
GAZİ ÜNİVERSİTESİ
Gazi
1926

YURTTA BARINAN ÖĞRENCİLERİN AYRILMA SEBEPLERİNİN VERİ MADENCİLİĞİ YÖNTEMLERİYLE ANALİZ EDİLMESİ

Berrin DEĞİRMENCİ

YÜKSEK LİSANS TEZİ
BİLGİSAYAR MÜHENDİSLİĞİ ANA BİLİM DALI

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

OCAK 2024

ETİK BEYAN

Gazi Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmasında;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmasında yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Berrin DEĞİRMENCİ

10/01/2024

YURTTA BARINAN ÖĞRENCİLERİN AYRILMA SEBEPLERİNİN VERİ MADENCİLİĞİ YÖNTEMLERİYLE ANALİZ EDİLMESİ

(Yüksek Lisans Tezi)

Berrin DEĞİRMENCİ

GAZİ ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

Ocak 2024

ÖZET

Barınma, tüm dünyada üniversite öğrencileri için önemli bir sorun oluşturmaktadır. Üniversite öğrencilerinin büyük bir bölümü öğrenim hayatlarının bir döneminde devlet yurtlarında kalmaktadır. Ülkemizde 2023 yılı verilerine göre devlet yurtları 900 bin yatak kapasitesine ulaşmış olmasına rağmen, halen barınma sorunu yaşayan öğrenciler bulunmaktadır. Bu noktada yurtların iyi organize edilmesi, mevcut kapasitenin en uygun şekilde kullanılması ve ihtiyaç sahibi öğrencilere ortam oluşturulması büyük önem arz etmektedir. Gerçekleştirilen bu tez çalışmasında, Kredi ve Yurtlar Genel Müdürlüğü (KYGM)'nde barınan öğrencilerin; yurttan ayrılma nedenlerinin veri madenciliği yöntemleriyle analiz edilmesi, yurt şartlarının iyileştirilmesi veya sosyal ortamın düzenlenmesi için öneriler sunulması, ayrılma nedenleri ve barınan öğrencilerin çeşitli özellikleri analiz edilerek; ayrılması baştan öngörülen öğrencilerin yurt başvuru ve yerleştirme kriterlerinin yeniden gözden geçirilmesinin sağlanması amaçlanmıştır. Mevcut veri üzerinde ilk aşamada veri ön işleme ve öznitelik analizi gerçekleştirilmiştir. Öznitelik seçimi, korelasyon tabanlı özellik seçimi ve bilgi kazancı kullanılarak yapılmıştır. Temel makine öğrenmesi yöntemleri ile modellerin aşırı uyumunu önlemek için 10-kat çapraz doğrulama kullanılmıştır. Ankara ilinde barınan öğrenci verileri ile oluşturulan yurt öğrenci veri kümesinde, karar tablosu (decision table) %86,95 doğruluk ve %89 AUC, adaboost %86,93 doğruluk ve %89 AUC, karar kütüğü (decision stump) %86,93 doğruluk ve %85 AUC ile iyi performans gösteren sınıflandırıcılar olmuşlardır. %91,6 ile en iyi AUC alanına sahip random forest %86,3 doğruluk oranı göstermiştir. Öncelikli öğrenci verileri dikkate alınan öncelikli öğrenci veri kümesinde ise logistBoost %81,89 doğruluk ve %88 AUC, karar ağacı (J48) %81,59 doğruluk ve %83 AUC oranı ile en iyi performans veren algoritmalardır. Sınıflandırıcılar; doğruluk, hatırlama, kesinlik, f-ölçüsü ve ROC eğrisi gibi metrikler kullanılarak değerlendirilmiştir. Sonuçlar; yurt öğrenci veri kümesinin decision table, öncelikli öğrenci veri kümesinin logistBoost algoritmalarını kullanarak daha iyi sınıflandırıldığını ortaya çıkarmıştır.

Bilim Kodu	:	92429
Anahtar Kelimeler	:	Yurt memnuniyeti, kümeleme, sınıflandırma, makine öğrenmesi, veri madenciliği
Sayfa Adedi	:	70
Danışman	:	Doç. Dr. Oktay YILDIZ

ANALYZING THE REASONS FOR LEAVING THE STUDENTS LIVING IN THE DORMITORY WITH DATA MINING METHODS

(M. Sc. Thesis)

Berrin DEĞİRMENCİ

GAZİ UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

January 2024

ABSTRACT

Accommodation is an important problem for university students all over the world. Most of the university students stay in state dormitories for a period of their education life. Although state dormitories have reached 900 thousand bed capacity in our country according to 2023 data, there are still students who have accommodation problems. At this point, it is of high importance to organise the dormitories efficiently, to use the existing capacity in the most optimum way and to create an environment for the students in need. In this thesis study, it is aimed to analyse the reasons for leaving the dormitories of the students accommodated in the General Directorate of Credit and Dormitories (KYGM) by using data mining methods, to provide suggestions for improving the dormitory conditions or organising the social environment, to ensure that the dormitory application and allocation criteria of the students who are initially predicted to leave are reviewed by analysing the reasons for leaving and various characteristics of the students accommodated. In the first stage, data pre-processing and attribute analysis were performed on the available data. Feature selection was performed using a correlation-based feature selection and information gain. In order to prevent overfitting of models with basic machine learning methods, 10-fold cross-validation was used. In the dormitory student dataset created with the data of students living in Ankara, decision table with 86.95% accuracy and 89% AUC, adaboost with 86.93% accuracy and 89% AUC, decision stump with 86.93% accuracy and 85% AUC were the best performing classifiers, random forest with the best AUC of 91.6% showed an accuracy rate of 86.3%. In the prioritised student dataset, logistBoost with 81.89% accuracy and 88% AUC, decision tree (J48) with 81.59% accuracy and 83% AUC were the best performing algorithms. The classifiers were evaluated using metrics such as accuracy, recall, precision, f-measure and ROC curve. The results revealed that the dormitory student dataset was better classified using decision table and the priority student dataset was better classified using logistBoost algorithms.

Science Code	:	92429
Key Words	:	Dormitory satisfaction, clustering, classification, machine learning, data mining
Page Number	:	70
Supervisor	:	Assoc. Prof. Dr. Oktay YILDIZ

TEŞEKKÜR

Yüksek lisans eğitimimde ve tez çalışmam boyunca bilgi ve tecrübesiyle bana yol gösteren, sabırla destekleyen danışmanım sayın Doç. Dr. Oktay YILDIZ'a, eğitimim boyunca bilgi ve akademik tecrübelerinden faydalandığım bütün hocalarıma, iş hayatımda eğitimime destek veren, birlikte mesai harcadığım yöneticilerime ve iş arkadaşlarıma, bütün zorlukların üstesinden birlikte geldiğimiz, emekleriyle ve sevgileriyle yanımda olan bana güvenen aileme ve canım arkadaşlarıma teşekkür ederim.

İÇİNDEKİLER

Sayfa

ÖZET iv

ABSTRACT v

TEŞEKKÜR vi

İÇİNDEKİLER vii

ÇİZELGELERİN LİSTESİ ix

ŞEKİLLERİN LİSTESİ xi

SİMGELER VE KISALTMALAR xiii

1. GİRİŞ 1

2. VERİ MADENCİLİĞİ 9

2.1. Veri Madenciliği Veri Kaynakları 10

2.1.1. Veritabanları 10

2.1.2. Veri ambarları 11

2.1.3. Veri kümeleri 11

2.2. Veri Ön İşleme 11

2.2.1. Veri temizleme 12

2.2.2. Veri entegrasyonu 12

2.2.3. Veri dönüştürme 12

2.2.4. Veri azaltma 13

2.3. Veri Madenciliği Yöntemleri 15

2.3.1. Sınıflandırma 15

2.3.2. Kümeleme 24

2.3.3. Birliktelik kuralları 24

2.4. Sınıflandırma Modellerinin Performans Değerlendirmesi 25

Sayfa

3. DENEYSEL ÇALIŞMA VE SONUÇLARI 29
3.1. Deneysel Çalışmada Kullanılan Veri Kümeleri 30
3.1.1. Yurt öğrenci veri kümesi 35
3.1.2. Öncelikli öğrenci veri kümesi 39
3.2. Gerçekleştirilen Çalışma 44
3.2.1. Veri ön işleme çalışması 45
3.2.2. Sınıflandırma çalışması 48
4. SONUÇ VE TARTIŞMA 57
KAYNAKLAR 61
EKLER 65
ÖZGEÇMİŞ 70

ÇİZELGELERİN LİSTESİ

Çizelge **Sayfa**

Çizelge 1.1. Dönem bazlı özel yurt ve kamu yurtlarının sayısı 2

Çizelge 2.1. Karışıklık matrisi 26

Çizelge 3.1. KYGM-Özel yurtların dönem bazlı kapasiteleri. 31

Çizelge 3.2. Dönem bazlı toplam ve örgün eğitimde öğrenim gören öğrenci sayısı 31

Çizelge 3.3 Dönem bazlı örgün eğitimde olan öğrenci sayısı ve KYGM kapasitesi..... 32

Çizelge 3.4. KYGM yurtlarında barınan öncelikli öğrenci sayısı 34

Çizelge 3.5. Yurt öğrenci veri kümesi öznitelikleri 35

Çizelge 3.6. Yurt öğrenci veri kümesi cinsiyet dağılımı 35

Çizelge 3.7. Yurt öğrenci veri kümesi yaş dağılımı..... 36

Çizelge 3.8. Yurt öğrenci veri kümesi yaş kategori dağılımı 36

Çizelge 3.9. Yurt öğrenci veri kümesi anne vefat durumu dağılımı 37

Çizelge 3.10. Yurt öğrenci veri kümesi vakıf üniversitesi durum dağılımı 37

Çizelge 3.11. Yurt öğrenci veri kümesi MEB okuyan kardeş sayısı dağılımı 38

Çizelge 3.12. Yurt öğrenci veri kümesi model tip dağılımı..... 38

Çizelge 3.13. Yurt öğrenci veri kümesi yurt bölüm uzaklık dağılımı 39

Çizelge 3.14. Yurt öğrenci veri kümesi bina kapasite dağılımı 39

Çizelge 3.15. Öncelikli öğrenci veri kümesi öznitelikleri 40

Çizelge 3.16. Öncelikli öğrenci veri kümesi cinsiyet dağılımı 40

Çizelge 3.17. Öncelikli öğrenci veri kümesi yaş kategori dağılımı 41

Çizelge 3.18. Öncelikli öğrenci veri kümesi yaş dağılımı 41

Çizelge 3.19. Öncelikli öğrenci veri kümesi engel türü dağılımı 42

Çizelge 3.20. Öncelikli öğrenci veri kümesi öncelik durumu dağılımı 42

Çizelge 3.21. Öncelikli öğrenci veri kümesi fakülte kategori dağılımı 43

Çizelge 3.22. Yurt öğrenci veri kümesi bilgi kazancı sıralaması ve korelasyon tabanlı özellik seçimi sonuçları 46

Çizelge **Sayfa**

Çizelge 3.23. Öncelikli öğrenci veri kümesi bilgi kazancı sıralaması ve korelasyon tabanlı özellik seçimi sonuçları 48

Çizelge 3.24. Deneysel çalışma karışıklık matrisi 50

Çizelge 3.25. Yurt öğrenci veri kümesi için sınıflandırıcıların performansı 51

Çizelge 3.26. Öncelikli öğrenci veri kümesi için sınıflandırıcıların performansı..... 53

ŞEKİLLERİN LİSTESİ

Şekil **Sayfa**

Şekil 1.1. Öğretim dönemine göre aktif üniversite sayısı 1
Şekil 1.2. Dönem bazlı özel yurt ve kamu yurtlarının sayısı 2
Şekil 2.1. Veri madenciliği süreçleri 10
Şekil 2.2. Veri ön işleme döngüsü 11
Şekil 2.3. Öznitelik seçim süreci 13
Şekil 2.4. Karar ağacı yapısı 16
Şekil 2.5. Destek vektör makinesi 17
Şekil 2.6. Yapay sinir ağı mimarisi 19
Şekil 2.7. Regresyon modelleri 21
Şekil 2.8. Bagging algoritmasının süreçleri 22
Şekil 2.9. Random forest algoritması süreçleri 23
Şekil 2.10. K-means algoritması 24
Şekil 2.11. Sınıflandırma değerlendirme metrikleri 25
Şekil 2.12. AUC – ROC Eğrisi 27
Şekil 3.1. Deneysel çalışma metodolojisi 29
Şekil 3.2. KYGM-Özel yurtların dönem bazlı kapasiteleri. 31
Şekil 3.3. Dönem bazlı toplam ve örgün eğitimde öğrenim gören öğrenci sayısı 32
Şekil 3.4. Dönem bazlı örgün eğitimde olan öğrenci sayısı ve KYGM kapasitesi 33
Şekil 3.5. KYGM yurtlarında barınan öncelikli öğrenci sayısı 34
Şekil 3.6. Yurt öğrenci veri kümesi cinsiyet dağılımı 36
Şekil 3.7. Yurt öğrenci veri kümesi baba vefat durumu dağılımı 37
Şekil 3.8. Yurt öğrenci veri kümesi bahçe durumu dağılımı 37
Şekil 3.9. Yurt öğrenci veri kümesi odadaki yatak sayısı dağılımı 38
Şekil 3.10. Öncelikli öğrenci veri kümesi cinsiyet dağılımı 40
Şekil 3.11. Öncelikli öğrenci veri kümesi adres ili dağılımı 42

Şekil **Sayfa**

Şekil 3.12. Öncelikli öğrenci veri kümesi puan türü dağılımı 43

Şekil 3.13. Öncelikli öğrenci veri kümesi barınma durumu dağılımı..... 44

Şekil 3.14. Öncelikli öğrenci veri kümesi yurt lokasyonu dağılımı 44

Şekil 3.15. Yurt öğrenci veri kümesi ROC eğrileri 52

Şekil 3.16. Öncelikli öğrenci veri kümesi ROC eğrileri..... 54

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Simgeler	**Açıklamalar**
Ghz	Gigahertz
GB	Gigabayt
s	Saniye

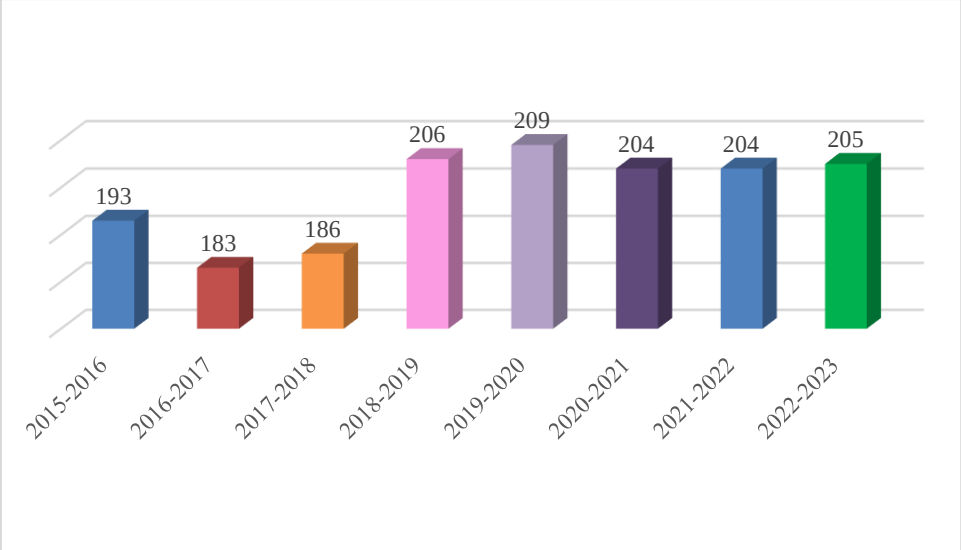
Kısaltmalar	**Açıklamalar**
ANN	Yapay Sinir Ağları - Artificial neural networks
AUC	ROC eğrisi altında kalan alan
CA	Kapsama Algoritması- Covering Algorithm
CCP	Müşteri Kayıp Tahmini
CPU	İşlemci
CRM	Müşteri İlişkileri Yönetimi - Customer Relationship Management
DF	Akıllı karar ormanı
EA	Kapsamlı Algoritma- Exhaustive Algorithm
FP	Yanlış Tahminlerin- False Positive
FT	İşlevsel Ağaçlar
GA	Genetik Algoritma
IBK	En yakın komşu algoritması
KNN	K-en yakın komşu
KYGM	Kredi ve Yurtlar Genel Müdürlüğü
LA	LEM2 Algoritması
LLM	Lojistik Model Ağacı
LMT	Logit yaprak modeli
LR	Lojistik Regresyon
LWL	Yerel Ağırlıklı Doğrusal Regresyon
MEB	Milli Eğitim Bakanlığı

Kısaltmalar	Açıklamalar
MLP	Çok katmanlı algılayıcı – Multilayer perceptron
NB	Naive Bayes
ÖSYM	T.C. Ölçme, Seçme Ve Yerleştirme Merkezi
RAM	Rastgele erişimli bellek
RBF	Radyal Tabanlı Fonksiyon -Radial Basis Function
RF	Rastgele Orman – Random Forest
ROC	Alıcı işlem karakteristiği
RST	Kaba Küme Teorisi
RT	Rastgele Tablo – Random Table
SMO	Karar destek vektörü çeşidi
TP	Doğru Tahminler- True Positive
YÖK	Yükseköğretim Kurumu

1. GİRİŞ

Öğrenci yurdu öğrencilerin eğitimleri süresince birlikte konakladıkları tesislerdir. Genellikle ailesinden ya da ikamet ettiği şehirden farklı bir şehirde eğitimine devam eden öğrenciler tarafından kullanılan yurtlarda çoğunlukla üniversite öğrencileri kalıyor olsa da yüksek lisans veya doktora öğrencileri tarafından da tercih edilmektedir.

Ülkemiz yükseköğretimde 2000'li yıllarda hızlı bir gelişme ve büyüme yaşanmaya başlamıştır bu da beraberinde öğrencilerin konaklama sorununu ortaya çıkarmıştır. Şekil 1.1. 'de öğretim dönemlerine göre üniversite sayısındaki artış gösterilmiştir [1]. Farklı şehirlerde açılan yeni üniversiteler ile birlikte öğrenciler ikamet ettikleri illerden öğrenim görecekleri illere taşınmaya başlamışlardır.



Şekil 1.1. Öğretim dönemine göre aktif üniversite sayısı

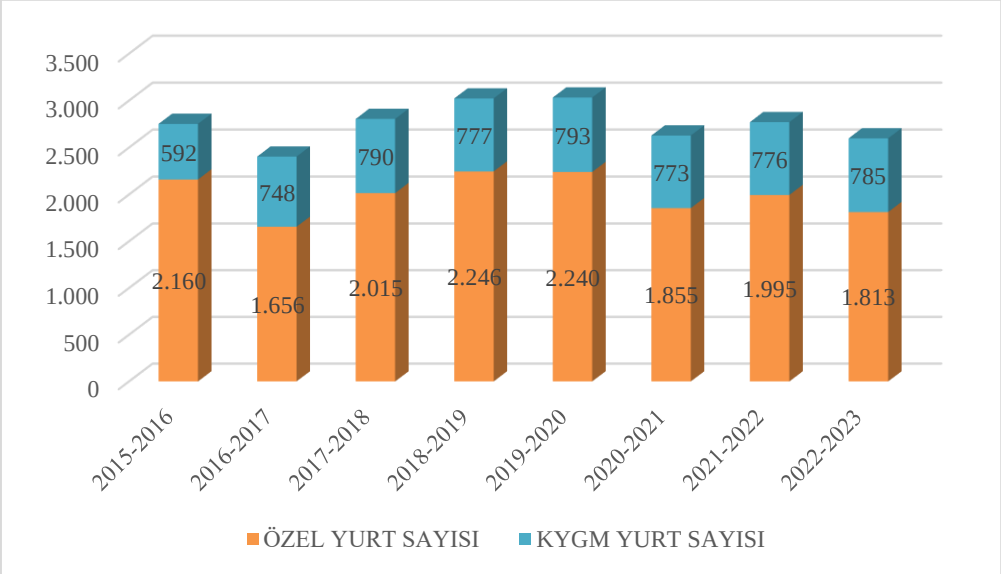
Üniversite öğrencilerinin barınma yerleri olarak; kamu, vakıf, dernek ve özel sektöre ait yurtları, apart evleri ve otelleri, özel pansiyonları, kiralık evleri ya da akrabalarının yanında kalmayı seçebilirler. Barınma yerlerinin seçilmesinde ise, öğrencinin kendisinin ve ailesinin sosyo-ekonomik durumu, güvenlik, kira fiyatları, okula yakınlık ve ulaşım kolaylığı, sosyal yaşam imkanları, üniversitenin bulunduğu ildeki konutların nitelik ve kapasitesi gibi birçok faktör etkili olmaktadır [2].

Yurt alternatifleri mukayese edildiğinde; kamu yurtlarının daha ucuz olduğu, özel yurtların odadaki kişi sayısıyla daha konforlu olduğu, üniversite yurtlarının ise üniversite kampüslerinde bulunmalarından dolayı açık alanların genişliği, rekreasyonel tesis ve imkanların çeşitliliği ve akademik çevre açısından diğer yurtlara göre daha üstün olduğu sayılabilir [3].

Çizelge 1.1. ve Şekil 1.2. de öğretim dönemlerine göre kamuya ait olan KYGM yurt ve özel yurt sayıları verilmiştir [4]. Öğrenciler belirtilen kriterlere göre barınacakları yurtları seçmektedirler.

Çizelge 1.1. Dönem bazlı özel yurt ve kamu yurtlarının sayısı

ÖĞRETİM DÖNEMİ	ÖZEL YURT SAYISI	KYGM YURT SAYISI
2015-2016	2 160	592
2016-2017	1 656	748
2017-2018	2 015	790
2018-2019	2 246	777
2019-2020	2 240	793
2020-2021	1 855	773
2021-2022	1 995	776
2022-2023	1 813	785



Şekil 1.2. Dönem bazlı özel yurt ve kamu yurtlarının sayısı

Yurtlarda barınarak üniversite öğrenimlerine devam eden öğrenciler, ailelerinden uzak yaşamlarına devam etmeleri sebebiyle barınma, beslenme, sağlık, ekonomik, sosyal gelişim vb. ile ilgili sorunlar yaşama riski ile karşı karşıya kalmaktadırlar. Ayrıca yurtta kalan

öğrencilerin yeni bir yaşam tarzı ile tanışmaları ve toplu kalınan bir ortama girmeleri nedeniyle başka problemler ile de karşılaşabilmektedirler. Bu nedenlerden yurttan ayrıldıkları varsayılmaktadır.

Gençlik ve Spor Bakanlığı Kredi ve Yurtlar Genel Müdürlüğünün yurtlar ile ilgili amaçları; yurt içinde ve dışında yükseköğrenim gören öğrenciler için yurtlar yapmak ve bunların işletmesini sağlamak, beslenme yardımı yapmak, çeşitli eğitimler ile gelişmelerine katkı sağlamak ve öğrencilerin yükseköğrenimlerini kolaylaştırmaktır. Yurtlarda barınan öğrencilerin, serbest zamanlarını en iyi şekilde değerlendirmek, kültürel yönden gelişmelerini sağlamak, zihinsel ve bedensel geliştirmeleri amacıyla sosyal, kültürel ve sportif faaliyetler düzenlenmektir. Bakanlık çeşitli spor federasyonları, vakıf, dernek, sivil toplum kuruluşları ile kamu kurum ve kuruluşlarına barınma için destek vermektedir [5]. Genel Müdürlüğün amacından da anlaşıldığı üzere öğrencilerin yurtlardan memnun kalması, kaliteli hizmet alması ve öğrenim hayatları boyunca yurttan ayrılmamaları istenmektedir. Tez çalışmamız ile kuruma yurttan ayrılması tahmin edilecek öğrenciler hakkında fikir verilebilir.

Yükseköğrenim gören öğrencileri barındırma Gençlik ve Spor Bakanlığının ana hizmetlerinden biridir. KYGM bu öğrencileri barındırmakla öğrenimlerine katkıda bulunmaktadır. Üniversitelerin örgün öğrenim programlarına kayıtlı öğrenciler devam ettikleri bölümün normal süresi kadar yurtlarda kalabilmektedirler. Yükseköğretimde yurtta barınmanın devamı öğrencilerin hizmet kalitesi algısıyla ilgilidir. “Öğrenci memnuniyeti” ve “kalite' terimler”, yükseköğretim otoritelerinin felsefesinin merkezinde yer almaktadır ve bunların önemi, geleceğin yükseköğretimi için yenilenmiş, öngörülebilir bir refah vaadiyle devam etmektedir [6].

Yapılan literatür araştırmasında yurt içi ve yurt dışında, öğrencilerin yurttan ayrılma sebeplerinin veri madenciliği yöntemleriyle analiz edildiği çok fazla çalışma bulunmamaktadır. Ancak konuya benzer müşteri kaybı araştırmaları mevcuttur. Literatürde kayıp tahmini; veri madenciliği, makine öğrenmesi ve hibrit yöntemler kullanılarak yapılmıştır. Bu teknikler, şirketlerin kayıp müşterileri belirlemesini, tahmin etmesini ve elde tutmasını sağlar. Karar ağaçları, müşteri kaybıyla ilgili problemlerin tahmin edilmesi için kullanılan en yaygın bilinen yöntemlerdir.

Büyük veri alanındaki son gelişmeler ile bu tür verileri analiz etmek için kullanılabilecek birçok veri madenciliği ve makine öğrenimi çözümü bulunmaktadır. Bu teknikler, verileri analiz eder ve müşteri kaybının ardındaki nedenleri belirler. CRM (Müşteri İlişkileri Yönetimi) bu teknikleri karlarını maksimize etmek için kullanabilir. Ayrıca, kaybedilecek müşterilerin oranını azaltmak için elde tutma stratejileri tasarlamak için kullanılabilir. CRM, veri madenciliği tekniklerini kullanarak doğru müşteri ihtiyaçlarını belirleyerek bir şirketin müşteriyi elde tutma hedefine ulaşabilir [7].

Ülkemizdeki yurt yerleştirme ve barınma hizmetine benzer yani merkezi yerleştirme ve yönetme yurt dışında görülmemektedir. Bu nedenle yurt dışındaki çalışmalarda; üniversite öğrencilerinin yurtta barınmasının lokal ve öğrenim gördükleri üniversite kampüsü ile sınırlı olduğu görülmüştür.

Yurt içindeki yurt yerleştirmesine benzer çalışmalar Frijters, Islam [8] ve Perach and Anily [9] makalelerinde bulunmaktadır. Kökalan'ın yaptığı çalışmada devlet, özel ve vakıf öğrenci yurtlarının hizmet kalitesi ölçülmüştür [10]. Kavuncu'nun yaptığı araştırmada Türkiye'de yükseköğrenimde öğrenim gören öğrenciler, yurt hizmetleri ve barınma sorunları hakkında bilgi verilmiştir [2].

Aslanargun'un çalışmasında Türkiye'deki üniversite öğrenci yurt hizmetleri ile ilgili memnuniyet anketi yapılmış ve sonuçlar değerlendirilmiştir. Çalışmada öğrencilerin sosyalleşme olanakları ve ücretlerden memnun oldukları ancak hijyen sorunları olduğunu belirttikleri bildirilmiştir. Yemek ve yatak hizmetlerindeki zayıf noktaların azaltılması için öğrencilerin de içinde bulunduğu iç denetim sistemiyle azaltılabileceği belirtilmiştir [11].

Bingöl ve Çavlin'in makalesinde Türkiye'de üniversite eğitimi için yer değiştirerek İstanbul'daki üniversitelerde eğitim alan öğrencilerin profilleri cinsiyet ve mesafe temelinde incelenmiştir [12].

Eşidir'in uzmanlık tezinde, Türkiye'de yükseköğretim öğrencilerinin barınma problemi; bölgesel dağılım, maliyet, finansman, insan-mekân ilişkisi, kapasite, işletme gibi birçok boyutta ele alınmıştır. Sorunlar ortaya konulmuş, barınma probleminin çözülmesi için öneriler yapılmıştır. Bu amaçla, ulusal ve uluslararası çalışmalardan faydalanılmış, sayısal analizler yapılarak rasyonel sonuçlara ve tahminlere ulaşılmaya çalışılmıştır [13].

Rave ve diğerlerinin önerdiği çerçeve, sınıf kalitesi ile öğrenci memnuniyeti arasındaki ilişkiyi incelemek için uygulanmıştır. İki hesaplama yaklaşımı (denetimli: makine öğrenimi ve denetimsiz: temel bileşen analizi) kullanılmıştır ve R dilini kullanan yarı otomasyonu sayesinde çerçeve verimli ve erişilebilirdir [14].

Ullah ve diğerleri; müşterileri sınıflandırmak için rastgele ağaç (RT), J48, rastgele orman (RF), karar kütüğü, adaboostM1 + karar kütüğü, torbalama + rastgele ağaç, naive bayes (NB), çok katmanlı algılayıcı (MLP), lojistik regresyon (LR) , IBK ve LWL algoritmalarını kullanmışlardır. En yüksek başarı %88,63 ile rastgele orman (RF)'dır. K-ortalama kümeleme tekniği kullanılarak müşteri profilleri çıkarılmıştır. Son adımda, model; her müşteri profili için elde tutma stratejileri önermiştir [7].

De Caigny ve diğerleri, logit yaprak modeli (LLM) diye adlandırılan hibrit bir model önermişlerdir. Logistik regresyon ve karar ağacının birleşimi bir modeldir. Bileşenleriyle mukayese edildiğinde daha başarılıdır. Müşteriler karar ağacı ile üç alt kümeye ayrılmış ve logistik regresyon ile sonuçlandırılmıştır [15].

Xu ve diğerleri çalışmalarında; özellik oluşturma, yığınlama modeli (logistik regresyon, karar ağacı, naive bayes) ve yumuşak oylamadan oluşan yeni bir müşteri kaybı tahmin sistemi önermişlerdir, %96,12 başarı elde edilmiştir [16].

Amin ve diğerleri çalışmalarında, müşteri kaybı ve kayıpsızlık ile ilgili önemli karar kurallarını çıkarmak için kaba küme teorisine (RST) dayanan akıllı bir kural tabanlı karar verme tekniği önermiştir. Kapsamlı simülasyon deneyleri, kapsamlı algoritma (EA), genetik algoritma (GA), kapsama algoritması (CA) ve LEM2 algoritması (LA) olmak üzere dört kural oluşturma mekanizması kullanılarak yapılmıştır en iyi performansı önerilen RST tabanlı CCP yaklaşımı göstermiştir [17].

Wu ve diğerlerinin önerdiği çerçeve, müşteri kaybını daha iyi yönetmek için telekomünikasyon operatörlerine eksiksiz bir kayıp analizi sağlamak için kayıp tahmini ve müşteri segmentasyon sürecini birleştirmiştir. Altı makine öğrenimi sınıflandırıcısı ile yapılan deneylerde üç veri seti kullanarak müşteri kayıp durumu tahmin edilmiştir. AdaBoost'un %77,19 doğruluk ile veri kümesi 1'de; random forest, %93,6 doğruluk ve ile

veri kümesi 2'de; random forest, doğruluk açısından %63,09 ile veri kümesi 3'te en iyi performansı göstermiştir [18].

Ahmad ve diğerlerinin geliştirdiği model, büyük veri platformunda makine öğrenimi tekniklerini kullanmıştır ve yeni bir özellik mühendisliği ve seçimi yöntemi oluşturmuştur. Modelin performansını ölçmek için eğri altındaki alan (AUC) standart ölçüsü benimsenmiştir ve elde edilen AUC değeri %93,3'tür [19].

Usman-Hamza ve diğerleri, müşteri kayıp tahminini (CCP) için akıllı karar ormanı (DF) modelleri geliştirmişlerdir. Lojistik model ağacı (LMT), rastgele orman (RF) ve işlevsel ağaçlar (FT), DF modelleri ile karşılaştırılmış ve önerilen DF modellerinin telekom endüstrisinde üstün tahmin performansı gösterdiği görülmüştür [20].

Lalwani ve diğerleri; telekom sektöründeki müşteri kaybının tahmini için rastgele orman, naive bayes, lojistik regresyon, destek vektör makinesi, karar ağaçları vb. makine öğrenmesi teknikleri uygulamışlardır. Son olarak test seti üzerinde elde edilen sonuçlar karışıklık matrisi ve AUC kullanılarak değerlendirilmiştir. Adaboost ve XGboost sınıflandırıcıların sırasıyla %81,71 ve %80,8 ile en yüksek doğruluğu verdiği tespit edilmiştir. %84'lük en yüksek AUC puanı, diğerlerinden daha iyi performans gösteren Adaboost ve XGBoost sınıflandırıcıları tarafından elde edilmiştir [21].

Suh çalışmasında; Kore'deki bir elektronik şirketindeki su arıtma cihazı kiralama hizmetinin müşteri davranış bilgilerini analiz etmiştir. Kaybın çok sık meydana geldiğini tespit etmiş ve kayıp tahmin modelini geliştirip doğrulamayı amaçlamıştır. Modelin performansını ölçmek için f-ölçüsü ve eğri altındaki alan (AUC) benimsenerek %93'lük bir F1 değeri ve %88'lik bir AUC elde edilmiştir [22].

Meghyasi ve Rad'in çalışmasında, telekomünikasyon endüstrilerinde müşteri kaybı tahminine genetik algoritma ve modüler sinir ağı temelli hibrit bir yöntem sunmuştur ve örnek olarak Irancell verilerini kullanmışlardır. Bu çalışmanın %95,5 olan doğruluk sonucu ile diğer yöntemlere göre daha başarılı olduğu gösterilmiştir [23].

Tezin literatüre katkısı

Ülkemizde son yıllarda yurt kapasitesi ve imkanlarında ciddi bir artış gözlemlenmektedir. Son 20 yılda Gençlik ve Spor Bakanlığına ait yurtların kapasitesi 4,7 kat artış göstermiş olup 870 bin yatak kapasitesine ulaşılmıştır [4]. Buna rağmen halen barınma sorunu yaşayan öğrenciler bulunmaktadır. Bu nedenle yurtların iyi organize edilmesi, mevcut kapasitenin en uygun şekilde kullanılması ve ihtiyaç sahibi öğrencilere ortam oluşturulması büyük önem arz etmektedir.

Gerçekleştirilen bu tez çalışmasında, Kredi ve Yurtlar Genel Müdürlüğü (KYGM)'nde barınan öğrencilerin; yurttan ayrılma nedenlerinin veri madenciliği yöntemleriyle analiz edilmesi, yurt şartlarının iyileştirilmesi veya sosyal ortamın düzenlenmesi için öneriler sunulması, ayrılma nedenleri ve barınan öğrencilerin çeşitli özellikleri analiz edilerek; ayrılması baştan öngörülen öğrencilerin yurt başvuru ve yerleştirme kriterlerinin yeniden gözden geçirilmesinin sağlanması amaçlanmıştır.

Bu çalışmada öğrencilerin yurttan ayrılma sebepleri veri madenciliği yöntemleri ile incelenmiş ve yurttan erken ayrılan öğrencilerin tespitine çalışılmıştır. Çalışmanın amacı, makine öğrenmesi ve veri madenciliğinde mevcut teknikleri araştırmak ve ayrılan öğrenci tahminleri için bir model önermek, ayrılma faktörlerini belirlemek ve elde tutma stratejileri sağlamaktır.

Yapılan deneysel çalışmalarda, çeşitli veri madenciliği algoritmaları kullanılarak yurttan ayrılan öğrenci tahmini gerçekleştirilmiştir. Bir sınıflandırıcının performansı, mevcut veri kümesine bağlıdır. Çalışmada iki veri kümesi kullanılmıştır; ilki Kredi ve Yurtlar Genel Müdürlüğünün Ankara ilinde bulunan yurtlarında barınan öğrencilerinin 2016-2022 yılları arasındaki verilerini içermektedir, ikinci veri kümesi ise Kredi ve Yurtlar Genel Müdürlüğünün yurtlarında 2016-2022 yılları arasında barınan öncelikli öğrenci verilerini içermektedir. Önemli öznitelikler, bilgi kazancı ve korelasyon tabanlı özellik seçimi gibi öznitelik seçme teknikleri kullanılarak seçilmiştir. Önerilen ayrılan öğrenci tahmin modeli, sınıflandırıcı doğrulama metrikleri kullanılarak değerlendirilmiştir. Doğruluk, TP oranı, FP oranı, kesinlik, hatırlama, f-ölçüsü ve ROC eğrisi kullanılarak modelin performansı ölçülür. Deneylerden, önerilen modelin yüksek doğruluk elde ederek, decision table algoritmasının diğer makine öğrenmesi algoritmalarına göre daha iyi doğruluk ürettiği belirlenmiştir.

Yapılan araştırmalar sonucunda, konu hakkında yurtiçinde pek fazla çalışma yapılmadığı görülmüştür. Çalışmanın yeni araştırmalar için öncü olacağı varsayılmaktadır.

Bu tez çalışmanın ilk kısmında yapılan literatür çalışmasıyla konuyla bağlantılı çalışmalar sunulmuştur, 2. bölümde çalışmada kullanılan yöntemler açıklanmıştır, 3. kısımda kullanılan veri kümesi ve yapılan deneysel çalışmalar gösterilmiştir, 4. Bölümde yapılan deneysel çalışmaların sonuçları tartışılmış ve değerlendirilmesi yapılmıştır.

2. VERİ MADENCİLİĞİ

Her gün büyük miktarda verinin toplandığı bir dünyada yaşamaktayız. Bu verilerin analiz edilmesi, sonuç çıkarılması, bilgi üretimi önemli bir ihtiyaçtır. Veri madenciliği verilerden bilgi keşfetmek için araçlar kullanarak bu ihtiyacı nasıl karşılayabileceğini ele almaktadır. Veri madenciliği, bir veri setinden en önemli bilgileri çıkarmayı ve bu bilgilerin daha sonra kullanılması için yapılandırılmasını sağlamaktadır. Veri madenciliği; makine öğrenmesi, veri tabanı ve istatistik sistemlerinden yöntemler içerir. Diğer istatistiksel çözümlere göre uygun maliyetli ve verimlidir. Veri madenciliği iş dünyasından, bilim ve mühendisliğe, tıptan günlük yaşamın her alanına yayılmaktadır.

Veri madenciliği kullanım alanları:

- Satış: veri madenciliği satışları artırmaya yardımcı olur. Hangi ürünlerin birlikte satıldığını ve hangi ürünlerin popüler olduğunu tespit eder.
- Pazarlama: işletmeler, pazarlama faaliyetlerini geliştirmek için veri madenciliğini kullanabilir. Hangi pazarlama stratejilerinin müşteriler için uygun olduğunu anlamalarına yardımcı olabilir.
- Üretim: veri madenciliği; zaman, maliyet, ham madde, stokların en verimli şekilde kullanılmasına yardımcı olabilir.
- Dolandırıcılık algılama: var olmaması gereken aykırı değerleri veya korelasyonları belirlemek için veri madenciliği kullanılabilir.
- İnsan kaynakları: veri madenciliği, çalışanların neden işten ayrıldığını ve işe alınanları kuruma katılmaya neyin ikna ettiğini analiz edebilir.
- Müşteri hizmetleri: veri madenciliği müşteri memnuniyeti için kullanılabilir.
- Müşteriyi elde tutma: şirketler, kayıp müşterilerin özelliklerini belirlemek için veri madenciliğini kullanabilir, aynı özelliklere sahip diğer müşterileri elde tutmak için stratejiler geliştirebilirler.
- Güvenlik: izinsiz giriş algılama teknikleri, ağ kesintileri gibi anormallikleri belirlemek için veri madenciliği kullanılabilir.
- Eğlence: yayın hizmetleri, kullanıcıların ne izlediğini veya dinlediğini analiz etmek ve alışkanlıklarına göre kişiselleştirilmiş önerilerde bulunmak için veri madenciliğinden faydalanabilir.

- Sağlık hizmeti: veri madenciliği, tıbbi durumların teşhis edilmesine, hastaların tedavi edilmesine ve tıbbi görüntüleme sonuçlarının analiz edilmesine yardımcı olabilir.

Şekil 2.1.' de veri madenciliği süreçleri görülmektedir. Bu süreçte yer alan adımlar: hedef tanımlama, veri toplama, veri ön işleme, model oluşturma, sonuç doğrulama, raporlama, analiz çıkarma, çıkan sonuçlara göre iyileştirme olarak açıklanabilir.



Şekil 2.1. Veri madenciliği süreçleri

2.1. Veri Madenciliği Veri Kaynakları

Veri, üzerinde işlem yapılmamış bilgi parçacıklarıdır. Veriler; araştırmalardan, deneylerden, programlardan ve internet ortamından elde edilebilir. Yapılandırılmış, yapılandırılmamış ve yarı yapılandırılmış olarak saklanırlar. Veri türleri; statik, dinamik, açık veri, büyük veri, kişisel veri, açık hükümet verisi gibi gruplara ayrılabilir. Verilerin depolanması, yönetilmesi, işlenmesi ve analiz edilmesi önemlidir. Kurumların verilerden değer elde etmesi, karar vermeyi ve kurum etkinliğinin artmasını sağlar.

2.1.1. Veritabanları

Veritabanı verilerin organize şekilde depolandığı, kolay ulaşılmasını ve yönetilmesini sağlayan bir sistemdir. Veritabanı yönetim sistemleri yazılımları, veritabanı yapılarını ve veri depolamayı tanımlamak için mekanizmalar sağlar; eşzamanlı, paylaşılan veya dağıtılmış veri erişimini belirlemek ve yönetmek için ve sistem çökmelerini veya yetkisiz

erişim girişimlerini engelleyerek depolanan bilgilerin tutarlılığını ve güvenliğini sağlamak için kullanılır [24].

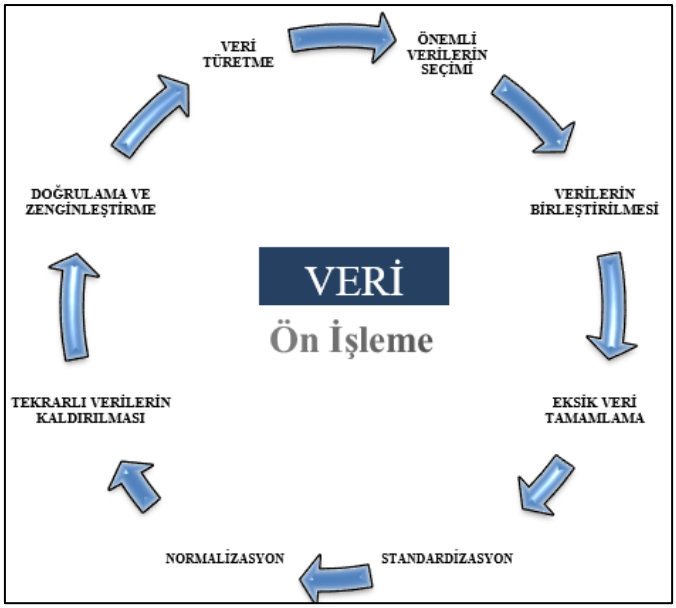
2.1.2. Veri ambarları

Veri ambarı, birden fazla kaynaktan gelen büyük miktarda veriyi toplayan ve birleştiren, hızlı sorgulama ve analiz amaçlarıyla tasarlanmış bir veritabanı türüdür. Veriler genellikle veri temizleme, dönüştürme süreçlerinden geçirilir, belirli bir zaman dilimine aittirler. Üzerlerinde güncelleme, silinme işlemleri yapılamaz.

2.1.3. Veri kümeleri

Özel amaçlar için oluşturulmuş genellikle araştırmacılar tarafından veri kaynağı olarak kullanılan elektronik dosyalardır. Veri görselleştirme ve analiz gereksinimleri için oluşturulurlar. Makine öğrenimi alanında sıklıkla kullanılırlar. Yapılandırılmamış veriler içerebildiklerinden veri madenciliği uygulamalarında kullanmak için ön işlemlerden geçirilirler.

2.2. Veri Ön İşleme



Şekil 2.2. Veri ön işleme döngüsü

Veri ön işleme, veri madenciliği yöntemleri uygulanmadan önce veri kümesi üzerinde yapılan; seçme, tekrarlanan verileri kaldırma, düzeltme, eksik veriyi tamamlama, temizleme, dönüştürme, boyut indirgeme, bütünleştirme, normalleştirme vb. işlemlerdir.

Bu bölümde, veri ön işlemede yer alan başlıca adımlar; veri temizleme, veri entegrasyonu, veri azaltma ve veri dönüşümü incelenmiştir.

2.2.1. Veri temizleme

Eksik değerler doldurulur, düzeltmeler yapılır, gürültülü veriler çıkarılır, aykırı değerler belirlenir veya kaldırılır böylece tutarsızlıklar giderilir [24]. Eksik veri için yapılabilecek işlemler:

- Eksik değerler veri kümesinden çıkarılır.
- Eksik değerler manuel doldurulur.
- Eksik değerler yerine sabit bir değer kullanılır.
- Tüm verilerin aritmetik ortalamasıyla eksik değerler doldurulur.
- Sınıflandırma yapılarak eksik verinin olduğu özniteliğin veri ortalaması alınıp eksik veri yerine kullanılır.
- Uygun bir tahminleme yöntemi ile eksik veri tespit edilir.

2.2.2. Veri entegrasyonu

Farklı veri kaynaklarından gelen veriler ile oluşturulan veri kümelerinde veri birleştirme işleminin yapılması gereklidir. Veri madenciliği sürecindeki fazlalıkların ve tutarsızlıkların azaltılmasına ve önlenmesine yardımcı olabilir.

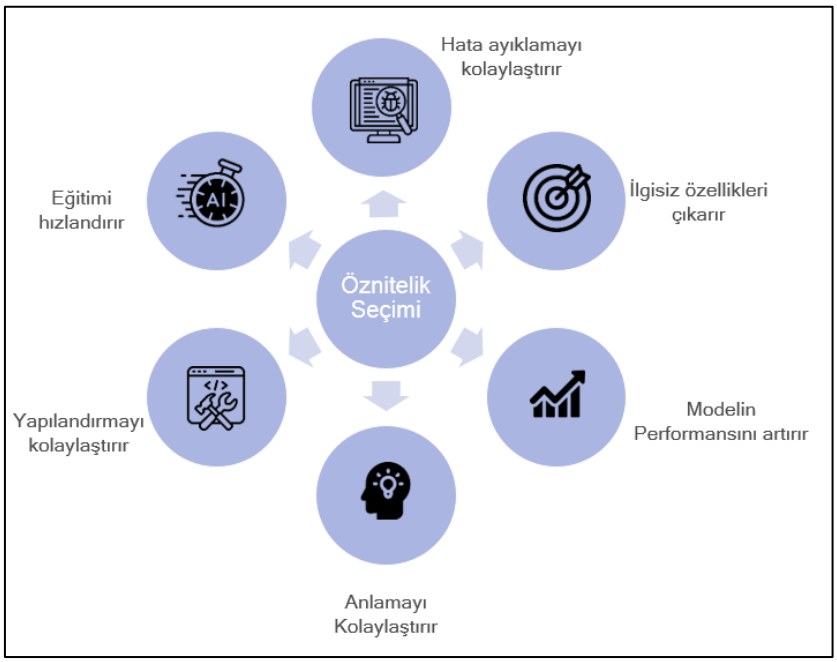
2.2.3. Veri dönüştürme

Bu ön işleme adımında veriler dönüştürülür veya birleştirilir, böylece ortaya çıkan madencilik süreci daha verimli ve bulunan modellerin anlaşılması daha kolay olur. Verilerin çok büyük ve çok küçük değerler içermesi durumunda normalizasyon ve standartlaştırma yapılır.

2.2.4. Veri azaltma

Karmaşık ve büyük bir veri setinde veri madenciliği yapmak uzun zaman alır. Bu da analizi pratik olmaktan veya uygulanabilir olmaktan çıkarır. Hacim olarak çok daha küçük olan ancak orijinalin bütünlüğünü koruyan veri seti madencilik açısından daha verimlidir. Bu işlemler öznitelik seçimi ile yapılabilir.

Öznitelik seçimi



Şekil 2.3. Öznitelik seçim süreci

Öznitelik seçimi, çözülmesi gereken problem için en önemli ve en yaralı özellikleri seçerek veri setindeki öznitelik sayısını azaltmayı amaçlar. İlgili olmayan ve gürültülü veriyi kaldırır, veri kalitesini artırır, özellik setinin boyutunu düşürerek algoritmanın hızını arttırır, veri setini görselleştirilebilir, anlaşılabilir ve basit bir şekilde tanımlanabilir hale getirir, elde edilen modelin başarısını arttırır [25]. Bu nedenle özellik seçimi önemlidir. Literatürde öznitelik seçimi için birtakım teknikler [8, 26-28] mevcuttur. Sıklıkla kullanılan iki yöntem açıklanmıştır.

Bilgi kazancı, Shannon'un bilgi teorisine dayanmaktadır. Bir veri kümesinin dönüşümünden kaynaklanan entropideki azalmayı hesaplar ve 0 ile 1 arasındadır. Entropy rastgeleliği,

belirsizliği ve beklenmeyen durumun ortaya çıkma olasılığını gösterir. Her değişken bilgi kazancını hedef değişken bağlamında değerlendirerek özellik seçimi için kullanabilir.

Genel entropi hesaplanır *Ent(A)*, her bir özniteliğin entropisi $Ent_B(A)$ ayrı ayrı hesaplanır, hesaplanan entropi değerleri genel entropi değerinden çıkartılarak hangi özniteliğin genel entropi değerini en çok azalttığı belirlenir (kazanç=Gain), entropiyi en çok azaltan öznitelik en çok kazanç sağlayandır [29].

Eş. 2.1'de m verinin miktarını ve p_i ise ilgili verinin olma olasılığını ifade etmektedir. Entropinin formuludür. Öznitelik entropisi Eş. 2.2.'de v verinin miktarını göstermektedir.

$$Ent(A) = -\sum_{i=1}^{m} p_i \, log_2 \, (p_i) \tag{2.1}$$

$$Ent_B(A) = \sum_{j=1}^{v} \frac{|A_j|}{|A|} \, Ent \, (A_j) \tag{2.2}$$

$$Gain = Ent \, (A) - \, Ent_A(B) \tag{2.3}$$

Korelasyon tabanlı özellik seçimi, korelasyon analizi kullanmaktadır. Korelasyon iki varlık arasındaki istatistiksel ilişki olarak tanımlanır. Makine öğreniminde korelasyon, iki veya daha fazla özellik veya niteliğin birbirleriyle ne kadar ilişkili olduklarını kontrol etmek için kullanılır. İki özellikten herhangi biri yüksek düzeyde ilişkiliyse veya her ikisi de aynı bilgiyi taşıyorsa bu durumda bunlardan biri gereksizdir aralarından seçim yapılır.

Korelasyon tabanlı özellik seçimi, özellik alt kümelerinin bilgi değerlerini ölçen bir fonksiyonun yanı sıra bir arama algoritması da kullanmaktadır. Her özelliğin sınıf etiketini tahmin ederken aralarındaki iç korelasyon değerlerini de dikkate almaktadır. İyi bir özellik altkümesi sınıfla yüksek derecede ilişkili ancak birbirleri ile düşük korelasyona sahip özelliklerden oluşur [30].

$$M_s = \frac{k\bar{r}_{ci}}{\sqrt{k + k(k-1)\bar{r}_{ii}}} \tag{2.4}$$

Eş. 2.4'te gösterilen eşitlikte, k altkümedeki özellik sayısı, $\bar{r}_{ci}$ dış değişken ile özellik arasındaki ortalama korelasyonu, $\bar{r}_{ii}$ özelliklerin birbirleri arasındaki ortalama iç korelasyonunu belirtmektedir.

2.3. Veri Madenciliği Yöntemleri

Veri madenciliği yöntemleri genel olarak sınıflandırma, kümeleme ve birliktelik analizi şeklinde üç grupta incelenebilir.

Sınıflandırma, kestirimci ya da diğer bir anlatımla tahmin yapan bir yöntemdir. Sınıflandırma bir veri kümesinin, içerdiği verinin ortak özelliğine göre belirli gruplara ayrılmasını sağlamaktadır. Sınıflandırma için birçok algoritma geliştirilmiştir. Başlıca algoritmalar; regresyon, karar ağaçları, k-en yakın komşu algoritması, bayes sınıflandırıcılar, regresyon ağaçları, rastgele orman şeklinde sıralanabilir [31].

Kümeleme, çeşitli uzaklık hesaplama yöntemleri kullanarak, bir veri kümesindeki birbirine benzer örnekleri kategorilere ayırmaktadır. K-ortalamalar, k-en yakın komşu ve k-en uzak komşu başlıca kümeleme teknikleridir.

Birliktelik kuralları, bir veri kümesindeki eylemlerin birlikte gerçekleşme durumlarını ortaya çıkarmaktadır. Apriori, en bilinen ve kullanılan algoritmadır.

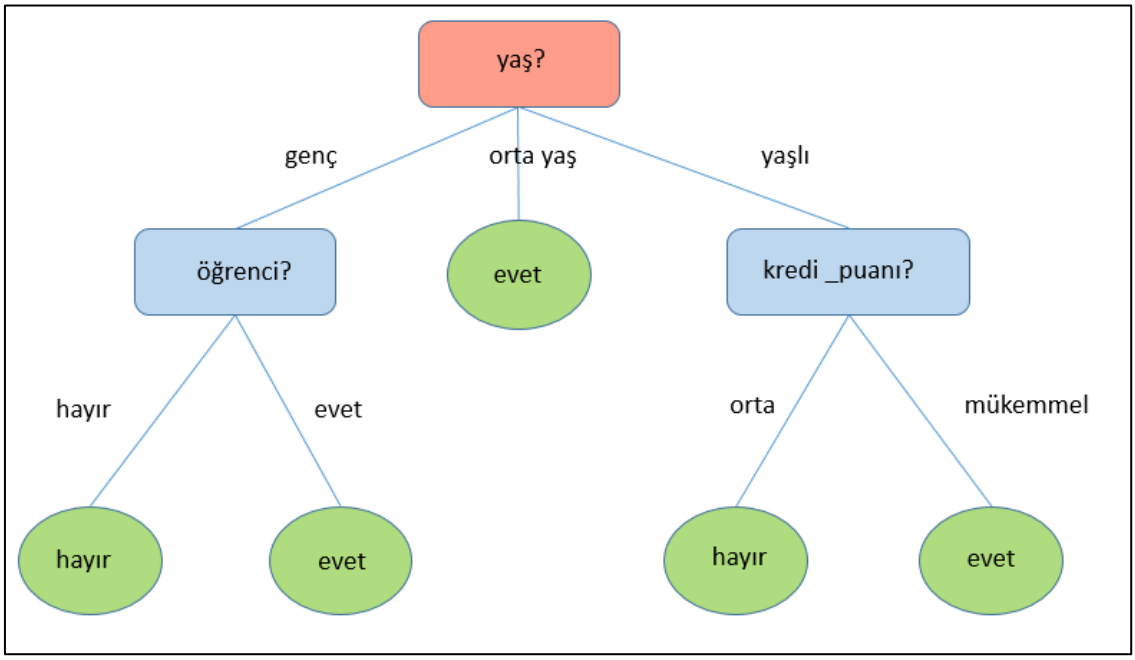
2.3.1. Sınıflandırma

Sınıflandırma analizi, en temel veri madenciliği yöntemlerindendir ve verileri farklı gruplara ayırmaya yarar. Çözülmesi gereken belirli bir soru veya zorluğa dayalı olarak veri noktalarının gruplara veya sınıflara atanmasını içerir. Güvenilirliği, verimliliği ve popülerliği dikkate alınarak, benzer problemlerin çözümü için kullanılan altı köklü makine öğrenmesi tekniği sunulur [21, 25, 32-34].

Karar ağaçları

Karar ağacı öğrenme yöntemleri, bir kararın olası her sonucunu göstermek için dallanma yöntemini kullanır. Ayrık değer nitelikleri ve sürekli değer nitelikleriyle de çalışabilirler.

Öğrenen ağaçlar daha sonra if-then kuralları şeklinde temsil edilir [25]. Karar ağaçları, belirli bir veri seti için sınıflandırma kuralları oluşturan karar kümelerini gösteren ağaç şeklindeki yapılardır. Müşteri kaybı probleminde, yüksek doğruluğa sahiptir. ID3, J48, C4.5, CHAID, CART en çok kullanılan karar ağacı algoritmalarıdır. Karar ağacının popüler olmasının nedeni; çok boyutlu verileri işleyebilmesi, edinilen bilginin ağaç şeklinde temsil edilebilmesi, kolay anlaşılabilir olmasıdır. Karar ağacının öğrenme ve sınıflandırma adımları basit ve hızlıdır. Genel olarak karar ağacı sınıflandırıcılarının doğruluğu yüksektir [24].



Şekil 2.4. Karar ağacı yapısı

Şekil 2.4'te teknoloji mağazası müşterisinin bilgisayar satın alma karar ağacı gösterilmiştir. Ağaç yapısında ilk bölünmenin başladığı yer köktür, verileri alt bölümlere ayıran özellikler düğümlerdir ve son kararın verildiği yer yapraklardır. Köke ve düğümlere karar vermek için çeşitli hesaplamalar yapılır. Bunlardan bazıları Eş. 2.1, 2.2 ve 2.3'te verilmiştir.

Karar ağaçlarında en çok bilinen algoritmalardan biri de Gini algoritmasıdır. Gini, Eş. 2.5 eşitliği ile ifade edilir.

$$Gini = 1 - \sum_{j} p^2 j \tag{2.5}$$

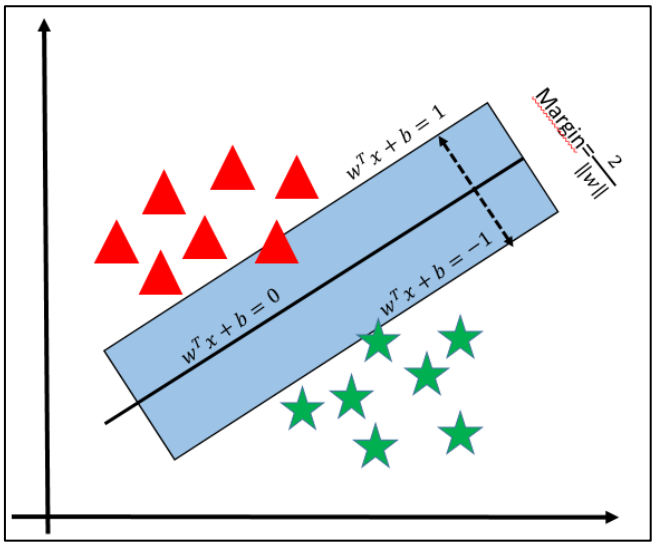
Pj, j sınıfının gerçekleşme olasılığıdır her sınıf için hesaplanır. Her bir sınıfın olasılıklarının karelerinin toplamının birden çıkarılmasıyla hesaplanır. Gini değeri 0'a ne kadar yakınsa o kadar iyi ayrım yapar.

Destek vektör makinesi

Destek vektör makineleri (SVM), Boser, Guyon ve Vapnik [35] tarafından geliştirilmiştir ve destek vektör ağları olarak da bilinirler. Regresyon analizi ve sınıflandırma işlemlerinde kullanılan, kalıpları tanıyan denetimli öğrenme yöntemidir. Performansı artırmak için çekirdek fonksiyonları kullanılır.

Temel destek vektör makinesi, verileri çok boyutlu bir girdi alanına haritalayarak ve alanı homojen bölümlere ayıran optimal bir hiperdüzlem oluşturarak örüntü tanıma problemlerini çözmektedir [36]. Yeni örneklerin tahminleri daha sonra bu bölümlere göre sınıflandırılır.

Şekil 2.5.'te gösterildiği gibi, bir düzlem üzerine yerleştirilmiş noktaları ayırmak için bir doğru çizer. Bu doğrunun, iki sınıfının noktaları için de maksimum uzaklıkta olmasını amaçlar. Ayrım için belirlenmesi gereken karar sınırı (decision boundary) sınıflar arasındaki en iyi boşluğu (margin) bularak çizgiyi (hyperplane | hiper düzlem) çeker. Karar sınırına en yakın noktalara veya değerlere destek (support) denir [37].



Şekil 2.5. Destek vektör makinesi

w ağırlık vektörü, x girdi vektörü, b bias olmak üzere iki sınıf hiperdüzleme pararlel iki margin ile ayrılır. $\|w\|$ w'nun oklid normu olarak ifade edilir.

$$Margin = \frac{2}{\|w\|} \tag{2.6}$$

Hiperdüzlem;

$$w^T x + b = 0 \tag{2.7}$$
$$w^T x + b = 1 \tag{2.8}$$
$$w^T x + b = -1 \tag{2.9}$$

Yeni bir değer için sonuç 0'dan küçükse Şekil 2.5.'te gösterilen yeşil yıldızlara yakın, 0'a eşit veya büyükse kırmızı üçgenlere yakın olacaktır.

Marginin optimize edilmesi için hesaplamalar lagrange fonksiyonu ile yapılmaktadır, Eş. 2.10 ile gösterilmiştir. α lagrange çarpanı, A aktif kısıtlar kümesini temsil eder.

$$\mathcal{L}(w, b; \alpha) = \frac{1}{2} w^T w - \sum_{i \epsilon A} \alpha_i \left(y_i (w^T x_i + b) - 1\right) \tag{2.10}$$

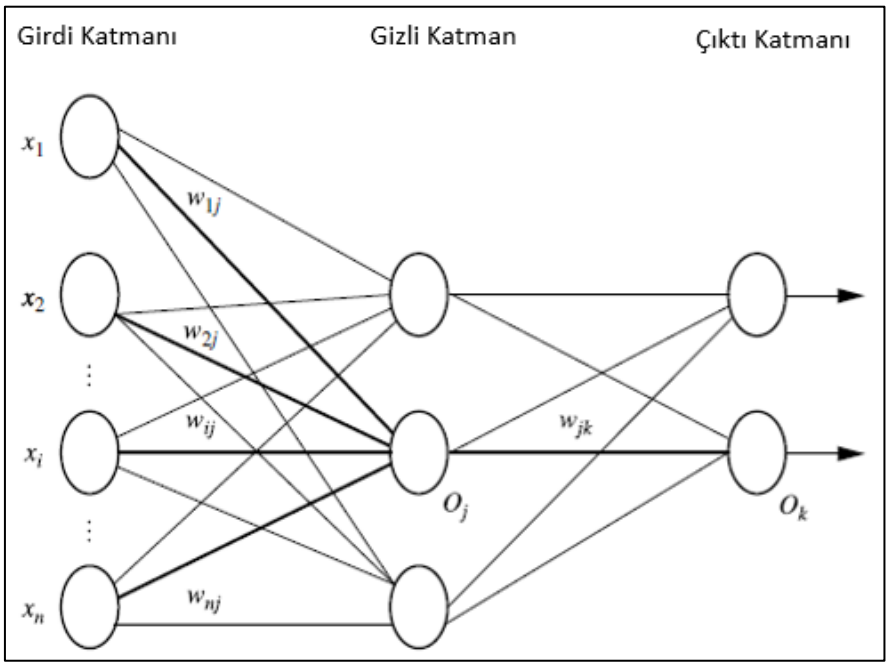
Doğrusal olmayan veriler düz bir hiperdüzlem ile birbirinden ayrılmayabilir. Bu durumda çekirdek fonksiyonları kullanılır ve en uygun ayırma düzlemine bu çekirdek fonksiyonları ile ulaşılır [38]. Literatürde kullanılan çekirdek fonksiyonları; doğrusal, polinom, gauss (RBF), laplace, anova, sigmoid, besseldir.

Yapay sinir ağları

Yapay sinir ağları, anlamlı veri temsilleri oluşturmak için bir öğrenme süreci gerçekleştiren çok sayıda ağırlıklı ara bağlantıya sahip süreç katmanlarında yer alan süreç birimlerinden (yani nöronlardan) oluşan yüksek derecede paralel bilgisayar sistemleridir [39].

Yapay sinir ağları (YSA), kayıp tahmini sorunu gibi karmaşık sorunları çözmek için popüler bir yaklaşımdır. Sinir ağları donanım tabanlı (nöronlar fiziksel bileşenlerle temsil edilir)

veya yazılım tabanlı (bilgisayar modelleri) olabilir ve çeşitli topolojiler ve öğrenme algoritmaları kullanabilir. Şekil 2.6. ‘da gösterildiği gibi temel yapay sinir ağı üç katmandan oluşur.



Şekil 2.6. Yapay sinir ağı mimarisi

Her katman birimlerden oluşur. Ağın girdileri, her eğitim grubu için ölçülen niteliklere karşılık gelir. Girişler, giriş katmanını oluşturan birimleri aynı anda besler. Bu girdiler girdi katmanından geçer ve daha sonra ağırlıklandırılır ve aynı anda gizli katman olarak bilinen "nöron benzeri" birimlerden oluşan ikinci bir katmanı besler. Gizli katman birimlerinin çıktıları başka bir gizli katmana girdi olarak verilebilir ve bu işlem tekrar edilir. Gizli katmanların sayısı isteğe bağlıdır, ancak pratikte genellikle yalnızca bir tanesi kullanılır. Son gizli katmanın ağırlıklı çıktıları, belirli veri grupları için ağın redüksiyonunu yayan çıktı katmanını oluşturan birimlere girdidir [24].

Naive bayes

Bayes sınıflandırıcısı, güçlü (naive) bağımsızlık varsayımıyla bayes teoremini uygulayan basit bir olasılıksal sınıflandırıcıdır. Bu sınıflandırıcı, bir sınıftaki belirli bir özelliğin başka bir özellik ile doğrudan ilişkili olmadığını varsayar, ancak o sınıfın özellikleri kendi aralarında bağımlılık yaptırabilir [32]. Naive bayes sınıflandırıcısı, uygulanabilirliğinin

kolay ve sonuca erişiminin hızlı olması nedeniyle birçok uygulamada kullanılmaktadır. Bayes sınıflandırıcıları büyük veritabanlarına uygulandığında yüksek doğruluk ve hız da sergilemektedir.

X bir veri grubudur. Bayes teoreminde *X,* bulgu olarak kabul edilir ve bir dizi *n* nitelik üzerinde yapılan ölçümlerle tanımlanır. *H* bir hipotez olmak üzere *X* veri grubu belirli bir *C* sınıfına aittir. $P(H|X)$, gözlemlenen *X* veri grubu göz önüne alındığında *H* hipotezinin geçerli olma olasılığıdır. Başka bir deyişle, *X*'in öznitelik tanımını bildiğimiz göz önüne alındığında, *X* grubunun *C* sınıfına ait olma olasılığı bulunur. Eş. 2.11 bayes teoremi formülüdür.

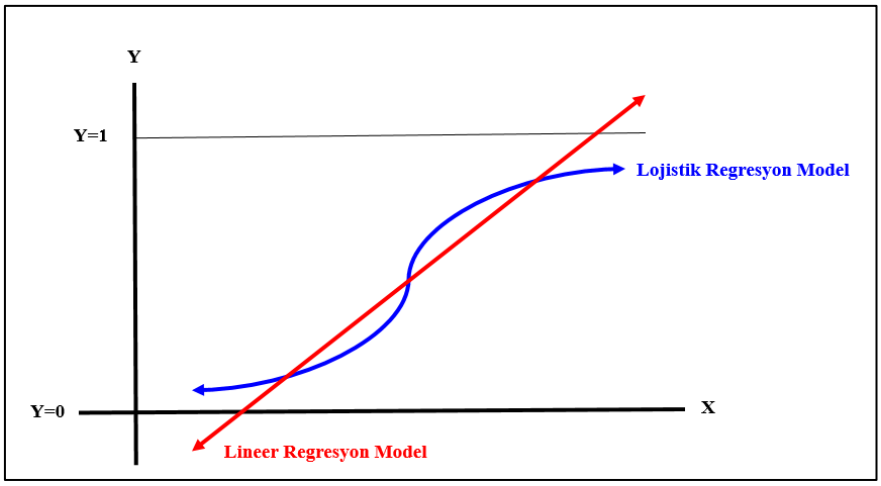
$$P(H|X) = \frac{P(X|H)\,P(H)}{P(X)} \tag{2.11}$$

Eşitlikteki; P(H) verisi problem durumunun girdi, P(X) verisi ise çıktı olasılığını, $P(H|X)$ verisi önceden gerçekleşmiş olan X verisine karşılık H çıktısının olma olasılığını, $P(X|H)$verisi yine önceden gerçekleşmiş H girdisine karşılık X çıktısının olma olasılığı durumunu ifade etmektedir.

Regresyon analizi

Regresyon analizi, aralarında sebep-sonuç ilişkisi bulunan değişkenler arasındaki ilişkiyi ortaya çıkarmayı ve bu ilişkiyi kullanarak tahminler yapmayı sağlar.

Regresyon, iki veya da daha fazla değişken arasındaki ilişkiyi tahmin etmeye yönelik istatistiksel bir yaklaşımdır. Bir bağımlı değişken ile bir veya daha fazla bağımsız değişken arasındaki ilişkiye odaklanıldığında, çeşitli değişkenlerin modellenmesi ve analiz edilmesi için birçok teknik içerir. Müşteri kaybı açısından regresyon analizi yaygın olarak kullanılmamaktadır ve bunun nedeni doğrusal regresyon modellerinin sürekli değerlerin tahmininde faydalı olmasıdır. Lojistik regresyon veya logit regresyon analizi (LR), olasılıksal istatistiksel sınıflandırma modelinin bir türüdür. Şekil 2.7. ‘de lineer ve lojistik regresyon modelleri gösterilmiştir.



Şekil 2.7. Regresyon modelleri

Eş. 2.12'de gösterilen lineer regresyon eşitliğinde *Y* tahmin edilen değerdir. Gerçek y değeri ile arasındaki fark ne kadar azsa tahmin o kadar gerçeğe yakındır. *x* bağımsız değişkendir. *a; x*=0 iken *y'*nin değeri, *b* çizilen doğrunun eğimidir. ε tesadüfi hata terimini ifade eder.

$$Y = a + bx + \varepsilon \tag{2.12}$$

Eş. 2.13 logistik regresyon formülüdür.

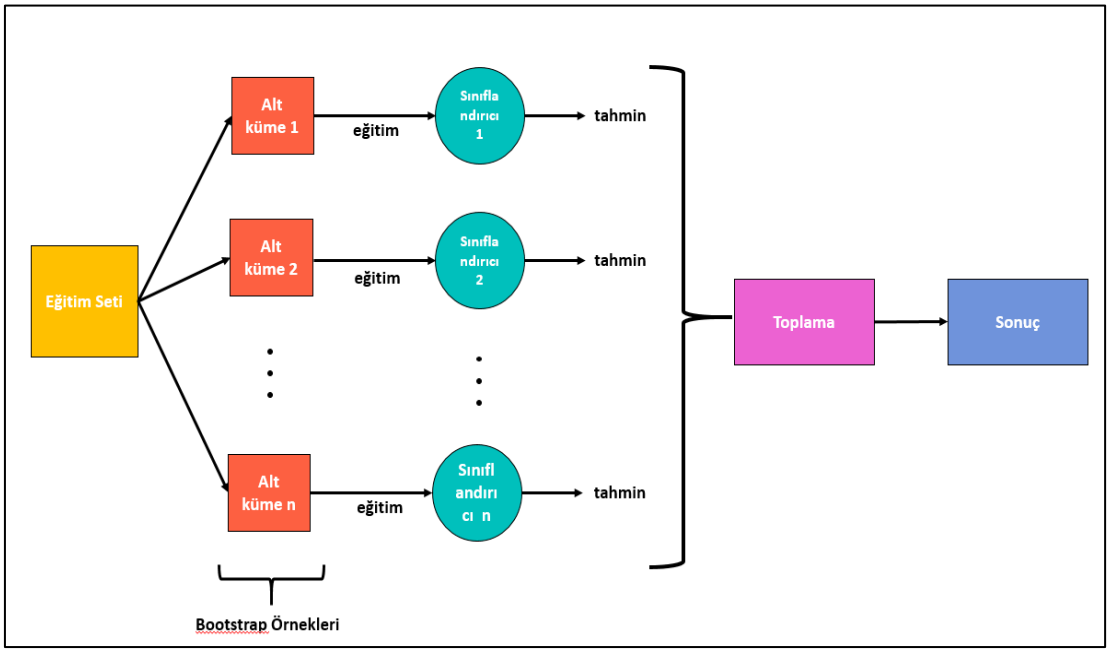
$$p = \frac{1}{1 + e^{-(b_0 + b_1 x)}} \tag{2.13}$$

Topluluk öğrenimi

Topluluk öğrenme, bir dizi temel sınıflandırıcı tarafından yapılan tahminleri birleştirir. Çoklu makine öğrenimi algoritmalarının kullanılması, bir sorun için bir dizi hipotez oluşturulmasını sağlar. Sınıflandırıcılar topluluğu, ortak bir sonuç üretmek için hipotezleri birleştirir [25]. Örneğin; rastgele orman, birden çok karar ağacının tahminlerini birleştiren topluluk sınıflandırıcılarından biridir. Topluluk öğrenmesinin popüler yöntemlerden bazıları torbalama, yükseltme, çoğunluk oylama ve istiflemedir.

Bagging - Bootstrap (torbalama)

Torbalama, yüksek varyansa sahip makine öğrenme algoritmalarının varyansını azaltmak için kullanılır. Torbalama, bootstrap örneklerini seçerek aynı eğitim veri kümesinin farklı örneklerini oluşturur. Aynı zamanda bootstrap, doğruluk tahmini için kullanılan yeniden örnekleme yöntemi olarak ifade edilir, küçük örneklem büyüklüğünü ele alır [40]. Bu yöntemde birçok eğitim verisi, ana veri setinden rastgele çekilir. "n" örnekli bir veri setinde "n" örneklemin rastgele seçilmesiyle bir bootstrap eğitim veri seti oluşturulur. Her defasında bir örnek seçilir, çekilen örnek eğitim setine tekrar eklenir. Böylelikle bir eğitim setinde aynı örneğin birden fazla sayıda seçilme olasılığı olduğu gibi aynı zamanda örneğin hiç çekilmemesi de olasıdır. Bu durumda her bir örnek 1/n olasılıkla seçilir. Şekil 2.8.'de bagging algoritmasının süreçleri gösterilmiştir.



Şekil 2.8. Bagging algoritmasının süreçleri

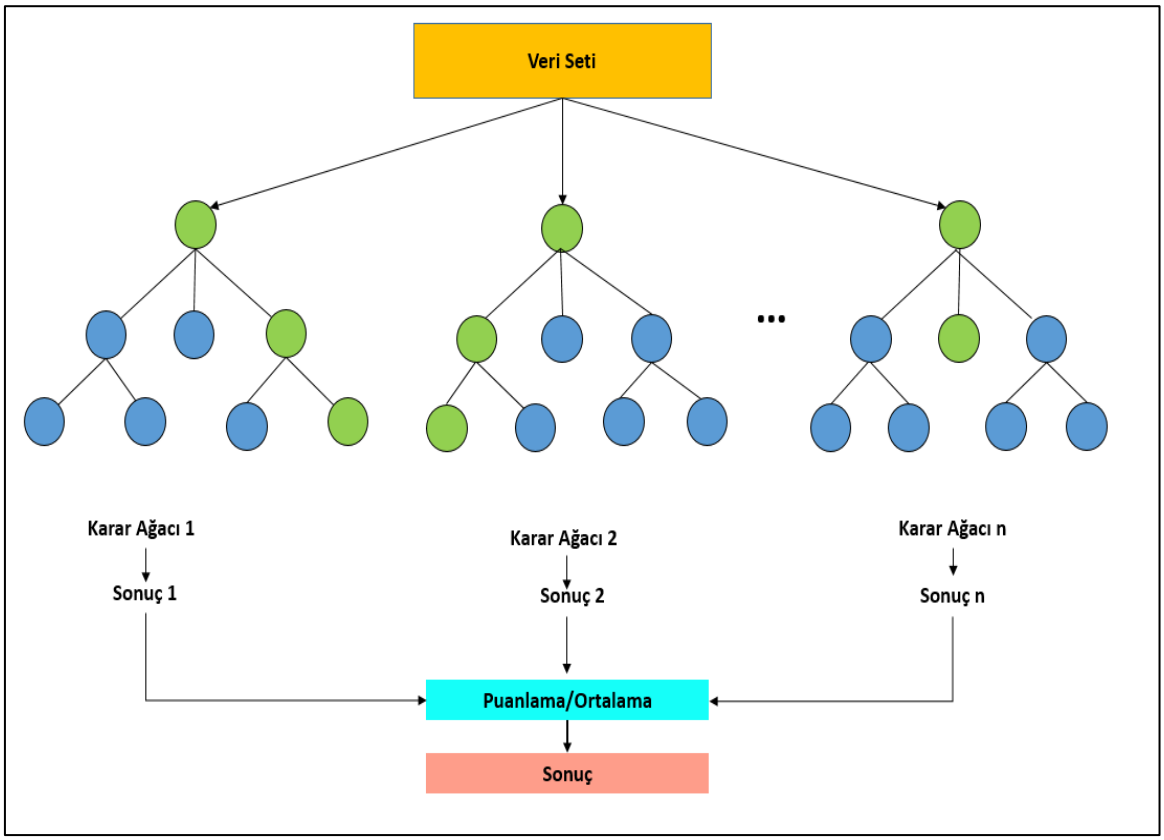
Boosting (yükseltme)

Yükseltmenin temel amacı, zayıf sınıflandırıcılar (veya zayıf öğrenme algoritmaları) olarak adlandırılan birçok sınıflandırma modelinden alınan kararların birleştirilmesi yoluyla sınıflandırma performansını iyileştirmektir. Boosting müşteri kaybının tahmin edilmesinde başarıyla uygulanmıştır. En çok kullanılan boosting yöntemi AdaBoost'tur.

Veri kümesi üzerinde eşit ağırlıklı örneklerle başlayarak, sınıflandırıcı eğitilir. Sınıflandırıcının performansı değerlendirilir ve yanlış sınıflandırılan örneklerin ağırlıkları artırılır. Bu, örnekleri doğru sınıflandırmaya yönlendirir. Diğer bir sınıflandırıcı eğitilir ve örneklerin ağırlıkları yeniden ayarlanır. Bu süreç, belirlenen sayıda sınıflandırıcı eğitilene kadar devam eder. Son adımda, her sınıflandırıcının ağırlığı, performansına göre hesaplanır ve bu ağırlıklar kullanılarak sınıflandırıcılar bir araya getirilir [41].

Random forest (rastgele orman)

Topluluktaki sınıflandırıcıların her biri bir karar ağacı sınıflandırıcısıdır ve sınıflandırıcılar topluluğu bir ormandır. Karar ağaçları, bölünmeyi belirlemek için her düğümdeki niteliklerin rastgele seçimi kullanılarak oluşturulur. Her ağaç, ormandaki tüm ağaçlar için bağımsız olarak ve aynı dağılımla örneklenen rastgele bir vektörün değerlerine bağlıdır. Sınıflandırma sırasında her ağaç oy verir ve en popüler sınıf döndürülür [24]. Şekil 2.9'da random forest algoritması gösterilmiştir.

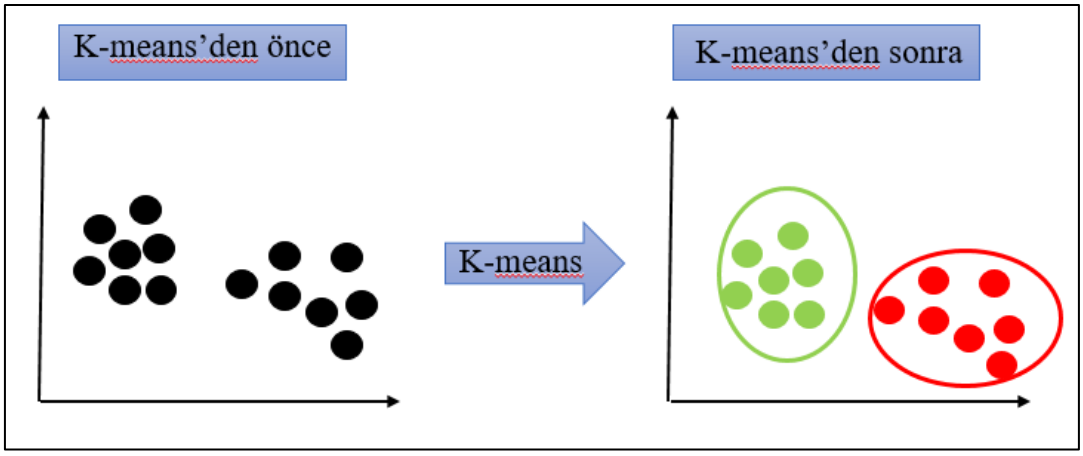


Şekil 2.9. Random forest algoritması süreçleri

2.3.2. Kümeleme

Kümeleme, ortak özellikler paylaşan veri noktalarını alt kümelere ayırarak bir veri kümesi içindeki benzerlikleri belirlemeyi amaçlar. Kümeleme yöntemlerinin çoğunda veri setindeki veriler arası uzaklık hesaplamaları kullanılır. Uzaklık hesaplamaları öklid, manhattan, minkowski yöntemlerinden biri ile yapılabilir. En popüler kümeleme algoritması k-ortalamalardır.

Şekil 2.10.'da k-ortalamalar (k-means) uygulanan veri setinin kümeleme sonucu gösterilmiştir. K-ortalamalarda, k değerindeki küme sayısı baştan belirlenir. Algoritmada rastgele k tane merkez noktası seçilir. Her veri ile rastgele belirlenen merkez noktaları arasındaki uzaklık hesaplanarak veri en yakın merkez noktasına göre bir kümeye atanır. Daha sonra her küme için yeniden bir merkez noktası seçilir ve yeni merkez noktalarına göre kümeleme işlemi yapılır. Bu durum sistem kararlı hale gelene kadar devam eder [42].



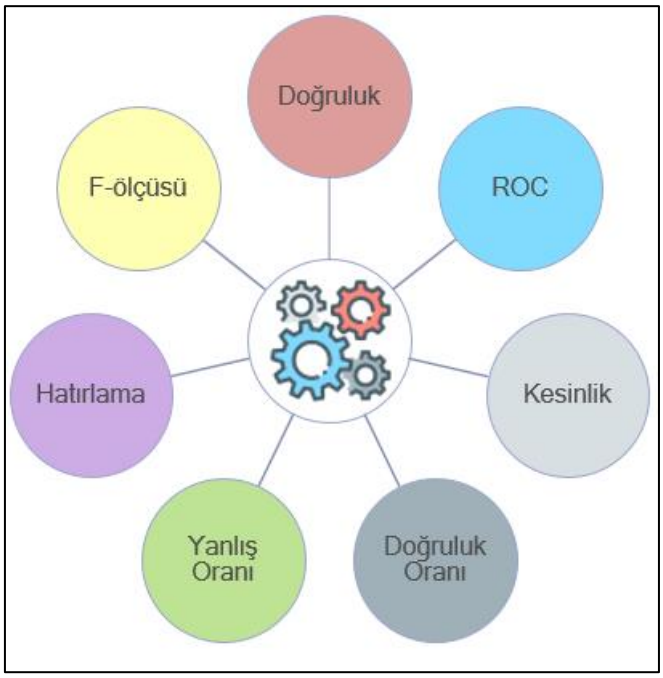
Şekil 2.10. K-means algoritması

2.3.3. Birliktelik kuralları

Birliktelik kuralları, birlikte meydana gelen olayları bulmak için büyük veri setlerindeki değişkenler arasındaki ilişkileri arar. Destek ve güven eşik değerlerinin yardımıyla veri kümesi içerisindeki ilişkiler bulunur. Pazar araştırmacıları arasında popülerdir. En fazla kullanılan yöntem apriori algoritmasıdır.

Apriori algoritması, büyük öğe kümelerini keşfetmek için veri üzerinde birden fazla kere geçiş yapar. İlk geçişte, tek tek öğelerin desteğini sayar ve hangilerinin büyük olduğunu, yani minimum desteğe sahip olduğunu belir. Sonraki her geçişte, bir önceki geçişte büyük olduğu tespit edilen öğe kümelerinden oluşan bir küme ile başlar. Bu kümeye aday öğe kümesi denir. Geçişin sonucunda hangi aday öğe kümesinin gerçekten büyük olduğunu belirler ve sonraki geçiş için başlangıç olur. Bu işleme yeni bir büyük öğe kümesi bulunmayana kadar devam eder. Daha sonra güven değeri hesaplanarak, verilen güven parametre değerinin üzerinde olan öğeler ile kurallar oluşturulur [43].

2.4. Sınıflandırma Modellerinin Performans Değerlendirmesi



Şekil 2.11. Sınıflandırma değerlendirme metrikleri

Modellerin performansları karışıklık matrisi çıkarıldıktan sonra, farklı ölçütler ile kıyaslanır. Bu metrikler başarı oranı(A), haturlama(R), kesinlik (P), özgüllük(S), F ölçüsü (F), AUC-ROC eğrileridir. Başarı oranı (A): Toplam veri içindeki doğru tahmin oranıdır, duyarlılık (R): Pozitif olarak tahmin edilmesi gereken örneklerin ne kadar başarılı tahmin edildiğinin göstergesidir, kesinlik (P): Pozitif olarak tahmin edilen örneklerin kaçının doğru tahmin edildiğinin oranıdır, F-ölçüsü (F): kesinlik ve hatırlamanın harmonik ortalamasıdır [32].

Çizelge 2.1. Karışıklık matrisi

		TAHMİN DEĞERLERİ	
		Pozitif	Negatif
GERÇEK DEĞERLER	Pozitif	TP	FN
	Negatif	FP	TN

Sınıflandırıcılar için TP oranı yüksek olmalıdır.

Hatırlama (Recall): Gerçek doğruların oranıdır (gerçek pozitif) ve Eş. 2.14 gibi hesaplanır:

$$Recall = \frac{Tp}{Tp + Fn} \tag{2.14}$$

Kesinlik (Precision): Pozitif sınıfın doğru tahmin edilme oranıdır ve Eş. 2.15 gibi hesaplanır:

$$Precision = \frac{Tp}{Tp + Fp} \tag{2.15}$$

Doğruluk (Accuracy): Tüm doğru tahminlerin oranıdır ve Eş. 2.16 gibi hesaplanır:

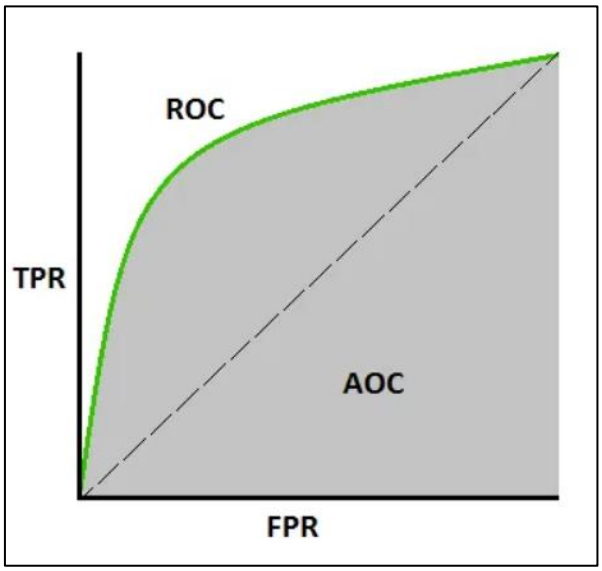
$$Accuracy = \frac{(Tp + Tn)}{(Tp + Fp + Tn + Fn)} \tag{2.16}$$

F-ölçüsü (F-measure): Kesinlik ve hatırlamanın harmonik ortalamasıdır ve Eş. 2.17 gibi hesaplanır:

$$F - measure = \frac{(2\ x\ Precision\ x\ Recall)}{(Precision + Recall)} \tag{2.17}$$

ROC eğrisi (Receiver operating characteristics): ROC eğrisi, FP ve FN arasındaki tüm olası maliyet oranlarına karşı ortalama performansı belirtir. ROC eğrisi değeri 1.0'e eşitse, bu mükemmel bir tahmindir. Şekil 2.12.'de ROC eğrisi gösterilmiştir [44]. Gerçek pozitif oranı (TPR) =Recall Eş. 2.14'te gösterilmiştir. Hatalı pozitif oranı (FPR) Eş. 2.18'deki gibi hesaplanır.

$$FPR = \frac{FP}{FP + TN} \tag{2.18}$$



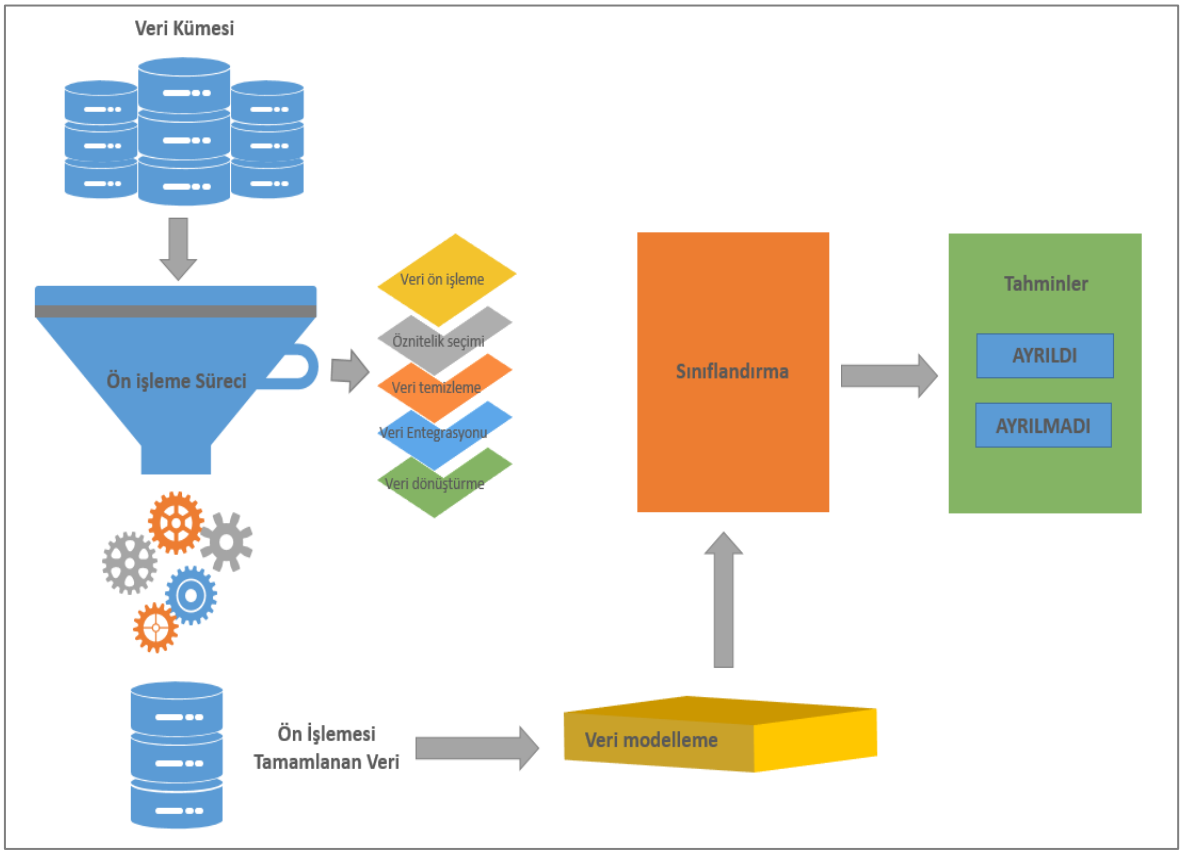
Şekil 2.12. AUC – ROC Eğrisi

AUC alanı (Area under the curve): Test setinin pozitif ve negatif sınıflarındaki model performansını ölçmek için AUC alanı sıklıkla kullanılmaktadır. AUC skorunun değeri ne kadar yüksekse, model hem pozitif hem de negatif sınıflarda o kadar iyi performans gösterir. Şekil 2.12.'de AUC alanı gösterilmiştir.

3. DENEYSEL ÇALIŞMA VE SONUÇLARI

Ayrılan öğrenci tahmin modeli; öğrenci kaybını öngörmeyi, uzun süreli barınacak öğrencileri elde tutmayı amaçlamaktadır. Öğrenci kaybının önlenmesi için uygun önlemlerin alınmasını sağlamaktır. Tez kapsamında gerçekleştirilen deneysel çalışmalarda yurt öğrenci veri kümesi ve öncelikli öğrenci veri kümesi olmak üzere iki veri kümesi kullanılmıştır.

Öğrencilerin barınma süreleri içinde ayrılıp ayrılmayacaklarının tahmini ikili bir sınıflandırma problemi olarak değerlendirilmiştir. Bu sorunun üstesinden gelebilmek için; daha önce yapılan çalışmalara dayanarak oldukça etkili teknikler olan; naive bayes, doğrusal regresyon, destek vektör makinesi, karar ağacı, rastgele ormanı vb. içeren makine öğrenimi algoritmaları [18, 21, 33]'de kullanılmıştır. Makine öğrenimi modellerinde, ön işlemede öznitelik seçiminin, sınıflandırma doğruluğunun iyileştirilmesinde önemli bir rol oynadığı görülmüştür.



Şekil 3.1. Deneysel çalışma metodolojisi

Şekil 3.1. tahmin modelini göstermekte ve adımlarını açıklamaktadır. İlk adımda; gürültü giderme için veri filtrelemeyi, dengesiz veri özelliklerinin kaldırılmasını ve verilerin normalleştirilmesini içeren veri ön işlemesi yapılır. Önemli özellikler, bilgi kazancı ve korelasyon tabanlı özellik seçimi kullanılarak verilerden çıkarılır. İkinci adımda öğrencileri; sınıflandırmak için farklı sınıflandırma algoritmaları uygulanır. Sınıflandırma algoritmaları arasında J48, rastgele orman (RF), karar kütüğü, karar tablosu, adaboost, logitboost, torbalama, naive bayes (NB), çok katmanlı algılayıcı (MLP), lojistik regresyon (LR) , SVM ve KNN bulunur. Son olarak öğrencilerin yurttan ayrılıp ayrılmayacağı tahmin edilir.

Deneysel çalışmalar, Weka 3.8.6 versiyonu ile yapılmıştır. Weka, Waikato Üniversitesi tarafından geliştirilmiştir. "Waikato Environment for Knowledge Analysis" sözcüklerinin baş harflerinden oluşmuş programın adıdır. Weka, veri madenciliği için makine öğrenimi algoritmaları koleksiyonudur. Veri hazırlama, kümeleme, regresyon, sınıflandırma, birliktelik kuralları ve görselleştirme için araçlar içerir [45].

Tez çalışmasındaki deneysel çalışmalar iki farklı bilgisayarla yapılmıştır, bilgisayarların özellikleri şu şekildedir:

Workstation - i9 CPU-2.80 GHz, 32 GB RAM, 64 bit işletim sistemi.
Notebook - i7 CPU- 2.80GHz, 16 GB RAM, 64 bit işletim sistemi.

3.1. Deneysel Çalışmada Kullanılan Veri Kümeleri

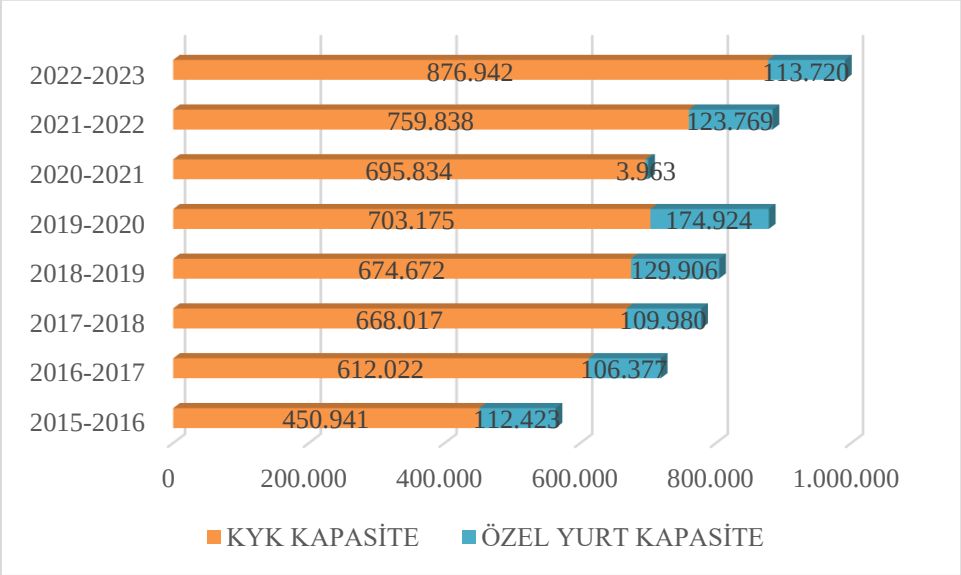
Gençlik ve Spor Bakanlığı Yurt Hizmetleri yönetmeliğinde belirtildiği gibi öğrencilerin yurtta barınması için;

- “Örgün eğitim veren bir yükseköğretim kurumuna kayıtlı olması.”
- “Öğrenim gördüğü yükseköğretim kurumu veya programının, öğrenci ve ailesinin ikamet ettiği il veya ilçe sınırları dışında bulunması” gerekir [46].

Çizelge 3.1. ve Şekil 3.2.’de dönem bazlı olarak KYGM ve özel yurtların kapasiteleri verilmiştir [4]. Öğrenciler öğrenim gördükleri il ve ilçedeki yurt kapasitesine göre yurt tercihlerini belirlerler.

Çizelge 3.1. KYGM-Özel yurtların dönem bazlı kapasiteleri.

DÖNEM	KYGM KAPASİTE	ÖZEL YURT KAPASİTE
2015-2016	450 941	112 423
2016-2017	612 022	106 377
2017-2018	668 017	109 980
2018-2019	674 672	129 906
2019-2020	703 175	174 924
2020-2021	695 834	3 963
2021-2022	759 838	123 769
2022-2023	876 942	113 720

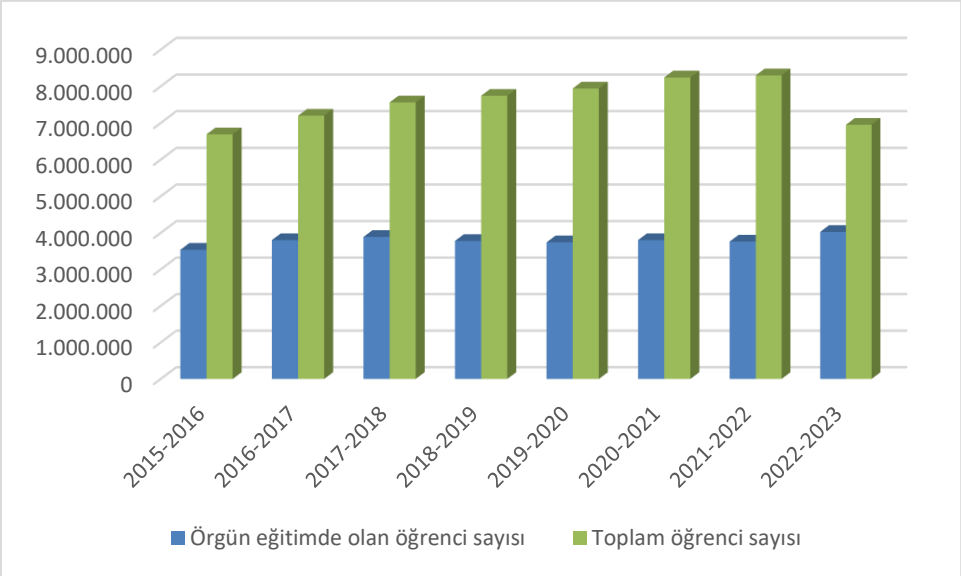


Şekil 3.2. KYGM-Özel yurtların dönem bazlı kapasiteleri.

Ülkemizde genç nüfusun fazla olması yükseköğretimdeki öğrenci sayısının yüksek olmasına yol açmaktadır. Çizelge 3.2. ve Şekil 3.3.'te yükseköğretimde dönem bazlı toplam öğrenci sayısı ve örgün eğitimde öğrenim gören öğrenci sayısı gösterilmiştir [1].

Çizelge 3.2. Dönem bazlı toplam ve örgün eğitimde öğrenim gören öğrenci sayısı

DÖNEM	ÖRGÜN EĞİTİMDE OLAN ÖĞRENCİ SAYISI	TOPLAM ÖĞRENCİ SAYISI
2015-2016	3 536 787	6 689 185
2016-2017	3 800 310	7 198 987
2017-2018	3 887 682	7 560 371
2018-2019	3 777 114	7 740 502
2019-2020	3 740 332	7 940 133
2020-2021	3 801 294	8 240 997
2021-2022	3 761 637	8 296 959
2022-2023	4 024 506	6 950 142

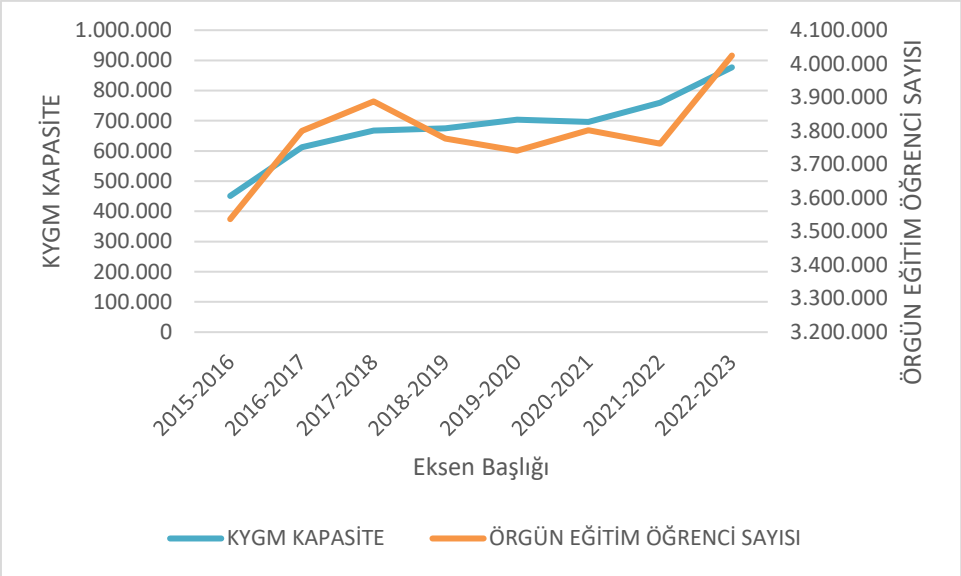


Şekil 3.3. Dönem bazlı toplam ve örgün eğitimde öğrenim gören öğrenci sayısı

Teze konu olan KYGM'de barınan yükseköğretim öğrencileri örgün eğitimde öğrenim görmeli ve öğrenim görülen bölüm ikamet il ve ilçesi dışında olmalıdır [46]. Çizelge 3.3. ve Şekil 3.4.'te yükseköğretimde örgün eğitim gören öğrenciler ve KYGM kapasitesinin artışı ve mukayesesi gösterilmiştir.

Çizelge 3.3 Dönem bazlı örgün eğitimde olan öğrenci sayısı ve KYGM kapasitesi

DÖNEM	ÖRGÜN EĞİTİM ÖĞRENCİ SAYISI	KYGM KAPASİTE
2015-2016	3 536 787	450 941
2016-2017	3 800 310	612022
2017-2018	3 887 682	668 017
2018-2019	3 777 114	674 672
2019-2020	3 740 332	703 175
2020-2021	3 801 294	695 834
2021-2022	3 761 637	759 838
2022-2023	4 024 506	876 942



Şekil 3.4. Dönem bazlı örgün eğitimde olan öğrenci sayısı ve KYGM kapasitesi

Yurt başvuruları; öğrenci ve ailesinin sosyal, ekonomik durumu, öğrencinin başarısı ve benzeri durumlar dikkate alınarak değerlendirilir. Veri kümesinde bu kıstaslar da kullanılmıştır. Çizelge 3.4. ve Şekil 3.5. 'de KYGM yurtlarında barınan öncelikli öğrenci sayısı 31.12.2022 tarihi itibariyle gösterilmiştir.

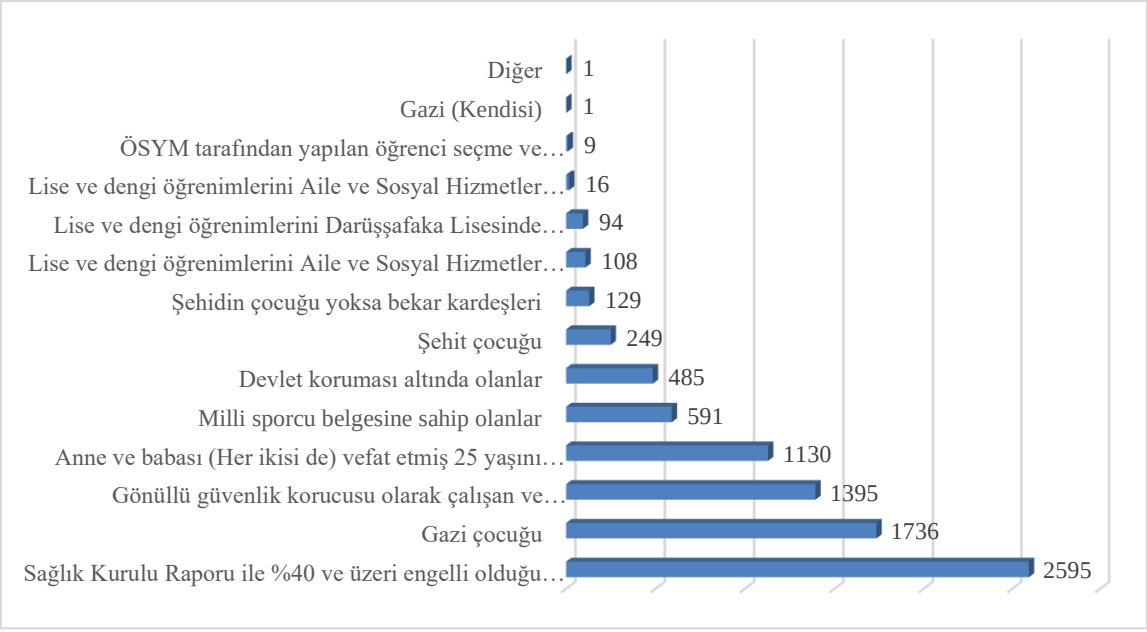
KYGM mevzuatına göre yurda öncelikli olarak yerleştirilen öğrenci kriterleri:

- "Şehit eşi ve çocukları, şehidin çocuğu yoksa bekar kardeşleri.
- Gazi ve gazi çocukları.
- %40 ve üzerinde engel durumu olanlar.
- Anne ve babası vefat etmiş olan yirmi beş yaşını geçmemiş öğrenciler.
- Lise ve dengi öğrenimlerini Aile ve Sosyal Hizmetler Bakanlığı bünyesinde barınarak tamamlayanlar.
- Devlet koruması altında olanlar.
- Lise ve dengi öğrenimlerini Darüşşafaka Lisesinde tamamlayanlar.
- Bakanlık tarafından yurtlarda barınmada öncelikli olacak milli sporcu belgesine sahip sporculara ilişkin belirlenen kriterleri taşıyanlar.
- Gönüllü güvenlik korucusu olarak çalışan, çalışırken vefat eden veya emekliye ayrılanların çocukları.

- Her öğretim yılında Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı tarafından yapılan öğrenci seçme ve yerleştirme sınavlarında her puan türünde, ham puan bazında ilk yüze giren öğrenciler.
- Bakanlık tarafından uygun görülenler ile ihtiyaç duyulması veya mücbir sebeplerin ortaya çıkması halinde öncelikli olarak yurtlarda barındırılacak diğer öğrenciler" [46].

Çizelge 3.4. KYGM yurtlarında barınan öncelikli öğrenci sayısı

ÖNCELİK DURUMU	ÖĞRENCİ SAYISI
Sağlık Kurulu Raporu ile %40 ve üzeri engelli olduğu tespit edilenler	2 595
Gazi çocuğu	1 736
Gönüllü güvenlik korucusu olarak çalışan ve çalışırken vefat edenlerin, gönüllü güvenlik korucusu olarak emekliye ayrılanların çocukları	1 395
Anne ve babası (Her ikisi de) vefat etmiş 25 yaşını geçmemiş olanlar	1 130
Milli sporcu belgesine sahip olanlar	591
Devlet koruması altında olanlar	485
Şehit çocuğu	249
Şehidin çocuğu yoksa bekar kardeşleri	129
Lise ve dengi öğrenimlerini Aile ve Sosyal Hizmetler Bakanlığı Yetiştirme Yurtlarında barınarak tamamlayanlar	108
Lise ve dengi öğrenimlerini Darüşşafaka Lisesinde tamamlayanlar	94
Lise ve dengi öğrenimlerini Aile ve Sosyal Hizmetler Bakanlığı Yetiştirme Yurtlarında barınarak tamamlayanlar ile Devlet koruması altında olanlar	16
ÖSYM tarafından yapılan öğrenci seçme ve yerleştirme sınavlarında her puan türünde ilk yüze girenler	9
Gazi (Kendisi)	1
Diğer	1



Şekil 3.5. KYGM yurtlarında barınan öncelikli öğrenci sayısı

3.1.1. Yurt öğrenci veri kümesi

Yurt öğrenci veri kümesi KYGM'nin Ankara ilinde bulunan yurtlarında 2016-2022 yıllarında barınan öğrenci verilerini içerir. Veriler,

- Örgün eğitim gören,
- TC vatandaşı olan,
- Öğrenim süresi dört yıl olan,
- Başvuru türü ÖSYM olan,
- Yurttan ayrılma nedeni, kendi isteğiyle yurttan ayrılan ve mezun olan

Öğrenciler içinden seçilmiştir. Çizelge 3.5.'te yurt öğrenci veri kümesi öznitelikleri gösterilmiştir. Yurt öğrenci veri kümesi 34 öznitelikli olup, 8783 veri örneği içermektedir. Sınıf dağılımı 4803 mezun, 3980 kendi isteğiyle yurttan ayrılan olacak şekildedir.

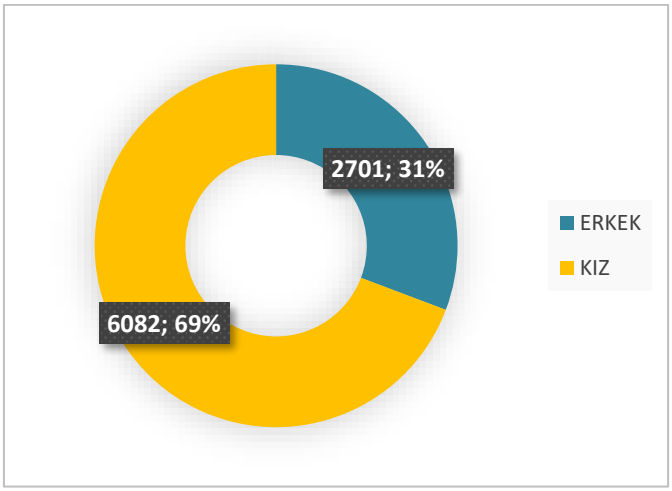
Çizelge 3.5. Yurt öğrenci veri kümesi öznitelikleri

No	Öznitelikler	No	Öznitelikler
1	CINSIYET	18	GELIR_PUAN
2	MEDENIDURUM	19	BASARI_PUAN
3	YURTTANAYRILDIGIYAS	20	OGRETIM_SEKLI
4	ADRES_IL_ID	21	PUAN_TURU
5	ENGEL_TURU_ID	22	VELAYET
6	ONCELIK_DURUM	23	BOLUM_OZEL_YETENEK
7	BASVURU_TURU	24	FAKULTE_GRUP
8	IL_DISI_UNI_KARDES	25	YURTLOKASYON
9	IL_ICI_UNI_KARDES	26	BAHCE
10	ANNE_VEFAT	27	MODELTIP
11	BABA_VEFAT	28	YATAKTIPI
12	ANNE_BABA_AYRI	29	YATAK_SAYI
13	ANNE_SAG_OZURLU_BEYAN	30	UCRET_TIPI
14	BABA_SAG_OZURLU_BEYAN	31	YURT_BOLUMUZAKLIK_GRUP
15	VAKIF_TAM_BURS_DEGIL	32	YURT_KAPASITE_GRUP_1
16	MEB_KARDES_SAYI	33	BINAKAPASITE_GRUP_1
17	SOSYAL_DURUM_PUAN	34	KALDIGI_AY_1

Yurt öğrenci veri kümesinin çeşitli özelliklerinin dağılımı Çizelge 3.6.'dan 3.14.'e ve Şekil 3.6.'dan 3.9.'a gösterilmiştir.

Çizelge 3.6. Yurt öğrenci veri kümesi cinsiyet dağılımı

CİNSİYET	ÖRNEK SAYISI
KIZ	2 701
ERKEK	6 082



Şekil 3.6. Yurt öğrenci veri kümesi cinsiyet dağılımı

Çizelge 3.7. Yurt öğrenci veri kümesi yaş dağılımı

YAŞ	ÖRNEK SAYISI
17	9
18	478
19	848
20	822
21	985
22	2 984
23	2 031
24	448
25	112
26	26
27	17
28	11
29	5
30	1
32	3
34	2
38	1

Çizelge 3.8. Yurt öğrenci veri kümesi yaş kategori dağılımı

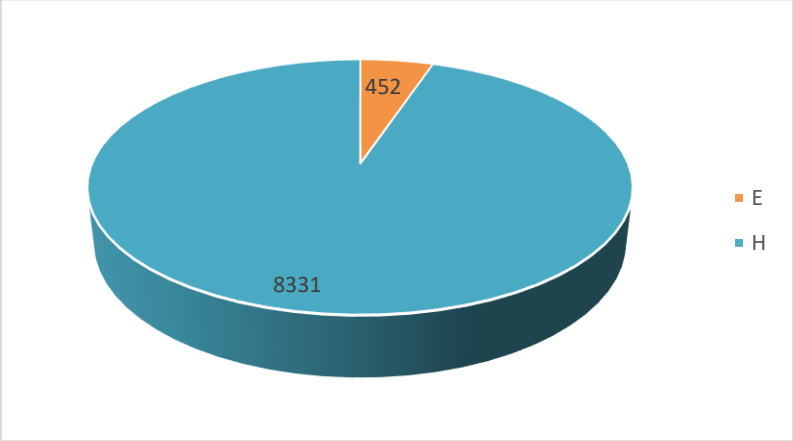
KATEGORİ	YAŞ ARALIĞI	ÖRNEK SAYI
1	17-18	487
2	19	848
3	20	822
4	21	985
5	22	2 984
6	23	2 031
7	>24	626

Çizelge 3.9. Yurt öğrenci veri kümesi anne vefat durumu dağılımı

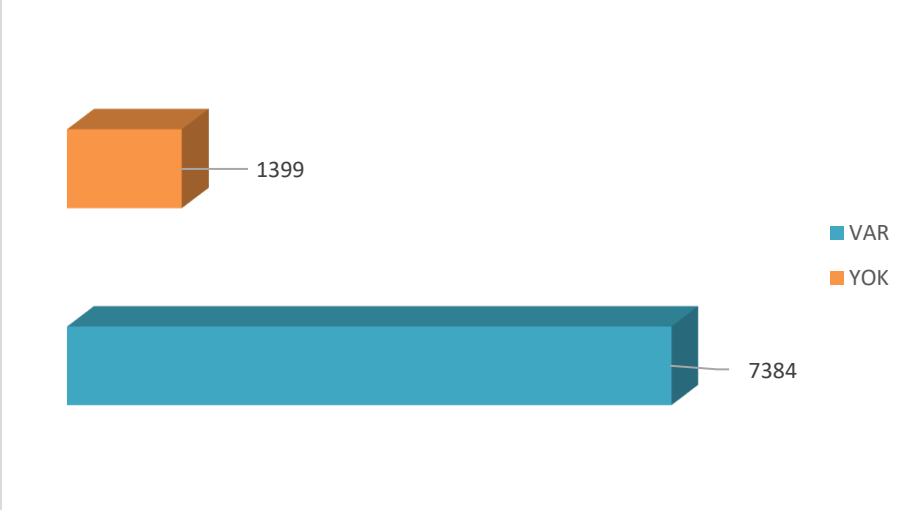
ANNE VEFAT DURUMU	ÖRNEK SAYISI
E	128
H	8 655

Çizelge 3.10. Yurt öğrenci veri kümesi vakıf üniversitesi durum dağılımı

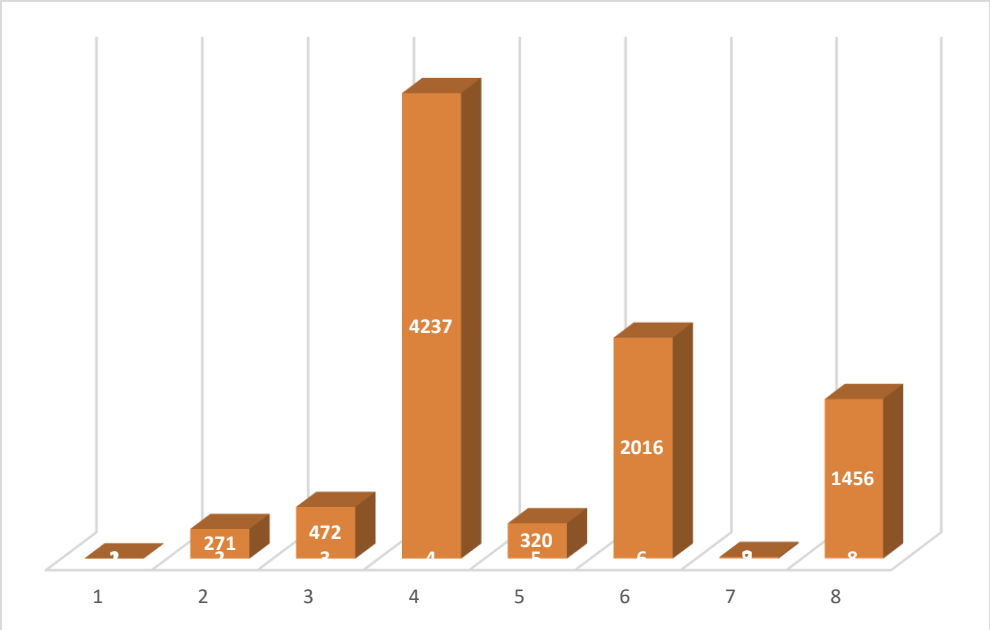
VAKIF ÜNİVERSİTESİ	ÖRNEK SAYISI
E	72
H	8 711



Şekil 3.7. Yurt öğrenci veri kümesi baba vefat durumu dağılımı



Şekil 3.8. Yurt öğrenci veri kümesi bahçe durumu dağılımı



Şekil 3.9. Yurt öğrenci veri kümesi odadaki yatak sayısı dağılımı

Çizelge 3.11. Yurt öğrenci veri kümesi MEB okuyan kardeş sayısı dağılımı

MEB KARDEŞ SAYISI	ÖRNEK SAYISI
0	3 619
1	3 288
2	1 362
3	374
4	104
5	26
6	8
7	2

Çizelge 3.12. Yurt öğrenci veri kümesi model tip dağılımı

MODEL ADI	MODEL ÖZELLİĞİ	ÖRNEK SAYISI
M1	Oda içinde banyo- wc	3 709
	Karyola/baza sisteminde	
	Oda içinde buzdolabı	
	Oda içinde çalışma masası	
M2	Oda içinde banyo-wc	952
	Ranza sisteminde	
	Katta/oda içinde buzdolabı	
M3	Katta banyo- wc	369
	Karyola sisteminde	
	Katta/oda içinde buzdolabı	
M4	Katta banyo- wc	3 378
	Ranza sisteminde	
	Katta/oda içinde buzdolabı	
M5	Tek katta ortak duş sistemi	375
	Ranza sisteminde	
	Katta/oda içinde buzdolabı	

Çizelge 3.13. Yurt öğrenci veri kümesi yurt bölüm uzaklık dağılımı

KATEGORİ	YURT BÖLÜM UZAKLIK (KM)	ÖRNEK SAYISI
1	0-3	2 552
2	4-6	2 667
3	7-10	782
4	11-15	818
5	16- 20	210
6	21- 25	1 234
7	26-30	347
8	31-40	45
9	41-50	69
10	>50	59

Çizelge 3.14. Yurt öğrenci veri kümesi bina kapasite dağılımı

KATEGORİ	KAPASİTE	ÖRNEK SAYISI
1	0-500	2 756
2	500-1000	3 967
3	1000>	2 060

3.1.2. Öncelikli öğrenci veri kümesi

Öncelikli öğrenci veri kümesi KYGM'nin yurtlarında 2016-2022 yıllarında barınan öncelikli öğrenci verilerini içerir. Veriler,

- Örgün eğitim gören,
- TC vatandaşı olan,
- Öğrenim süresi dört yıl olan,
- Başvuru türü ÖSYM olan,
- Türkiye'deki yurtlarda barınan,
- Öncelik durumu olan,
- Yurttan ayrılma nedeni kendi isteğiyle yurttan ayrılan ve mezun olan

Öğrenciler içinden seçilmiştir. Çizelge 3.15. 'te öncelikli öğrenci veri kümesi öznitelikleri gösterilmiştir. Veri kümesi 35 öznitelikli olup, 3048 veri örneği içermektedir. Sınıf dağılımı 1604 mezun, 1444 kendi isteğiyle yurttan ayrılan olacak şekildedir.

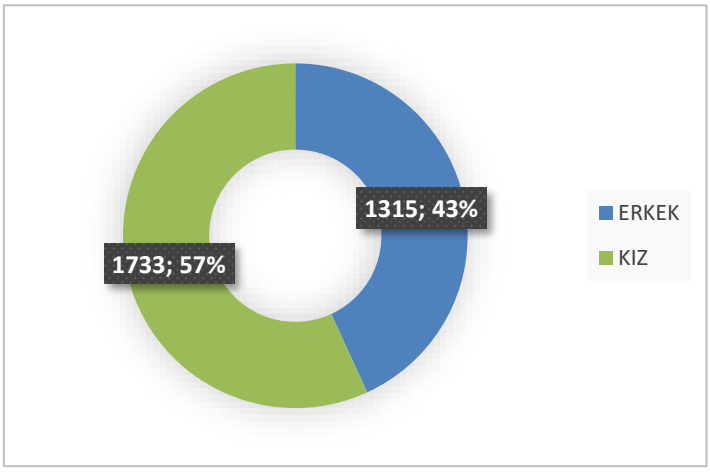
Çizelge 3.15. Öncelikli öğrenci veri kümesi öznitelikleri

No	Öznitelikler	No	Öznitelikler
1	CINSIYET	19	BASARI_PUAN
2	MEDENIDURUM	20	OGRETIM_SEKLI
3	YURTTANAYRILDIGIYAS	21	PUAN_TURU
4	ADRES_IL_ID	22	VELAYET
5	ENGEL_TURU_ID	23	BOLUM_OZEL_YETENEK
6	ONCELIK_DURUM	24	FAKULTE_GRUP
7	BASVURU_TURU	25	YURTLOKASYON
8	IL_DISI_UNI_KARDES	26	BAHCE
9	IL_ICI_UNI_KARDES	27	MODELTIP
10	ANNE_VEFAT	28	YATAKTIPI
11	BABA_VEFAT	29	YATAK_SAYI
12	ANNE_BABA_AYRI	30	UCRET_TIPI
13	ANNE_SAG_OZURLU_BEYAN	31	YURT_BOLUMUZAKLIK_GRUP
14	BABA_SAG_OZURLU_BEYAN	32	YURT_KAPASITE_GRUP_1
15	VAKIF_TAM_BURS_DEGIL	33	BINAKAPASITE_GRUP_1
16	MEB_KARDES_SAYI	34	KALDIGI_AY_1
17	SOSYAL_DURUM_PUAN	35	BARINMADURUM
18	GELIR_PUAN		

Öncelikli öğrenci veri kümesinin çeşitli özellikleri dağılımı Çizelge 3.16.'dan 3.21.'e ve Şekil 3.10.'dan 3.14.'e gösterilmiştir.

Çizelge 3.16. Öncelikli öğrenci veri kümesi cinsiyet dağılımı

CİNSİYET	ÖRNEK SAYISI
ERKEK	1 315
KIZ	1 733



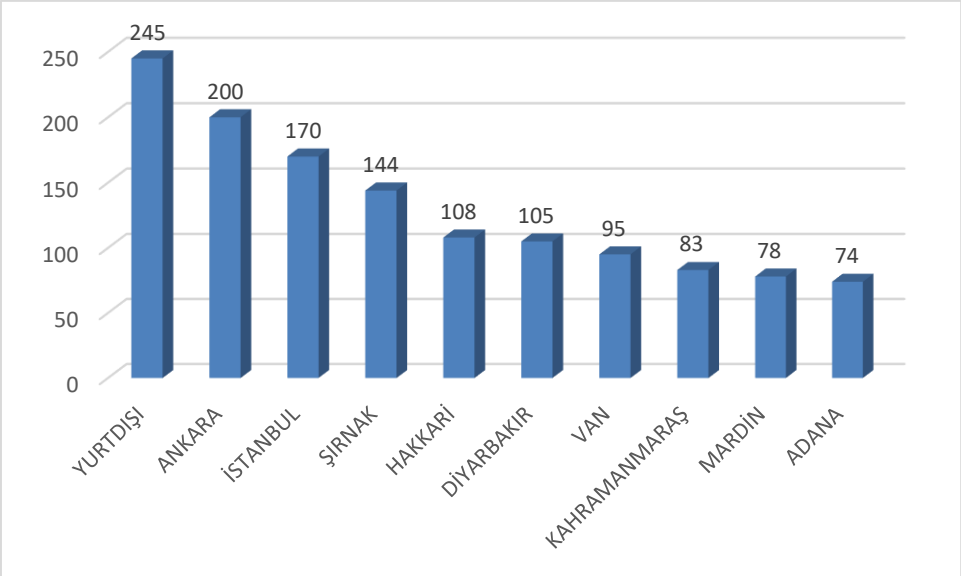
Şekil 3.10. Öncelikli öğrenci veri kümesi cinsiyet dağılımı

Çizelge 3.17. Öncelikli öğrenci veri kümesi yaş kategori dağılımı

KATEGORİ	YAŞ ARALIĞI	ÖRNEK SAYISI
1	17-18	99
2	19	230
3	20	323
4	21	365
5	22	811
6	23	717
7	>24	503

Çizelge 3.18. Öncelikli öğrenci veri kümesi yaş dağılımı

YAŞ	ÖRNEK SAYISI
17	3
18	96
19	230
20	323
21	365
22	811
23	717
24	287
25	108
26	44
27	25
28	12
29	9
30	3
31	2
32	1
33	2
34	4
35	1
38	1
39	2
52	1
55	1



Şekil 3.11. Öncelikli öğrenci veri kümesi adres ili dağılımı

Şekil 3.11.'de örnek sayısı en fazla olan 10 il gösterilmiştir.

Çizelge 3.19. Öncelikli öğrenci veri kümesi engel türü dağılımı

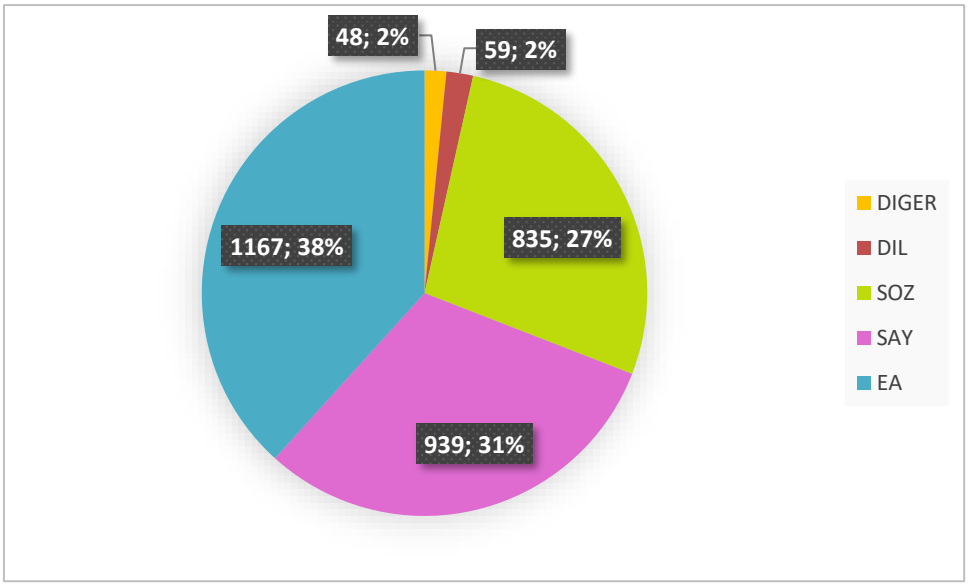
ENGEL TÜRÜ	ÖRNEK SAYISI
Görme engelli	99
Görme kaybı	44
İşitme engelli	40
İşitme kaybı	24
Konuşma/İletişim bozukluğu	1
Ortopedik engelli (Eller/Kolların kullanımında kısıtlılık)	118
Ortopedik engelli (tekerlekli sandalye/baston/koltuk değneği kullanımı v.b.)	25
Kronik sağlık sorunu (diyabet, Crohn hastalığı v.b.)	49
Psikolojik/ psikiyatrik hastalıklar (anksiyete, depresyon v.b.)	8
Nörolojik hastalıklar (epilepsi, travmatik beyin hasarı v.b.)	14
Yaygın gelişimsel bozukluklar (PDD) (otizm, Asperger v.b.)	1
Diğer	103

Çizelge 3.20. Öncelikli öğrenci veri kümesi öncelik durumu dağılımı

ÖNCELİK DURUMU	ÖRNEK SAYISI
ÖSYM tarafından yapılan öğrenci seçme ve yerleştirme sınavlarında her puan türünde ilk yüze girenler	1
Gazi (Kendisi)	2
Şehidin çocuğu yoksa bekar kardeşleri	4
Sivil Şehit Çocukları	9
Lise ve dengi öğrenimlerini Aile ve Sosyal Hizmetler Bakanlığı Yetiştirme Yurtlarında barınarak tamamlayanlar	14
Lise ve dengi öğrenimlerini Darüşşafaka Lisesinde tamamlayanlar	40
Devlet koruması altında olanlar	42

Çizelge 3.20. (devam) Öncelikli öğrenci veri kümesi öncelik durumu dağılımı

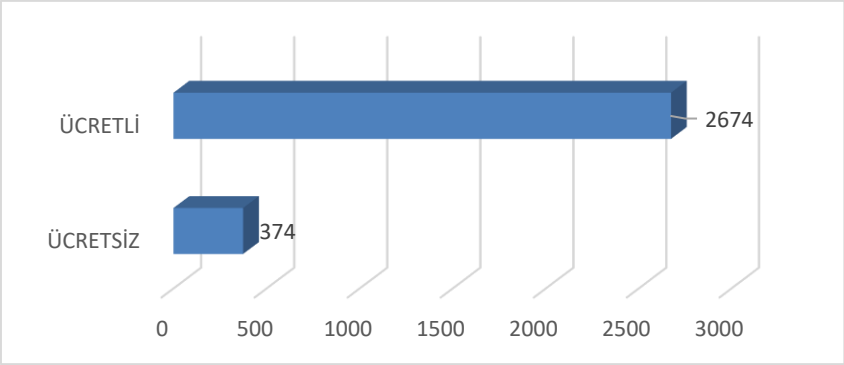
Depremzede Öğrenci	50
Şehit çocuğu	83
Milli sporcu belgesine sahip olanlar	102
Sivil Gazi Çocukları	110
Lise ve dengi öğrenimlerini Aile ve Sosyal Hizmetler Bakanlığı Yetiştirme Yurtlarında barınarak tamamlayanlar ile Devlet koruması altında olanlar	146
Diğer	196
Anne ve babası (Her ikisi de) vefat etmiş 25 yaşını geçmemiş olanlar	354
Gazi çocuğu	561
Sağlık Kurulu Raporu ile %40 ve üzeri engelli olduğu tespit edilenler	602
Gönüllü güvenlik korucusu olarak çalışan ve çalışırken vefat edenlerin, gönüllü güvenlik korucusu olarak emekliye ayrılanların çocukları	732



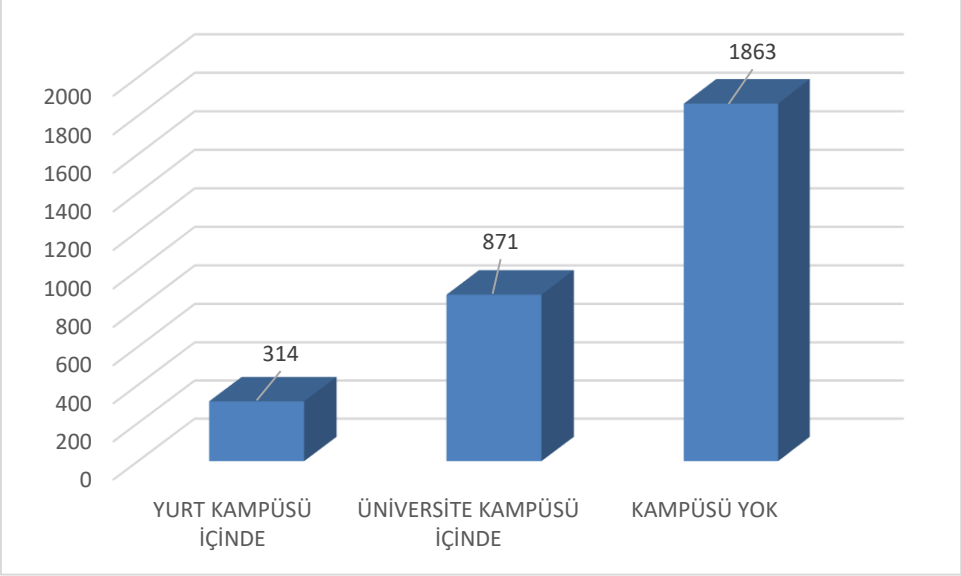
Şekil 3.12. Öncelikli öğrenci veri kümesi puan türü dağılımı

Çizelge 3.21. Öncelikli öğrenci veri kümesi fakülte kategori dağılımı

FAKÜLTE KATEGORİ	FAKÜLTE ADI	ÖRNEK SAYISI
DENIZ	DENİZCİLİK FAKÜLTESİ	2
HAVA	HAVACILIK FAKÜLTESİ	7
GSF	GÜZEL SANATLAR FAKÜLTESİ	34
ZF	ZİRAAT FAKÜLTESİ	38
TF	TURİZM FAKÜLTESİ	50
SPOR	SPOR YÖNETİMİ FAKÜLTESİ	58
ILAHIYAT	İLAHİYAT FAKÜLTESİ	67
HUKUK	HUKUK FAKÜLTESİ	101
IF	İLETİŞİM FAKÜLTESİ	113
SF	SAĞLIK HİZMETLERİ FAKÜLTESİ	342
MUH	MÜHENDİSLİK MİMARLIK FAKÜLTESİ	346
EF	EĞİTİM FAKÜLTESİ	537
IIBF	İDARİ İKTİSADİ BİLİMLER FAKÜLTESİ	594
FEF	FEN EDEBİYAT FAKÜLTESİ	759



Şekil 3.13. Öncelikli öğrenci veri kümesi barınma durumu dağılımı



Şekil 3.14. Öncelikli öğrenci veri kümesi yurt lokasyonu dağılımı

3.2. Gerçekleştirilen Çalışma

Tez çalışmasında kullanılan iki veri kümesi olan yurt öğrenci ve öncelikli öğrenci veri kümesine eksik veri giderme, gürültülü verileri çıkarma, tekrarlı verileri kaldırma, veri standardizasyonu, veri türetme, verileri kategorik hale dönüştürme ve öznitelik seçimi ile veri azaltma işlemleri uygulanarak veri ön işleme çalışması yapılmıştır.

Veri kümelerini sınıflandırmak için naive bayes, lojistik regresyon, adaboost, bagging (torbalama), logitboost, karar tablosu, karar kütüğü, J48, rastgele orman, çok katmanlı algılayıcı (MLP), SVM ve KNN algoritmaları kullanılmıştır.

Sınıflandırıcıların performansları Çizelge 3.25. ve Çizelge 3.26.'da gösterilmiştir.

3.2.1. Veri ön işleme çalışması

Yurt öğrenci veri kümesi ve öncelikli öğrenci veri kümesi için binlerce öğrenci kaydından konuyla ilgili olanlar seçilmiştir, verilerin veri madenciliği yöntemleriyle analiz edilebilmesi için tarih ya da metin içeren özellikler sayısallaştırılmıştır. Bazı özellikler kategorik hale getirilmiştir, verilerden yeni veriler üretilmiştir, gürültülü veriler çıkarılmış ve eksik veriler tamamlanmıştır. Veri ön işleme tamamlandığında Çizelge 3.5.'te gösterilen 34 öznitelikli 8783 veri kaydı ve Çizelge 3.15.'te gösterilen 35 öznitelikli 3084 veri kaydı sınıflandırmaya hazır hale getirilmiştir.

Öznitelik seçimi, alan bilgisine dayalı olarak; bir veri kümesinden ilgili özellikleri seçmek için çok önemli bir adımdır. Bu çalışmada; WEKA araç setini kullanılmıştır, özellik seçimi bilgi kazancı ve korelasyon tabanlı özellik seçimi tekniklerini ile yapılmıştır. Böyle yüksek boyutlu veri kümelerinde, bazı nitelikler performans ölçümünü iyileştirir ve karar verme süreci için faydalıdır, diğerleri ise daha az önemli niteliklerdir. Öznitelik seçiminde bu norm göz önüne alınmıştır.

Yurt öğrenci veri kümesi veri ön işleme süreçleri

Yurt öğrenci veri kümesi için yapılan veri ön işleme çalışmaları aşağıda belirtilmiştir. Grup sayısı belirlenirken her seferinde sınıflandırma yapılmıştır ve başarı göz önüne alınmıştır.

- Eksik verilerin birçoğu veri kaynağına gidilerek tamamlanmıştır. Bulunmayan veri örneği veri kümesinden çıkarılmıştır.
- Gürültülü veriler veri kümesinden çıkarılmıştır.
- Tekrarlı veriler veri kümesinden çıkarılmıştır.
- Yurttanayrildigiyas, verisinin hesaplanması için yurttan ayrılma tarihi- dogum_tarihi çıkarılmıştır. Veriler kategorik hale getirilmiş ve 7 gruba ayrılmıştır.
- Gelir_puan, ondalıklı kısımlar yuvarlanmış tam sayıya dönüştürülmüştür, veriler kategorik hale getirilmiş ve 10 gruba ayrılmıştır.

- Basari_puan, ondalıklı kısımlar yuvarlanmış tam sayıya dönüştürülmüştür, veriler kategorik hale getirilmiş ve 5 gruba ayrılmıştır.
- Puan_turu, ayni manaya gelen ama farklı yazılan türler aynı ifadelere dönüştürülmüştür.
- Fakulte_grup, fakülte verileri kısaltılmış ve simgeler verilmiştir, veriler kategorik hale getirilmiş ve 22 gruba ayrılmıştır.
- Yatak_sayi, odadaki yatak sayısı hesaplanarak bulunmuştur.
- Yurt_bolumuzaklik_grup, veri sistemde hazır halde bulunmamaktadır. Öğrenci okul bölümü ile yurt arasındaki mesafe hesaplanarak bulunmuştur. Veriler kategorik hale getirilmiş ve 10 gruba ayrılmıştır.
- Yurt_kapasite_grup, kategorik hale getirilmiş ve 3 gruba ayrılmıştır.
- Binakapasite_grup, kategorik hale getirilmiş ve 3 gruba ayrılmıştır.
- Kaldıgı_ay, verisinin hesaplanması için yurttan_ayrılma tarihi- yurda_giris_tarihi çıkarılmış ve ay hesabı yapılmıştır.

Yurt öğrenci veri kümesi Çizelge 3.5.'te gösterilen 34 öznitelikten oluşmaktadır, öznitelik seçimi için kullanılan bilgi kazancı ve korelasyon tabanlı özellik seçimi sonuçları Çizelge 3.22.'de gösterilmiştir. Bilgi kazancı Eş. 2.3, korelasyon ise Eş. 2.4 ile hesaplanmıştır. Sonuçlara göre yüksek sıralama değerlerine sahip olan toplam 34 özellikten ilk 27'sı seçilmiştir. KALDIGI_AY_1 sınıflandırmayı direkt etkilediğinden çıkarılmıştır. Bu çalışmada kullanılan yurt öğrenci veri kümesi 26 öznitelik içermektedir.

Çizelge 3.22. Yurt öğrenci veri kümesi bilgi kazancı sıralaması ve korelasyon tabanlı özellik seçimi sonuçları

No	Öznitelik	Bilgi Kazancı	Korelasyon	No	Öznitelik	Bilgi Kazancı	Korelasyon
34	KALDIGI_AY_1	0,6425	0,7266	29	YATAK_SAYI	0,0036	0,0097
3	YURTTANAYRILDIGIYAS	0,4923	0,6709	33	BINAKAPASITE_GRUP_1	0,0034	0,0639
1	CINSIYET	0,0497	0,2617	32	YURT_KAPASITE_GRUP_1	0,0027	0,0562
24	FAKULTE_GRUP	0,0483	0,0799	19	BASARI_PUAN	0,0017	0,0246
31	YURT_BOLUMUZAKLIK_GRUP	0,0371	0,2060	11	BABA_VEFAT	0,0011	0,0395
26	BAHCE	0,0265	0,1913	8	IL_DISI_UNI_KARDES	0,0008	0,0337
27	MODELTIP	0,0188	0,1073	23	BOLUM_OZEL_YETENEK	0,0007	0,0262
28	YATAKTIPI	0,0187	0,0097	10	ANNE_VEFAT	0,0003	0,0191
16	MEB_KARDES_SAYI	0,0119	0,1275	13	ANNE_SAG_OZURLU_BEYAN	0,0003	0,0189
30	UCRET_TIPI	0,0087	0,0900	14	BABA_SAG_OZURLU_BEYAN	0	0,0017
22	VELAYET	0,0067	0,0877	7	BASVURU_TURU	0	0,0126
12	ANNE_BABA_AYRI	0,0066	0,0959	2	MEDENIDURUM	0	0,0006
21	PUAN_TURU	0,0051	0,0315	20	OGRETIM_SEKLI	0	0,0097
18	GELIR_PUAN	0,0050	0,0419	9	IL_ICI_UNI_KARDES	0	0,0021
4	ADRES_IL_ID	0,0041	0,0366	5	ENGEL_TURU_ID	0	0,0005
17	SOSYAL_DURUM_PUAN	0,0039	0,0219	6	ONCELIK_DURUM	0	0,0090
15	VAKIF_TAM_BURS_DEGIL	0,0037	0,0694	25	YURTLOKASYON	0	0,0243

Öncelikli öğrenci veri kümesi veri ön işleme süreçleri

Öncelikli öğrenci veri kümesi için yapılan veri ön işleme çalışmaları aşağıda belirtilmiştir. Grup sayısı belirlenirken her seferinde sınıflandırma yapılmıştır ve başarı göz önüne alınmıştır.

- Eksik verilerin birçoğu veri kaynağına gidilerek tamamlanmıştır. Bulunmayan veri örneği veri kümesinden çıkarılmıştır.
- Gürültülü veriler veri kümesinden çıkarılmıştır.
- Tekrarlı veriler veri kümesinden çıkarılmıştır.
- Yurttanayrildigiyas, verisinin hesaplanması için yurda giriş tarihi -yurttan ayrılma tarihi- dogum_tarihi çıkarılmıştır. Veriler kategorik hale getirilmiş ve 7 gruba ayrılmıştır.
- Gelir_puan, ondalıklı kısımlar yuvarlanmış tam sayıya dönüştürülmüştür, veriler kategorik hale getirilmiş ve 10 gruba ayrılmıştır.
- Basari_puan, ondalıklı kısımlar yuvarlanmış tam sayıya dönüştürülmüştür, veriler kategorik hale getirilmiş ve 5 gruba ayrılmıştır.
- Puan_turu, ayni manaya gelen ama farklı yazılan türler aynı ifadelere dönüştürülmüştür.
- Fakulte_grup, fakülte verileri kısaltılmış ve simgeler verilmiştir, veriler kategorik hale getirilmiş ve 14 gruba ayrılmıştır.
- Yatak_sayi, odadaki yatak sayısı hesaplanarak bulunmuştur.
- Yurt_bolumuzaklik_grup, veri sistemde hazır halde bulunmamaktadır. Öğrenci okul bölümü ile yurt arasındaki mesafe hesaplanarak bulunmuştur. Veriler kategorik hale getirilmiş ve 8 gruba ayrılmıştır.
- Yurt_kapasite_grup, kategorik hale getirilmiş ve 3 gruba ayrılmıştır.
- Binakapasite_grup, kategorik hale getirilmiş ve 3 gruba ayrılmıştır.
- Kaldıgı_ay, verisinin hesaplanması için yurttan_ayrılma tarihi- yurda_giris_tarihi çıkarılmış ve ay hesabı yapılmıştır.

Öncelikli öğrenci veri kümesi Çizelge 3.15.'te gösterilen 35 öznitelikten oluşmaktadır, öznitelik seçimi için kullanılan bilgi kazancı ve korelasyon tabanlı özellik seçimi sonuçları Çizelge 3.23'te gösterilmiştir. Bilgi kazancı Eş. 2.3, korelasyon ise Eş. 2.4 ile hesaplanmıştır. Sonuçlara göre yüksek sıralama değerlerine sahip olan toplam 35 özellikten ilk 24'ü

seçilmiştir. KALDIGI_AY_1 sınıflandırmayı direkt etkilediğinden çıkarılmıştır. Bu çalışmada kullanılan öncelikli öğrenci veri kümesi 23 öznitelik içermektedir.

Çizelge 3.23. Öncelikli öğrenci veri kümesi bilgi kazancı sıralaması ve korelasyon tabanlı özellik seçimi sonuçları

No	Öznitelik	Bilgi Kazancı	Korelasyon	No	Öznitelik	Bilgi Kazancı	Korelasyon
34	KALDIGI_AY_1	0,5657	0,5838	12	ANNE_BABA_AYRI	0,0020	0,0529
3	YURTTANAYRILDIGIYAS	0,3676	0,6000	13	ANNE_SAG_OZURLU_BEYAN	0,0005	0,0255
1	CINSIYET	0,0594	0,2853	14	BABA_SAG_OZURLU_BEYAN	0,0004	0,0238
5	ENGEL_TURU_ID	0,0502	0,0155	15	VAKIF_TAM_BURS_DEGIL	0,0003	0,0212
6	ONCELIK_DURUM	0,0467	0,0609	10	ANNE_VEFAT	0,0003	0,0194
24	FAKULTE_GRUP	0,0345	0,0582	11	BABA_VEFAT	0,0002	0,0171
4	ADRES_IL_ID	0,0219	0,0186	28	YATAKTIPI	0,0000	0,0405
19	BASARI_PUAN	0,0101	0,0767	31	YURT_BOLUMUZAKLIK_GRUP	0,0000	0,0469
21	PUAN_TURU	0,0089	0,0152	33	BINAKAPASITE_GRUP_1	0,0000	0,0331
25	YURTLOKASYON	0,0088	0,1145	2	MEDENIDURUM	0,0000	0,0102
23	BOLUM_OZEL_YETENEK	0,0084	0,1007	30	UCRET_TIPI	0,0000	0,0267
29	YATAK_SAYI	0,0073	0,0961	17	SOSYAL_DURUM_PUAN	0,0000	0,0471
16	MEB_KARDES_SAYI	0,0057	0,0902	7	BASVURU_TURU	0,0000	0,0010
32	YURT_KAPASITE_GRUP_1	0,0054	0,0363	22	VELAYET	0,0000	0,0230
18	GELIR_PUAN	0,0054	0,0300	27	MODELTIP	0,0000	0,0376
35	BARINMADURUM	0,0052	0,0845	8	IL_DISI_UNI_KARDES	0,0000	0,0453
26	BAHCE	0,0039	0,0731	9	IL_ICI_UNI_KARDES	0,0000	0,0234
20	OGRETIM_SEKLI	0,0020	0,0531				

3.2.2. Sınıflandırma çalışması

Yurt öğrenci ve öncelikli öğrenci veri kümesinde iki tür öğrenci vardır. Birincisi; mezun öğrenciler, yurtta okuldan mezun olana kadar kalırlar. İkinci tip ise yurttan kendi isteğiyle vaktinden önce ayrılan öğrencilerdir. Yapılan çalışma, yurttan ayrılan öğrenciyi hedefliyor ve ayrılışın arkasındaki nedenleri tanımlıyor. Bu çalışmada, öğrenci verilerini sınıflandırmak için bir dizi makine öğrenme tekniği kullanılmıştır. Algoritmalardan hangisinin öğrencileri en iyi şekilde sınıflandırdığını değerlendirmektir. 10 kat çapraz doğrulama ile naive bayes (NB), J48, karar kütüğü, karar tablosu, adaboost, logitboost, torbalama, çok katmanlı

algılayıcı (MLP), lojistik regresyon (LR), SVM, rastgele orman (RF) ve KNN kullanıldı. Sınıflandırma işlemi WEKA 3.8.6 araç takımı kullanılarak gerçekleştirilmiştir. WEKA araç takımında kullanılan bazı algoritmaların açıklamaları verilmiştir.

Multilayer perceptron -MLP (çok katmanlı algılayıcı): Örnekleri sınıflandırmak ve çok katmanlı bir algılayıcı öğrenmek için geri yayılımı kullanır. Yapay sinir ağı modelidir. Bu ağdaki düğümlerin tümü sigmoiddir [47].

SMO: SMO algoritması temel olarak destek vektör makinesi yöntemlerini kullanır.

IBK (k-nearest neighbor – k-en yakın komşu): k-en yakın komşu sınıflandırıcısıdır.

Decision table (Karar tablosu): Basit bir karar tablosu ile çoğunluk sınıflandırıcısı oluşturmak için kullanılan bir sınıflandırıcıdır. Karar tabloları, daha yüksek seviye bir tablodaki her bir girişin, başka bir tablo oluşturmak için bir çift ek özniteliğin değerleri ile parçalandığı hiyerarşik bir tablodan oluşur.

Desicion stump (Karar kütüğü): Bir karar güdüsü oluşturma ve kullanma sınıfıdır. Genellikle bir yükseltme algoritması ile birlikte kullanılır. Regresyon (ortalama kare hatasına göre) veya sınıflandırma (entropiye dayalı) yapar [47].

J48: C4.5 algoritması temeline dayanan bir karar ağacı algoritmasıdır.

WEKA araç setinde kullanılan parametler; Adaboost algoritması uygulanırken seçilen parametrelerden biri olan sınıflandırıcı için decision stump kullanılmıştır ve iterasyon 10 olarak belirlenmiştir. Bagging için sınıflandırıcı olarak RepTree ve iterasyon sayısı 10 olarak seçilmiştir. Logitboost algoritmasında sınıflandırıcı için decision stump kullanılmıştır ve iterasyon 10 olarak belirlenmiştir. Decision table'da özellik seçimi sıralaması için bestfirst seçilmiştir. Random forest iterasyonu 100 ile sınırlandırılmıştır. SMO'da çekirdek fonksiyonu olarak polinom seçilmiştir. IBk'de arama algoritması olarak lineerNNsearch kullanılmıştır.

Sınıflandırma algoritmalarının performansını k-kat çapraz doğrulama ile değerlendirmek popülerdir. Güvenilir bir doğruluk tahmini nispeten küçük bir varyansa sahip olacaktır ve bu

nedenle birçok çalışma k-kat çapraz doğrulamayı gerçekleştirmeyi önermiştir [48]. Bu tez çalışmasında, 10 kat çapraz doğrulama kullanılmıştır veri seti 10 parçaya bölünür, 9 parça eğitim verisi olarak kullanılırken 1 parça test verisi olarak kullanılır. Bu işlem 10 kez yinelenir ve her defasında başka bir parça doğrulama için kullanılır. Başarı değerlendirmesi için ise doğruluk oranlarının ortalaması alınır.

Sınıflandırıcıların, tahminindeki performansını uygun parametrelerle değerlendirmek amacıyla, Çizelge 3.24.'te gösterilen karışıklık matrisinin içeriğinden hesaplanan kesinlik, hatırlama, doğruluk ve F-ölçüsü ölçümleri kullanılmıştır [21].

Çizelge 3.24. Deneysel çalışma karışıklık matrisi

	Ayrılanlar (Kendi İsteğiyle Yurttan Ayrılan Öğrenciler)(Pozitif)	Ayrılmayanlar (Mezun olan Öğrenciler) (Negatif)
Ayrılanlar (Pozitif)	TP	FN
Ayrılmayanlar (Negatif)	FP	TN

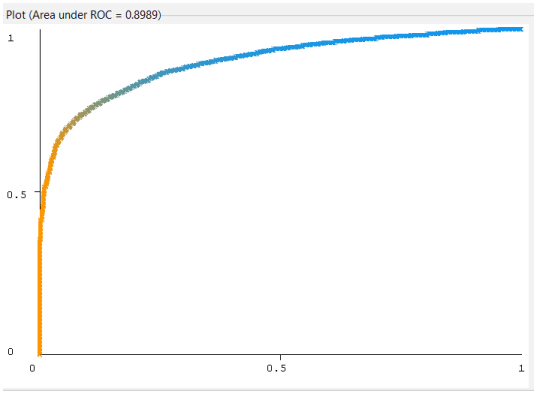
Deneysel çalışmaların sonuçları, Çizelge 3.25. ve Çizelge 3.26.'da gösterilmiştir. Doğru tahmin oranı, yanlış tahmin oranı, kesinlik, hatırlama, f-ölçüsü, ROC eğrisi değerlerine göre sınıflandırıcıların başarıları belirlenmiştir. ROC eğrileri ve AUC alanları performansları Şekil 3.15. ve Şekil 3.16.'da gösterilmiştir.

Yurt öğrenci veri kümesi için sınıflandırıcıların performansı

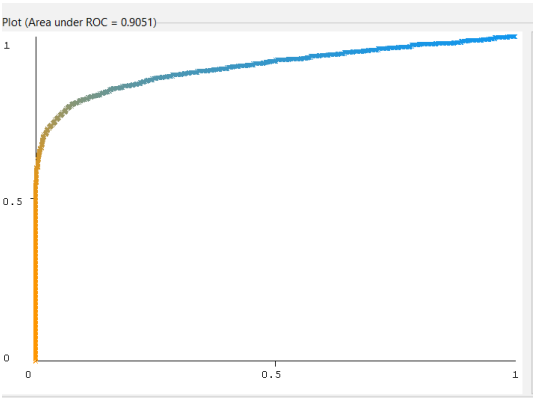
Yurt öğrenci veri kümesi için sınıflandırıcıların performansı Çizelge 3.25'te gösterilmiştir. Decision table %86,95 doğruluk ve 0,89 AUC ile en iyi sonucu vermiştir. AdaBoost ve decision stump %86,92 başarı göstermiştir. Sonuçlar topluluk öğrenimi algoritmalarının daha iyi sınıflandırma yaptığını göstermiştir.

Çizelge 3.25. Yurt öğrenci veri kümesi için sınıflandırıcıların performansı

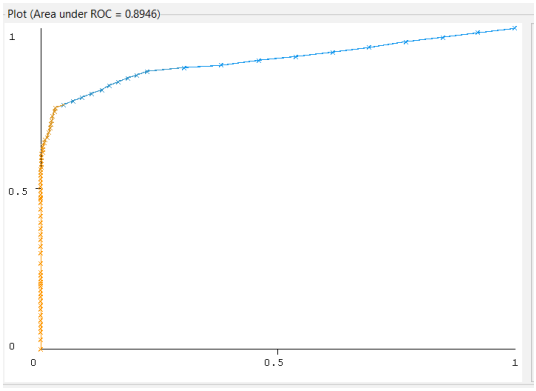
Sınıflandırıcılar	Accuracy %	TP Rate	FP Rate	Precision	Recall	F-Measure	AUC
Naive Bayes	82,93	0,77	0,12	0,84	0,77	0,80	0,90
LogisticRegresyon	86,27	0,79	0,08	0,89	0,79	0,84	0,91
AdaBoost	86,93	0,75	0,03	0,95	0,75	0,84	0,90
Bagging	86,36	0,77	0,06	0,92	0,77	0,84	0,91
LogitBoost	86,80	0,75	0,04	0,94	0,75	0,84	0,92
Decision Table	86,95	0,75	0,03	0,95	0,75	0,84	0,89
Decision Stump	86,93	0,75	0,03	0,95	0,75	0,84	0,85
J48	86,55	0,78	0,06	0,91	0,78	0,84	0,87
RandomForest	86,30	0,78	0,07	0,90	0,78	0,84	0,92
SVM	86,50	0,77	0,06	0,92	0,77	0,84	0,86
KNN	72,62	0,66	0,22	0,71	0,66	0,69	0,72
ANN (Multiplayer Perceptron)	84,41	0,79	0,11	0,85	0,79	0,82	0,90



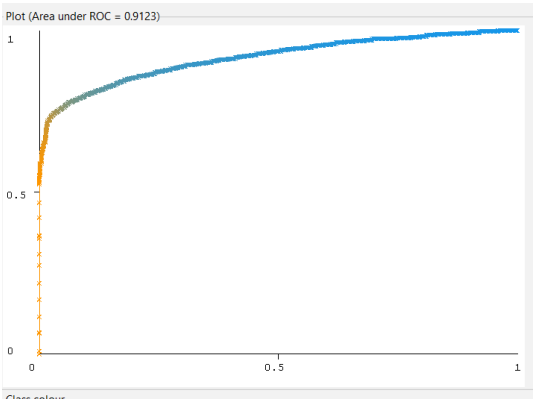
(a) Naive bayes ROC eğrisi



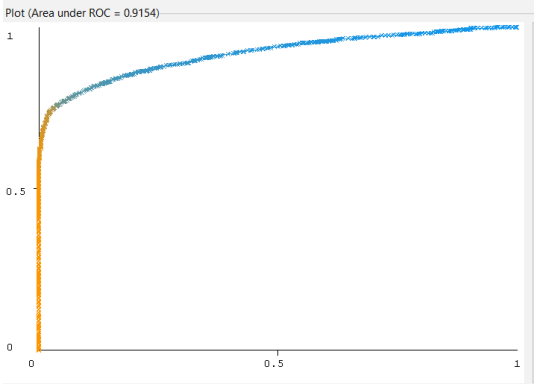
(b) Logistic regresyon ROC eğrisi



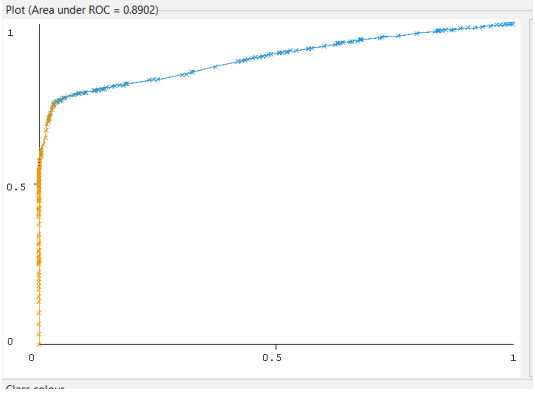
(c) AdaBoost ROC eğrisi



(ç) Bagging ROC eğrisi

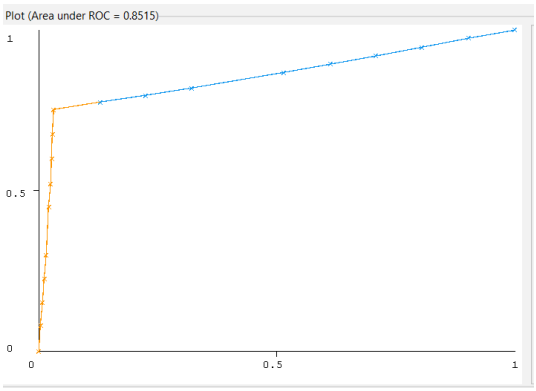


(d) LogitBoost ROC eğrisi

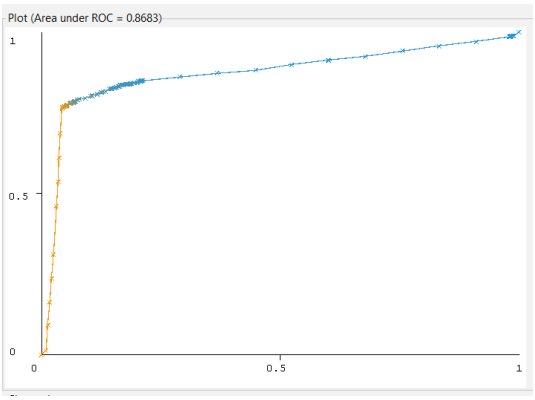


(e) Decision table ROC eğrisi

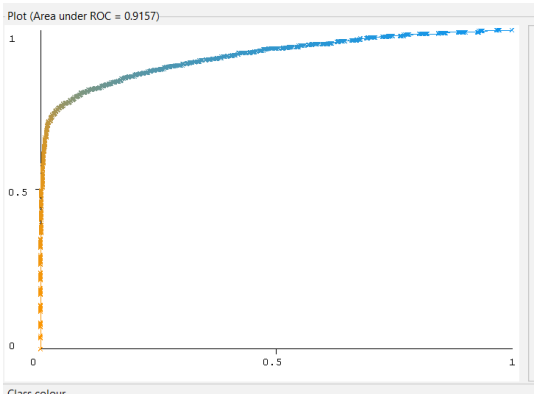
Şekil 3.15. Yurt öğrenci veri kümesi ROC eğrileri



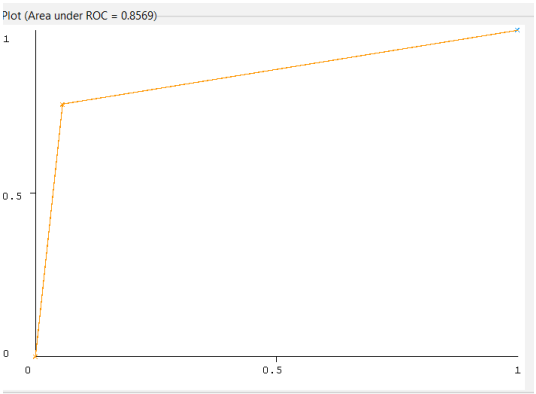
(f) Decision stump ROC eğrisi



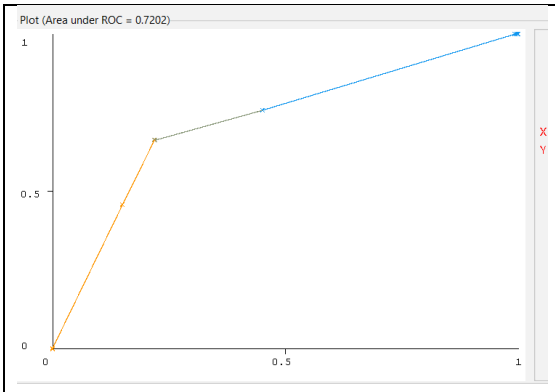
(g) J48 ROC eğrisi



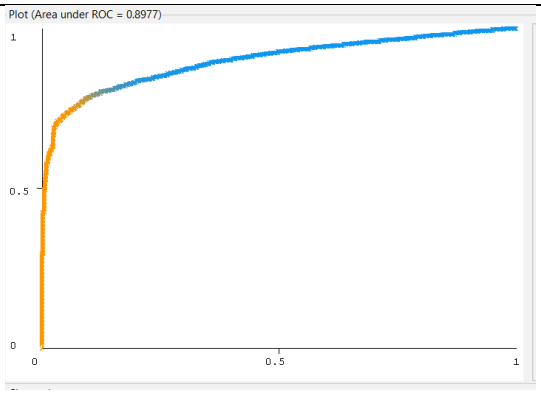
(ğ) Random forest ROC eğrisi



(h) SVM ROC eğrisi



(ı) KNN ROC eğrisi



(i) MLP ROC eğrisi

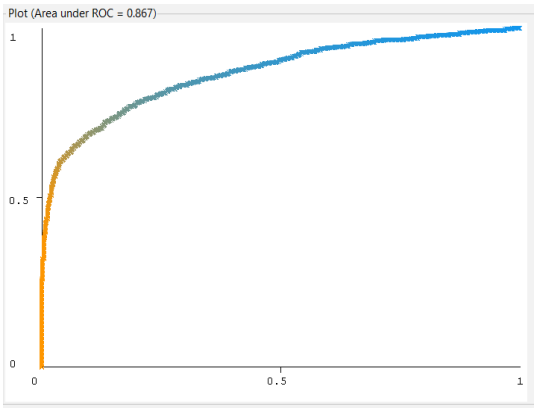
Şekil 3.15. (devam) Yurt öğrenci veri kümesi ROC eğrileri

Öncelikli öğrenci veri kümesi için sınıflandırıcıların performansı

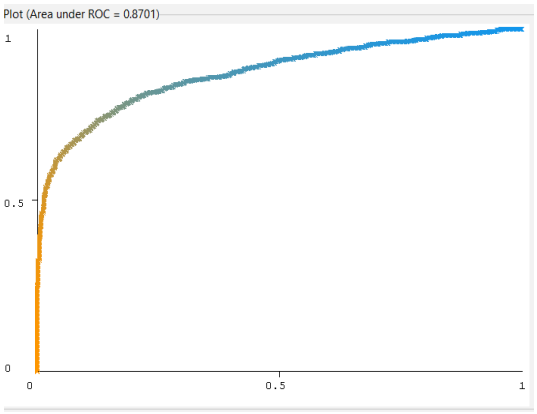
Öncelikli öğrenci veri kümesi için sınıflandırıcıların performansı çizelge 3.26'da gösterilmiştir. LogitBoost %81,89 doğruluk ve 0,88 AUC alanı ile en iyi sonucu vermiştir. J48 %81,6 ve bagging %81,43 başarı göstermiştir.

Çizelge 3.26. Öncelikli öğrenci veri kümesi için sınıflandırıcıların performansı

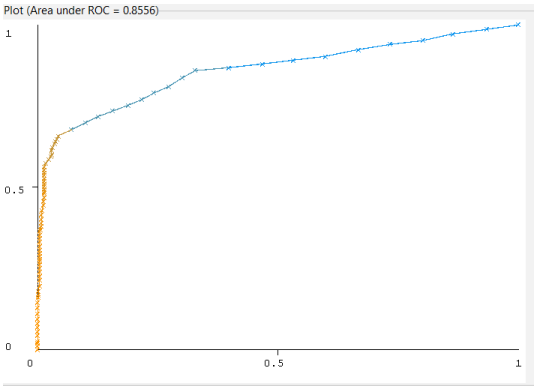
Sınıflandırıcılar	Accuracy %	TP Rate	FP Rate	Precision	Recall	F-Measure	AUC
Naive Bayes	79,40	0,71	0,13	0,83	0,71	0,77	0,87
LogisticRegresyon	80,22	0,75	0,15	0,82	0,75	0,78	0,87
AdaBoost	81,40	0,66	0,04	0,93	0,66	0,77	0,86
Bagging	81,43	0,72	0,10	0,87	0,72	0,79	0,86
LogitBoost	81,89	0,70	0,07	0,90	0,70	0,79	0,88
Decision Table	80,94	0,69	0,09	0,88	0,69	0,78	0,87
Decision Stump	81,40	0,66	0,04	0,93	0,66	0,77	0,80
J48	81,59	0,72	0,10	0,87	0,72	0,79	0,83
RandomForest	81,30	0,72	0,10	0,86	0,72	0,79	0,87
SVM	80,15	0,72	0,13	0,84	0,72	0,78	0,80
KNN	63,65	0,58	0,31	0,63	0,58	0,60	0,63
ANN (Multiplayer Perceptron)	75,33	0,72	0,22	0,75	0,72	0,73	0,82



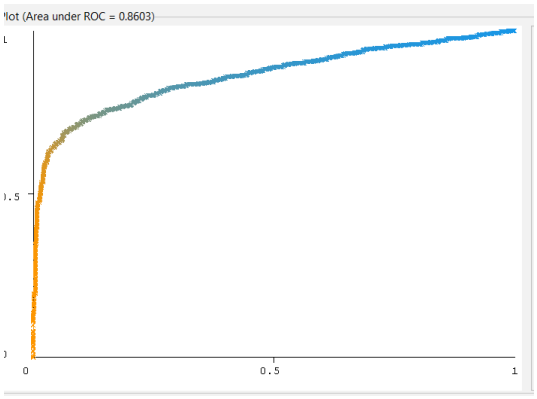
(a) Naive bayes ROC eğrisi



(b) Logistic regresyon ROC eğrisi

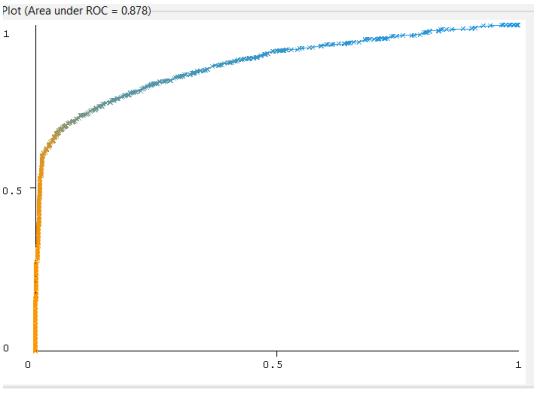


(c) AdaBoost ROC eğrisi

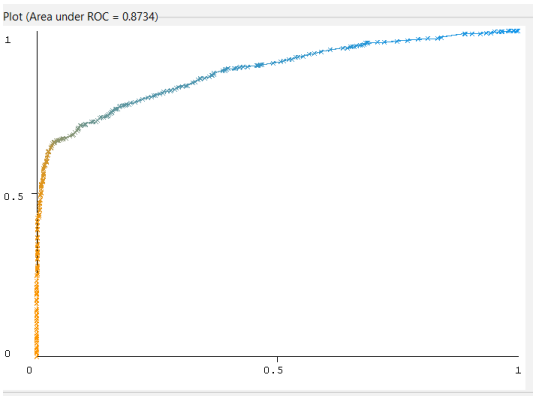


(ç) Bagging ROC eğrisi

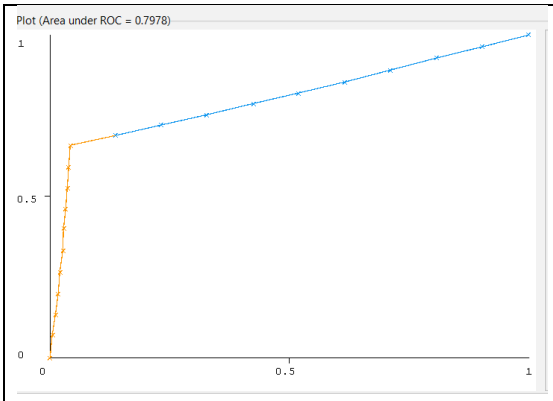
Şekil 3.16. Öncelikli öğrenci veri kümesi ROC eğrileri



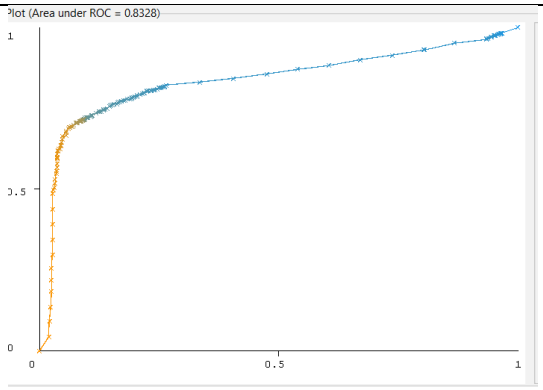
(d) LogitBoost ROC eğrisi



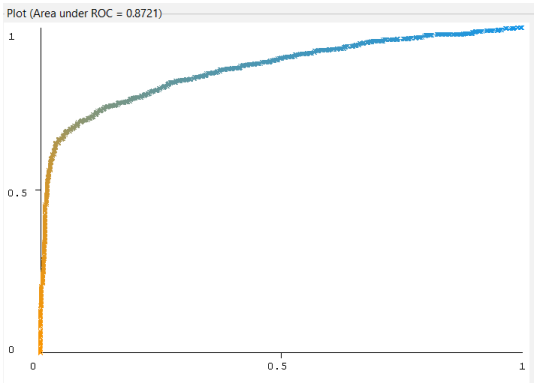
(e) Decision table ROC eğrisi



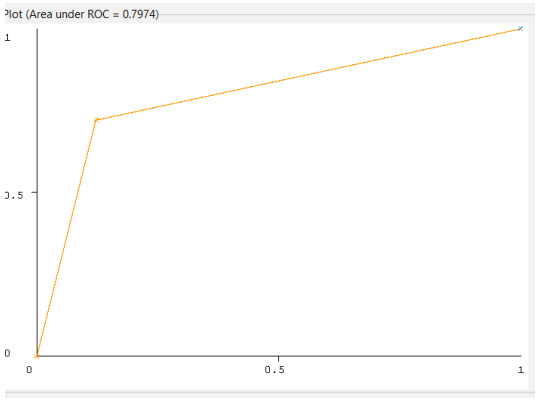
(f) Decision stump ROC eğrisi



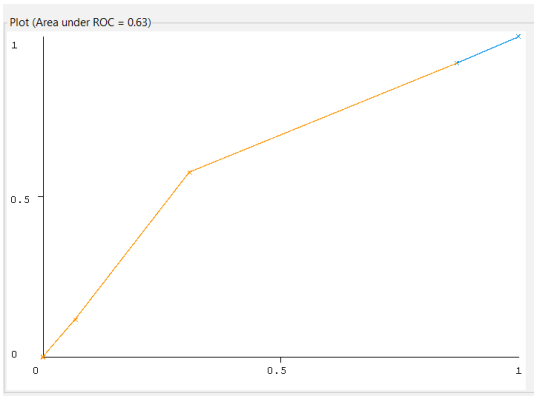
(g) J48 ROC eğrisi



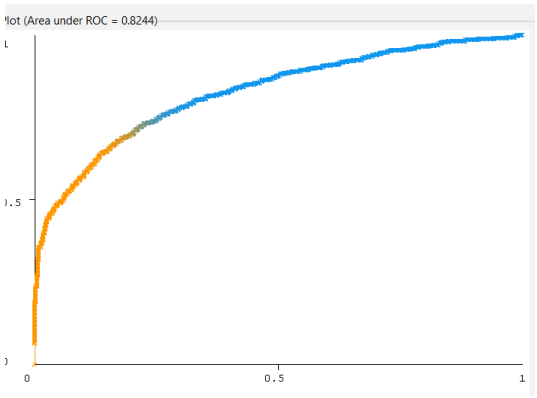
(ğ) Random forest ROC eğrisi



(h) SVM ROC eğrisi



(ı) KNN ROC eğrisi



(i) MLP ROC eğrisi

Şekil 3.16. (devam) Öncelikli öğrenci veri kümesi ROC eğrileri

4. SONUÇ VE TARTIŞMA

Ülkemizde devlet yurt kapasitesi artmasına rağmen artan genç nüfusla beraber üniversite öğrencileri arasında barınma sorunu yaşanmaktadır. Bu nedenle yurtların iyi organize edilmesi, mevcut kapasitenin en uygun şekilde kullanılması ve ihtiyaç sahibi öğrencilere ortam oluşturulması büyük önem arz etmektedir. Gerçekleştirilen bu tez çalışmasında, KYGM'nde barınan öğrencilerin yurttan ayrılma nedenlerinin veri madenciliği yöntemleriyle analiz edilebilmesi için ayrılma nedenleri ve barınan öğrencilerin çeşitli özellikleri kullanılmıştır ayrılması baştan öngörülen öğrencilerin tespitine çalışılmıştır. Böylece yurt kapasitesi en verimli şekilde kullanılabilecek ve öğrencilerin iyi hizmet alması sağlanacaktır.

Literatür taraması yapılmış yurttan ayrılma nedenlerinin analiz edilmesine benzer pek fazla kaynak olmadığı görülmüştür. Benzer çalışmalar incelenmiştir, kullanılan yöntemlerin zayıf ve güçlü yönleri ele alınmıştır. Sorunun çözümüne uygun bir model önerilmiştir.

Tez çalışmasında iki veri kümesi kullanılmıştır; ilki KYGM'nün Ankara ilinde bulunan yurtlarında barınan öğrencilerin 2016-2022 yılları arasındaki verilerini, ikinci veri kümesi ise KYGM'nün yurtlarında 2016-2022 yılları arasında barınan öncelikli öğrenci verilerini içermektedir. Veri kümelerinde öğrencilerin bilgileri ile barındıkları yurtların fiziksel özellikleri bulunmaktadır. Yapılan deneysel çalışmalar neticesinde topluluk öğrenimi algoritmalarının daha başarılı sonuçlar verdiği tespit edilmiştir. Yurt öğrenci veri kümesi, karar tablosu (decision table) %86,95 doğruluk ve %89 AUC, adaboost %86,93 doğruluk ve %89 AUC, karar kütüğü (decision stump) %86,93 doğruluk ve %85 AUC ile iyi performans gösteren sınıflandırıcılar olmuşlardır. %91,6 ile en iyi AUC sahip random forest %86,3 doğruluk oranı göstermiştir. Öncelikli öğrenci veri kümesi ise logistBoost %81,89 doğruluk ve %88 AUC, karar ağacı (J48) %81,59 doğruluk ve %83 AUC oranı ile en iyi performans veren algoritmalardır.

Yapılan çalışmanın başarısının değerlendirilebilmesi için uzman görüşüne başvurulmuştur. 5 yurt müdürü, 2 yurt yönetim memuru, 3 genel müdürlük personeli olmak üzere 10 personele; yurtta barınan bütün öğrencilerin ve öncelikli öğrencilerin yurttan ayrılma nedenleri ile ilgili görüşleri sorulmuştur. Uzman görüşüne göre, yurtta barınan bütün öğrenciler için sırasıyla:

- Yurdun model tipi (yatak tipi, tuvalet banyonun yeri, buzdolabı olup olmadığı gibi)
- Odadaki yatak sayısı
- Öğrencinin öğrenim gördüğü bölümün yurda olan uzaklığı
- Cinsiyet
- Yaş
- Öğrencinin öğrenim gördüğü fakülte/bölüm
- Yurdun ücreti
- Öğrencinin velayet durumu
- Öğrencinin kardeş sayısı

Yurttan ayrılmayı etkileyen önemli durumlardır.

- Yurdun model tipi içinde bulunan tuvalet ve banyonun konumu ve yatak tipi vb. değerlendirildiğinde, öğrencilerin odanın içinde tuvalet ve banyonun olmasını tercih ettikleri, yatak tipinin baza ya da karyola olmasını istediklerini,
- Öğrencilerin odanın kalabalık olmasını istemedikleri 3-4 kişilik odaların ideal olduğunu,
- Yurdun okula yakın olmasının önemli olduğunu hem ulaşım maliyetinin düştüğü hem de zamandan tasarruf edilebileceğini,
- Kız öğrencilerin erkek öğrencilere göre daha uzun süre yurtta kaldığını,
- Yaş ilerledikçe yurttan ayrılma eğiliminin arttığı, yaşı daha büyük olanların yurt kurallarına uymasının zorlaştığını,
- Öğrencinin öğrenim gördüğü bölümün yurttan ayrılma sebebi olabileceğini bazı bölümlerin ders çalışma ekipmanlarının farklılığı, çalışma süresinin farklı olması, sosyal ortamın etkisinin olabileceğini,
- Yurt ücretinin değişmesinin etkili olabileceğini,
- Öğrencinin anne babasının ayrı olmasının velayet durumunun öğrenciyi etkileyebileceğini,
- Kardeş sayısı fazla olan öğrencilerin kalabalık ortama daha kolay adapte olduğu ancak tek çocuk olanların uyum sorunu yaşadığını belirttiler.

Öncelikli öğrenciler için sırasıyla:

- Engel türü

- Odadaki yatak sayısı
- Öncelik durumu
- Yurdun kampüs alanı içinde olup olmaması
- Öğrencinin öğrenim gördüğü fakülte/bölüm
- Cinsiyet
- Yaş
- Öğrencinin başarı puanı
- Öğrencinin özel yetenek bölümünde öğrenim görmesi (güzel sanatlar fakültesi, spor akademisi vb.)

Yurttan ayrılmayı etkileyen önemli durumlardır.

- Öğrencinin engel türünün örneğin görme engelli, işitme engelli, ortopedik engeli bulunan öğrenciler vb. engelli odalarında kalmalarına rağmen güçlüklerle karşılaşabileceklerini,
- Öğrencilerin odanın kalabalık olmasını istemediklerini 3-4 kişilik odaların ideal olduğunu,
- Öğrencinin öncelik durumu ÖSYM ilk yüz öğrencisi, milli sporcular, yetiştirme yurdundan gelenler ile engelli öğrencilerin yurttan ayrılma nedenlerinin farklı olacağı ayrıca öncelik durumuna göre yurt ücreti ödeme durumunun değiştiğini,
- Yurdun, okul kampüsü içinde olmasının önemli olduğunu hem ulaşım maliyetinin düştüğü hem de engelli öğrencilerin ulaşımın daha kolay olacağını,
- Öğrencinin öğrenim gördüğü bölümün yurttan ayrılma sebebi olabileceğini bazı bölümlerin ders çalışma ekipmanlarının farklılığı, çalışma süresinin farklı olması, sosyal ortamın etkisinin olabileceğini,
- Kız öğrencilerin erkek öğrencilere göre daha uzun süre yurtta kaldığını,
- Yaş ilerledikçe yurttan ayrılma eğiliminin arttığı, yaşı daha büyük olanların yurt kurallarına uymasının zorlaştığını,
- Başarılı öğrencilerin yurtta daha uzun süre kaldığını,
- Özel yetenek bölümü öğrencilerin örneğin enstrüman çalan, heykel yapan vs. yurt koşullarından daha çok etkileyebileceğini belirttiler.

Yapılan deneysel çalışmada yurt öğrenci veri kümesi sınıflandırmasını etkileyen öznitelikler incelendiğinde; yaş, cinsiyet, fakülte, yurdun bölüme uzaklığı, bahçe durumu, model tipi, yatak sayısı, kardeş sayısı, ücret tipi, velayet, anne babanın ayrılık durumu, puan türü, gelir puanı, adres ili, sosyal puanın daha etkili olduğu görülmüştür. Bu durumun uzman görüşü ile örtüştüğü başarılı bir sonuç alındığı tespit edilmiştir.

Yapılan deneysel çalışmada kullanılan ikinci veri kümesi olan öncelikli öğrenci veri kümesinde ise sınıflandırmayı etkileyen özniteliklerin önemli görülenleri; yaş, cinsiyet, engel türü, öncelik durumu, fakülte, adres ili, başarı puanı, yurt lokasyonu, bölümün özel yetenek olup olmadığı, yatak sayısı, kardeş sayısı, yurt kapasitesi, gelir puanı, barınma durumudur. Sonuçların uzman görüşü ile uyumlu olduğu görülmüştür.

Deneysel çalışmalarda kümeleme de kullanılmış ancak alınan sonuçlarda veriler belirgin bir şekilde kümelere ayrılmamıştır. Kullanılan veri kümelerinin, kümelemeye uygun olmadığı tespit edilmiştir. İlk deneysel çalışmalarda Ankara ilinde barınan öğrenciler, öğrenci veri kümesinde kullanılmıştır ancak son yıllarda artan öğrenci sayısı nedeniyle yurtların fiziksel özelliklerinde çeşitli değişlikler yapıldığı bunun da veri dağılımını olumsuz etkilediği görülmüştür. Diğer il verileri kullanıldığında daha başarılı sonuçlar alınabileceği ön görülmektedir.

Çalışmasının neticesinde öğrencilerin özelliklerine göre yurttan ayrılma nedenleri tespit edilmiş, erken dönemde yurttan ayrılan öğrenciler başarılı şekilde bulunmuştur. Böylece yurt koşullarının iyileştirilmesi ya da sosyal ortamın düzenlenmesi ve yurttan ayrılmanın asgari düzeye indirgenmesi hedeflenmektedir. İlk birkaç ayda yurttan ayrılması öngörülen öğrenci profilinin yurt başvuru ve yerleştirme kriterlerinin yeniden gözden geçirilmesi önerilecektir.

Kurum tarafından yurttan kendi isteğiyle ayrılan öğrencilerin ayrılma nedeni detaylı olarak kayıt altına alınırsa farklı öğrenci ve yurt verileri kullanılarak daha başarılı sonuçlar alınabilir.

Bu tez çalışmasında Ankara ilinde yurtta barınan ve dört yıl öğrenim gören öğrenciler ile öncelikli öğrenci verileri kullanılmıştır. Gelecek çalışmalarda kullanılan verilerin diğer illere ve öğrencilere de yaygınlaştırılması, dezavantajlı gruplar için çalışmanın derinleştirilmesi planlanmaktadır.

KAYNAKLAR

1. İnternet: YÖK. (t.y). *Resmi istatistikler*. Web: https://istatistik.yok.gov.tr/ Son Erişim Tarihi:14/06/2023.

2. Kavuncu, E. (2014). Türkiye'de üniversitelerde öğrenci-yurt hizmetleri ve barınma sorunu. *Eğitime Bakış Eğitim-Öğretim ve Bilim Araştırma Dergisi*, 30, 48-50.

3. Önder, M. ve Özdemir, İ. (2016). Gözardı edilen bir işletme türü olarak üniversite öğrenci yurtlarının profili. *Uluslararası Alanya İşletme Fakültesi Dergisi*, 8(1), 79-87.

4. İnternet: Milli Eğitim Bakanlığı [MEB]. (t.y). *Resmi istatistikler*. Web: https://sgb.meb.gov.tr/www/resmi-istatistikler/icerik/64 Son Erişim Tarihi:14/06/2023.

5. İnternet: Kredi ve Yurtlar Genel Müdürlüğü. (t.y). *Kredi ve Yurtlar Genel Müdürlüğü'nün genel bilgileri.* Web: https://kygm.gsb.gov.tr/Sayfalar/2434/2389/GenelBilgiler, Son Erişim Tarihi:14/08/2023.

6. Nadiri, H., Kandampully, J. and Hussain, K. (2009). Students' perceptions of service quality in higher education. *Total Quality Management*, 20(5), 523-535.

7. Ullah, I., Raza, B., Malik, A. K., Imran, M., Islam, S. U. and Kim, S. W. (2019). A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecom sector. *IEEE Access*, 7, 60134-60149.

8. Frijters, P., Islam, A. and Pakrashi, D. (2019). Heterogeneity in peer effects in random dormitory assignment in a developing country. *Journal of Economic Behavior & Organization*, 163, 117-134.

9. Perach, N. and Anily, S. (2022). Stable matching of student-groups to dormitories. *European Journal of Operational Research*, 302(1), 50-61.

10. Kökalan, Ö. and Korkmaz, C. (2019). Measurement of service quality in public and private/foundation university dormitories: comparison of Turkish and foreign students. *Yüksek Öğretim Dergisi*, 9(3), 22-36.

11. Aslanargun, E. (2014). Türkiye'de yükseköğretim ve öğrenci yurt hizmetleri memnuniyeti. *Eğitime Bakış Eğitim-Öğretim ve Bilim Araştırma Dergisi*, 10(30), 30-34.

12. Çavlin, A. and Bingöl, Ş. (2022). Gender difference in internal educational migration: distance analysis between hometowns and universities of Istanbul. *Sosyoloji Araştırmaları Dergisi*, 25(3), 432-449.

13. Eşidir, Y. (2017). *Yükseköğretimde barınma politikalarının değerlendirilmesi ve Türkiye için öneriler*. Ankara: Kalkınma Bakanlığı, 48-52.

14. Perez Rave, J. I., Jaramillo Álvarez, G. P. and Gonzalez Echavarria, F. (2022). A psychometric data science approach to study latent variables: a case of class quality and student satisfaction. *Total Quality Management & Business Excellence*, 33(13-14), 1547-1572.

15. De Caigny, A., Coussement, K. and De Bock, K. W. (2018). A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research*, 269(2), 760-772.

16. Xu, T., Ma, Y. and Kim, K. (2021). Telecom churn prediction system based on ensemble learning using feature grouping. *Applied Sciences*, 11(11), 4742.

17. Amin, A., Anwar, S., Adnan, A., Nawaz, M., Alawfi, K., Hussain, A. and Huang, K. (2017). Customer churn prediction in the telecommunication sector using a rough set approach. *Neurocomputing*, 237, 242-254.

18. Wu, S., Yau, W.-C., Ong, T.-S. and Chong, S.-C. (2021). Integrated churn prediction and customer segmentation framework for telco business. *IEEE Access*, 9, 62118-62136.

19. Ahmad, A. K., Jafar, A. and Aljoumaa, K. (2019). Customer churn prediction in telecom using machine learning in big data platform. *Journal of Big Data*, 6(1), 1-24.

20. Usman-Hamza, F. E., Balogun, A. O., Capretz, L. F., Mojeed, H. A., Mahamad, S., Salihu, S. A., Akintola, A. G., Basri, S., Amosa, R. T. and Salahdeen, N. K. (2022). Intelligent decision forest models for customer churn prediction. *Applied Sciences*, 12(16), 8270.

21. Lalwani, P., Mishra, M. K., Chadha, J. S. and Sethi, P. (2022). Customer churn prediction system: a machine learning approach. *Computing*, 104, 271-294.

22. Suh, Y. (2023). Machine learning based customer churn prediction in home appliance rental business. *Journal of Big Data*, 10(1), 41.

23. Meghyasi, H. and Rad, A. (2020). Customer churn prediction in telecommunication industry using data mining methods. *Innovaciencia*, 8(1), 1-8.

24. Han, J., Kamber, M., and Pei, J. (2012). *Data mining: Concepts and techniques*. Massachusetts: Morgan Kaufmann, 9-383.

25. Mishra, P., Varadharajan, V., Tupakula, U. and Pilli, E. S. (2018). A detailed investigation and analysis of using machine learning techniques for intrusion detection. *IEEE Communications Surveys & Tutorials*, 21(1), 686-728.

26. Uğuz, H. (2011). A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm. *Knowledge-Based Systems*, 24(7), 1024-1032.

27. Niu, K., Jiao, H., Gao, Z., Jia, G., Yang, G. and Cheng, C. (2017). *A developed feature selection method for classification based on united information gain*. 2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), San Francisco, 1-4.

28. Kowshalya, A. M., Madhumathi, R. and Gopika, N. (2019). Correlation based feature selection algorithms for varying datasets of different dimensionality. *Wireless Personal Communications*, 108, 1977-1993.

29. İnternet: Kotan, M. (t.y). *Veri madenciliği*. Web: https://mkotan.sakarya.edu.tr/sites/mkotan.sakarya.edu.tr/file/EKO469_VM_H10_Siniflandirma.pdf, Son Erişim Tarihi:28/12/2023.

30. Budak, H. (2018). Özellik seçim yöntemleri ve yeni bir yaklaşım. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 22, 21-31.

31. Çınar, A. (2019). Veri madenciliğinde sınıflandırma algoritmalarının performans değerlendirmesi ve R dili ile bir uygulama. *Marmara Üniversitesi Öneri Dergisi*, 14(51), 90-111.

32. Uddin, S., Khan, A., Hossain, M. E. and Moni, M. A. (2019). Comparing different supervised machine learning algorithms for disease prediction. *BMC Medical Informatics and Decision Making*, 19(1), 1-16.

33. AL-Najjar, D., Al-Rousan, N. and AL-Najjar, H. (2022). Machine learning to develop credit card customer churn prediction. *Journal of Theoretical and Applied Electronic Commerce Research*, 17(4), 1529-1542.

34. Rausch, T. M., Derra, N. D. and Wolf, L. (2022). Predicting online shopping cart abandonment with machine learning approaches. *International Journal of Market Research*, 64(1), 89-112.

35. Boser, B. E., Guyon, I. M. and Vapnik, V. N. (1992). *A training algorithm for optimal margin classifiers*. Proceedings of the fifth annual workshop on Computational learning theory, 144-152.

36. Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20, 273-297.

37. İnternet: Şener, Y. (t.y.). *Destek vektör makineleri (support vector machine | svm) çalışma mantığı ve python uygulaması*. Web: https://yigitsener.medium.com/destek-vektor-makineleri-support-vector-machine-svm-calisma-mantigi-ve-python-uygulamasi-992163ff3eec, Son Erişim Tarihi:26/04/2023.

38. Tuncer, E. ve Bolat, E. D. (2022). Destek vektör makineleri ile EEG sinyallerinden epileptik nöbet sınıflandırması. *Politeknik Dergisi*, 25(1), 239-249.

39. Jain, A. K., Mao, J. and Mohiuddin, K. M. (1996). Artificial neural networks: A tutorial. *Computer*, 29(3), 31-44.

40. Karoğlu, T. T. ve Hayrettin, O. (2019). Türkiye'de üniversiteye yerleşme başarısının bootstrap örnekleme yöntemi kullanılarak yapılan hiyerarşik kümeleme analizi ve iki yönlü kümeleme analiziyle incelenmesi. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 23(3), 925-935.

41. İnternet: Bilişim Kitabı. (t.y.). *AdaBoost algoritması*. Web: https://bilisimkitabi.com/adaboost-algoritmasi, Son Erişim Tarihi:12/12/2023.

42. İnternet: Medium. (t.y). *Machine learning-clustering (kümeleme)-k-means algorithm-hierarchical clustering (hiyerarşik kümeleme)-part 13*. Web: https://medium.com/@ekrem.hatipoglu/machine-learning-clustering-kumeleme-k-means-algorithm-part-13-be33aeef4fc8 Son Erişim Tarihi:26/11/2023.

43. Agrawal, R. and Srikant, R. (1994). *Fast algorithms for mining association rules*. Proc. 20th int. conf. very large data bases, VLDB, 487-499.

44. İnternet: Towards Data Science. (t.y). *Understanding AUC - ROC Curve*. Web: https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5, Son Erişim Tarihi:12/11/2023.

45. İnternet: WEKA. (t.y). *Genel*. Web: https://www.cs.waikato.ac.nz/ml/weka/index.html, Son Erişim Tarihi:22/08/2023.

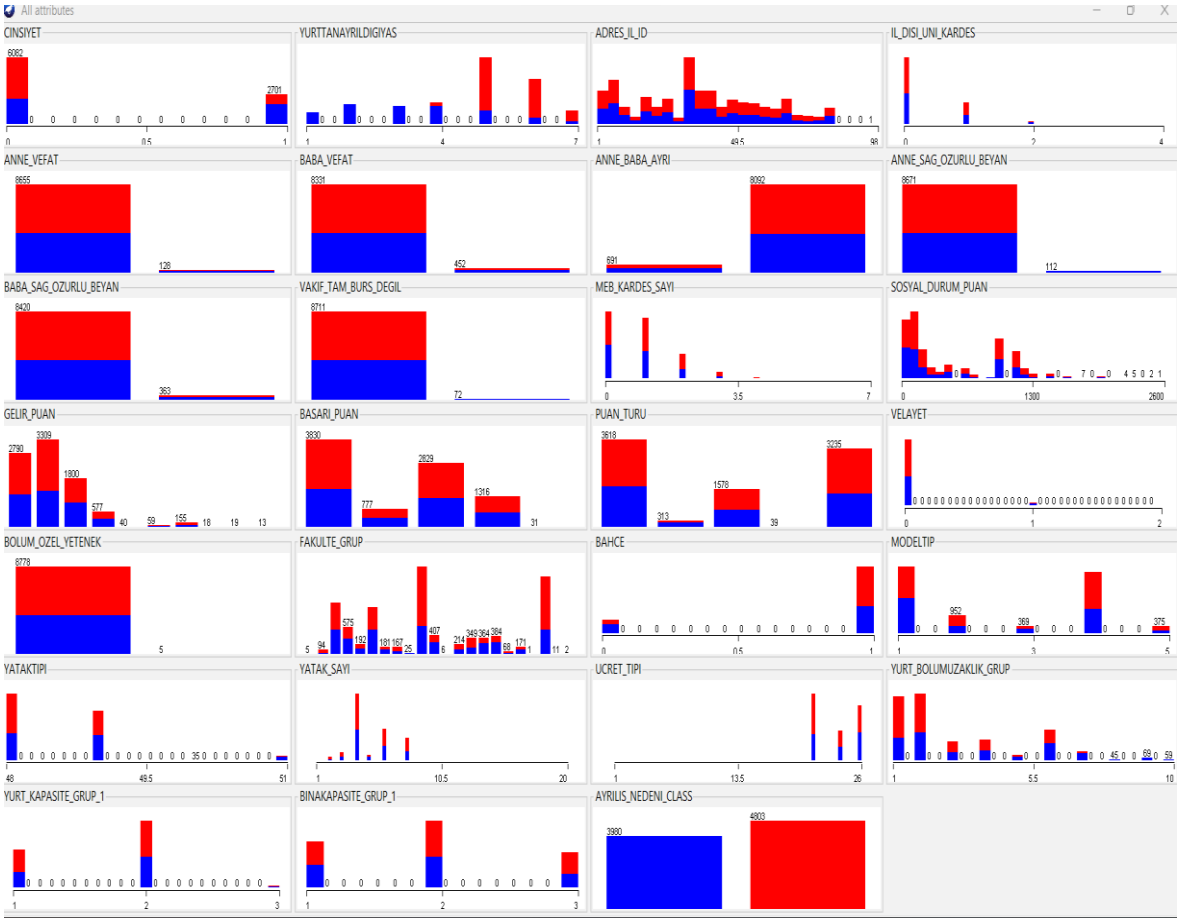
46. İnternet: Gençlik ve Spor Bakanlığı. (t.y.). *Yurt Hizmetleri Yönetmeliği, Barınma hizmetleri*. Web: https://www.mevzuat.gov.tr/mevzuat?MevzuatNo=38510&MevzuatTur=7&MevzuatTertip=5, Son Erişim Tarihi:26/09/2023.

47. Kara, Ş. E. ve Şamlı, R. (2021). Yazılım projelerinin maliyet tahmini için WEKA'da makine öğrenmesi algoritmalarının karşılaştırmalı analizi. *Avrupa Bilim ve Teknoloji Dergisi*(23), 415-426.

48. Wong, T.-T. and Yeh, P.-Y. (2020). Reliable accuracy estimates from k-fold cross validation. *IEEE Transactions on Knowledge and Data Engineering*, 32(8), 1586-1594.

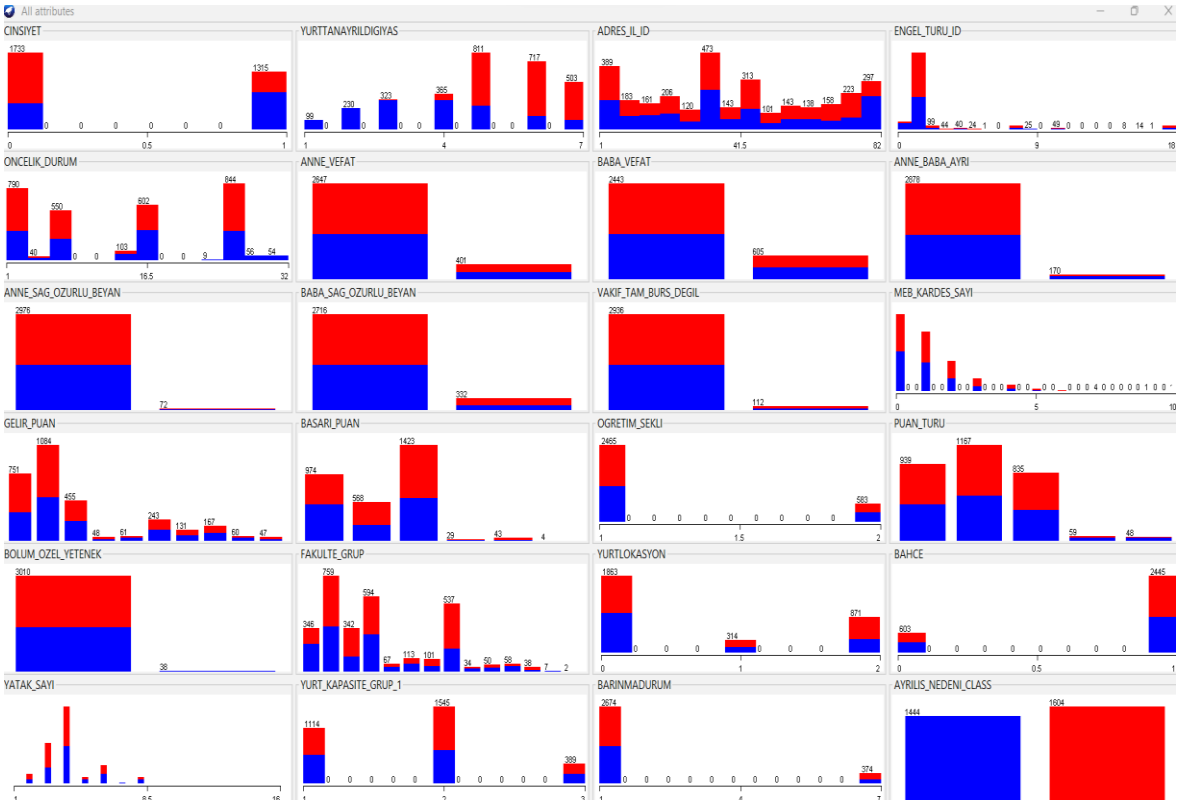
EKLER

EK-1. WEKA yurt öğrenci veri kümesi öznitelikleri dağılımı ekran görüntüsü



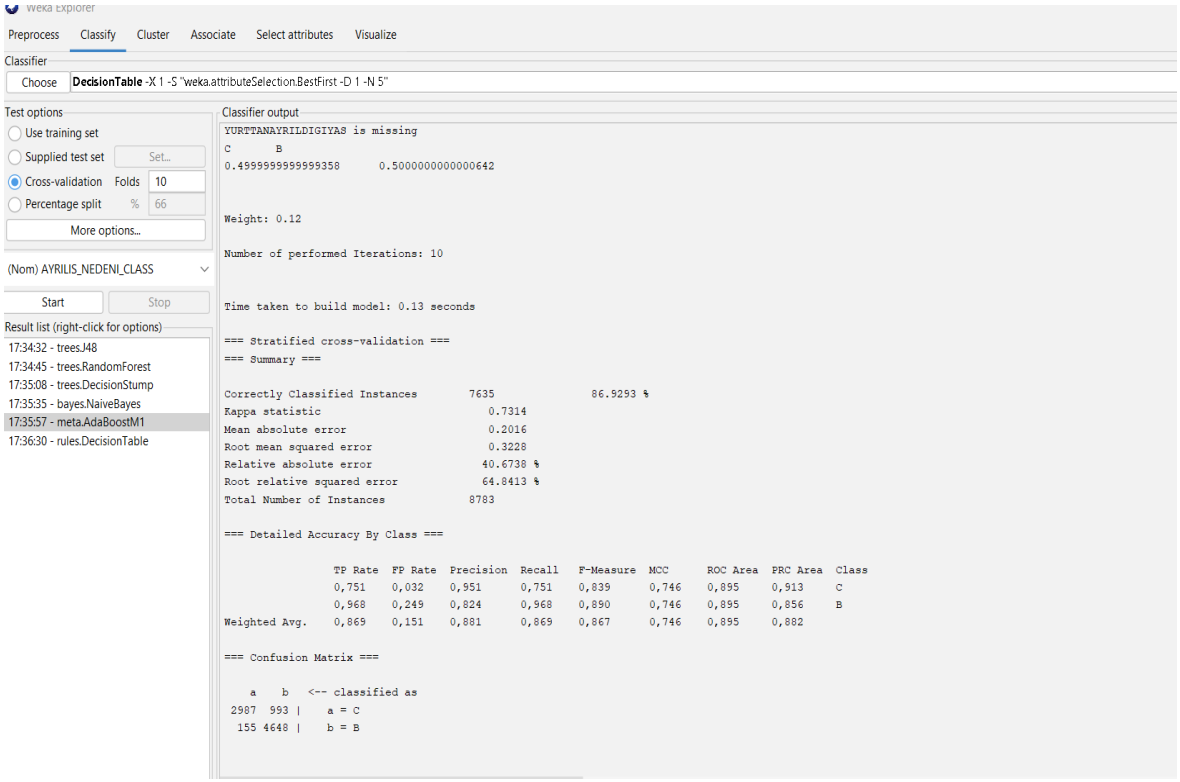
Şekil 1.1. WEKA yurt öğrenci veri kümesi öznitelikleri dağılımı ekran görüntüsü

EK-2. WEKA öncelikli öğrenci veri kümesi öznitelikleri dağılımı ekran görüntüsü



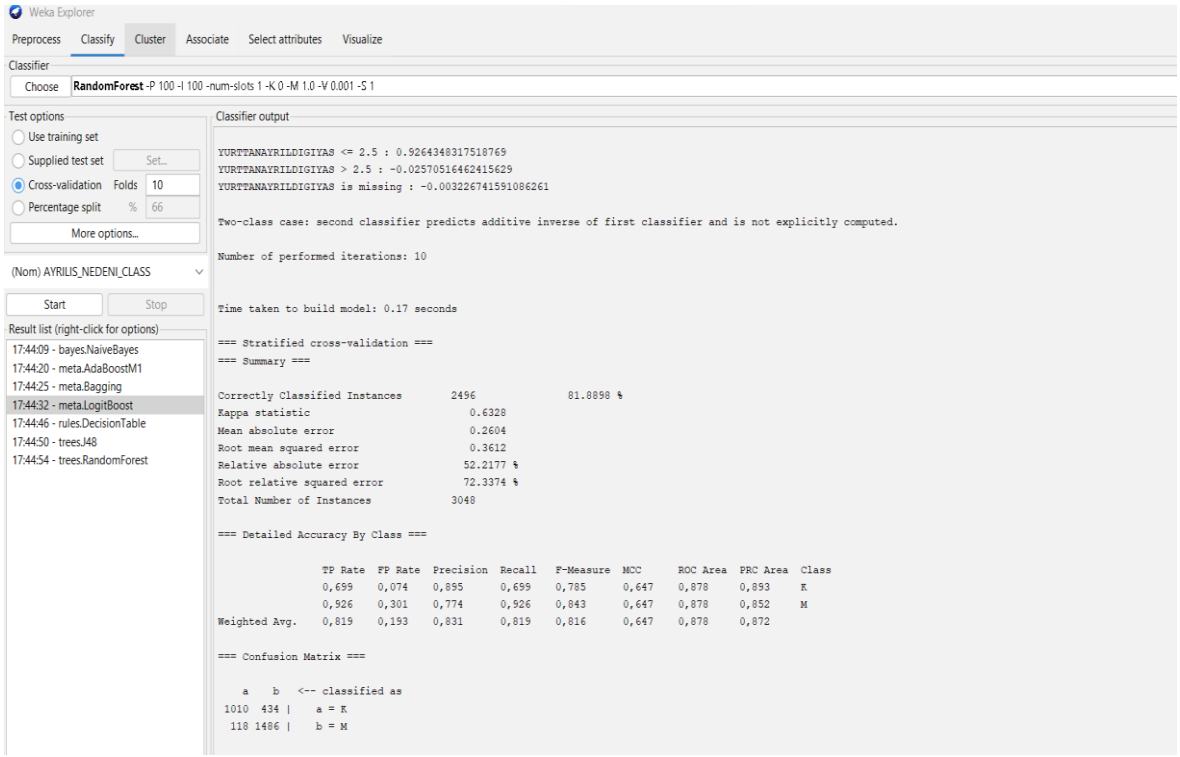
Şekil 2.1. WEKA öncelikli öğrenci veri kümesi öznitelikleri dağılımı ekran görüntüsü

EK-3. WEKA yurt öğrenci veri kümesi sınıflandırma ekran görüntüsü



Şekil 3.1. WEKA yurt öğrenci veri kümesi sınıflandırma ekran görüntüsü

EK-4. WEKA öncelikli öğrenci veri kümesi sınıflandırma ekran görüntüsü



Şekil 4.1. WEKA öncelikli öğrenci veri kümesi sınıflandırma ekran görüntüsü



Gazili olmak ayrıcalıktır…