve veri hakkında bilgi edinme ile devam eder. Bu aşamalardan sonra veriyi hazırlama ve modelleme aşamaları gelir. Modelleme kısmında alınan sonuç yetersiz görülürse tekrardan veriyi hazırlama kısmına geçilebilir. Sonra değerlendirme ve üretim aşamaları gelir. Değerlendirme kısmı tatminkar ise model ürün haline dönüştürülebilir [29].

Bu tez çalışmasında da CRISP-DM süreci tercih edilmiş, Şekil 3.3'te gösterilen adımlarına uyarak çalışma ortaya koyulmuştur.



Şekil 3.3 CRISP-DM Metodolojisi Süreç Akışı [30]

Bu çalışmanın uygulama kısmında veri ön işleme, modelleme, test gibi aşamaları kolaylıkla yapılmasını sağlayan *sci-kit learn* paketi ile *numpy* ve *pandas* kütüphanesi

kullanılmıştır. Pandas kütüphanesi genellikle veri yapılandırma, filtreleme, dönüştürme işlemlerinde kullanılır. Sci-kit learn paketi içerisinde model oluşturma ve test kısmında kullanılan birçok kütüphane bulunmaktadır. Özellikle sınıflandırma yöntemlerini kullanırken, veriyi normalleştirirken, test için bölerken, eğitirken ve tahmin işlemleri yaparken kullanılır [31].

### **3.3. Tavsiye Sistemleri**

Al-Badarenah ve Alsakran'a göre [32], tavsiye sistemleri gerek ticari ve gerek akademik anlamda giderek önemli hale gelmektedir. Tavsiye sistemleri özellikle dijital dünyada kişilerin bir ürün ya da bir davranış için tercih yapmak durumunda kaldıklarında karar vermelerine yardımcı olan sistemlerdir. Tavsiye sistemlerinin kullanıcıya ürün ya da davranış önermek için kullandığı ana adımlar, kullanıcı verilerinin analiz edilmesi, faydalı bilgiler çıkarımı ve kullanıcılar için optimal öğeyi ön görmeyi içerir.

Tavsiye sistemleri genel olarak, ürün derecelendirmeleri, ürün özellikleri, satın alma geçmişi, kullanıcı demografik bilgileri, metin yorumları ve kullanıcı verilerini kullanarak kullanıcıya en iyi öneriyi sunmayı hedeflemektedir. Tavsiye sistemleri farklı sınıflandırmalara sahip olsa da genel olarak dört ana başlık altında sınıflandırılmıştır:

#### **3.3.1. İçerik Tabanlı Tavsiye Metotları**

İçerik temelli öneriler, henüz görülmemiş öğelerin içeriği ile kullanıcının tercihleri (kullanıcının geçmişte beğendikleri öğeler) arasındaki benzerliklere dayanarak, hedef kullanıcıya öğeler (örneğin makaleler, müzikler, görüntüler) önerir. Örneğin, sistem bir web sayfasındaki anahtar kelimelerin oluşumunu kullanıcının ilgisiyle ilişkilendirmeye çalışabilir. Kullanıcı tarafından derecelendirilen öğelerin içeriği analiz edilerek, kullanıcının ilgi alanlarının bir profili oluşturulur. Daha sonra, kullanıcı, öneri sistemiyle etkileşime girdiğinde, önerilen ürünler kullanıcının profilindeki öğelere benzer öğelerden oluşur. Saf içerik tabanlı bir tavsiye sistemi bazı dezavantajlara sahiptir. Özellikle, içerik özelliği çıkarma yöntemleri sadece metin verilerini analiz etmeye uygundur. Öte yandan, metin tabanlı olmayan verilere içerik tabanlı sistemler uygulamak, örneğin multimedya (video ve ses akışları), zor bir süreçtir. İkinci bir dezavantaj, içerik tabanlı sistemlerin

yalnızca kullanıcının profili ile yakından ilişkili öğeleri önerebilmesidir; bu nedenle, kullanıcı yalnızca kullanıcının ilgilendiği daha önceki öğelere benzer öğeleri görecektir. Bu sorunun üstesinden gelmenin yollarından biri, kullanıcı profiline çok benzeyen yüksek alaka düzeyindeki öğeleri filtrelemektir. Bu problemin üstesinden gelmenin diğer bir yolu, benzerliğin yanı sıra, tavsiye için öğeleri sıralamak için çeşitliliği kullanmaktır. Son olarak, boş profili olan yeni bir kullanıcı, güvenilir bir öneri almak için yeterli sayıda maddeye girdi sağlamak zorundadır. Bu genellikle yeni kullanıcı problemi olarak adlandırılır, çünkü çok az derecelendirmeye sahip olan yeni bir kullanıcı çoğu zaman uygun öneriler alamaz.

### **3.3.2. İşbirlikçi Tavsiye Metotları**

İşbirlikçi tavsiye sistemleri, bir hedef kullanıcının geçmiş tercihleri ile diğer kullanıcılar veya nesnelere arasındaki benzerliklerini baz alarak tavsiye veren sistemlerdir. İçerik tabanlı tavsiye sistemlerinin aksine, işbirlikçi tavsiye sistemleri kullanıcının derecelendirmelerine dayanarak önerilerde bulunur. Kullanıcılar veya nesnelere tercihlerine göre kümeleme gibi yöntemler ile gruplandırılır. Kümeler oluşturulduktan sonra, hedef kullanıcıyla en güçlü ilişkiye sahip kümeler, kullanıcıya tavsiye olarak sunulur. İçerik tabanlı sistemlerden farklı olarak, işbirlikçi öneri sistemi, çeşitli türde içeriklerle (metin, resim, müzik, video klip vs.) başa çıkabilir; tavsiyelerin kalitesini artırabilir. İşbirlikçi tavsiye metodunda genellikle hafıza bazlı ve model bazlı olmak üzere iki teknik kullanılır.

İşbirlikçi tavsiye metodu, erken puanlayıcı problemi, seyreklik problemi ve gri koyun problemi gibi bazı problemlerden muzdariptir. Erken puanlama problemi (soğuk başlatma problemi), henüz derecelendirilmemiş yeni bir ürün için veya sistemi yeni kullanmaya başlayan bir kullanıcı için tavsiyelerin gerekli olduğu durumları ifade eder. Bu durumda, işbirlikçi metot iyi bir şekilde çalışmaz, çünkü ne nesne ne de kullanıcı hakkında fazla derecelendirme bilgisi yoktur. Seyreklik problemi, mevcut ürün derecelendirmelerinin kullanıcılar arasındaki benzerlikleri belirlemek için yetersiz kalması ve kötü bir öneriye yol açmasıyla ortaya çıkmaktadır. Bu problem aynı zamanda azaltılmış kapsam olarak da bilinir. Gri koyun sorunu, sıradışı tercihleri olan kararsız kullanıcılara kötü bir tavsiyede bulunabilir.

### **3.3.3. Bilgiye Dayalı Tavsiye Metodu**

Bilgiye dayalı öneri yaklaşımları, özellikle çözümlerin tanınması ve doğrulanması için derin ürün bilgisinin gerekli olduğu alanlar için uygundur (finansal hizmetler). Bilgiye dayalı tavsiye veren sistemler, genellikle yapay zekâ tekniklerini benimseyen, alıcı ile derecelendirme yapmak yerine bir ürünün özelliklerine dayanarak bir alıcı ile bir ürün arasında bir eşleşme elde etmek için çıkarımlar yapılırken kullanır [32].

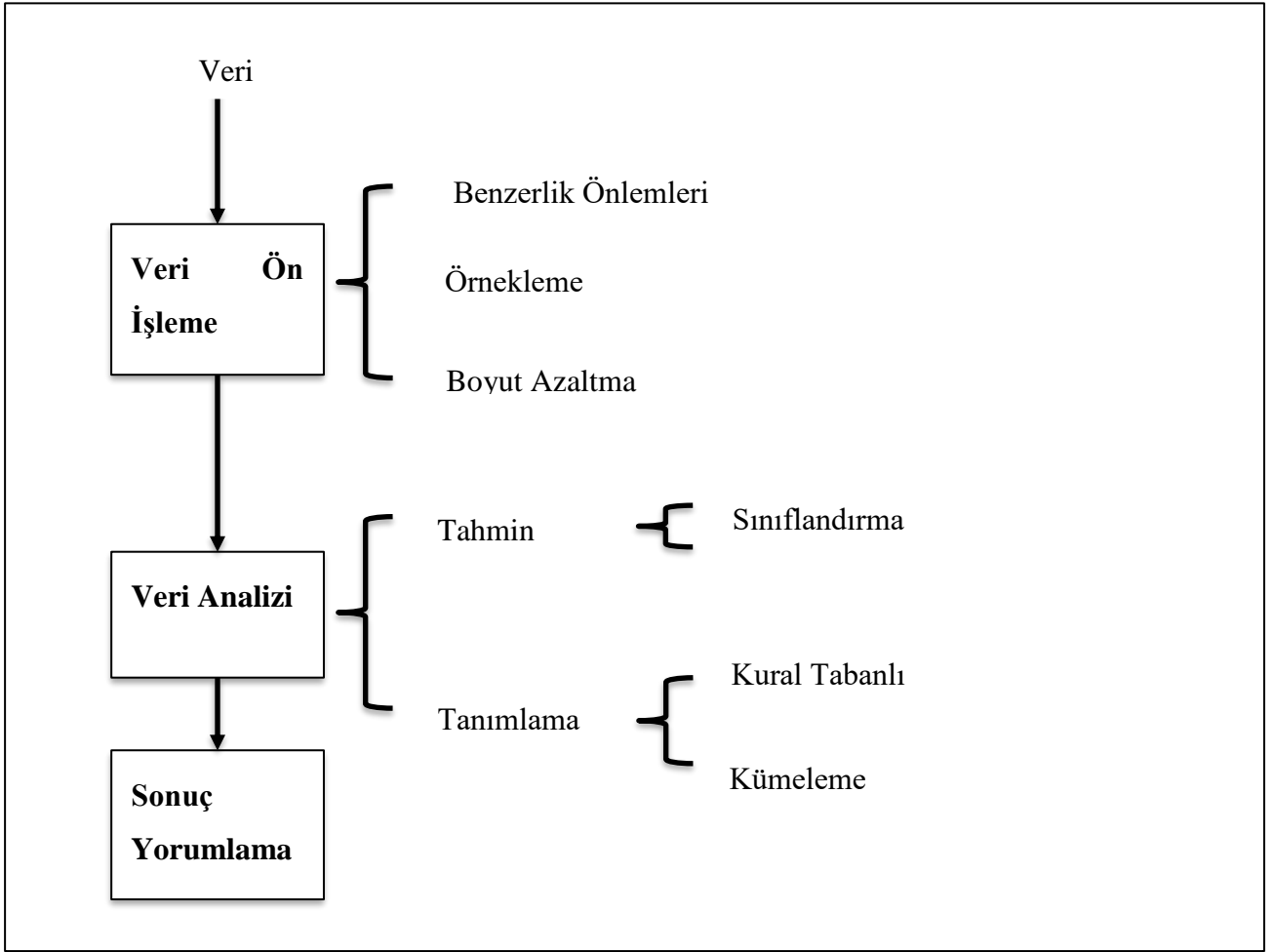
### **3.3.4. Hibrit Tavsiye Metodu**

Hibrit tavsiye metotları, geleneksel tavsiye sistemlerinin doğruluk oranlarını arttırmak amacıyla bilgiye dayalı, içerik tabanlı ve iş birlikçi tavsiye metotlarından oluşan bir yöntemdir. Ağırlıklandırma, dönüştürme ve karıştırma gibi uygulama aşamasında kullanılan teknikleri vardır.

Bu çalışmada hedef kullanıcıya en uygun küme tavsiye edildiği için işbirlikçi tavsiye metodu ile ilerlenmiştir.

### **3.3.5. Tavsiye Sistemlerinde Veri Analizi Süreci**

Amatriain ve arkadaşları [33] , tavsiye sistemlerinin çoğunun özünde bir veri madenciliği tekniğinin özel bir örneği olarak anlaşılabilir bir algoritmaya sahip olduğunu belirtmişlerdir. Veri madenciliği süreci Şekil 3.4'te belirtildiği gibi art arda gerçekleştirilen üç adımdan oluşur: Veri Ön İşleme, Veri Analizi ve Sonuç Yorumlama.



Şekil 3.4 Veri Analizi Süreçleri

### 3.3.5.1. Veri Ön İşleme

Veriler, bir nesnenin niteliği veya özelliği olarak tanımlandığı nesnelere topluluğu ve nitelikleri olarak tanımlanır. Gerçek hayat verilerinin, analiz adımıdaki makine öğrenme teknikleri tarafından kullanılması için önceden işlenmesi (örneğin temizlenmiş, filtrelenmiş, dönüştürülmüş) gerekmektedir. Birlikte ya da ayrı ayrı kullanılacak üç farklı veri ön işleme yöntemi vardır.

#### 1. Benzerlik Önlemleri

İşbirlikçi tavsiye metodu için tercih edilen yaklaşımlardan biridir. Bu veri ön işleme yöntemi çoğu sınıflandırıcı ve kümeleme tekniği gibi uygun bir benzerlik veya mesafe



ölçüsü tanımlamaya son derece bağlıdır. Bu mesafeyi tanımlarken aşağıdaki algoritmalarından yararlanır:

- Öklid uzaklığı
- Minkowski uzaklığı
- Manhattan uzaklığı
- Cosine benzerliği
- Pearson korelasyonu

## *2.Örnekleme*

Örnekleme, büyük bir veri setinden ilgili verilerin bir alt kümesini seçmek için veri madenciliğinde kullanılan ana tekniktir. Hem ön işleme hem de son veri yorumlama adımlarında kullanılır. Tüm veri setinin işlenmesi hesaplama açısından çok uğraştırıcı olduğu için örnekleme yöntemi kullanılabilir. Ayrıca eğitim ve test veri setlerini oluşturmak için de kullanılabilir. Örneklemenin ana konusu temsil edilen orijinal veri setinin bir altkümesini bulmaktır. En basit örnekleme tekniği, rastgele örneklemedir; burada herhangi bir öğeyi seçmede eşit olasılık vardır. Ancak, daha sofistike yaklaşımlar mümkündür. Örneğin, tabakalı örneklemede veriler, belirli bir özelliğe dayalı olarak birkaç bölüme bölünür, ardından her bölüm üzerinde rastgele örnekleme yapılır.

## *3.Boyut Azaltma*

Öneri Sistemleri yalnızca çok boyutlu bir alanı tanımlayan özelliklere sahip bir veri setine sahip olmakla kalmaz, aynı zamanda o alanda çok seyrek bilgi barındırır. Yani, nesne başına sınırlı sayıda özellik için değerler vardır. Kümeleme ve aykırı değer tespiti için kritik olan noktalar arasındaki yoğunluk ve mesafe kavramları çok boyutlu alanlarda daha az anlamlı hale gelir. Boyut azaltma teknikleri, orijinal çok boyutlu alanı daha düşük bir boyutluluğa dönüştürerek bu sorunun üstesinden gelmeye yardımcı olur. En sık kullanılan boyut azaltma yöntemleri temel bileşen analizi ve tekil değer ayrışımıdır. Temel Bileşen Analizi (TBA), yüksek boyutlu veri kümelerindeki kalıpları bulmak için kullanılan klasik bir istatistiksel yöntemdir. TBA, verilerden en az hata cinsinden en büyük varyans miktarını hesaba katan düzenli bir bileşen listesi elde etmeyi sağlar: İlk bileşen tarafından

yakalanan varyans miktarı, ikinci bileşen üzerindeki varyans miktarından daha büyüktür. Tekil Değer Ayrışımı boyutsallığın azaltılması için güçlü bir tekniktir. matris faktörleştirme yaklaşımının özel bir uygulamasıdır ve bu nedenle TBA ile de ilgilidir. Bir tekil değer ayrışımı, yeni özelliklerin “kavramları” temsil ettiği ve koleksiyon bağlamındaki her bir kavramın gücünün hesaplanabilir olduğu daha düşük boyutlu bir özellik alanı bulmaktır.

### **3.3.5.2. Veri Analizi**

Veri analizinde uygulanabilecek birçok farklı yöntem bulunmaktadır. Sınıflandırma, bu yöntemlerden biridir. Sınıflandırıcı, bir özellik alanı ile özellikleri sınıflandırılacak öğelerin özelliklerini temsil ettiği ve etiketlerin sınıfları temsil ettiği bir etiket alanının arasındaki eşlemedir. Örneğin bir restoran Öneri Sistemi, restoranları tanımlayan bir dizi özelliğe dayanarak, restoranları iki kategoriden (iyi, kötü) biri olarak sınıflandıran bir sınıflandırıcı ile uygulanabilir. Sınıflandırma yöntemleri denetimli ve denetimsiz olmak üzere ikiye ayrılır.

Bu çalışmada denetimli sınıflandırma yöntemleri açıklanmış, kümeleme ve kural tabanlı yöntemlere kullanılmadığı için değinilmemiştir. Sınıflandırma yöntemleri aşağıdaki gibidir:

#### *1. En Yakın Komşular (Nearest Neighbors)*

Eğitim verilerini kendi içerisinde kümeleyerek sınıflandırma sonucu bilinmeyen bir hedef geldiği zaman onu en ilgili olan kümenin içine dahil ederek tavsiyede bulunur. Bu sınıflandırıcı, tüm eğitim programını hafızaya alır ve sadece yeni kaydın özelliklerinden biri eğitim örneklerinden biriyle tam olarak eşleşirse sınıflandırır.

#### *2. Karar Ağacı (Decision Tree)*

Karar ağaçları, bir ağaç yapısı biçimindeki bir hedef niteliğinde düğümler içerir. Bu düğümler sınıflandırılacak gözlemler veya niteliklerden ve bunların hedef değerlerinden oluşur. Ağacın düğümleri şunlar olabilir:

a) Karar düğümleri, bu boğumlarda alt ağacın hangi dalının geçerli olduğunu belirlemek için tek bir öznitelik değeri test edilir.

b) Yaprak düğümleri, hedef özniteliğinin değeri test edilir.

Karar ağaçları, öneri sistemleri için model tabanlı bir yaklaşımda kullanılabilir. Kullanıcı tercihlerinde yer alan tüm değişkenleri modelleyen bir karar şeması oluşturmak için içerik özelliklerini kullanır. Tahmin edileceği üzere, karar alma sürecinde yer alan tüm değişkenleri açıklamaya çalışan bir karar şeması oluşturmak çok zordur. Ancak, karar şemaları, sistemin belirli bir bölümünü modellemek için de kullanılabilir.

### *3. Kural Tabanlı Sınıflandırma*

Kural tabanlı sınıflandırma, “if-else” öznitelik bağlantılarından yapılan bir sınıflandırmadır. Sonuç olarak ortaya çıkan kural olumlu veya olumsuz bir sınıflandırmadır. Örneğin bir hedef öznitelikleri kural koşulunu yerine getiriyorsa, bir kuralın belirli bir örneğini kapsadığını söyleriz. Bir kuralın kapsamını, öncülüğünü sağlayan kayıtların kesri olarak tanımlarız. Öte yandan, doğruluğunu hem öncülüğü hem de sonucu veren kayıtların kesri olarak tanımlarız. Kural temelli sınıflandırıcılar sembolik olmaları nedeniyle son derece etkileyici olmaları ve herhangi bir dönüşüm olmadan verilerin öznitelikleriyle çalışması gibi avantajlara sahiptirler.

### *4. Bayes sınıflandırma*

Bayes sınıflandırma, sınıflandırma problemlerini çözmek için olasılık barındıran bir yapıya sahiptir. Koşullu olasılık tanımına ve Bayes teoremine dayanır. Sınıflandırıcı, özelliklerin olasılıksal bağımsızlığını varsayar; yani, belirli bir özelliğin varlığı veya yokluğu, diğerlerinin varlığı veya yokluğu ile ilgili değildir. Ancak, bağımsızlık varsayımı bazı özellikler için geçerli olmayabilir. Bu durumda, olağan yaklaşım Bayesian Belief Networks (BBN) denilen bir yöntem kullanmaktır. BBN’ler, öznitelikler arasındaki bağımlılığı ve her düğümü kendi içinde ilişkilendiren bir olasılık tablosunu kodlamak için dolamsız bir çizge kullanır. BBN’ler, grafiksel bir model kullanarak bir alandaki önceki bilgileri yakalamanın bir yolunu sağlar.

## 5. Yapay Sinir Ağları

Yapay Sinir Ağı (YSA) biyolojik beynin mimarisinden ilham alınan birbirine bağlı düğümlerin ve ağırlıklı bağlantıların bir araya getirilmesiyle oluşan bir ağıdır. Bir YSA herhangi bir sayıda katmana sahip olabilir. Bir YSA'daki katmanlar üç türe ayrılır: giriş katmanı, gizli katman ve çıkış katmanı. Giriş katmanındaki birimler, ağı besleyen verilere yanıt verir. Gizli birimler girdi birimlerinden ağırlıklı çıktıları alır. Ve çıkış birimleri, gizli birimlerden gelen ağırlıklı çıktıya cevap verir ve ağın son çıktısını oluşturur. YSA'nın temel avantajlarından bazıları aktivasyon fonksiyonuna bağlı olarak doğrusal olmayan sınıflandırma görevlerini yerine getirebilmeleri ve paralel yapıları nedeniyle ağın bir kısmı başarısız olduğunda verimli bir şekilde çalışabilmeleridir. YSA, birkaç öneri modülünden veya veri kaynağından girişi birleştirmek için kullanılabilir.

## 6. Destek Vektör Makinesi (SVM)

Bir SVM sınıflandırıcısının amacı, verilerin aralıklarını maksimize edecek şekilde ayıran doğrusal bir hiper düzlem (karar sınırı) bulmaktır. SVM'nin ardındaki mantık, aralığı en yükseğe çıkararak karar sınırı bulunursa, gelecekte bilinmeyen öğeleri yanlış sınıflandırma olasılığımızın düşük olmasıdır. Maddeler doğrusal olarak ayrılmazsa, bir değişkeni tanıtarak SVM'yi yumuşak bir aralık sınıflandırıcısına dönüştürmeye karar verebiliriz. Öte yandan, karar sınırı doğrusal değilse, verileri daha yüksek boyutlu bir alana dönüştürmemiz gerekir. Bu durum, Kernel numarası olarak bilinen matematiksel bir dönüşüm sayesinde gerçekleştirilir. SVM'ler son zamanlarda birçok ortamda performansları ve etkinlikleri ile popülerlik kazanmıştır. SVM'ler ayrıca tavsiye sistemlerinde gelecek vaat eden sonuçlar ortaya koymuştur.

## 7. Sınıflandırıcı Grupları

Sınıflandırıcı gruplarının kullanımının arkasındaki temel fikir, eğitim verilerinden bir grup sınıflayıcı oluşturmak ve tahminlerini bir araya getirerek sınıf etiketlerini tahmin etmektir. Sınıflayıcı grupları, sınıflandırıcıların bağımsız olduğunu varsaydığımızda çalışır. Topluluk oluşturmak için çeşitli yaklaşımlar mümkündür. En yaygın iki teknik "Bir Araya Getirme" (Bagging) ve "Yükseltmedir" (Boosting). Bagging'de, her önyükleme örneğinde sınıflayıcıyı oluşturarak, yerine örnekleme yapılır. Boosting'de daha önce yanlış sınıflandırılmış kayıtlara daha fazla odaklanılarak eğitim verilerinin

dağılımını uyarlamalı olarak değiştirmek için tekrarlı bir prosedür kullanılır. Başlangıçta, tüm kayıtlara eşit ağırlık verilir. Ancak, Bagging'den farklı olarak, her Boosting turunun sonunda ağırlıklar değişebilir: Yanlış sınıflandırılmış kayıtlar ağırlıklarını artırırken, doğru sınıflandırılmış kayıtlar ağırlıklarını azaltacaktır. Deneysel sonuçlar, toplulukların izolasyondaki herhangi bir sınıflandırıcıdan daha iyi sonuçlar verebileceğini göstermektedir.

#### 8. Sınıflandırıcıları Değerlendirme

Öneri Sistemleri için en yaygın olarak kabul edilen değerlendirme ölçüsü, “Derecenin Ortalama Hatası” veya “Karekök Ortalama Hatasıdır”. Bu önlemler, Öneri Sistemlerinin amacına yönelik herhangi bir varsayımda bulunmaksızın doğruluğu hesaplar. Bir modeli değerlendirmek için genellikle şu önlemleri göz önünde bulundurulur:

Gerçek Pozitifler (TP): gerçekten A sınıfına ait A sınıfına ait olarak sınıflandırılan örneklerin sayısı,

Gerçek Negatifler (TN): A sınıfına ait olmayan ve aslında A sınıfına ait olmayan sınıflandırmaların sayısı,

Yanlış Pozitifler (FP): A sınıfı olarak sınıflandırılan ancak A sınıfına ait olmayan örneklerin sayısı,

Yanlış Negatifler (FN): v sınıfına ait olarak sınıflandırılmayan ancak aslında A sınıfına ait olan örnekler,

Doğruluk oranı ise;

$$(T P + T N)/(T P + T N + F P + F N)$$

formülü ile elde edilir.

### **3.3.5.3. Sonu Yorumlama**

Bu adımda ortaya ıkan sonular deęerlendirilir. Eęer sonular tutarlı deęilse veri madencilięi srelerinde bir nceki adımların zerinden geilir.

## 4. ÖNERİLEN KARIYER DESTEK SİSTEMİ

Çalışmanın bu bölümünde, kariyer planlama aşamasında olan lisans öğrencilerine yönelik bir kariyer destek sistemi oluşturulacak ve uygulamanın aşamaları tanıtılacaktır.

Uygulamada, 3. Bölümde bahsedilen veri madenciliği yöntemlerinden CRISP-DM yaklaşımı ile ilerlenmiştir. CRISP-DM'in ilk adımı olan problem tanımı, bu çalışmanın başında kurgulanmış olan iki araştırma sorusunun ekseninde, *“mevcut mezun verilerinin makine öğrenmesi yöntemleri ile değerlendirilmesi sonucunda, lisans öğrencilerinin kariyerlerine ilişkin öneri oluşturulabilir mi?”* ve ikincil olarak *“öneri oluşturulabildiği durumlarda bu önerinin doğruluk oranı nedir?”* biçiminde tanımlanmıştır.

Problem belirlendikten sonra veri toplanma ve veriyi anlama aşamasına geçilmiştir. Bu aşamalarda yapılan çalışma ve çıkarımlar müteakip başlıklarda sunulmuştur. Bu başlıklar altında yapılmış olan çalışmalar birinci araştırma sorusunun cevabını vermektedir.

### 4.1. Veri Toplama

Uygulamada üniversite lisans öğrencilerine yönelik bir öneri sistemi kurulması amaçlanmış, vaka çalışması olarak da Hacettepe Üniversitesi Endüstri Mühendisliği Bölümü lisans mezunları seçilmiştir. Bu mezun grubunun seçilmesindeki temel neden, bölüm tarafından halihazırda bir mezun takip sisteminin aktif olarak kullanılıyor olmasıdır. Çalışmanın amacı ve kapsamına uygun olarak belirlenen veriler Hacettepe Üniversitesi Etik Kurulu'ndan alınan izin ile birlikte Hacettepe Üniversitesi Öğrenci İşleri Daire Başkanlığından ve Hacettepe Üniversitesi Endüstri Mühendisliği Bölümü Mezun Takip Sisteminden (MyEMU) alınarak bir araya getirilmiştir. Öğrenci İşleri Daire Başkanlığından alınan verilerde;

- Özlük bilgileri hariç diğer kişisel bilgiler (okul başlangıç yılı, okula giriş tipi, yaşı, mezuniyet yılı, doğum yeri),
- 2019 yılına kadar mezun olmuş tüm öğrencilerin Hazırlık Programı geçme notu,

Lisans öğreniminde almış olduğu tüm derslerin notları bulunmaktadır.

MyEMU'den alınan verilerde ise öğrencinin;

- Mezuniyet ortalaması (2,0 ile 4,0 arasında bir değer),
- Erasmus Programından yararlanıp yararlanmadığı (Var, Yok),
- Çalıştığı şirket, şirketteki birim ve pozisyon, işe başlama ve ayrılma tarihleri (eğer bu işte çalışmaya devam ediyorsa ayrılma tarihi boş gelecektir)
- Staj yaptığı şirket, şirketteki birim ve pozisyon, staja başlama ve bitirme tarihleri
- Her şirketin başlıca faaliyet gösterdiği sektör,
- Yetenekleri ve yeteneklerinin derecelendirilmesi (yabancı dil bilgisi, programlama becerisi, bilgisayar yetenekleri, kişisel yetenekler vb.),
- Girmiş olduğu sınav türleri (YDS, ALES, KPSS, TOEFL, GRE VERBAL, GRE-QUANTITATIVE, GRE-ANALYTICAL, IELTS (Academic, General)) ve aldığı puanı,
- Cinsiyeti (Kadın, Erkek),
- Öğrenci topluluklarında görev alıp almadığı ve almışsa görev bilgisi (Başkan, Yönetim Kurulu Üyesi, Üye ) yer almaktadır.

#### 4.2. Veriyi Anlama

Veri anlama ve ön işleme aşamasında, öncelikli olarak tüm değişkenleri eşsiz bir şekilde ilişkilendirebilecek bir parametre belirlenmiştir. Bahse konu ilişkilendirmenin sağlanması için MyEMU ve Öğrenci İşleri Daire Başkanlığından alınan veriler eşsiz değer olan öğrenci numaralarına göre birleştirilmiştir.

Birden fazla şirkete, staja, Erasmus'a, yeteneğe, öğrenci topluluğuna sahip olan mezunların modeli eğitirken göstereceği etki diğerlerinden daha fazla olacağından, veri kartezyen veri tipinde oluşturulmuştur. Örneğin, bir mezun iki şirkette çalışmış, üç



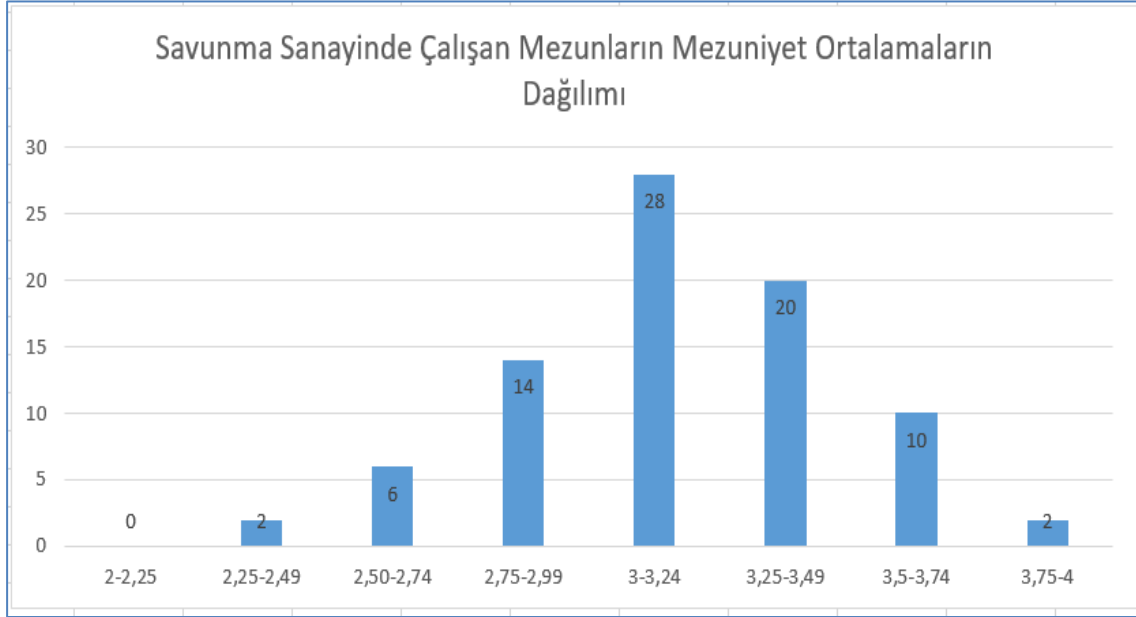
yetenek bilgisi ve iki sınav notu belirtmiş ise, bu öğrenci toplamda 12 kere (2 x 3 x 2) modeli beslemesi sağlanmıştır.

MyEMU'ye veri girişi yapmış olan mezunların toplamda 93 farklı sektörde faaliyet gösteren 968 farklı şirkette çalışmış ve/veya staj yaptığı görülmektedir. Tüm sektörlerin kullanımı modelin başarımlarını etkileyeceğinden, 93 farklı sektör arasından Endüstri Mühendisliği öğrencilerinin çoğunlukla çalıştığı veya çalışmak istediği ilk altı sektör belirlenmiş ve çalışmada kullanılmıştır. Bu doğrultuda mezunların çalıştığı sektör verileri aşağıdaki şu altı sektör başlığı altında toplanmıştır:

- “Savunma Sanayi”,
- “Otomotiv”,
- “Elektronik”,
- “Bilişim”,
- “Hava Yolları”,
- “Danışmanlık”

Bunlar haricinde kalan şirketler de “Diğer” sektör başlığı altında kategorilendirilmiştir. Bu sektörler ve bu sektörde faaliyet gösteren şirketlerde çalışan mezunlar hakkında başlıca bilgiler aşağıda verilmektedir.

**Savunma Sanayi:** Silah ve askeri teknolojileri üreten ve pazarlayan, uzay arařtırmaları yapan řirketlerin birleřtirilmesinden oluřmuřtur. Savunma sanayinde alıřan mezunların not ortalamaları 4 tam not üzerinden incelendiđinde, ortalamaların 3,0 ile 3,24 aralıđında en yüksek frekansa sahip olduđu grlmřtr. Őekil 4.1’de ayrıca, her not aralıđındaki mezun đrenci frekansını belirtmektedir.



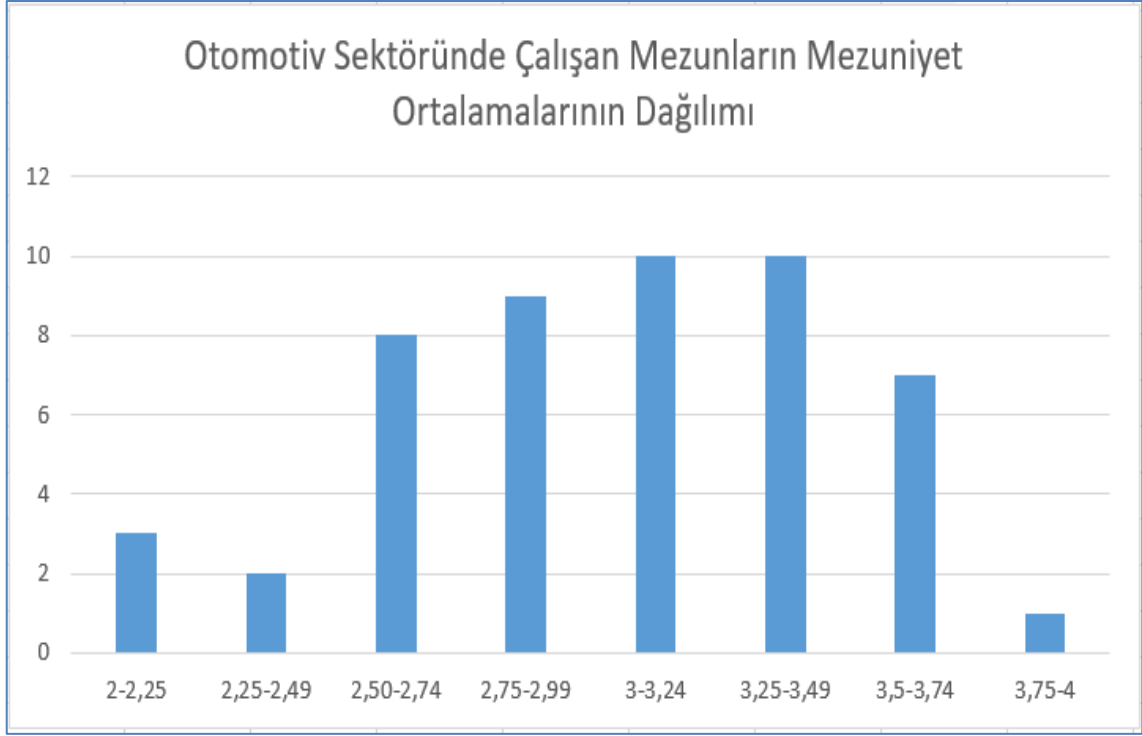
Őekil 4.1 Savunma Sanayinde alıřan Mezunların Not Ortalamalarının Dađılım Grafiđi

Ayrıca bu sektrde alıřan mezunların verilerinin Cinsiyet, Not Ortalaması, Erasmus ve Topluluk Bilgisi gibi analizleri izelge 4.1’ de gsterilmiřtir.

izelge 4.1 Savunma Sanayinde alıřan Mezun Verilerinin Temel zellikleri

Cinsiyet		Not Ortalaması			Erasmus Bilgisi		Topluluk Bilgisi		
Kadın	Erkek	Maks	Min	Ort	Var	Yok	Bařkan	YK	ye
49	34	3,79	2,26	3,15	15	68	4	4	9

**Otomotiv:** Otomobil, otobüs, tır vb. ulaşım araçlarını üreten veya onarım yapan ve yedek parça imalatı servisi sunan firmaların içinde bulunduğu sektördür. Şekil 4.2’de mezunların not ortalamalarının dağılımı ve Çizelge 4.2’de verilerin analizi yapılmıştır.

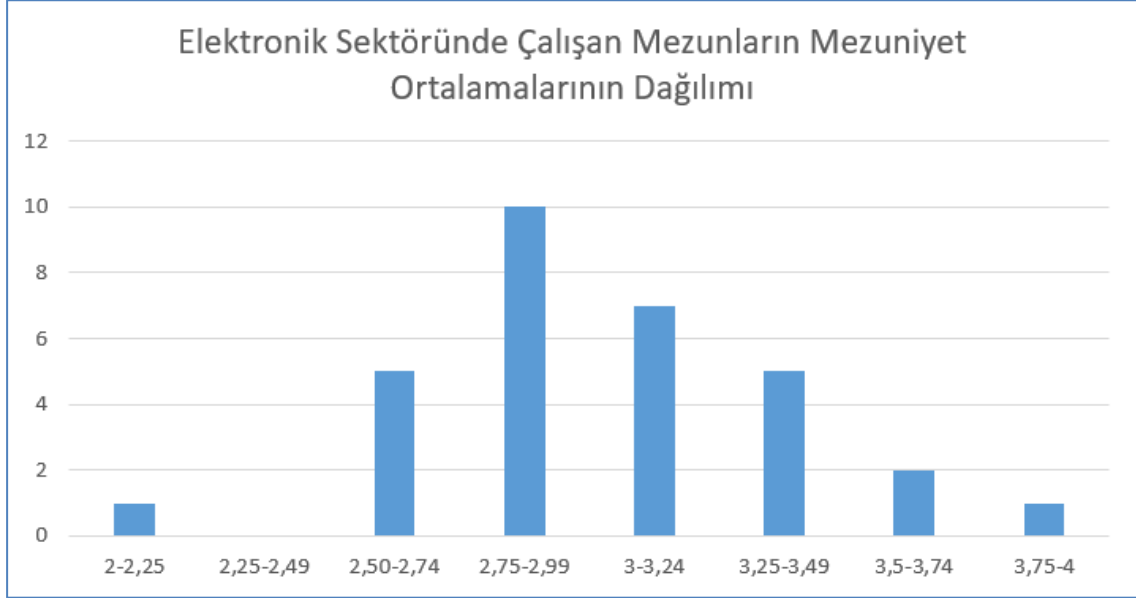


Şekil 4.2 Otomotiv Sektöründe Çalışan Mezunların Not Ortalamalarının Dağılım Grafiği

Çizelge 4.2 Otomotiv Sektöründe Çalışan Mezunların Verilerinin Analizi

Cinsiyet		Not Ortalaması			Erasmus Bilgisi		Topluluk Bilgisi		
Kadın	Erkek	Maks	Min	Ort	Var	Yok	Başkan	YK	Üye
32	18	3,78	2,1	3,04	11	39	0	3	2

**Elektronik:** Tüketici elektronik cihazları ve bileşenleri üreten, elektrik dağıtımını ile ilgili şirketlerin bulunduğu sektördür. Bu sektörde çalışan mezunların not ortalamalarının dağılımına bakıldığında (Şekil 4.3), ağırlığın 2,75 ile 2,99 arasında not ortalamaları bulunan mezunlarda olduğu gözlemlenmektedir.



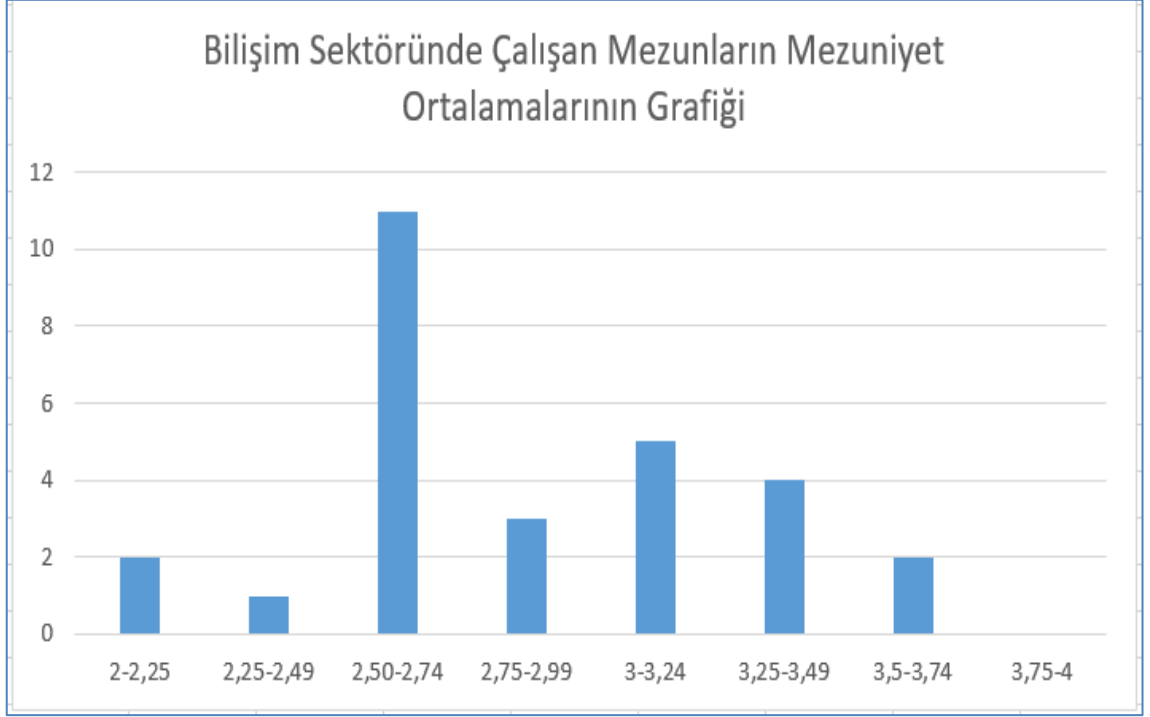
Şekil 4.3 Elektronik Sektöründe Çalışan Mezunların Not Ortalamalarının Dağılım Grafiği

Çizelge 4.3’de elektronik sektörü verilerinin analizi yapılmıştır.

Çizelge 4.3 Elektronik Sektöründe Çalışan/Çalışmış Mezunların Verilerinin Analizi

Cinsiyet		Not Ortalaması			Erasmus Bilgisi		Topluluk Bilgisi		
Kadın	Erkek	Maks	Min	Ort	Var	Yok	Başkan	YK	Üye
22	10	3,82	2,18	3,01	4	28	0	2	1

**Bilişim:** Yazılım, bilgi teknolojileri ve telekomünikasyon işleri yapan şirketlerin bulunduğu sektördür. Not ortalamaları grafiği Şekil 4.4'te verilmiştir.



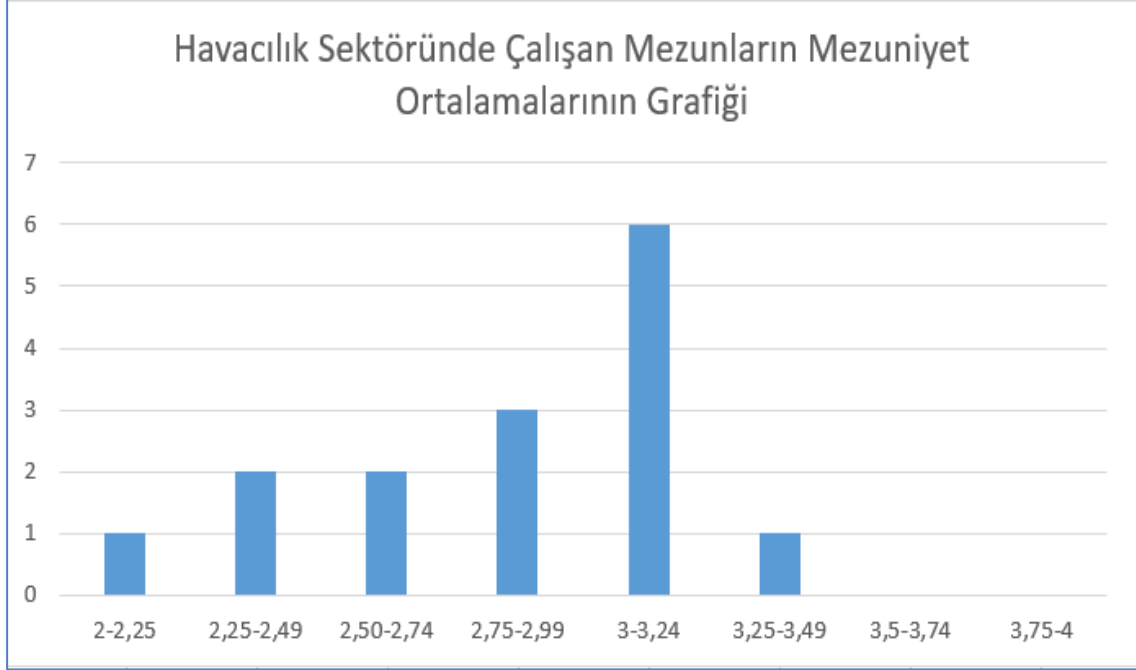
Şekil 4.4 Bilişim Sektöründe Çalışan Mezunların Not Ortalamalarının Dağılım Grafiği

Çizelge 4.4'te bilişim sektöründe çalışan veya çalışmış mezunların verilerinin analizleri yapılmıştır.

Çizelge 4.4 Bilişim Sektöründe Çalışan Mezun Verilerinin Analizi

Cinsiyet		Not Ortalaması			Erasmus Bilgisi		Topluluk Bilgisi		
Kadın	Erkek	Maks	Min	Ort	Var	Yok	Başkan	YK	Üye
16	12	3,73	2,14	2,86	4	24	1	0	1

**Hava Yolları:** Hava yollarında yolcu veya kargo taşımacılığı, hava alanı hizmeti kısaca havacılık işleri sürdüren firmaları içeren sektördür. Bu sektörde çalışan mezunların not ortalamalarının dağılımı Şekil 4.5’te, analizi ise Çizelge 4.5’te verilmiştir.

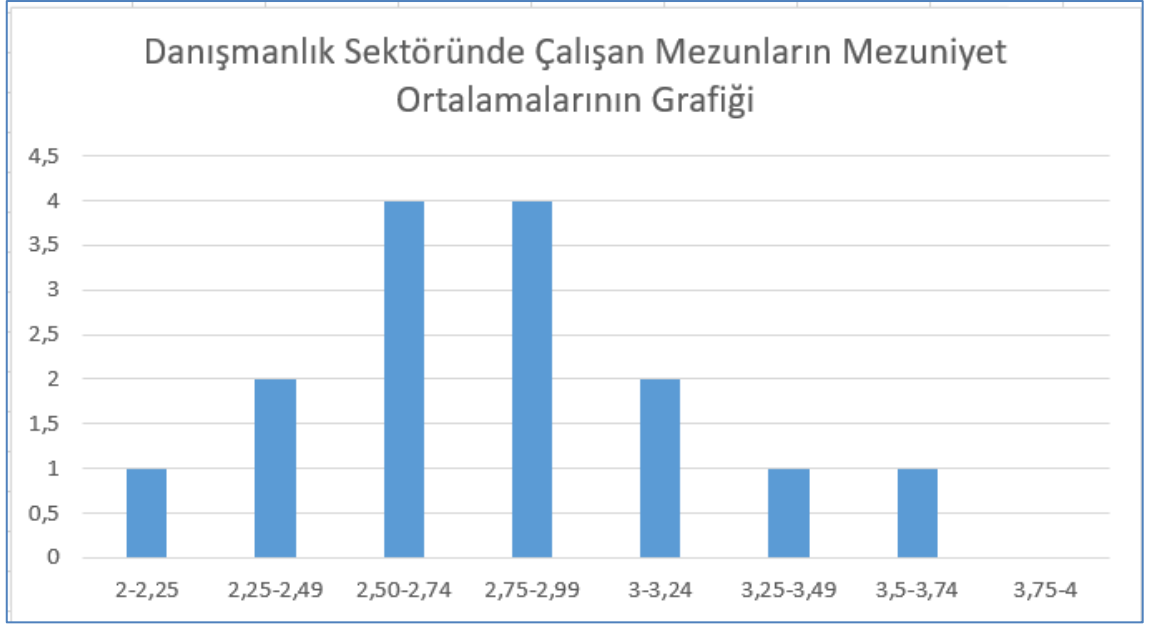


Şekil 4.5 Hava Yolları Sektöründe Çalışan Mezunların Not Ortalamaları Dağılım Grafiği

Çizelge 4.5 Hava Yolları Sektöründe Çalışan Mezunların Verilerinin Analizi

Cinsiyet		Not Ortalaması			Erasmus Bilgisi		Topluluk Bilgisi		
Kadın	Erkek	Maks	Min	Ort	Var	Yok	Başkan	YK	Üye
7	8	3,43	2,22	2,88	3	12	0	0	0

**Danışmanlık:** Genelde finansal denetim, vergi ve danışmanlık işleri yapan firmaların içinde yer aldığı sektördür. Bu sektörde çalışan mezunların not ortalamalarının dağılım grafiği Şekil 4.6’da ve verilerinin analizi Çizelge 4.6’da gösterilmiştir.

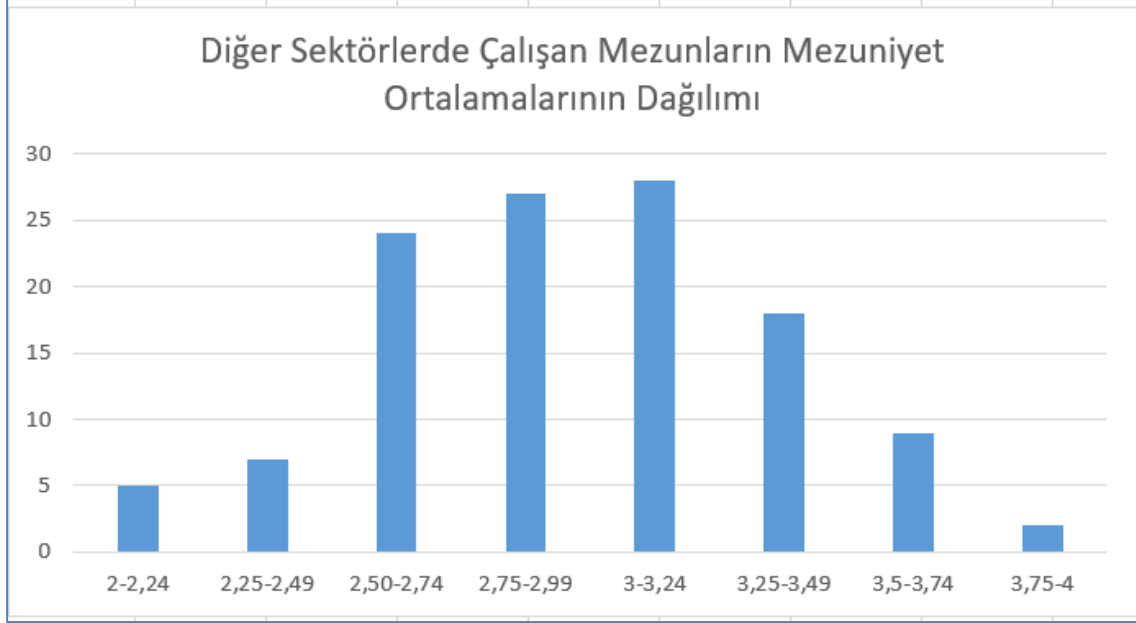


Şekil 4.6 Danışmanlık Sektöründe Çalışan Mezunların Not Ortalamalarının Dağılım Grafiği

Çizelge 4.6 Danışmanlık Sektöründe Çalışan Mezunların Verilerinin Analizi

Cinsiyet		Not Ortalaması			Erasmus Bilgisi		Topluluk Bilgisi		
Kadın	Erkek	Maks	Min	Ort	Var	Yok	Başkan	YK	Üye
7	8	3,53	2,14	2,81	2	13	1	0	1

**Diğer:** Yukarıda bahsedilen sektörlerin dışında istihdam edilen mezunların not ortalamalarının dağılım grafiği Şekil 4.7’de ve verilerinin analizleri Çizelge 4.7’de gösterilmiştir.



Şekil 4.7 Diğer Kategorisinde Değerlendirilen Firmaların Not Ortalamalarının Dağılım Grafiği

Çizelge 4.7 Diğer Kategorisinde Yer Alan Mezunların Verilerinin Analizi

Cinsiyet		Not Ortalaması			Erasmus Bilgisi		Topluluk Bilgisi		
Kadın	Erkek	Maks	Min	Ort	Var	Yok	Başkan	YK	Üye
60	60	3,89	2,11	2,96	15	105	1	2	4



### 4.3. Veri Çerçevesi Hazırlama

Veri çerçevesi hazırlama aşamasında Hacettepe Üniversitesi Öğrenci İşleri Daire Başkanlığı ve MyEMU olmak üzere iki farklı kaynaktan alınan veri setleri incelenirken, 9 mezunun bilgilerinde eksiklikler olduğu görülmüş, bu mezunlar veri setinden çıkartılmış, 295 mezun bilgisi üstünden devam edilmiştir.

Modelin öğrenme sürecini etkileyecek ancak kendi içinde tutarsızlıkları veya süreç içerisinde nitelik değişimleri gösteren ders bilgileri veri setinden çıkarılmıştır. Öğrenme sürecinde çıkartılan ders çeşitleri şunlardır:

- Seçmeli dersler,
- Başka üniversitelerden alınan dersler,
- Kodu değişmiş dersler,
- Seçmeli ders iken zorunlu derse veya zorunlu ders iken seçmeli derse çevrilmiş tüm dersler.

Veri setinde sadece bu tez çalışması tamamlanana kadar kodu ve zorunluluğu aynı kalan ders bilgileri dikkate alınmıştır. Bu sayede veri seti içerisinde ders bilgileri kısmında modelin öğrenmesine değer katmayacak verilerden arındırılması sağlanmıştır.

Mezuniyet not ortalamaları dörtlük sistemden çıkartılıp Hacettepe Üniversitesi yüzlük not skalasına çevrilmiş, veri setine böyle dahil edilmiştir.

Öğrenci notları ham veri setinde A1, A2, A3, B1, B2, B3, C1, C2, C3, D1, D2, D, ve M kodlarıyla yer almış, analiz aşamasında bu notlara numerik karşılıklar tanımlanmıştır. Ancak, 2014 yılında Hacettepe Üniversitesi notlandırma sisteminde yapılan değişiklik sebebiyle numerik düzenleme yapılmadan evvel not sınıflarının yeniden gruplandırılmasına ihtiyaç duyulmuştur. Ham veri setinde yer alan her bir not koduna karşılık bir numerik değer atama yapılması durumunda yapılan bu değişikliğin ağırlıklandırmayı etkileyebileceği öngörüldüğünden tüm notlar kategorik yapıya

çevrilmiştir. Bu işlem sonucunda veri setinde yalnızca A, B, C, D ve M not sınıfları kalmıştır.

Sonrasında bu kategorik notlar Çizelge 4.8’de sunulan şekilde; ordinal değerlerle değiştirilmiş ve numerik hale getirilmiştir. M notunun sayısal değerinin NaN (“not a number”) olmasının sebebi eksik verileri düzeltme aşamasında tüm sütunun ortalamasının NaN değerlere yansımastır.

Çizelge 4.8 Ders Notlarının Veri Dönüşümü

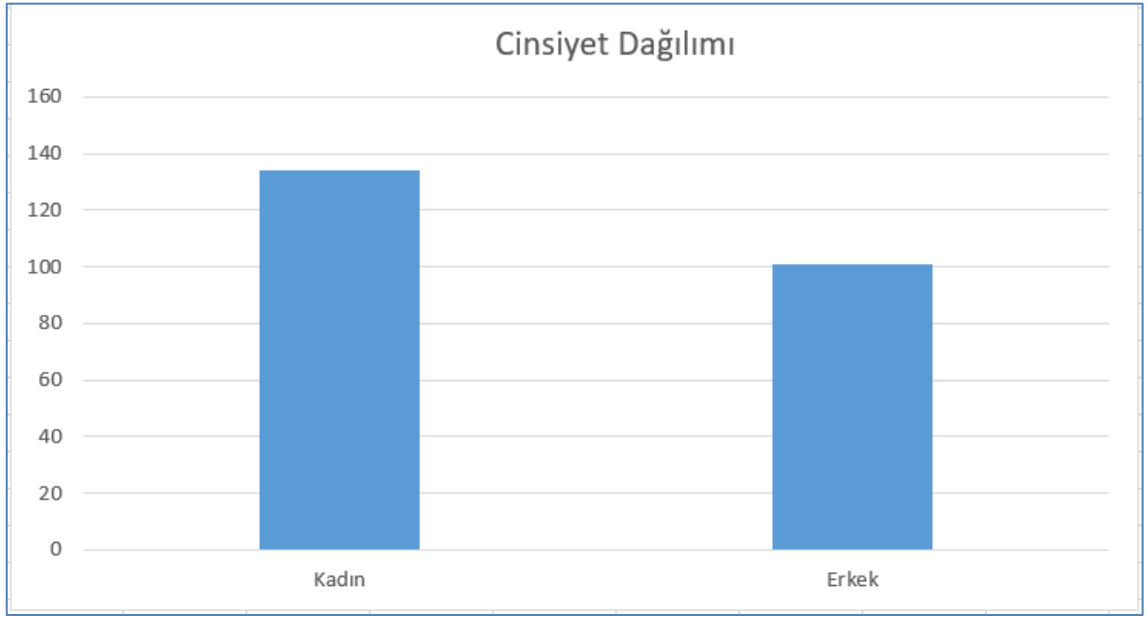
Ham Veri Seti	Kategorik Karşılık	Ordinal Karşılık
A1, A2, A3	A	4
B1, B2, B3	B	3
C1, C2, C3	C	2
D1,D2, D	D	1
M	M	NaN

Yine veri setinde bulunan eksik veriler 3. Bölümde bahsedilen *sklearn* kütüphanesi ile yerlerine ortalamalarını alacak şekilde düzenlenmiştir.

- Öğrencilerin cinsiyet bilgileri Erkek ‘0’ ve Kadın ‘1’ olmak üzere ikili hale getirilmiştir (LabelEncoder).
- Erasmus bilgisi Erasmus programından yararlanan öğrenciler için ‘1’, yararlanmayanlar için ‘0’ olarak değiştirilmiştir.
- Öğrenci topluluk bilgisi ise; eğer herhangi bir toplulukta görev almamışsa ‘0’, bir toplulukta normal üye olarak görev almış ise ‘1’, yönetim kurulu üyesi olarak görev almışsa ‘2’, topluluk yönetim kurulu başkanı olarak görev almış ise ‘3’ olarak kodlanmıştır.
- Öğrencinin girmiş olduğu sınav bilgileri (IELTS, TOEFL vs.) kategorik ve nominal bir veri olduğu için 3. bölümde bahsedilen *sklearn* paketinin *OneHotEncoder* kütüphanesi kullanılarak sayısal veriye çevrilmiştir.

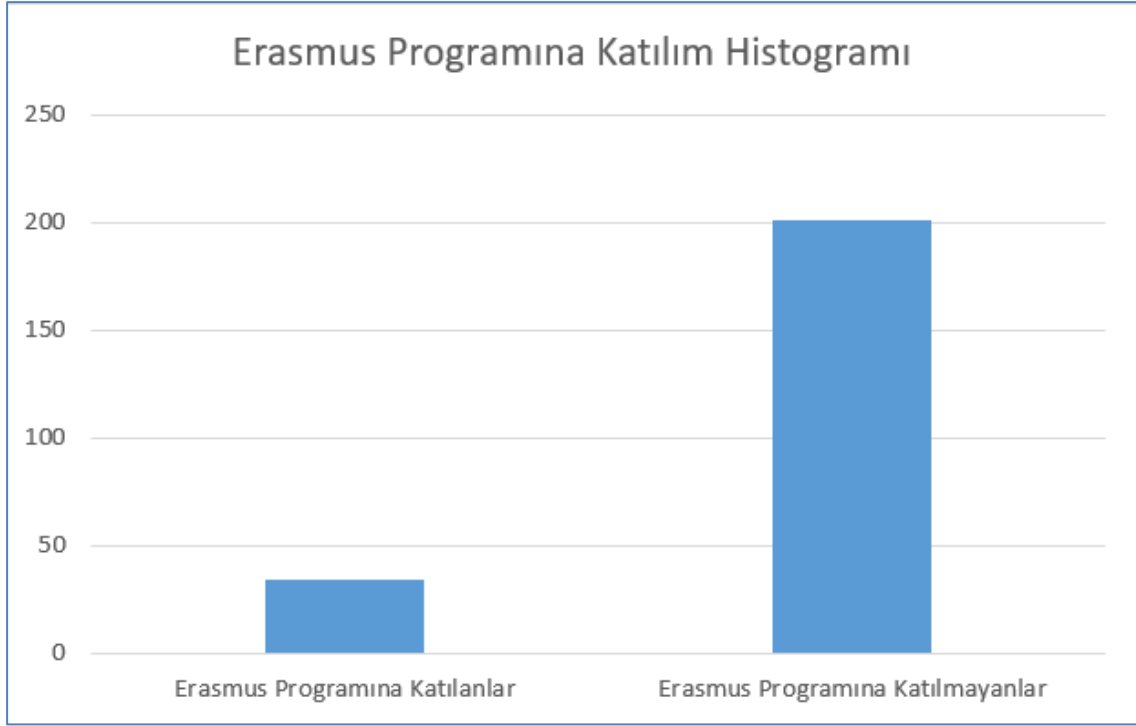
Veri seti içerisinde 298 mezun içerisinde sektör bilgisi bulunmayan mezunların bilgileri veri setinden kaldırılmıştır. Toplam mezun sayısı veri silme işleminden sonra 235'e ve 3857 satır olan veri çerçevesi ise 3301'e düşmüştür.

Bu veri çerçevesine göre verimizdeki 235 öğrenci içerisinde cinsiyet durumu Şekil 4.8'de gösterilmiştir. 134 kadın, 101 erkek mezun sayısı ile veri seti beslenmiştir.



Şekil 4.8 Cinsiyet Dağılımı

Bu veri setinde mezunların Erasmus programına katılım durumu Şekil 4.9'da belirtilmiştir 34 mezun Erasmus programına katılmışken, 201 mezun katılmamıştır.



Şekil 4.9 Erasmus Programına Katılım Durumu Dağılımı

Modelin girdisi olan 37 bağımsız değişken, diğer bir deyişle veri setindeki özellikler,

- Cinsiyet
- Erasmus Programı
- Mezuniyet Ortalaması
- Hazırlık Notu
- TOEFL
- Öğrenci Topluluğu Bilgisi
- YDS
- YDS(FR)
- ALES
- KPSS
- GRE-VERBAL
- GRE-QUANTITATIVE
- GRE-ANALYTICAL
- IELTS (GENERAL)
- IELTS (AKADEMİK)

- ECO135 İktisada Giriş I
- ECO136 İktisada Giriş II
- EMU101 Endüstri Mühendisliğine Giriş
- EMU102 Bilgisayar Destekli Teknik Çizim
- EMU111 Mühendislikte Bilgisayar Uygulamaları
- EMU112 Bilgisayar Programlama
- EMU221 Optimizasyona ve Modellemeye Giriş
- EMU222 Deterministik Yöneylem Araştırması
- EMU231 Mühendisler İçin Olasılık ve İstatistik I
- EMU232 Mühendislik İçin Olasılık ve İstatistik II
- EMU242 İş Analizi ve Tasarımı
- EMU321 Stokastik Yöneylem Araştırması
- EMU322 Simülasyonla Modelleme ve Analiz
- EMU331 İstatistiksel Kalite Kontrol
- EMU341 Üretim Planlama ve Kontrol I
- EMU342 Üretim Planlama ve Kontrol II
- EMU363 Mühendislik Ekonomisi
- EMU464 Proje Yönetimi
- KIM127 Temel Kimya
- MAT123 Matematik I
- MAT124 Matematik II
- MAT245 Doğrusal Cebir

olarak sıralanmaktadır.

Hazırlanan veri setinde bağımlı değişken olarak sadece *Sektör Bilgisi* bulunmaktadır. Sektör bilgisi mezunların çalışmış olduğu veya halen çalışmakta olduğu şirket bilgisinin alınması, ve devamında da o şirketin faaliyet gösterdiği başlıca sektörün elde edilmesinden oluşmaktadır. Bir mezun birden fazla sektörde yer alsa dahi bu bilgi veri setine dahil edilmiştir. Yukarıda bahsedildiği gibi yedi ayrı sektör kategorisi altında toplanmıştır. Sektör bilgisi kategorik ve nominal bir veridir.

#### 4.4. Model Oluřturma

CRISP-DM metodolojisinin veri ön hazırlığı ve ön işlemleri sonrasındaki adımı olan model adımıda öncelikle hazırlanan veri seti Bölüm 3'te bahsedilen *Pandas* kütüphanesi aracılığıyla modele aktarılmıştır. Devamında, sayısal veriler içerisinde boş olan hücreler için *sklearn* kütüphanesi kullanılarak o sütunun ortalaması boş olan hücrelere yazılmıştır. Bağımsız değişkenlerin bir örneği Çizelge 4.9'da verilmiştir.

Çizelge 4.9 Bağımsız Değişkenlerin Veri Setine Benzer Örneği

<i>Cinsiyet</i>	<i>GPA</i>	<i>Erasmus</i>	<i>HAZ</i>	<i>ECO135</i>	<i>EMU101</i>	<i>EMU102</i>	<i>YDS</i>	<i>Toefl</i>	<i>KPSS</i>	<i>IELTS</i>	<i>ALES</i>	<i>Topluluk</i>
1	2,23	0	3	2	4	1	70	0	80	0	85	3
0	3,40	1	4	3	2	4	85	0	0	0	90	1

Ardından bu veriler eğitim (%66) ve test verisi (%33) olarak ikiye bir oranında *sklearn* altındaki *train\_test\_split* kütüphanesi kullanılarak bölünmüştür. Bağımsız değişkenler *StandardScaler* aracılığı ile standardize edilmiştir.

Hangi özelliğin model için anlamlı olup olmadığı, her bir özelliğin çözüme olan katkısı, hangi özelliklerin öğrenmeye negatif yönde etkili olduğunu bulmak için özellik seçimi yöntemlerinden ileriye doğru eleme yöntemi modele gönderilen veri seti üstüne işlenmiştir. Gözetimli makine öğrenmesinde sınıflandırma yönteminde ileriye doğru özellik eleme yönteminde en yaygın kullanılan yöntem Ki-Kare yöntemi seçilmiştir.

Ki-kare yöntemi bağımlı değişken ile bağımsız değişkenler arasındaki ilişkiyi inceler. Ki-kare yöntemini *sklearn* paketinin *featureSelection* kütüphanesi aracılığıyla veri setine uygulanmıştır. Bu kütüphane istenen sayıdaki bağımlı değişkeni en çok etkileyen veri setini döndürür.

Modele gönderilen 39 özelliğin arasında hangi özelliklerin modeli daha fazla etkilediğini bulmak için sırasıyla ilk özellikten son özelliğe kadar bu kütüphane çalıştırılmıştır. Bu kütüphane aşağıdaki 26 özelliğin birlikte kullanıldığı takdirde modelin en yüksek sonucu vereceğini göstermiştir.

- Mezuniyet Not Ortalaması
- TOEFL
- Öğrenci Topluluğu Bilgisi
- YDS
- YDS(FR)
- ALES
- KPSS
- GRE-QUANTITATIVE
- IELTS (GENERAL)
- IELTS (AKADEMİK)
- ECO135 İktisada Giriş I
- ECO136 İktisada Giriş II
- EMU111 Mühendislikte Bilgisayar Uygulamaları
- EMU112 Bilgisayar Programlama
- EMU221 Optimizasyona ve Modellemeye Giriş
- EMU222 Deterministik Yöneylem Araştırması
- EMU231 Mühendisler İçin Olasılık ve İstatistik I
- EMU232 Mühendislik İçin Olasılık ve İstatistik II
- EMU242 İş Analizi ve Tasarımı
- EMU321 Stokastik Yöneylem Araştırması
- EMU322 Simülasyonla Modelleme ve Analiz
- EMU341 Üretim Planlama ve Kontrol I
- EMU363 Mühendislik Ekonomisi
- MAT123 Matematik I
- MAT124 Matematik II
- MAT245 Doğrusal Cebir

Bu 26 özellik dışında kalan Erasmus Programı bilgisi, Cinsiyet bilgisi, Hazırlık Not bilgisi, GRE-VERBAL ve GRE-ANALYTICAL bilgileri, EMU101 Endüstri Mühendisliğine Giriş, EMU102 Bilgisayar Destekli Teknik Çizim, EMU331 İstatistiksel Kalite Kontrol, EMU342 Üretim Planlama ve Kontrol II, EMU464 Proje Yönetimi, KIM127 Temel Kimya bilgileri veri setinden çıkartılarak uygulama çalıştırılmıştır.

*Sklearn* ile uygulanan özellik seçim sürecinden alınan sonuçlardan, bölüm müfredatında öğrencilerin alması zorunlu olan 16 dersin 11'inin (%68.75) modelde etkili olduğu ve bağımlı değişkeni (iş sektörü) etkilediği görülmüştür. Bu derslerin bilgisi bölüm başkanlığı ile paylaşılmıştır.

Veriler sınıflandırma modellerinde çalışacak hale hazırlandıktan sonra yukarıda Ki-kare yönteminden sonra elde edilen 26 özellik ile birlikte Bölüm 3'de bahsedilen; Lojistik regresyon, K-NN, SVM, Naive Bayes, Karar Ağacı, Rassal Orman sınıflandırma modellerinde çalıştırılmıştır.

Veri setini test ve eğitim verisi diye bölerken bu bölme işlemi rastgele yapılmadan önce (*sklearn.crossvalidation* kütüphanesi ile kullanılırken *random\_state* parametresinin var olması) model bir defa çalıştırılmış çıkan doğruluk oranının bu 26 özellik içerisinde hangisi olduğu bulunmuştur. Çıkan sonuçlara göre en yüksek oranı %70.33 ile En Yakın Komşuluk vermiştir. (Doğruluk oranı *scikitlearn* kütüphanesinin *score* fonksiyonu tarafından elde edilmiştir.) Diğer sonuçlar ise Lojistik Regresyon %62.8, Destek Vektör Makinesi %65.74, Gaussian Naive Bayes %38.47, Karar Ağacı %68.68 ve Rassal Orman ise %68.04 doğruluk oranı ile çalışmıştır.

Makine öğrenmesi yöntemlerinin bu model sonuçları için karşılaştırılması Bölüm 5'de yapılmıştır.



#### 4.5. Değerlendirme

Değerlendirme adımında, bu çalışmanın ikinci araştırma sorusunun “Tavsiye sisteminde kullanılmış olan alternatif algoritmaların doğruluk oranlarının karşılaştırılması nedir? Hangi algoritma ne kadar başarılı olmuştur?” cevabı aranmıştır. Bu araştırma sorusunun cevabı için, genellikle makine öğrenmesi sınıflandırma modellerinde kullanılan Çizelge 4.11’de tablosu verilen karmaşıklık matrisinden yola çıkan üç farklı başarı kriteri kullanılmıştır. Bunlar;

1.Precision (Kesinlik): Sonucun ne kadar doğru olduğunu gösterir.

$$\text{Precision} = \text{True Positives} / (\text{True Positives} + \text{False Positives})$$

2.Recall (Duyarlılık): Doğru örnekleri bulma yeteneğini gösterir.

$$\text{Recall} = \text{True Positives} / (\text{True Positives} + \text{False Negatives})$$

3.F1 Score: Precision ve Recall’un harmonik ortalamasıdır.

$$\text{F1\_Score} = (2 * \text{Precision} * \text{Recall}) / \text{Precision} + \text{Recall}$$

Çizelge 4.11 Karmaşıklık Matrisi Tablosu

	<b>Olumlu Tahmin Edilen</b>	<b>Olumsuz Tahmin Edilen</b>
<b>Olumlu Gerçek Değer</b>	TP(True Positive)	TN(True Negative)
<b>Olumsuz Gerçek Değer</b>	FP(False Positive)	FN(False Negative)

Yukarıdaki bilgiler doğrultusunda değerlendirme kriterleri sklearn kütüphanesi aracılığıyla hesaplanmış, 5. Bölüm Sonuç ve Tartışma kısmında verilmiştir.

#### 4.6. Konuşlandırma

Bu çalışma sonucunda ortaya çıkarılan model, her yıl mezun olacak öğrencilere tavsiyelerde bulunabilecek şekilde genişletilebilir ve uygulama olarak kullanılabilir. Yeni mezun öğrenci, modelde etkili olduğu belirlenen özellikleri uygulamaya girdi olarak

verdikten sonra, uygulamanın kendisine bir sektör tavsiye etmesini isteyebilir. Bu şekilde, yeni mezun iş başvurularında şansının en yüksek olduğu tavsiye sistemi tarafından belirtilmiş olan sektörden iş arayışlarına başlayabilir. Ayrıca modelin her altı ayda bir güncel çalışma verileri ile tekrarlanması ve bu modellemenin bir plan dahilinde sürekli olarak yapılması bu modellemenin değerli ve kullanılabilir bir bilgiye dönüşmesine yardımcı olacaktır.

## 5. SONUÇ VE TARTIŞMA

Bu çalışmada Hacettepe Üniversitesi Endüstri Mühendisliği Bölümü'nün ilk mezunlarından itibaren 2018 yılına kadar olan verileri kullanılmış olup, yeni mezun olacak olan öğrencilere iş bulma konusunda yardımcı olabilecek sektör önerisi yapmak amaçlanmıştır. Hacettepe Üniversitesi ÖİDB ve Hacettepe Üniversitesi Endüstri Mühendisliği Bölümünden alınan bu veriler temizlenmiş, dönüştürülmüş, birleştirilmiş ve modele girecek şekilde anlam kazandırılmıştır.

Sektör tavsiyesi için oluşturulan modelde farklı sınıflandırma algoritmaları kullanılmış ve hangisinin doğruluk oranının yüksek olduğu karşılaştırılmıştır. Bu karşılaştırma, çalışmanın başında konmuş olan ikinci araştırma sorusunun da cevabını teşkil etmektedir. Bölüm 4'te bahsedilen karşılaştırılmada random\_state durumu var olduğu için modelin orada bir kere çalıştırılması yeterli olmuştur. Fakat random\_state durumu yok olduğu zaman kurulan modeldeki algoritmaların her çalıştırıldığında test verisi değişiklik gösterdiği için farklı doğruluk oranları ortaya koyduğu görülmüştür. Bu doğruluk oranlarından ortalama bir değer elde etmek ve doğruluk oranlarının nasıl ilerlediğini görmek için model 100, 200, 500 ve 1000 kez çalıştırılmış, elde edilen sonuçların ortalamaları Çizelge 5.1'de gösterilmiştir.

Çizelge 5.1 Doğruluk Oranlarının Farklı Tekrarlarda Karşılaştırılması

Algoritmalar	Tekrar Sayıları			
	100	200	500	1000
Lojistik Regresyon	60.36	60.38	60.44	60.42
Rassal Orman	67.39	67.52	67.52	67.46
SVM	65.07	65.04	64.92	64.87
Karar Ağacı	67.19	67.27	67.20	67.21
Gaussian Naive Bayes	34.91	35.06	34.94	34.99
En Yakın Komşuluk	66.61	65.94	66.02	66.35

Çizelge 5.1’de görüldüğü üzere model `random_state=0` olduğu durumda bir kere çalıştırıldığında en yakın komşuluk algoritması en yüksek değeri alırken, `random_state` durumunun kontrol edilmediği durumda en yüksek doğruluk oranı ortalamasını rassal orman göstermiştir. Bunun nedeni en yakın komşuluk algoritmasının bir kez çalıştırılan durumda ortaya çıkan eğitim verisinde en iyi sonucu vermesidir. Diğer durumda ise tekrar sayısı arttıkça %67.46 doğruluk oranına yakınsamıştır. En iyi değeri veren algoritmanın rassal orman çıkmasının sebebi ise Bölüm 4’te bahsedilen Ki-kare yöntemi ile özellik seçimine gitmemiz olabilir. Çünkü rassal orman algoritmaları genelde özellik seçiminde en önemli özellikleri tanımlamak içinde kullanılır.

Doğruluk kriterleri 1000 defa çalıştırıldıktan sonra alınan sonuçların ortalaması Çizelge 5.2’de, 500 defa çalıştırıldıktan sonra alınan sonuçların ortalaması Çizelge 5.3’te, 200 defa çalıştırıldıktan sonra alınan sonuçların ortalaması Çizelge 5.4’te ve 100 defa çalıştırıldıktan sonra alınan sonuçların ortalaması Çizelge 5.5’te verilmiştir.

Çizelge 5.2 Doğruluk Kriterlerinin Karşılaştırılması (1000 Çalıştırma)

<b>Algoritmalar</b>	<b>Precision</b>	<b>Recall</b>	<b>F1 Score</b>
Lojistik Regresyon	0.609	0.500	0.525
En Yakın Komşuluk	0.686	0.655	0.646
SVM	0.770	0.540	0.592
Gaussian Naive Bayes	0.435	0.458	0.289
Karar ağacı	0.664	0.641	0.632
Rassal Orman	0.684	0.632	0.636

Çizelge 5.3 Doğruluk Kriterlerinin Karşılaştırılması (500 Çalıştırma)

<b>Algoritmalar</b>	<b>Precision</b>	<b>Recall</b>	<b>F1 Score</b>
Lojistik Regresyon	0.610	0.500	0.524
En Yakın Komşuluk	0.686	0.658	0.646
SVM	0.772	0.542	0.596
Gaussian Naive Bayes	0.434	0.438	0.288
Karar ağacı	0.664	0.640	0.632
Rassal Orman	0.686	0.632	0.638

Çizelge 5.4 Doğruluk Kriterlerinin Karşılaştırılması (200 çalıştırma)

<b>Algoritmalar</b>	<b>Precision</b>	<b>Recall</b>	<b>F1 Score</b>
Lojistik Regresyon	0.605	0.500	0.520
En Yakın Komşuluk	0.685	0.660	0.645
Destek Vektör Makinesi	0.770	0.540	0.595
Gaussian Naive Bayes	0.435	0.455	0.285
Karar ağacı	0.660	0.645	0.630
Rassal Orman	0.685	0.635	0.635

Çizelge 5.5 Doğruluk Kriterlerinin Karşılaştırılması (100 Çalıştırma)

<b>Algoritmalar</b>	<b>Precision</b>	<b>Recall</b>	<b>F1 Score</b>
Lojistik Regresyon	0.604	0.500	0.524
En Yakın Komşuluk	0.683	0.657	0.646
Destek Vektör Makinesi	0.769	0.543	0.594
Gaussian Naive Bayes	0.437	0.455	0.289
Karar ağacı	0.661	0.644	0.632
Rassal Orman	0.680	0.634	0.637

Doğruluk kriterleri incelendiği zaman Precision değeri genelde en yüksek karar vektör makinesi algoritmasında çıkarken, Recall ve F1 Score kriterlerinde en yüksek değeri en yakın komşuluk algoritması vermiştir.

Tavsiye sistemleri kullanıcılara farklı seçenekler arasından kullanıcıların bilgilerine göre en uygun olanlarını tavsiye ederler. Bu çalışmada mezunların Bölüm 4'te bahsedilen 26 özelliğine bakılarak, mezun olacak öğrencilerin de bu bilgilerini modele dahil ederek %67.5 doğruluk oranıyla sektör tavsiye edilmiştir. Bu çalışma 235 mezunun bilgisi üzerinden yürütülmüştür. Bu mezun bilgileri arttıkça tavsiye sisteminin doğruluk oranı artabilir. Dolayısıyla, çalışmanın daha fazla katılımcı sayısı ile tekrarlanması, kariyer destek sisteminde kullanılacak algoritmaların performansının daha güvenilir bir şekilde değerlendirilmesini sağlayacaktır.

4.6 Konuşlandırma bölümünde belirtilmiş olan ve geliştirilmiş tavsiye sistem modelinin uygulamaya dönüştürülmesine ek olarak birkaç gelecek çalışma planlanmıştır. İlk olarak, oluşturulmuş olan bu tavsiye sisteminin, mevcut çalışmada kullanılmış olan özelliklere ek yeni özellikler ile çalıştırılması öngörülmektedir. Bunlar arasında başlıca olarak lisansüstü eğitim bilgileri, işe girme şekli ve yandal / çift anadal bilgileri düşünülmektedir. Ayrıca, bu çalışmada kullanılmamış olan alan içi (teknik) seçmeli derslerin ve alan dışı (teknik olmayan) seçmeli derslerin de ileriki çalışmalarda modelin bağımsız değişkenleri olarak kullanılması değerli bulgular sunabilir. Modele girdi olarak

kullanılacak kiři sayısı arttıkça, daha fazla sektör kategorisi veya sektör yerine řirket kategorileri ile çalıřtırılması dūřünülmektedir. Bu planlanan çalıřmalar haricinde, geliřtirilen karar destek sisteminin farklı programlarda ve üniversitelerde öğrenim gören öğrenci verileri kullanılarak çalıřtırılmasının, devamında da elde edilen sonuçların birbiri ile karşılaştırılmasının deęerli olacağı dūřünülmektedir.

## 6. KAYNAKLAR

- [1] A. Symons, A. H. Bryan, E. Sammartino, E. Wright, Book Reviews: Film Stud 19, 1 (2019) 100–106.
- [2] W. S. McCulloch, P. Walter, A logical calculus of the ideas immanent in nervous activity: The bulletin of mathematical biophysics 5.4 (1943) 115-133.
- [3] J. McCarthy, M.L. Minsky, N. Rochester, C.E. Shannon, A proposal for the dartmouth summer research project on artificial intelligence, 31 Ağustos 1955, AI magazine, Dartmouth, 27.4 (2006) 12-12.
- [4] S.J. Russell, P. Norvig. Artificial intelligence: a modern approach. Malaysia; Pearson Education Limited, (2016).
- [5] A. Newell, J. C. Shaw, H. A. Simon. Report on a general problem solving program. IFIP congress 256 (1959).
- [6] J. McCarthy, B. Liskov, P. Abrahams Lisp Session: Hist. Program. Lang., (1981) 173–197.
- [7] E.A. Feigenbaum, B. G. Buchanan, J. Lederberg, On generality and problem solving: A case study using the DENDRAL program (1970).
- [8] L. M. Fagan, E. H. Shortliffe, B. G. Buchanan, Decision Making : From 106, 1980 (1982) 97–106.
- [9] J. McDermott, R1: A rule-based configurer of computer systems: Artificial intelligence 19.1 (1982) 39-88.
- [10] D. Yarowsky, Word-sense disambiguation using statistical models of Roget’s categories trained on large corpora (1992) 454.
- [11] J. Hays, A. A. Efros, Scene completion using millions of photographs Commun: ACM, 51, 10 (2008) 87–94.
- [12] A. L Samuel, Programming computers to play games, Advances in Computers, 1 (1960) 165-192.
- [13] S. Das, A. Dey, A. Pal, N. Roy Applications of Artificial Intelligence in Machine Learning: Review and Prospect: Int. J. Comput. Appl. 115, 9 (2015) 31–41.
- [14] Z. Chen, C. Li, W. Sun, Bitcoin price prediction using machine learning: An approach to sample dimension engineering: J. Comput. Appl. Math. 365, (2020) 112-395.
- [15] J. Y. Chung, S. Lee, Dropout early warning systems for high school students using machine learning: Child. Youth Serv. Rev. 96, November 2018, (2019) 346–353.



- [16] N. Tomasevic, N. Gvozdenovic, S. Vranes, An overview and comparison of supervised data mining techniques for student exam performance prediction: *Comput. Educ.* (2019).
- [17] C. F. Chien, L. F. Chen, Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry: *Expert Syst. Appl.* 34, 1 (2008) 280–290.
- [18] C. Vialardi, J. Bravo, L. Shafti, A. Ortigosa, Recommendation in Higher Education Using Data Mining Techniques, *Educational Data Mining*, (2009) 190 – 199.
- [19] I. Paparrizos, B. B. Cambazoglu, A. Gionis, Machine learned job recommendation, (2011) 325.
- [20] T. R. Razak, M. A. Hashim, N. M. Noor, I. H. A. Halim, N. F. F. Shamsul, Career path recommendation system for UiTM Perlis students using fuzzy logic, 2014 5th Int. Conf. Intell. Adv. Syst. Technol. Converg. Sustain. Futur. ICIAS 2014 - Proc.(2014).
- [21] J. Zhang, Q. Peng, S. Sun, C. Liu, Collaborative filtering recommendation algorithm based on user preference derived from item domain features: *Phys. A Stat. Mech. its Appl.* 396, (2014) 66–76.
- [22] N. D. Almalis, G. A. Tsihrintzis, N. Karagiannis, A. D. Strati, FoDRA - A new content-based job recommendation algorithm for job seeking and recruiting, IISA 2015 - 6th Int. Conf. Information, Intell. Syst. Appl., (2016).
- [23] A. R. Deshpande, M. Emmanuel, Recommendation System Overview, *Int. J. Innov. Adv. Comput. Sci. IJIACS* ISSN, 6, 12, (2017) 2347–8616.
- [24] B. Patel, V. Kakuste, M. Eirinaki, CaPaR: A career path recommendation framework, Proc. - 3rd IEEE Int. Conf. Big Data Comput. Serv. Appl. Big Data Service (2017) 23–30.
- [25] E. Alpaydin, *Introduction to machine learning*. MIT press, (2009).
- [26] S. B. Kotsiantis, I. Zaharakis, P. Pintelas, Supervised machine learning: A review of classification techniques, *Emerging artificial intelligence applications in computer engineering*, 160, (2007) 3-24.
- [27] scikitlearn-knn, <https://scikit-learn.org/stable/modules/neighbors.html> (Erişim tarihi: 4 Ekim 2019).
- [28] C. Cortes, V. Vapnik, Support vector machine: *Machine learning*, 20.3 (1995) 273-297.
- [29] A. Azevedo, M. F. Santos, KDD , Semma And Crisp-Dm : A Parallel Overview,

- IADIS Eur. Conf. Data Min., **(2008)** 182–185.
- [30] R. Wirth, CRISP-DM : Towards a Standard Process Model for Data Mining, Proc. Fourth Int. Conf. Pract. Appl. Knowl. Discov. Data Min., 24959, **(2000)** 29–39.
- [31] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, Scikit-Learn: Machine Learning In Python, Matthieu Perrot: J. Mach. Learn. Res.,12, **(2011)** 2825–2830.
- [32] A. Al-Badarenah, J. Alsakran, An Automated Recommender System for Course Selection, Int. J. Adv. Comput. Sci. Appl., 7, 3, **(2016)** 166–175.
- [33] X. Amatriain, A. Jaimes, N. Oliver, J. M. Pujol, Recommender Systems Handbook, **(2015)**.



T.C.  
HACETTEPE ÜNİVERSİTESİ  
Rektörlük

Sayı : 35853172-300  
Konu : Muhammet AKGÜN Hk. (Etik Komisyon)

FEN BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

İlgi : 29.04.2019 tarihli ve 23154132-300/00000569038 sayılı yazınız.

Enstitünüz Endüstri Mühendisliği Anabilim Dalı Tezli Yüksek Lisans öğrencisi **Muhammet AKGÜN**'ün Doç. Dr. **Özlem Müge AYDIN TESTİK** danışmanlığında yürüttüğü "**Kariyer Planlama için Bulanık Karar Destek Sistemi**" başlıklı tez çalışması Üniversitemiz Senatosu Etik Komisyonunun **14 Mayıs 2019** tarihinde yapmış olduğu toplantıda incelenmiş olup, etik açıdan uygun bulunmuştur.

Bilgilerinizi ve gereğini saygılarımla rica ederim.

e-imzalıdır  
Prof. Dr. Rahime Meral NOHUTCU  
Rektör Yardımcısı

*lr.*

Evrakın elektronik imzalı suretine <https://belgedogrulama.hacettepe.edu.tr> adresinden 42a3290c-be2f-43f7-9f82-7d8d89b855e6 kodu ile erişebilirsiniz.  
Bu belge 5070 sayılı Elektronik İmza Kanunu'na uygun olarak Güvenli Elektronik İmza ile imzalanmıştır.

Hacettepe Üniversitesi Rektörlük 06100 Sıhhiye-Ankara  
Telefon:0 (312) 305 3001-3002 Faks:0 (312) 311 9992 E-posta:yazimd@hacettepe.edu.tr İnternet  
Adresi: www.hacettepe.edu.tr

Duygu Didem İLFRİ



*Dracife İmza*



HACETTEPE ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ  
YÜKSEK LİSANS/~~DOKTORA~~ TEZ ÇALIŞMASI ORJİNALLİK RAPORU

HACETTEPE ÜNİVERSİTESİ  
FEN BİLİMLER ENSTİTÜSÜ  
ENDÜSTRİ MÜHENDİSLİĞİ ANABİLİM DALI BAŞKANLIĞI'NA

Tarih: 08/10/2019

Tez Başlığı: Kariyer Planlama İçin Karar Destek Sistemi

Yukarıda başlığı gösterilen tez çalışmamın a) Kapak sayfası, b) Giriş, c) Ana bölümler d) Sonuç kısımlarından oluşan toplam 53 sayfalık kısmına ilişkin, 08/10/2019 tarihinde tez danışmanım tarafından *Turnitin* adlı intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan orijinallik raporuna göre, tezimin benzerlik oranı % 3 'tür.

Uygulanan filtrelemeler:

- 1- Kaynakça hariç
- 2- Alıntılar hariç/dâhil
- 3- 5 kelimededen daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Fen Bilimleri Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esasları'nı inceledim ve bu Uygulama Esasları'nda belirtilen azami benzerlik oranlarına göre tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

Tarih ve İmza

Adı Soyadı: Muhammet AKGÜN

Öğrenci No: N16123506

Anabilim Dalı: Endüstri Mühendisliği

Programı: Endüstri Mühendisliği Tezli Yüksek Lisans

Statüsü:  Y.Lisans  Doktora  Bütünleşik Dr.

08-10-2019

*Uğur*

**DANIŞMAN ONAYI**

UYGUNDUR.

*Özlem Müge*

Prof. Dr. Özlem Müge TESTİK



HACETTEPE UNIVERSITY  
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING  
THESIS/DISSERTATION ORIGINALITY REPORT

HACETTEPE UNIVERSITY  
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING  
TO THE DEPARTMENT OF INDUSTRIAL ENGINEERING

Date: 08/10/2019

Thesis Title: A Decision Support System for Career Recommendation

According to the originality report obtained by my thesis advisor by using the *Turnitin* plagiarism detection software and by applying the filtering options stated below on 08/10/2019 for the total of 53 pages including the a) Title Page, b) Introduction, c) Main Chapters, d) Conclusion sections of my thesis entitled as above, the similarity index of my thesis is 3 %.

Filtering options applied:

1. Bibliography/Works Cited excluded
2. Quotes excluded / included
3. Match size up to 5 words excluded

I declare that I have carefully read Hacettepe University Graduate School of Science and Engineering Guidelines for Obtaining and Using Thesis Originality Reports; that according to the maximum similarity index values specified in the Guidelines, my thesis does not include any form of plagiarism; that in any future detection of possible infringement of the regulations I accept all legal responsibility; and that all the information I have provided is correct to the best of my knowledge.

I respectfully submit this for approval.

Date and Signature

Name Surname: Muhammet AKGÜN  
Student No: N16123506  
Department: Industrial Engineering  
Program: Master of Science in Industrial Engineering Programme  
(with Thesis)  
Status:  Masters  Ph.D.  Integrated Ph.D.

08.10.2019

**ADVISOR APPROVAL**

APPROVED.

Prof. Dr. Özlem Müge TESTİK

## ÖZGEÇMİŞ

Adı Soyadı : Muhammet Akgün  
Doğum yeri : Ankara  
Doğum tarihi : 20.09.1991  
Medeni hali : Bekar  
Yazışma adresi : Turgut Özal Mah. Beyazgül Sit. 8B Blok No:7 Batıkent  
Yenimahalle ANKARA  
Telefon : +905058328956  
Elektronik posta adresi : ce.muhammetakgun@gmail.com  
Yabancı dili : İngilizce

### EĞİTİM DURUMU

Lisans : Hacettepe Üniversitesi Bilgisayar Mühendisliği  
Yüksek Lisans :  
Doktora :

### İş Tecrübesi

2014- 2016 Uz Consulting International – Yazılım Mühendisi  
2016-2018 Bilişim Bilişim A.Ş. – Yazılım Mühendisi  
2018- ... Icterra – Kıdemli Yazılım Mühendisi



