

Yurtta Barınan Öğrencilerin Ayrılma Sebeplerinin Veri Madenciliği Yöntemleriyle Analiz Edilmesi

Berrin Değirmenci*, Oktay Yıldız²

¹Bilgisayar Mühendisliği/Fen Bilimleri Enstitüsü, Gazi Üniversitesi, Ankara, Türkiye

²Bilgisayar Mühendisliği/Fen Bilimleri Enstitüsü, Gazi Üniversitesi, Ankara, Türkiye
*berrin.degirmenci@gazi.edu.tr

Abstract – Ülkemizde Üniversite öğrencilerinin büyük bir bölümü öğrenim hayatlarının bir döneminde devlet yurtlarında kalmaktadır. Bu çalışmanın amacı, Kredi ve Yurtlar Genel Müdürlüğü (KYGM)'nde barınan öğrencilerin; yurttan ayrılma sebeplerinin veri madenciliği yöntemleriyle analiz edilmesidir. Yapılan analizler sonucunda 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 sağlanmasıdır. İlk olarak veri ön işleme ve öznitelik analizi gerçekleştirilmiştir. Öznitelik seçimi, bilgi kazancı ve korelasyon öznitelik sıralama filtresi kullanılarak yapılmıştır. Modellerin aşırı uyumunu önlemek için eğitim seti üzerinde k-katlı çapraz doğrulama kullanılarak, veri setine popüler tahmin modelleri olan; lojistik regresyon, naive bayes, destek vektör makinesi, karar ağaçları, KNN, ANN vb. uygulanmıştır. Önerilen kayıp tahmin modeli; doğruluk, kesinlik, hatırlama, f-ölçüsü ve karar verici etkinliği (ROC) alanı gibi metrikler kullanılarak değerlendirilmiştir. En başarılı sonuçları; Karar Tablosu (DecisionTable) algoritması %86,95 doğruluk ve %89 ROC alanı, Adaboost algoritması %86,93 doğruluk ve %89 ROC alanı ve Karar Kütüğü (Decision Stump) algoritmasının %86,93 doğruluk ve %85 ROC alanı ile göstermiştir.

Keywords – Yurt Memnuniyeti, Sınıflandırma, Kümeleme, Makine Öğrenmesi, Veri Madenciliği

I. GİRİŞ

Yükseköğrenim öğrencileri barınma yerleri olarak; kamu/vakıf/dernek ve özel sektöre ait yurtları, apart evleri, özel pansiyonları, kiralık evleri ya da akrabalarının yanında kalmayı seçebilirler. Barınma yerlerinin seçiminde ise, öğrencinin kendisinin ve ailesinin sosyo-ekonomik durumu, güvenlik, kiralar, okula yakınlık ve ulaşım kolaylığı, üniversite öğreniminin sürdürüldüğü şehirdeki konutların kapasite ve durumları gibi birçok faktör etkili olmaktadır [1].

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 yurttan 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.

KYGM amaçları incelendiğinde öğrencilerin yurtlardan memnun kalması, kaliteli hizmet alması ve öğrenim hayatları boyunca yurttan ayrılmamaları istenmektedir. Çalışmamız ile Kuruma yurttan ayrılması tahmin edilecek öğrenciler hakkında fikir verilebilir.

Yurt dışındaki çalışmalarda; üniversite öğrencilerinin yurttan barınmasının lokal ve öğrenim gördükleri üniversite kampüsü ile sınırlı olduğu görülmüştür, ülkemizdekine benzer yurt yerleştirme çalışmaları [2], [3] bulunmaktadır. Öğrenci

yurtlarının hizmet kalitesinin ölçüldüğü bir çalışma yapılmıştır [4]. Yükseköğrenimde yurt hizmetleri ve barınma hakkında bir araştırma yapılmıştır [1].

Yurttan ayrılma nedenlerinin analiz edilmesi ile ilgili ülkemizde çok fazla çalışma bulunmamaktadır; ancak konuya benzer müşteri kaybı araştırmaları mevcuttur. Literatürde makine öğrenmesi, veri madenciliği ve hibrit teknikler kullanılarak kayıp tahmini yapılmıştır. Bu teknikler, şirketlerin kayıp müşterileri belirlemesini, tahmin etmesini ve elde tutmasını sağlar.

Literatürdeki benzer çalışmalardan bazı örnekler aşağıda verilmiştir. Müşterileri sınıflandırmak için çeşitli algoritmaları kıyaslayan çalışmada en yüksek başarı %88,63 ile Rastgele Orman (RF)'dir [5]. Başka bir makalede Logit yaprak modeli (LLM) diye hibrit bir model önerilmiştir. Logistik regresyon ve karar ağacının birleşimi bir modeldir. Bileşenleriyle mukayese edildiğinde daha başarılıdır [6]. Bu çalışma; özellik oluşturma, yığınlama modeli (logistik regresyon, karar ağacı, naive bayes) ve yumuşak oylamadan oluşan yeni bir müşteri kaybı tahmin sistemi önermektedir, %96.12 başarı elde edilmiştir [7]. Diğer bir çalışma, müşteri kaybı ve kayıpsızlık ile ilgili önemli karar kurallarını çıkarmak için kaba küme teorisine (RST) dayanan akıllı bir kural tabanlı karar verme tekniği önermektedir. 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

[8]. Önerilen çerçeve, müşteri kaybını daha iyi yönetmek için telekomünikasyon operatörlerine eksiksiz bir kayıp analizi sağlamak için kayıp tahmini ve müşteri segmentasyon sürecini birleştirir. Altı makine öğrenimi sınıflandırıcısı ile yapılan deneylerde üç veri seti kullanarak müşteri kayıp durumu tahmin edilir. 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 [9]. Müşteri kayıp tahminini (CCP) için akıllı karar ormanı (DF) modelleri geliştirmişlerdir. Lojistik model ağacı (LMT), rastgele orman (RF) ve İşlevsel Ağaçlar (FT), DF modelleri ile karşılaştırılmış ve önerilen DF modellerinin telekom endüstrisinde üstün tahmin performansı gösterdiği görülmüştür [10]. Yapılan diğer bir çalışmada Telekom sektöründeki müşteri kaybının tahmini için çeşitli makine öğrenmesi teknikleri uygulamışlardır. Adaboost ve XGboost Sınıflandırıcının sırasıyla %81,71 ve %80,8 ile en yüksek doğruluğu verdiği tespit edilmiştir [11].

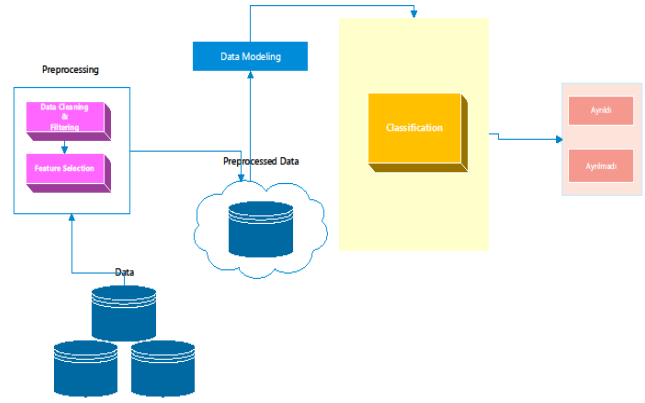
Yaptığımız çalışmada öğrencilerin yurttan ayrılma sebepleri veri madenciliği yöntemleri ile incelenmiş ve yurttan erken ayrılan öğrencilerin tespitine çalışılmıştır. Çalışmanın amacı, makine öğrenmesi ve veri madenciliğinde mevcut teknikleri araştırmak ve ayrılan öğrenci tahminleri için bir model önermek,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

Bu makalede, çeşitli veri madenciliği algoritmaları kullanan ayrılan öğrenci tahmin modeli önerildi. Bir sınıflandırıcının performansı, mevcut veri setine bağlıdır. Veri seti olarak Kredi ve Yurtlar Genel Müdürlüğü'nün Ankara ilinde bulunan yurtlarında barınan öğrencilerinin 2016-2022 yılları arasındaki verileri kullanılmıştır. Önemli öznitelikler, bilgi kazancı ve korelasyon öznitelik sıralama filtresi 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 alanı kullanılarak modelin performansı ölçülmüştür. Deneylerden, önerilen modelin yüksek doğruluk elde ederek, DecisionTable 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.

II. METARYALLER VE YÖNTEM

Ayrılan öğrenci tahmin modeli; öğrenci kaybını öngörmeyi ve uzun süreli barınacak öğrencileri elde tutmayı amaçlamaktadır. Öğrenci kaybının önlenmesi için uygun önlemlerin alınmasını sağlamaktır.

Bu, ayrılıp ayrılmayacakların tahmin edildiği ikili bir sınıflandırma problemidir. Bu sorunun üstesinden gelebilmek için; daha önce yapılan çalışmalara dayanarak oldukça etkili teknikler olan; doğrusal regresyon, destek vektör makinesi, naive bayes, karar ağacı, rastgele ormanı vb. içeren makine öğrenimi algoritmaları [9], [11], [12] kullanılır. Makine öğrenimi modellerinde, ön işleme sonrası özellik seçimi, sınıflandırma doğruluğunun iyileştirilmesinde önemli bir rol oynar.



Şekil 1. Ayrılan Öğrenci Analizi Çerçevesi

Şekil 1 Ayrılan Öğrenci Analizi çerçevesini göstermekte ve adımlarını açıklamaktadır. İlk adımda; gürültü giderme, veri filtreleme, dengesiz veri özellikleri kaldırma ve verilerin normalleştirilmesini içeren veri ön işleme yapılır. Önemli Özniteliklerin seçimi bilgi kazancı ve korelasyon öznitelik sıralama filtresi kullanılarak yapılmıştır. Sorunun çözümü için naive bayes, lojistik regresyon, geliştirme (boosting), torbalama, karar tablousu, karar kütüğü, karar ağacı, rastgele orman, destek vektör makinesi, , en yakın komşu, yapay sinir ağları (MLP), vb. makine öğrenmesi teknikleri kullanılmıştır. Tahminler yapılmış ve sınıflandırıcıların başarıları; doğruluk, kesinlik, hatırlama f-ölçüsü ve karar verici etkinliği (ROC) alanı gibi metrikler kullanılarak değerlendirilmiştir.

A. Veri Ön İşleme

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

Yapılan çalışmada, Yurt Öğrenci veri seti içindeki 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 Tablo 1'de gösterilen 34 öznitelikli 8783 veri kaydı sınıflandırmaya hazır hale getirilmiştir.

Tablo 1. Seçim Yapılacak Öznitelikler

	Öznitelikler		Öznitelikler
1	CINSİYET	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	MODELTIPI
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

B. Öznitelik Seçimi

Öznitelik seçimi, çalışılan problem için en önemli ve en yararlı özellikleri seçerek veri setindeki özellik sayısının azaltılmasını amaçlamaktadır. Algoritma hızını artırır, ilgisi olmayan ya da ilgisi düşük olan veriyi ortadan kaldırır, veri kalitesini artırır, veri setini daha basit ve anlaşılabilir hale getirir, önerilen modelin başarısını artırır [13]. Bu nedenle öznitelik seçimi önemlidir. Literatürde öznitelik seçimi için bir takım teknikler [2], [14], [15], [16] mevcuttur.

Bilgi kazancı, bir veri kümesinin dönüşümünden kaynaklanan entropideki azalmayı hesaplar. Her değişken bilgi kazancını hedef değişken bağlamında değerlendirerek özellik seçimi için kullanılabilir. Korelasyon ise iki varlık arasındaki istatistiksel ilişki olarak tanımlanır. Makine öğreniminde korelasyon, iki veya daha fazla özelliğin 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.

Bu çalışmada öznitelik seçimi, Bilgi Kazancı ve Korelasyon Nitelikleri Sıralama Filtresi teknikleri WEKA araç setini kullanılarak yapıldı. Yurt Öğrenci veri setinde, her iki tekniğin Tablo 2’de gösterilen sonuçlarına göre yüksek sıralama değerlerine sahip olan toplam 34 özelliğten ilk 27’si seçildi. KALDIGI_AY_1 sınıflandırmayı direkt etkilediğinden çıkarılmıştır. Bu çalışmada kullanılan veri seti 26 öznitelik içermektedir. Böyle yüksek boyutlu bir veri setinde, bazı nitelikler performans ölçümünü iyileştirir ve karar verme süreci için faydalıdır, diğerleri ise daha az önemli niteliklerdir. Öznitelik seçiminde bu norm göz önüne alınmıştır.

Tablo 2. Bilgi Kazancı ve Korelasyon Öznitelik Seçim filtresi sıralama sonuçları

No	Öznitelik	Bilgi Kazancı Sıralaması	Korelasyon Sıralaması
34	KALDIGI_AY_1	0.64249692	0.726607
3	YURTTANAYRILDIGIYAS	0.49232142	0.670929
1	CINSIYET	0.04970366	0.261717
24	FAKULTE GRUP	0.04834594	0.079863
31	YURT_BOLUMUZAKLIK_GRUP	0.03705054	0.20598
26	BAHCE	0.02653318	0.191283
27	MODELTIPI	0.01884151	0.107257
28	YATAKTIPI	0.01873054	0.009715
16	MEB_KARDES_SAYI	0.01191664	0.127498
30	UCRET TIPI	0.00871423	0.089996
22	VELAYET	0.00671395	0.087683
12	ANNE_BABA_AYRI	0.00661807	0.09589
21	PUAN_TURU	0.00512664	0.031489
18	GELIR_PUAN	0.00503488	0.041905
4	ADRES_IL_ID	0.00409062	0.036573
17	SOSYAL_DURUM_PUAN	0.00389283	0.021928
15	VAKIF_TAM_BURS_DEGIL	0.00369065	0.069434
29	YATAK_SAYI	0.0036464	0.009715
33	BINAKAPASITE GRUP 1	0.00344194	0.063874
32	YURT_KAPASITE GRUP 1	0.00269516	0.05618
19	BASARI_PUAN	0.00165298	0.024623
11	BABA_VEFAT	0.00112052	0.039521

8	IL_DISI_UNI_KARDES	0.00084824	0.033667
23	BOLUM_OZEL_YETENEK	0.00065037	0.026218
10	ANNE_VEFAT	0.00026113	0.01908
13	ANNE_SAG_OZURLU_BE_YAN	0.00025486	0.01885
14	BABA_SAG_OZURLU_BE_YAN	0.00000212	0.001715
7	BASVURU_TURU	0	0.012634
2	MEDENIDURUM	0	0.000649
20	OGRETIM_SEKLI	0	0.009714
9	IL_ICI_UNI_KARDES	0	0.002055
5	ENGEL_TURU_ID	0	0.000454
6	ONCELIK_DURUM	0	0.009035
25	YURTLOKASYON	0	0.024252

C. Sınıflandırma Algoritmaları

Sınıflandırma; tahmin edici bir yöntemdir, denetimli bir öğrenme algoritmasıdır. Sınıflandırıcılar bir veri setinin, içerdiği ortak özelliklerine göre belirli kategorilere ayrılmasını sağlamaktadır. Bu amaçla çeşitli algoritmalar geliştirilmiştir [17].

Yurt Öğrenci veri setinde iki tür öğrenci vardır. Birincisi, mezun öğrenciler; yurttan mezun olana kadar kalırlar. İkinci tip ise yurttan kendi isteğiyle vaktinden önce ayrılan öğrencilerdir. Önerilen model, yurttan ayrılan öğrenciyi hedefliyor ve ayrılışın arkasındaki nedenleri tanımlıyor. Bu çalışmada, öğrenci verilerini sınıflandırmak için bir dizi makine öğrenme tekniği kullanılmıştır. Algoritmalarından hangisinin öğrencileri en iyi şekilde sınıflandırdığını değerlendirmektir. 10-k çapraz doğrulama ile J48, Karar Kütüğü, Karar Tablosu, Adaboost, LogitBoost, Torbalama, Naive Bayes (NB), Rastgele Orman (RF), Çok Katmanlı Algılayıcı (MLP), Lojistik Regresyon (LR), SVM ve KNN kullanıldı. Sınıflandırma işlemi WEKA 3.8.6 araç takımı kullanılarak gerçekleştirilmiştir.

Aşağıda; güvenilirliği, verimliliği ve popülerliği dikkate alınarak, makine öğrenmesi için kullanılan altı köklü teknik sunulur [11], [12], [13], [18], [19].

Karar Ağaçları, belirli bir veri kümesi için sınıflandırma kuralları oluşturabilen karar kümelerini temsil eden ağaç şeklindeki yapılardır. Müşteri kaybı probleminde, yüksek doğruluğa sahiptir. Destek vektör makinesi, verileri çok boyutlu bir girdi alanına haritalayarak ve alanı homojen bölümlere ayıran optimal bir hiperdüzlem oluşturarak örüntü tanıma problemlerini çözmektedir [20]. Yeni örneklerin tahminleri daha sonra bu bölümlere göre sınıflandırılır. Yapay sinir ağları, insan özelliklerinden biri olan öğrenme ile yeni bilgiler türetme, oluşturma gibi yetenekleri gerçekleştiren bilgisayar sistemleridir. Bayes sınıflandırıcısı, Bayes teoremini güçlü (naive) bağımsızlık varsayımlarıyla uygulamaya dayanan 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 [18]. Naive Bayes sınıflandırıcısının birçok uygulamada faydalı olduğu görülmektedir. Regresyon analizi değişkenler arasındaki ilişkiyi ölçmek için kullanılan istatistiksel bir süreçtir. Bir bağımlı değişkenin birden fazla bağımsız değişken arasındaki ilişkiye odaklanıldığında, çeşitli değişkenlerin modellenmesi ve analiz edilmesi için birçok teknik içerir. Topluluk öğrenme,

bir dizi temel sınıflandırıcı tarafından yapılan tahminleri birleştirir. Çoklu makine öğrenimi algoritmalarının kullanılması, bir sorun için bir dizi hipotez oluşturulmasına sağlar. Sınıflandırıcılar topluluğu, ortak bir sonuç üretmek için hipotezleri birleştirir [13]. Örneğin; rastgele orman, birden çok karar ağacının tahminlerini birleştiren topluluk sınıflandırıcılarından biridir. Topluluk öğrenmesinin popüler yöntemlerinden bazıları torbalama, güçlendirme, çoğunluk oylama ve istiflemedir.

D. Model Başarısının Değerlendirme Yöntemi

Sınıflandırıcıların, tahminindeki performansını uygun parametrelerle değerlendirmek amacıyla, Tablo 3'de gösterilen karışıklık matrisinin içeriğinden hesaplanan kesinlik, hatırlama, doğruluk ve F-ölçüsü ölçümlerini kullanıyoruz [11].

Tablo 3. Karışıklık Matrisi

	Ayrılanlar (Kendi İsteğiyle Yurttan Ayrılan Öğrenciler)	Ayrılmayanlar (Mezun olan Öğrenciler)
Ayrıldı	Tp	Fn
Ayrılmadı	Fp	Tn

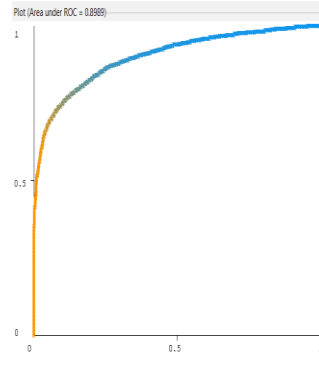
Sınıflandırıcıların performansları karışıklık matrisi çıkarıldıktan sonra, farklı ölçütler ile kıyaslanmıştır. Bu ölçütler doğruluk oranı (Tp Rate), yanlışlık oranı (Fp Rate), başarı oranı(A), hatırlama(R), kesinlik (P), F-ölçüsü (F), AUC-ROC eğrileridir. Başarı oranı (A): Toplam veri içindeki doğru tahmin oranıdır, duyarlılık (R): Pozitif olarak tahmin edilmesi gerekenlerin, kaçının doğru tahmin edildiğini gösteren metriktir, kesinlik (P): Pozitif olarak tahmin edilen örneklerin kaçının doğru tahmin edildiğinin oranıdır, F-Ölçüsü (F): kesinlik ve duyarlılığın harmonik ortalamasıdır [18].

III. SONUÇ VE TARTIŞMA

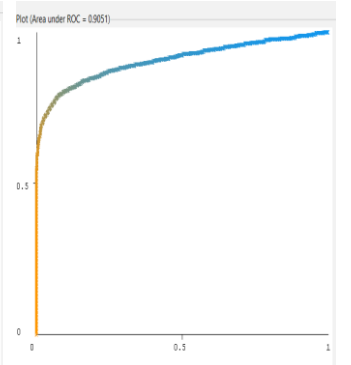
Deneyisel çalışmanın sonuçları, tablo 4'te gösterilmiştir. Doğru tahmin oranı, yanlış tahmin oranı, kesinlik, hatırlama, f-ölçüsü, ROC alanı değerlerine göre sınıflandırıcıların başarıları belirlenmiştir. Karar Tablosu (Decision Table) en yüksek doğruluk oranına sahiptir. ROC alanı değeri 1.0'e eşitse, bu mükemmel bir tahmindir.

Tablo 4. Sınıflandırıcıların Performansları

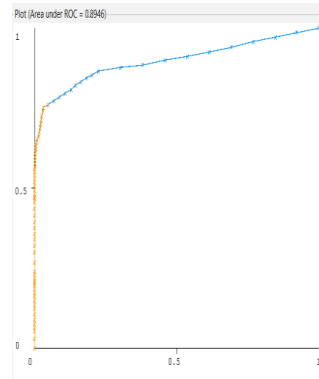
Classifier	Accurac y	TP Oran	FP Oranı	Precisi on	Recall	F-Measu re	ROC Area
Naive Bayes	82,9329	0,766	0,118	0,843	0,766	0,803	0,899
LogisticRegre syon	86,2689	0,79	0,077	0,894	0,79	0,839	0,905
AdaBoost	86,9293	0,751	0,032	0,951	0,751	0,839	0,895
Bagging	86,36	0,766	0,055	0,92	0,766	0,836	0,912
LogitBoost	86,8041	0,754	0,038	0,943	0,754	0,838	0,915
Decision Table	86,9521	0,753	0,034	0,949	0,753	0,839	0,89
Decision Stump	86,9293	0,751	0,032	0,951	0,751	0,839	0,851
J48	86,5536	0,778	0,062	0,912	0,778	0,84	0,868
RandomFores t	86,3031	0,783	0,071	0,902	0,783	0,838	0,916
SVM	86,4966	0,771	0,057	0,918	0,771	0,838	0,857
KNN	72,6176	0,663	0,221	0,713	0,663	0,687	0,72
ANN (Multiplayer Perceptron)	84,4131	0,792	0,112	0,854	0,792	0,822	0,898



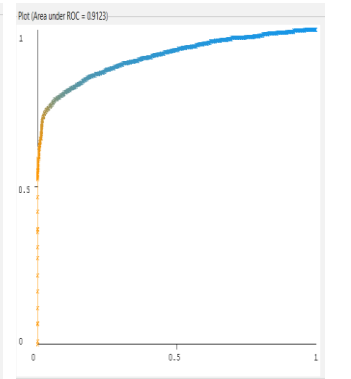
Şekil 2 Naive Bayes ROC Alanı



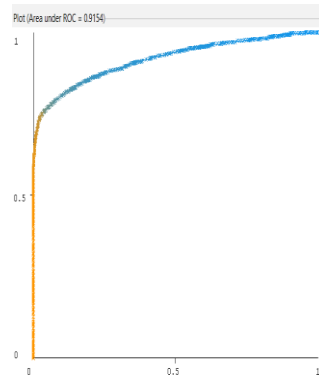
Şekil 3 LogisticRegresyon ROC Alanı



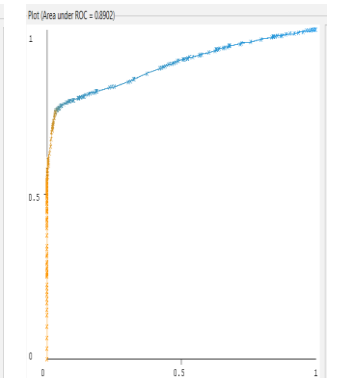
Şekil 4 AdaBoost ROC Alanı



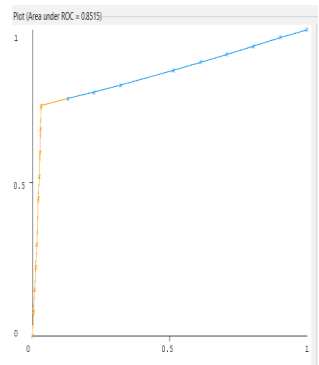
Şekil 5 Bagging ROC Alanı



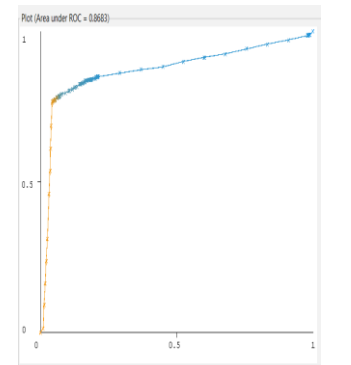
Şekil 6 LogitBoost ROC Alanı



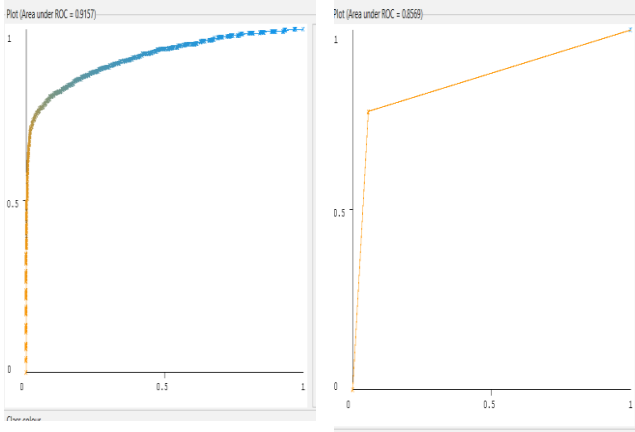
Şekil 7 Decision Table ROC Alanı



Şekil 8 Decision Stump ROC Alanı

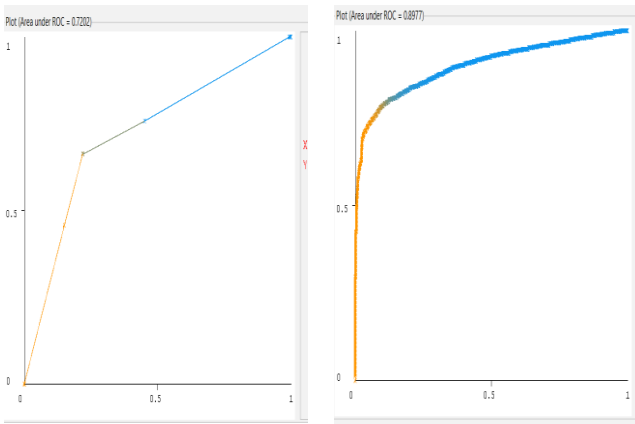


Şekil 9 J48 ROC Alanı



Şekil 10 Random Forest ROC Alanı

Şekil 11 SVM ROC Alanı



Şekil 12 KNN ROC Alanı

Şekil 13 MLP ROC Alanı

Literatür taraması yapılmış yurttan ayrılma nedenlerinin analiz edilmesine benzer pek fazla çalışma olmadığı görülmüştür ancak bu probleme benzer müşteri kaybı analizleri yapılmıştır. Bu çalışmalar incelenmiştir, kullanılan yöntemlerin zayıf ve güçlü yönleri ele alınmıştır. Sorunun çözümüne uygun bir model önerilmiştir.

Yaptığımız deneysel çalışmalar neticesinde Tablo 4'te gösterildiği gibi topluluk öğrenimi algoritmalarının daha başarılı sonuçlar verdiği tespit edilmiştir. Karar Tablosu (DecisionTable) algoritmasının %86,95 doğruluk ve %89 ROC alanı, Adaboost algoritmasının %86,93 doğruluk ve %89 ROC alanı ve Karar Kütüğü (Decision Stump) algoritmasının %86,93 doğruluk ve %85 ROC alanı ile iyi performans gösterdiği görülmüştür. %91,6 ile en iyi ROC alanına sahip Random Forest %86,3 doğruluk oranı göstermiştir.

Çalışmasının sonucunda öğrencilerin özelliklerine göre yurttan ayrılma nedenleri tespit edilmiş, erken dönemde yurttan ayrılan öğrenciler başarılı şekilde bulunmuştur. Çıkan sonuçların değerlendirilmesi ile 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.

Gelecek çalışmalarda Ankara ilinde yurttan ayrılan ve dört yıl öğrenim gören öğrenciler için alınan verilerin diğer illere ve öğrencilere de yaygınlaştırılması, dezavantajlı grup için çalışmanın derinleştirilmesi planlanmaktadır.

KAYNAKLAR

- [1] E. Kavuncu, "Türkiye'de Üniversitelerde Öğrenci-Yurt Hizmetleri ve Barınma Sorunu", *Eğitim-Öğretim ve Bilim Araştırma Dergisi*, sayı:30, pp. 48-50, 2014.
- [2] P. Frijters et al., "Heterogeneity in peer effects in random dormitory assignment in a developing country", *Journal of Economic Behavior and Organization*, 163, pp.117-134,2019.
- [3] N. Perach et al. "Stable matching of student-groups to dormitories", *European Journal of Operational Research*, 302 (1) , pp.50-61, 2022.
- [4] Ö. Kökalan et al., "Measurement of Service Quality in Public and Private/Foundation University Dormitories: Comparison of Turkish and Foreign Students", *Yüksek Öğretim Dergisi*, 9(3), pp.263-278, 2019
- [5] I. Ullah et al., "A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector", *IEEE Access*, vol:7, pp: 60134 – 60149, 2019.
- [6] A. De Caigny et al., "A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees", *European Journal of operational research*, vol:269 (2), pp.760-772,2018.
- [7] Tianpei Xu. et al., "Telecom Churn Prediction System Based on Ensemble Learning Using Feature Grouping", *Applied Sciences-Basel*, 11(11), doi: 10.3390/app11114742, 2021.
- [8] A. Amin et al., "Customer churn prediction in the telecommunication sector using a rough set approach", *Neurocomputing*, vol:237, pp.242-254, 2017.
- [9] S. Wu et al., "Integrated Churn Prediction and Customer Segmentation Framework for Telco Business", *IEEE ACCESS*, Vol. 9, 62118-62136, 2021.
- [10] F. Usman-Hamza et al., "Intelligent Decision Forest Models for Customer Churn Prediction", *Applied Sciences*, Vol. 12,8270, 2022.
- [11] P. Lalwani et al. "Customer churn prediction system: a machine learning approach", *Computing*, Vol. 104, No:2, pp.271-294,2021.
- [12] D. Al-Najjar et al., "Machine Learning to Develop Credit Card Customer Churn Prediction", *Journal of Theoretical and applied electronic commerce research*, Vol. 17, Issue:4,pp:1529-1542,2022
- [13] P. Mishra et al., "A Detailed Investigation and Analysis of Using Machine Learning Techniques for Intrusion Detection", *IEEE Communications Surveys & Tutorials*, vol.21(1), pp. 686-728,2019
- [14] H. Uğuz, "A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm", *Knowledge-Based Systems*, Vol:24, pp:1024-1032,2011
- [15] K. Niu et al., "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),2017
- [16] AM. Kowshalya et al., "Correlation Based Feature Selection Algorithms for Varying Datasets of Different Dimensionality", *Wireless Personan Communications*, Vol.108, pp:1977-1993, 2019
- [17] A. Çınar, "Performance Evaluation Of Classification Algorithms In Data Mining And An Application With The R Language", *Marmara Üniversitesi Öneri Dergisi*, vol. 14(51), pp. 90-111,2019
- [18] S. Uddin et al. (2019), "Comparing different supervised machine learning algorithms for disease prediction", *BMC Medical Informatics and Decision Making*, 19(1), 2019
- [19] TM. Rausch et al., "Predicting online shopping cart abandonment with machine learning approaches", *International Journal of Market Research*, Vol:64, pp:89-112, 2022.
- [20] C. Cortes and V. Vapnik, *Support-Vector Networks. Machine Learning*, 20, 273-297. <http://dx.doi.org/10.1007/BF00994018>,1995