

**T.C.
GEBZE TEKNİK ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ**

**GOOGLE PLAY STORE'DAKİ MOBİL OYUNLARIN İNDİRİLME
SAYILARININ, DEĞERLENDİRME PUANLARININ VE YORUM
SAYILARININ MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE
TAHMİNİ**

**NİL ARICI
YÜKSEK LİSANS TEZİ
İŞLETME ANABİLİM DALI**

**GEBZE
2022**

T.C.
GEBZE TEKNİK ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ

**GOOGLE PLAY STORE'DAKİ MOBİL
OYUNLARIN İNDİRİLME SAYILARININ,
DEĞERLENDİRME PUANLARININ VE
YORUM SAYILARININ MAKİNE
ÖĞRENMESİ YÖNTEMLERİ İLE TAHMİNİ**

Nil ARICI
YÜKSEK LİSANS TEZİ
İŞLETME ANABİLİM DALI

Tez Danışmanı
Doç. Dr. Farid HUSEYNOV

GEBZE
2022

	YÜKSEK LİSANS JÜRİ ONAY FORMU
---	--------------------------------------

GTÜ Sosyal Bilimler Enstitüsü Yönetim Kurulu'nun 23/03/2022 tarih ve 2022/1309 sayılı kararıyla oluşturulan jüri tarafından 05/04/2022 tarihinde tez savunma sınavı yapılan Nil Arıcı'nın tez çalışması İşletme Anabilim Dalında YÜKSEK LİSANS tezi olarak kabul edilmiştir.

JÜRİ

ÜYE

(TEZ DANIŞMANI) : Doç. Dr. Farid HUSEYNOV

ÜYE

: Dr. Öğr. Üyesi Ceren ÇUBUKÇU ÇERASI

ÜYE

: Doç. Dr. İlkay KARADUMAN

ONAY

Gebze Teknik Üniversitesi Sosyal Bilimler Enstitüsü Yönetim Kurulu'nun
...../...../..... tarih ve/..... sayılı kararı.

İMZA/MÜH

ÖZET

Günümüzde her kesimden ve yaştan insan mobil cihaz sahibidir. Mobil cihazlarda oynanan android işletim sistemiyle uyumlu oyunlar ise oldukça yaygındır. Mobil oyun sektörü gün be gün büyümektedir ve milyon dolarlık bir pazar halini almıştır. Google Play Store en çok mobil oyun indirilen uygulama marketlerinden biridir. Google Play Store’da oyunların çoğu ücretsiz oyunlardan oluşmaktadır. Genel olarak oyunlar kategorisinde gelir modeli olarak freemium gelir modeli uygulanmaktadır. Yani ücretsiz bir oyunun içinde premium özellikte ücretli öğelere yer verilmektedir. Uygulama içi satın alma ögesiyle beraber reklam içerme de bir seçenektir. Reklam içerme durumunda gelir reklamdan sağlanmaktadır. Uygulama içi satın alma ile kullanıcı reklam izlemekten kurtulabilir. Bu oyunlarda puan sayısının önemi büyüktür. Ne var ki değerlendirme puanı tek başına yeterli bir ölçüt olamaz. Kullanıcıların sosyal kanıt arayışı, yüksek puanlı oyunlarda yüksek indirme ve yorum sayısı beklentisini de doğurmaktadır. Bu sebeple bir uygulama mağazaya yüklenmeden önce değerlendirme puanını tahmin etmek kadar indirme sayısını ve yorum sayısını tahmin etmek de önemlidir. Bu çalışmada Google Play Store Türkiye’den elde edilmiş yaklaşık 15.000 eşsiz veriyle mobil oyun sektörünün güncel durumu anlaşılmaya çalışıldı. Veriler Python programlama diline ait Selenium kütüphanesi kullanılarak veri kazıma yöntemi ile elde edildi. Daha sonra bilimsel hesaplama kütüphaneleri olan Numpy ve Pandas kütüphaneleri kullanılarak istatistiksel analizlere uygun hale getirildi. Mağazaya oyunlar yüklenmeden önce indirme sayısı, değerlendirme puanı ve yorum sayısı gibi bilgileri tahmin edilmeye çalışıldı. Çalışma kapsamında ikili sınıflandırma ve regresyon modelleri kullanıldı. İkili sınıflandırmada Random Forest, Naive Bayes, Decision Tree, SVC gibi algoritmalara yer verildi. Regresyon modeli olarak ise Lineer Regresyon kullanıldı. Çalışma kapsamında tanımlayıcı istatistiklere de yer verildi. Bununla birlikte oyunların ücretli olup olmadığı, reklam içerip içermediği, uygulama içi satın alma içerip içermediği gibi ikili durumların frekanslarına yer verildi. Google Play Store oyunlar kategorisiyle ilgili detaylı bilgiler içeren bu çalışma konuyla ilgilenen sektör profesyonelleri ve akademik araştırmacılar için önemli bulgular sunmaktadır.

Anahtar Kelimeler: Google Play Oyunlar, Makine Öğrenmesi, Lineer Regresyon, İkili sınıflandırma, Python, Veri kazıma

SUMMARY

Today, people from all walks of life and age own mobile devices. Games played on mobile devices, compatible with the android operating system, are quite common. The mobile game industry is growing day by day and has become a billion-dollar market. Google Play Store is one of the most used app stores for mobile game downloads. Most of the games in the Google Play Store are free games. In general, the freemium revenue model is applied as the revenue model in the games category. In other words, a free game includes paid items with premium features. Advertising content is also an option alongside in-app purchases. In case of including ads, income is provided from obligatory ad viewing. Through in-app purchase, the user can avoid obligatory advertising watching. In these games, the review score given by the user is of great importance. However, the review score alone cannot be a sufficient criterion. Users' search for social proof also creates expectations for high downloads and comments on high-rated games. That's why it's important to estimate the number of downloads and comments, as well as the rating score, before an app is uploaded to the store. In this study, it was tried to understand the current situation of the mobile game industry with approximately 15,000 unique data obtained from Google Play Store Turkey. The data were obtained by data scraping using the Selenium library of the Python programming language. Later, it was made ready for statistical analysis by using the scientific calculation libraries Numpy and Pandas. Before the games are uploaded to the store, information such as the number of downloads, rating points and comments were tried to be estimated. Within the scope of the study, binary classification and regression models were used. Algorithms such as Random Forest, Naive Bayes, Decision Tree, and SVC were included in the binary classification. Linear Regression was used as the regression model. Descriptive statistics were also included within the scope of the study. In addition, the frequencies of binary situations such as whether the games are paid, whether they contain advertisements, whether they contain in-app purchases are included. This study, which contains detailed information about the Google Play Store games category, provides important findings for industry professionals and academic researchers interested in the subject.

Keywords: Google Play Games, Machine Learning, Linear Regression, Binary classification, Python, Data scraping

TEŞEKKÜR

Çalışmaktan keyif aldığım bu projede desteklerini esirgemeyen danışmanım sayın Doç. Dr. Farid HUSEYNOV'a ve bu süreçte her zaman yanımda olan değerli aileme teşekkür ederim.



İÇİNDEKİLER

ÖZET	i
SUMMARY	ii
TEŞEKKÜR	iii
İÇİNDEKİLER	iv
ŞEKİLLER DİZİNİ	vi
TABLOLAR DİZİNİ	vii
1. GİRİŞ	1
1.1. Çalışmanın Arka Planı	1
1.2. Problemin Tanımı	2
1.3. Çalışmanın Amacı	3
1.4. Araştırma Soruları	3
1.5. Tezin Akışı	4
2. LİTERATÜR	5
2.1. Google Play Store Hakkında	5
2.2. Daha önce yapılmış akademik çalışmalar	6
2.2.1. Tanımlayıcı İstatistiklere Dayalı Çalışmalar	6
2.2.2. Makine Öğrenmesi Uygulamalarına Yönelik Çalışmalar	8
2.2.3. Literatürdeki Diğer Makine Öğrenmesi Çalışmaları	9
3. METOT	15
3.1. Veri Kaynağı ve Değişken Özellikleri	15
3.2. Verilerin Elde Edilmesi	15
3.3. Verilerin Analize Hazırlanması	16
3.4. Eksik Verilerin Çıkarılması	16
4. BULGULAR	17
4.1. Frekanslar	17
4.2. Tanımlayıcı İstatistikler	22
4.3. Kruskal Wallis Testleri	23
4.4. İkili Sınıflandırma	28
4.4.1. Tahmin Modelleri	28
A. Decision Tree (Karar Ağacı)	28
B. Random Forest (Rastgele Orman)	28

C. Naïve Bayes	29
D. KNN (K En Yakın Komşu Yöntemi)	29
E. SVC (Destek Vektör Makineleri)	30
4.4.2. İkili Sınıflandırma Sonuçları	31
4.5. Regresyon Analizi	33
5. Sonuç ve Tartışma	37
5.1. Kısıtlamalar ve Öneriler	39
KAYNAKÇA	41
ÖZGEÇMİŞ	44



ŞEKİLLER DİZİNİ

<u>Sekil No:</u>	<u>Sayfa</u>
2.1 Erişilebilir Uygulamaların Aralık 2009'dan Eylül 2021'e Kadar Olan Gösterimi	5
2.2 Yıllara Göre Artan Geliştirici Sayısı	6
4.1 Oyun Türlerine Göre Frekansları Gösteren Sütun Grafik	17
4.2 Oyunların İçerik Derecelendirmeye Göre Frekanslarını Gösteren Sütun Grafik	18
4.3 Oyunların Ücretli Olmasına Göre Frekanslarını Gösteren Sütun Grafik	19
4.4. Reklam İçerme Durumunun Frekanslarını Gösteren Sütun Grafik.	20
4.5. Uygulama İçi Satın Alma İçeren Oyunların Frekanslarını Gösteren Sütun Grafik	20
4.6. Editörün Seçimi Olma Durumunun Frekanslarını Gösteren Sütun Grafik	21
4.7. Karar Ağacının Şematik Gösterimi	28
4.8. Rastgele Orman Algoritmasının Şematik Gösterimi	29
4.9. KNN Algoritmasının Şematik Gösterimi	30
4.10. SVC Algoritmasının Şematik Gösterimi	31

TABLolar DİZİNİ

<u>Tablo No:</u>	<u>Sayfa</u>
4.1. Oyun Türlerine Göre Frekanslar ve Oranlar	17
4.2. İçerik Derecelendirmeye Göre Oyunların Frekansları ve Yüzdeleri	18
4.3. Ücretli ve Ücretsiz Oyunların Frekansları ve Yüzdeleri	19
4.4. Mağazadaki Oyunların Ne Kadarının Reklam İçerdiğini Gösteren Frekanslar ve Yüzdeler	20
4.5. Uygulama İçi Satın Alma İçeren Oyunların Frekansları ve Yüzdeleri	21
4.6. Editörün Seçimi Olan Uygulamaların Sıklıkları ve Oranları	21
4.7. Sayısal Değişkenlere Ait Tanımlayıcı İstatistikler	22
4.8. Farklı Oyun Türlerine Göre Yorum Sayılarının Karşılaştırılmasına İlişkin Kruskal Wallis Testi Sonucu	23
4.9. Farklı Oyun Türlerine Göre Boyutların Karşılaştırılmasına İlişkin Kruskal Wallis Testi Sonucu	24
4.10. Farklı Oyun Türlerine Göre İndirilme Sayılarının Karşılaştırılmasına İlişkin Kruskal Wallis Testi Sonucu	25
4.11. Farklı Oyun Türlerine Göre Derecelendirme Puanlarının Karşılaştırılmasına İlişkin Kruskal Wallis Testi Sonucu	26
4.12. İkili Sınıflandırma İçin Seçilmiş Karar Değişkenleri	32
4.13. İkili Sınıflandırma Doğruluk Skorları	33
4.14. Tahmin Edici Regresyon Modelleri İçin Seçilen Karar Değişkenleri	34
4.15. Lineer Regresyon Modelleri İçin Performans Metrikleri	35

1. GİRİŞ

1.1. Çalışmanın Arka Planı

Gelişen teknoloji ile mobil cihazların yalnızca iletişim için kullanıldığı yıllar oldukça geçmişte kaldı. Mobil oyunlara yönelik ilgi her geçen gün artarken, mobil uygulama mağazalarına yüklenen oyunların sayıları da giderek artmaktadır. Mobil oyun sektöründe rekabet kıyasıya sürmektedir. Her yaştan kitleye hitap eden oyunlar, uygulama marketlerinde yayınlanırken uygulama içi ürünler de oyun meraklılarına sunulmaktadır. Ücretli oyunlar, oyun içi ürünler, oyun içi reklamlar gibi pek çok etmenle büyük bir ekonomi yaratan bu pazara ev sahipliği yapan en önemli platformlardan biri Google Play Store'dur. Google Play Store oyunlar dahil pek çok uygulamayı barındıran bir platformdur. Android işletim sistemine yönelik uygulamaları içeren Google Play Store'da mobil oyunların çoğu ücretsiz oyunlardan oluşmaktadır (Arıcı ve Huseynov, 2021). Geliştiricilerin çoğunun ücretsiz oyun yayınladığı bu platformda şüphesiz ki indirme sayıları ve değerlendirme puanları çok büyük önem arz etmektedir. Nitekim oyun puanları ne kadar yüksek olursa ve “sosyal kanıtı” yani yorum sayıları ve indirme oranları yüksek olursa oyun içi ürün satılma olasılığı artacaktır. Reklam gelirleri de bu durumdan olumlu olarak etkilenecektir. Sosyal kanıt, ürün popülerliğini ve talebini gösteren bir sosyal etki taktiğidir (Cialdini and Cialdini, 2007). Belirsizlik durumlarında, insanlar kendi eylemleri için rehberlik elde etmek adına bir bilgi kaynağı olarak sosyal kanıttan yararlanır (Cialdini and Trost, 1998). Genellikle firmalar, insanların belirsizlik durumlarında birbirlerinin davranışlarını alışkanlıkla takip ettikleri gerçeğinden yararlanmak için sosyal kanıt kullanırlar. E-ticaret web siteleri, kullanıcıların endişelerini azaltmak için sosyal kanıt kullanır ve bu nedenle güvenilirliği oluşturmak adına sosyal kanıt ipuçlarını uygular (Roethke et al., 2020).

Çalışmanın arka planında veri madenciliği yöntemlerine yer verildi. Veri madenciliği ya da günümüzdeki kullanımıyla veri bilimi, veriden faydalı bilgi özütünü çıkarma ile ilgilenir. Çalışma kapsamında veri kazıma yöntemleri kullanıldı. Python programlama diline ait “Selenium” kütüphanesi yardımıyla Google Play Store oyunlar kategorisinden 17 farklı kategorideki oyunların bilgilerine erişildi. Burada temel amaç olabildiğince çok oyuna erişilerek oyun kategorileri ile ilgili bilgi özütüne erişmektir. Makine öğrenmesi yöntemleri ile verinin davranışı çözümlenmek istendi.

Veri bilimi ana bilim dalı, temel olarak bilgisayar bilimi ve araştırma alanı olarak istatistiksel analizlere dayanmaktadır. Ayrıca ekonomi, finans, veri yapısı, veri bilgi işlem gibi alan bilgisi gerektirmektedir. Veri bilimi uzmanlığı yeteneklerine ek olarak güçlü mesleki beceriler, ticari iletişim becerileri, yenilik yetenekleri, büyük veri uygulamaları için matematik ve bilgisayar bilimi, veri analizi ve işleme gibi yetenekler gerekmektedir (Qiang et al., 2019).

1.2. Problemin Tanımı

Mobil oyun endüstrisindeki hızlı rekabetin karşısında geliştiriciler büyük bir baskı altındadırlar. Mağazaya sürekli yeni oyunlar yüklenirken kendi oyunlarının beğenilip beğenilmeyeceği önemli bir merak konusudur. En gözde örneklerden biri olan Google Play Store'daki durumu çözümlemek geliştiriciler açısından oldukça önemlidir.

Google Play Store fazlasıyla rekabetçi bir ortamdır. Birbirine benzer veya birbirinden farklı birçok oyun çeşidini içermektedir. Milyonlarca kullanıcının olduğu platformda kullanıcının ilgisini cezbedebilmek gerçekten önemlidir. Aslında bir kullanıcıya ulaşmanın önünde pek çok aşama vardır. Öncelikle oyunun kullanıcıya ulaşması gerekmektedir. Önerilenler listesinde bulunması bir oyun için önemli bir etkidir. Oyun en çok tercih edilenler listesinde değilse, kullanıcının karşısında çıkma ihtimali düşecektir. Oyuna erişmek için ya doğrudan oyunun ismi aranmalı ya da oyunun anahtar kelimeleri kullanılmalıdır. Tüm bunlar oyunun indirilmesi için yeterli değildir. Oyun kullanıcının karşısında çıktıktan sonra, reklam içerip içermemesi, ücretli olup olmaması gibi unsurlarla da kullanıcının süzgecinden geçer. Eğer oyun piyasaya sürülmeden ne kadar indirilebileceği tahmin edilirse veya mevcut bir oyunun indirme sayıları, yorum sayıları ve değerlendirme puanlarının nasıl arttırılacağı tahmin edilebilirse bu çok büyük bir avantaj olacaktır.

Ayrıca literatürde bu konuda yeterli sayıda çalışmanın olmaması bu alanda bir bilgi açığını meydana getirmektedir. Veri bilimi kullanılarak elde edilmiş Google Play Store verilerine bu alanı aydınlatmak için ihtiyaç bulunmaktadır. Editörün seçimi olma, ücretli olma, reklam içerme, uygulama içi satın alma içerme gibi konuların oyunlar içerisindeki frekanslarını bilmek mağazanın genel atmosferini anlamak açısından faydalı olacaktır.

Oyun türleri arasında indirme sayıları, boyut, derecelendirme puanları, yorum sayıları gibi değişkenlerde anlamlı bir farklılık olup olmadığını bilmek geliştiriciler açısından önemli olabilir. Bu sayede hangi türden oyun geliştirmenin daha karlı olacağı sorusu cevaplanmış olabilir.

1.3. Çalışmanın Amacı

Mobil oyun sektörü giderek önem kazanmaktadır ve büyük bir ekonomik değere sahiptir. Her geçen gün büyüyen ve gelişen mobil oyun sektörüne Türkiye’den geliştiricilerin katılması da büyük bir ekonomik fırsat olabilir. Çalışma kapsamında mobil oyun sektörüne dair açıklayıcı ve tahmin edici bulgulara ulaşılması amaçlanmaktadır. Google Play Store’dan elde edilmiş yaklaşık 14.000 veri ile mobil oyun sektörünün güncel durumu analiz edildi. Çalışma kapsamında çeşitli makine öğrenmesi algoritmaları kullanılarak birden fazla hedef değişkenin (İndirilme sayısı, derecelendirme puanı ve yorum sayısı) incelenmesi ve tahmin edilmesi amaçlanmıştır. Çalışmanın sonucu olarak çeşitli şartlar altında indirilme, puanlanma ve yorum sayısı gibi değişkenlerin nasıl etkilendiğine dair durumun aydınlatılması hedeflendi. Elde edilen bulgular her algoritma için ayrıca değerlendirilerek en uygun model belirlenmesi istenmektedir. Ayrıca değişkenlerin tanımlayıcı istatistiklerine de yer verilmesi amaçlandı. Değişkenlerin birbirleri üzerindeki etkilerinin istatistiksel olarak anlamlı olup olmadığı tespit edilmesi amaçlandı.

Çalışma kapsamında Google Play Store verilerine bakılarak, henüz uygulama mağazaya yüklenmeden indirme sayılarının, değerlendirme puanlarının ve yorum sayılarının tahmin edilebilmesi amaçlanmaktadır. Değerlendirme puanlarının, indirme sayılarının ve yorum sayılarının önceden tahmini, geliştirilen uygulamaya ayrılan zamanın değip değmeyeceği konusunda fikir verebilir. Giderek önem kazanan mobil oyun sektörüne dair yapılan bu ayrıntılı çalışma, oyun geliştiricileri ve bu konuyla ilgilenen araştırmacılar için önemli bulgular sunmayı hedeflemektedir.

1.4. Araştırma Soruları

1. Oyun mağazaya yüklenmeden önce indirilme sayısı tahmin edilebilir mi?

2. Yayınlanmadan önce bir oyunun ne kadar reaksiyon alacağı yani yorum sayısı tahmin edilebilir mi?
3. Değerlendirilme puanı oyun mağazaya yüklenmeden önce tahmin edilebilir mi?
4. Mağazadaki oyunların, editörün seçimi olma, ücretli olma, reklam içerme, uygulama içi satın alma içerme vb. frekansları nelerdir?
5. Oyun türlerine göre yorum sayılarının, indirme sayılarının, derecelendirme puanlarının ve oyun boyutlarının arasında anlamlı bir farklılık var mıdır?

Bu sorular, çalışma içerisinde cevaplanarak detaylı analizlere yer verildi.

1.5. Tezin Akışı

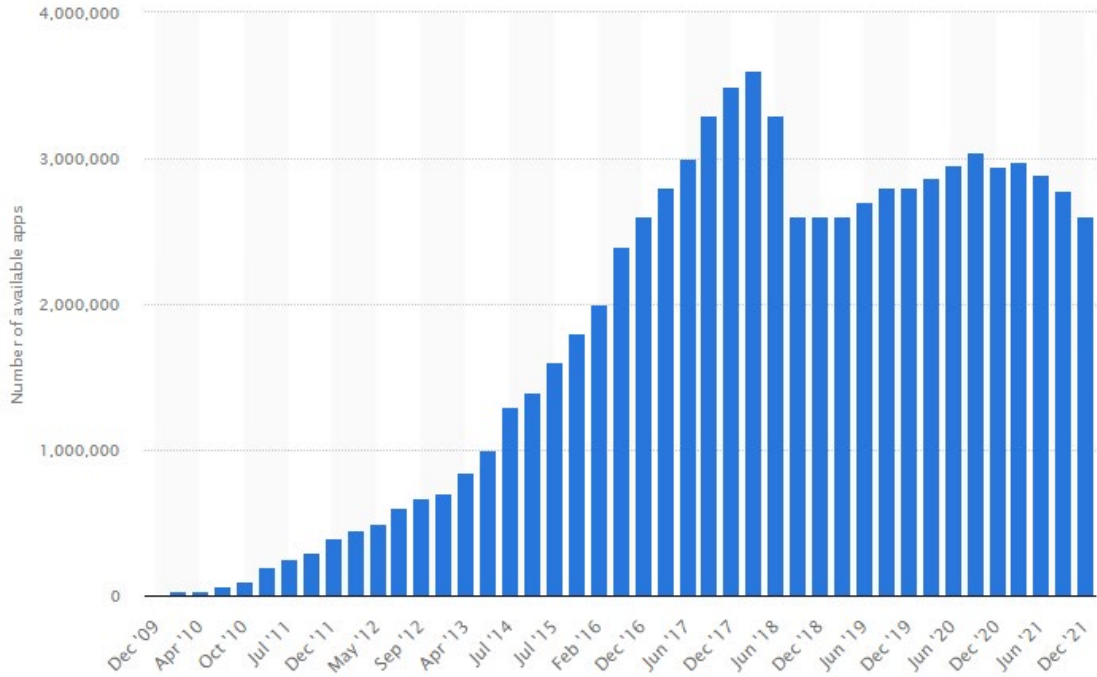
Giriş bölümünde tezin genel amacına, araştırma sorularına ve çalışmanın kapsamına yer verildi. Literatür bölümünde konuyla ilgili açıklayıcı bilgilere yer verildi. Bu bölüm Google Play Store'un tarihi, konuyla ilgili makine öğrenmesi, konu dışı makine öğrenmesi ve tanımlayıcı istatistik çalışmalarını içermektedir. Ayrıca bu tezdeki veri seti ve veri elde etme metodu kullanılarak elde edilmiş çalışmalara da yer verildi. Metod bölümünde verilerin nasıl elde edildiği, veri işleme süreçlerini nasıl gerçekleştiği, değişken özellikleri gibi bilgilere yer verildi. Bulgular bölümünde araştırmada elde edilen bulgulara detaylı bir şekilde yer verildi. Sonuç bölümünde elde edilen bulgular yorumlanarak önerilerde bulunuldu. Ayrıca çalışmaya ait kısıtlardan da bahsedildi.

2. LİTERATÜR

2.1. Google Play Store Hakkında

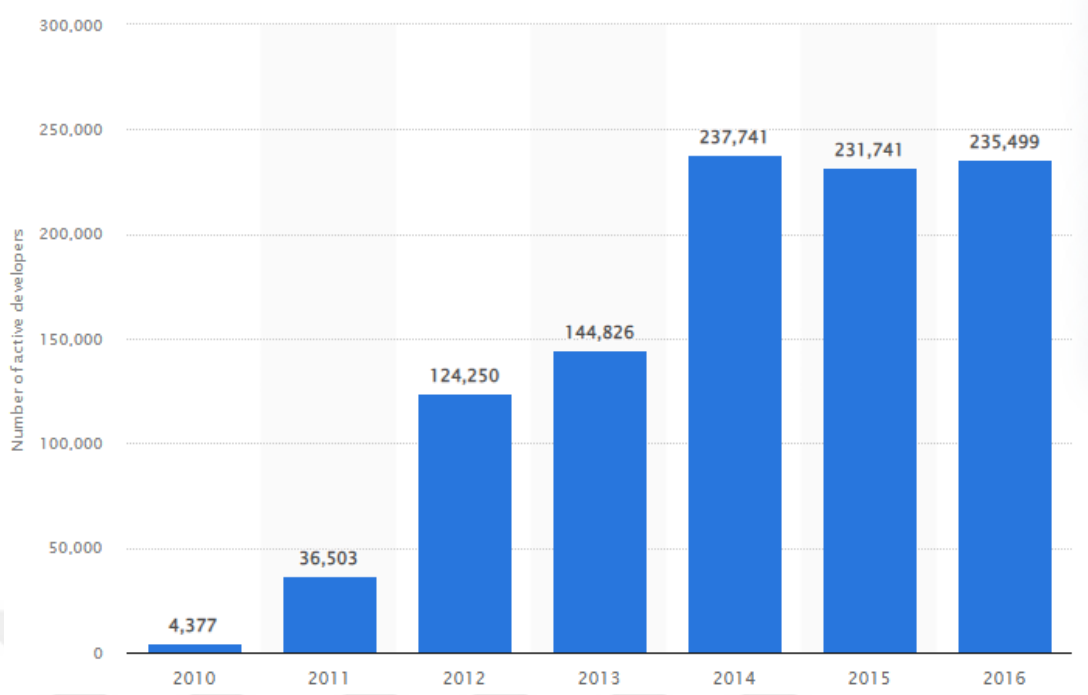
Eski adı “Android Market” olan Google Play Store, 2012 yılında içerisinde pek çok kategoriyi bulundurur halde piyasaya çıktı (Web 1, 2012). 2012 yılından bu yana her geçen gün yeni uygulamaların yüklendiği Google Play Store’a uygulama yüklemek için bir geliştirici hesabı açılması gerekir. Bir defaya mahsus olmak üzere 25 Dolar’lık bir ücret ödenir ve daha sonra sınırsız uygulama yayınlanabilir (Web 2, 2012). Google Play Store, en çok mobil uygulama yüklenen mağazalardan biridir (Reddy and Nallabolu, 2019).

Çalışma kapsamında Google Play Store’daki mobil oyunlar kategorisiyle ilgilenilmektedir. Mobil oyunlar 17 farklı türden oluşmaktadır. Bunlar, “Yarış, Strateji, Spor, Rol Yapma, Müzik, Simülasyon, Masa Oyunları, Macera, Kumarhane, Klasik, Kelime, Kağıt, Eğlencelik Bilgi, Eğitici, Bulmaca, Arcade, Aksiyon” olarak belirlendi.



Şekil 2.1 Erişilebilir Uygulamaların Aralık 2009’dan Eylül 2021’e Kadar Olan Gösterimi (Web 3, 2022a).

Şekil 2.1.’de görüldüğü gibi Google Play Store’a eklenen uygulamaların sayısı en çok Aralık 2017 ile Haziran 2018 arasında artmıştır. Daha sonra bir düşüş yaşansa da pandemi ile tekrar yükseldi.



Şekil 2.1 Yıllara Göre Artan Geliştirici Sayısı (Web 4, 2022b).

Şekil 2.2.'de görüldüğü gibi yıllar geçtikçe Google Play Store'a katılan geliştirici sayısı artmıştır. Bu durum geliştiriciler arasında bir rekabet ortamı oluşmasına sebep olabilir.

Google Play Store içerisinde pek çok kategoriden uygulamayı barındıran bir platformdur. Bir uygulamaya bakıldığında şu bilgilere sahip olduğu görülmektedir: Uygulama logosu, uygulama adı uygulama kategorisi, derecelendirme puanı, indirme sayısı, yorum sayısı, gereken android sürümü, mevcut android sürümü, içerik derecelendirmesi vb.

2.2. Daha önce yapılmış akademik çalışmalar

2.2.1. Tanımlayıcı İstatistiklere Dayalı Çalışmalar

Bir çalışmada (Reddy and Nallabolu, 2020) Kaggle adlı platformdan elde ettikleri bir veri seti ile Google Play Store'un geneline dair makine öğrenimi tabanlı bir tanımlayıcı istatistik çalışması yapıldı. Tanımlayıcı analize göre, eğitimle ilgili uygulamalara sıklıkla rastlanırken müzikli oyunların nadir olduğu kanısına varıldı. "Kitaplar" yüksek, "Haritalar ve Navigasyonlar" en düşük puanlıdır. "Yarış Oyunları" sıklıkla indirilirken ve "Etkinlikler" nadiren indirilir. "Aksiyon oyunları" sıklıkla yorumlanırken ve "Etkinlikler" nadiren yorum almıştır. "Rol yapma

oyunları” yüksek fiyatlıdır ve “Etkinlikler” daha az fiyatlıdır. “Rol yapma oyunlarının” boyutu çok büyüktür ve “Araçların” boyutu küçüktür.

Oyun kategorilerinin incelendiği bir çalışmada (Latif ve ark. 2019), Google oyun kazıyıcı kullanarak her oyun kategorisinden en az 550 uygulama ücretsiz ve ücretli olarak toplandı. Oyunlarda tüm kategorilerdeki 3600 ücretli uygulama ve 10.000 ücretsiz uygulama kümülatif olarak kazındı. Çalışma, ücretsiz uygulamaların %66’sının uygulama içi satın alma ürünü sunduğunu ortaya çıkardı. Bu oran ücretli uygulamalar içinse %17’ydi. Ücretsiz oyunların çoğuna kullanıcıların 5 üzerinden 4.5 verdiği tespit edildi. Bununla birlikte ücretsiz oyun kategorisinde en çok “Kart oyunları” kategorisinin tercih edildiği sonucuna varıldı. Ücretli oyunlarda ise kullanıcıların seçimi “Bulmaca” oyunlarından yanaydı.

Bu tez kapsamında elde edilen veriler kullanılarak dört farklı bilimsel yayın yapıldı. Bu yayınlardan ilki oyun türlerine ait frekansları paylaşan çalışmaydı (Arıcı ve Huseynov, 2021a). Bu çalışmada oyun türlerine göre indirme sayılarına ve derecelendirme puanlarına detaylı bir şekilde yer verildi. Çalışma kapsamında oyunlara ait uygulama içi ürünlerin en ucuz ve en pahalılarına da yer verildi. Çalışmanın sonuçlarına göre en yüksek puanları kumarhane oyunları, kelime oyunları ve bulmaca oyunları almıştı. En düşük puanlar ise macera, masa ve simülasyon oyunlarındaydı. İndirilme sayısı en çok yarış oyunlarında iken en az eğitici oyunların indirildiği kaydedilmişti.

Bir diğer yayında ücretli oyunlar ve ücretsiz oyunlar kıyaslandı (Arıcı ve Huseynov 2021b). Ücretli oyunlara yapılan yorum sayıları daha fazla çıktı. Bununla birlikte ücretli oyunlara yapılan güncellemelerin ücretsiz oyunlara göre daha geç yapıldığı ortaya çıktı. Oyun içi en pahalı ürünlerin eğlencelik bilgi oyunları ve kumarhane oyunlarına ait olduğu paylaşıldı. Ücretsiz oyunların ücretli oyunlardan daha fazla indirildiği bilgisi de şaşırtıcı değildi.

Diğer bir yayında ise editörün seçimi olma durumu incelendi (Arıcı ve Huseynov, 2021c). Çalışma bulgularına göre editörler tarafından seçilen oyunlarda en çok arcade oyun türü indirildi. Önerilen oyunların %28’i eğitici oyunlardır. Bununla birlikte önerilen oyunların boyut, indirme sayısı, yorum sayısı, derecelendirme puanı gibi metrikleri daha yüksek çıktı. Bu çalışmanın bulgularına göre geliştiricilerin editörün seçimi kategorisine girmesi en çok indirme sayısı açısından kritiktir. “Editörün seçimi” olan oyunlarda en çok %24.5 ile arcade oyunlar indirildi. İkinci en yüksek indirilen oyun türü %16.8 ile aksiyon oyunları oldu.

Üçüncü sırada ise %13.3 ile spor oyunları vardır. “Editörün seçimi” olan oyunlar içerisinde en çok yorumu %21.8 ile strateji oyunları aldı. İkinci sırada %18.6 ile aksiyon oyunları vardır. Üçüncü sırada ise %16.6 ile arcade oyunlar bulunmaktadır.

İlgili diğer yayında ise oyunların reklam içerip içermeme durumundaki farkları inceledi(Arııcı ve Huseynov, 2022). Çalışmaya göre mağazadaki oyunların %83’ü reklam içermekteydi. Reklam içeren oyunların içerisinde en çok indirilen oyun türü ise yarış oyunları oldu. Reklam içeren oyunların içerisinde en çok yorumu alan ise aksiyon oyunlarıdır. Mann-Whitney U testi sonuçlarına göre reklam içeren oyunlar daha fazla yorum almıştır ve bu oyunların aldığı yorumlar yani reaksiyonlar reklam içermeyen oyunlardan daha fazladır. Yalnızca reklam içeren oyunların puanları 0.1 puan daha düşük çıktı. Reklam içeren oyunlarda en çok %11.1 ile yarış oyunları indirildi. Reklam içeren oyunlarda ikinci en çok indirilen oyun türü %10.5 ile arcade oyunlar oldu. Üçüncü sırada ise %9.2 ile aksiyon oyunları vardır. Reklam içeren oyunlarda en az %2.4 ile kumarhane oyunları indirildi. En az indirilen diğer oyunlar ise %3.7 ile rol yapma ve kağıt oyunlarıdır. Reklam içeren oyunlar içerisinde en çok yorumu %12.6 ile aksiyon oyunları aldı. İkinci sırada %9.9 ile yarış oyunları vardır. Üçüncü sırada ise %9.2 ile spor oyunları bulunmaktadır. Reklam içeren oyunlarda en az yorum alan oyun kategorileri ise %0.9 ile eğitici, %2.7 ile macera, %3.3 ile masa oyunlarıdır.

2.2.2. Makine Öğrenmesi Uygulamalarına Yönelik Çalışmalar

Bir araştırmada (Lengkong and Maringka, 2020) yüksek puan alan uygulamaların özelliklerini belirlemek için Google Play Store'daki boyut, yükleme, inceleme, tür (ücretsiz / ücretli), derecelendirme, kategori, içerik derecelendirmesi ve fiyat özelliklerini aydınlatmaya çalışıldı. Bu araştırma, Google Play Store'daki yüksek dereceli uygulamalardaki en önemli özellikleri belirlemek için bir Random Forest sınıflandırıcısı kullanıldı. Bununla birlikte araştırmada Google Play Store'daki yüksek dereceli uygulamalardaki en etkili özellikleri belirlemek için Gradient Boost Algoritmasını kullanılmıştır. 10 katlı çapraz doğrulama yöntemi ile %99.93 doğruluk elde edildi.

267.000 benzersiz uygulama verisi kullanılan bir çalışmada (Bashir ve ark, 2019) Random Forest, K- En Yakın Komşu ve Destek Vektörü gibi bazı makine öğrenme algoritmaları kullanıldı. Çalışmada Google Play Store’da herhangi bir uygulamayı yüklemeyen önce kullanıcı derecelendirmesini ve indirme sayısını

tahmin etme hedeflenmekteydi. Farklı algoritmalar kullanılarak tahmin edilen değerlendirme puanı ve indirme sayıları değerlerinin karşılaştırılmasına göre, KNN ve SVM algoritmalarının başarı oranını Random Forest'tan daha doğru tahmin edebileceğine karar verdiler.

Detaylı literatür taraması sonucunda bu alanda çalışmaların sınırlı olduğu görüldü. Bu çalışmanın bulguları literatüre olumlu katkılar sunacaktır.

2.2.3. Literatürdeki Diğer Makine Öğrenmesi Çalışmaları

Şimdiye kadar bahsedilen çalışmalar Google Play Store'daki verilerle uygulanan makine öğrenmesi çalışmalarıydı. Oysaki literatürde çeşitli konularda pek çok makine öğrenmesi çalışması mevcuttur. Bunlardan biri veri bilimi uygulamalarını kalp hastalıkları teşhisinde kullanan çalışmadır (Göktaş ve Yağanoğlu, 2020). Çalışma kapsamında dört farklı algoritma kullanılarak sınıflandırma işlemi yapıldı. Bununla birlikte algoritmaların yeni hastalık teşhisi tahminindeki başarı oranları incelendi. Test örneklem büyüklüğü %25 olarak belirlendi. KNN algoritmasında, test verisi büyüklüğü %33 olarak kararlaştırıldı. Metrik olarak minkowski tekniği uygulandı. Random Forest algoritmasında test verisi büyüklüğü %30 olarak belirlendi. Son olarak SVC'de ise test boyutu %33 olarak belirlenerek lineer yöntem kullanılmıştır. C4.5 Decision Tree algoritması hastalık tahmini konusunda %83 oranında başarı elde edildi. C4.5 Decision Tree algoritmasını %81 başarı oranıyla RF, %75 başarı oranıyla SVC ve %68 başarı oranıyla KNN algoritmaları takip etti.

Bir diğer araştırmada restoran, film ve ürün yorumlarını içeren çeşitli veri setlerinden yararlanılarak veri bilimi sınıflandırma algoritmaları aracılığıyla duygu analizi yapılmıştır (Çelik vd., 2021). Duygu analizi yapmak amacıyla SVC, KNN, Naive Bayes, Decision Tree ve Random Forest sınıflayıcılarından yararlanıldı. Veri çeşitliliği ile boyutunu çoğaltmak için yarısı olumlu yarısı olumsuz olmak üzere toplamda 1000 adet yorum içeren üç ayrı veri seti bir araya getirildi. Sonuçlar SVC'nin duygu analizi konusunda farklı metotlara nazaran daha iyi sonuç çıkardığını ortaya koydu. Bu araştırmada duygu analizi için beş ayrı veri bilimi sınıflandırma algoritmasının verimliliği kıyaslandı. Bütün algoritmaların her bir veri setine özel olarak çalıştırılmasındansa bütün veri setlerinden meydana gelen tek bir veri seti ile eğitime sokulmasının verimlilik açısından daha iyi olduğu saptandı. Bu durum veri büyüklüğü ile çeşitliliğinin önemini göstermektedir. SVC algoritmasının

birleştirilmiş veri setinin çeşitliliğine ve büyüklüğüne karşın %85 değerinde bir accuracy (doğruluk) skoru oranıyla diğer algoritmalara nazaran en efektif analiz yöntemini oluşturduğu görüldü (Çelik vd., 2021).

Makine öğrenmesi yöntemleri, satış tahmini alanında da sıkça kullanılmaktadır. Bu alanda yapılan bir tahmin çalışmasında “satıldı”, “satılmadı” şeklinde ikili bir sınıflandırma yapıldı (Dilkı, 2020). Firmalar, yeni bir geçmiş verilerden bazı analizler yaparak yeni bir ürünü piyasaya sunmadan önce geçmiş verilerden yararlanarak geleceğe dönük tahminler yapabilmektedirler. Bu şekilde, en uygun sayıda üretim yaparak hem işgücü ile hammadde tutarını hem de lojistik gibi maliyetlerin meydana getirebileceği zararı minimum düzeyde tutmayı hedeflerler. Üretilen ürününe ait hedef kitleyi saptayarak bu doğrultuda satış hamleleri geliştirme fırsatı edinirler. Bu sebeple çalışma kapsamında, makine öğrenmesi problemlerinden sınıflandırma problemleri uygulanmıştır. Satılmak istenen ürün ile özellikleri aynı olan, daha önce piyasaya sunulmuş ürünlerin bilgileri alınarak, “satılmadı” ve “satıldı” olarak ikili sınıflandırma çalışması yapıldı. KNN, Naive Bayes ve SVC kullanılan araştırmada, veri seti test seti ve eğitim seti olarak ikiye bölündü. Çalışma sonucu olarak, 0,71 accuracy (doğruluk) ve KNN algoritmaları ile en yüksek doğruluk elde edildi. Bu alandaki çalışmalarda KNN algoritması nispeten daha iyi sonuçlar vermektedir.

Makine öğrenimi yöntemleri hayvancılıkta da kullanılmaktadır. Bir araştırmada, arı kanatları üzerindeki kavşak noktalarına bakılarak arı türlerinin sınıflandırılması amaçlanmıştır (Demir vd., 2018). Türlerin sınıflandırması amacıyla beş ayrı ilden alınan arı kanat fotoğrafları üzerindeki kavşak noktaları tespit edildi. Arı kanatlarında bulunan kavşak noktalarının en az yanılma payıyla saptanması için yeni bir algoritma tavsiye edildi. -27 farklı biçimsel özellik kavşak noktaları kullanılarak çıkarıldı. Bu nitelikler normalize edilerek sınıflandırılmaya tabi tutuldu. SVC, Yapay Sinir Ağları, K-Mean ve KNN sınıflandırma metodu olarak kullanıldı ve en iyi sonucu Yapay Sinir Ağları verdi.

Tarım alanında da makine öğrenmesi çalışmaları kullanılabilmektedir. Makine öğrenmesi algoritmaları kullanılarak, Türkiye’de tüketimi ve üretimi fazla olan kayısı iç çekirdeğinin, acı ya da tatlı olarak kategorize edilmesi bir araştırmanın konusunu meydana getirmektedir (Şenel, 2020). İki türden kayısı iç çekirdeği için de fazlasıyla talep vardır. Kayısı iç çekirdekleri, depolama şartları gibi sebeplerle birbirine karışabilmektedir. Tüketiciler açısından bu durum olumsuz karşılanır. Ne

yazık ki kayısı iç çekirdeğini gözle ayırt edebilmek her zaman mümkün değildir. Bu sınıflandırma çalışmasının amacı insan unsurunu devre dışı bırakabilecek bir metod geliştirmektir. Bu sınıflandırma işlemi için Çok Katmanlı Algılayıcı, Random Forest, Decision Tree, SVC, KNN, Gaussian Naive Bayes, Adaptive Boosting ve algoritmaları kullanıldı. Algoritmaların yarıdan fazlası yeterli sayıda karar değişkeni ile sınıflandırma işlemini %100 başarı ile elde edebilmektedir. En az sayıda karar değişkeni kullanarak en yüksek başarıyı Random Forest algoritması elde etti. Sonuçlar göre, kayısı iç çekirdeklerinin sınıflandırılması veri bilimi yöntemleri ile başarılı bir şekilde gerçekleştirilebilmektedir (Şenel, 2020)

Makine öğrenmesi yöntemleri meteoroloji alanında da kullanılmaktadır. Bir çalışmada SVM kullanılarak yağış akış ilişkisinin tahmini yapıldı (Demirci, 2019). Afetlerin engellenmesi ve havza yönetimi, su yapılarının inşasının planlaması ve su kaynaklarının daha verimli kullanılması amacıyla akış ve yağış verilerinin tahmin edilmesinin önemi büyüktür. Bu çalışmada Amerika Birleşik Devletleri'nde yer alan bir rezervuar nehirdeki 731 günlük sıcaklık, yağış ve akış bilgilerini içeren verilerden faydalanılarak modeller meydana getirildi. Bu veriler M5 Decision Tree (M5T) ve SVM yöntemlerinde kullanıldı ve yağış akış ilişkisi tahmin edildi. Her iki tahmin yöntemiyle elde edilen sonuçlar gerçek ölçüm sonuçları ile karşılaştırılmıştır. Bunun sonucunda M5 Decision Tree modellerinin akış tahmininde daha iyi performans gösterdiği ortaya çıkmıştır.

Bir çalışmada, makine öğrenmesi yöntemleri ile Amazon ürün yorumları veri seti üzerinde duygu analizi sınıflandırma verimlilikleri analiz edildi (Göçgün ve Onan, 2021) Çalışmada yapılan yorumların ne türden duygular içerdiği ya da duygu içerip içermediği belirlendi. Çalışmada, Python dili kullanılarak Jupyter Lab üzerinde 4 farklı makine öğrenmesi tahmini yapıldı ve Kaggle Amazon ürün yorumları içeren veri seti edinildi. Çapraz doğrulama yöntemi k=10 katlı olacak şekilde kullanılmıştır. Kullanılan yöntemler içerisinde Decision Tree sınıflandırma algoritmasının yüzde 94 doğruluk oranı ile daha başarılı olduğu belirlendi.

Mali başarısızlık alanında yazılan bir çalışmanın amacı, farklı yoğunluklarda teknoloji kullanımına sahip imalat firmalarının başarısızlıklarını SVM algoritmaları ile tahmin ederek çekirdek fonksiyonlarının performanslarını karşılaştırmak ve farklı teknoloji yoğunluğu kullanıma sahip imalat firmalarının finansal açıdan başarısız olmalarını etkileyen önemli faktörlerin neler olduğunu araştırmaktır (Kuzu ve Yakut, 2020). Çalışmanın uygulamasında Borsa İstanbul'a kayıtlı, imalat sektöründe faaliyet

gösteren 153 firmanın 2012-2015 dönemine ait 11 mali oranı ve üretimde kullandıkları teknoloji yoğunluğunu gösteren teknoloji sınıfı kategorik değişkeni kullanıldı. Yapılan analizler sonucunda finansal başarısızlığın tahmini uygulamalarında SVC ve tüm çekirdek fonksiyonlarının kullanılabileceği görüldü. Firmaların başarısız olmasını etkileyen önemli değişkenlerin belirlenmesinde Karar Ağaçları algoritmalarından J48 kullanıldı. Tüm teknoloji sınıflarında finansal başarısızlık üzerinde en çok etkisi olan değişken grubunun kârlılık oranları olduğu tespit edildi.

Diğer bir ilgi çekici çalışmada köşe yazılarındaki etkisiz kelime kullanımı ve noktalama işaretleri gibi özellikler sayesinde yazarı tanımanın mümkün olduğu ortaya koyuldu (Uyar vd., 2021). Cumhuriyet gazetesinde sıklıkla köşe yazısı yazan yazarlardan altı tanesi seçilerek her birine ait son 120 yazı çalışmaya dahil edildi. Bu yazılara ait etkisiz kelime kullanımı ve noktalama işaretlerinin sıklığı gibi dokuz farklı değişken elde edildi. Sekiz makine öğrenimi algoritması eğitildikten sonra yazıya ait yazarı tanıma başarısı önışlemden geçirildi ve ön işlemsiz veri kümelerinde tek tek ölçülerek, en az %82 ve en fazla %92 olmak üzere yüksek doğruluğa sahip sonuçlar alındı. Decision Tree temelli Random Forest algoritmasında değişken önem analizi yapılarak cümle başına ortalama kelime sayısının ve virgül kullanma sıklığının en ayırıcı değişkenler olduğu tespit edildi.

Bir çalışmada makine öğrenmesi yöntemleri finans alanında uygulandı (Sel,2020). Yatırımcılar için borsa endekslerinin tahmini endeksin etkilendiği çok fazla değişken olmasından dolayı zor olduğu kadar ihtiyaç duyulan bir konudur (Sel,2020). Yatırımcıların elde ettikleri karı maksimize etmek için doğru tahminin önemi büyüktür. Dalgalanmalar pandemi sürecinde tüm yatırım araçlarında ve piyasada oldukça fazla bir şekilde gözlemlendi. Bu araştırmada, altın fiyatları üzerinde pandemi dönemindeki kripto paraların etkileri araştırıldı. Makine öğrenimi uygulananlar Python programlama dili kullanılarak yapıldı. Makine öğrenme modellerinden çok değişkenli, Decision Tree, SVM ve Random Forest regresyon modelleri bir arada kullanıldı. Sonuç olarak; en yüksek açıklama oranına ($R^2=0,91$) sahip olan SVM yardımıyla altın fiyatları üzerinde Bitcoin +0,755, EOS -0,596, Tether -0,122 ve Tron + 0,220 etkisi olduğu ortaya çıktı. Elde edilen sonuçlar piyasada oynaklığın fazla olduğu dönemlerde yatırımların yönlendirilmesinde karar vericilere yardımcı olabileceği düşünülmektedir (Sel, 2020).

Veri bilimi yöntemleri ile iş yaşam dengesinde yıpranma tahmini yapan bir çalışma da makine öğrenmesi yöntemlerinin iş hayatında da kullanılabileceğini göstermektedir (Çelik, 2019). Çelik'e göre, iş yaşam dengesinde çalışanların yıpranma durumunun tahmini, şirketlerde insan kaynakları departmanı için çalışan performansını anlamaya yönelik önemli bir göstergedir. Bu tür göstergeler için klasik istatistik analizleri yerine veri bilimi ile tahmin yapmak daha etkili sonuçlar verecektir. Bu çalışmada iş hayatında yıpranma ve dolayısıyla işi bırakma durumunun tahmini veri bilimi yöntemlerinden Decision Tree ve SVC yöntemleri ile gerçekleştirilerek performans sonuçları karşılaştırılmıştır. Analiz için örnek veri seti IBM şirketi Watson Analytics programı kapsamında sunulan bir veri tabanından elde edildi. Veri seti 1470 adet çalışanın 35 farklı özneliği içermektedir. Çalışmada yıpranmayı etkileyen faktörler ortaya çıkarıldı ve Decision Tree ile SVC yöntemlerinin tahmin performansında sırasıyla %84.09 ve %91.36 doğruluk oranları tespit edildi.

Bir başka çalışmada bebek ağlamaları veri bilimi metotları ile sınıflandırıldı(Kop ve Bayındır, 2021).Çalışmada, bebek ağlamalarından, bebeğin herhangi bir hastalığının olup olmadığını anlayabilmek ve ihtiyaçlarını anlayabilmek için yapay öğrenme algoritmaları kullanıldı. İki farklı veri seti bu amaçla kullanıldı. Veri setleri öğrenme, eğitim ve test aşamalarından geçtikten sonra ağlama türlerinin tespit edilmesi amaçlandı. Ağlama seslerinin öznitelikleri çıkarılıştır. Öznitelikleri çıkarılan ses sinyallerinin sınıflandırılma aşamasında makine öğrenme algoritmalarından KNN algoritması, Çok Katmanlı Algılayıcılar, Decision Tree ve Random Forest yöntemleri kullanıldı ve başarı oranları kıyaslandı. Sonuçta Çok Katmanlı Algılayıcılar yöntemi kullanılarak %93 başarı oranı elde edildi.

Diğer bir çalışmada kalp yetmezliği hastalarının hayatta kalımlarını sınıflandırma algoritmaları ile tahmin edilmesi amaçlandı (Potur ve Erginel, 2021). Kalp yetmezliği ölüm oranı yüksek, son yıllarda çok yaygınlaşan kronik bir hastalıktır. Hastaların hayatta kalma oranı meme kanseri, bağırsak kanseri ve prostat kanseri gibi kanser türlerine göre daha düşüktür. Kalp yetmezliği hastalarının hayatta kalımlarını tahmin etmenin önemi büyüktür. Hastalığın erken aşamada teşhisi hayatta kalım tahmini ve en önemli risk faktörlerinin belirlenmesi ile sağlanabilir. Bu çalışmada kalp yetmezliği hastalarının hayatta kalımlarının tahmin edilmesi amacıyla Naive Bayes, lojistik regresyon, çok katmanlı algılayıcı, SVC ve J48 Decision Tree

sınıflandırma yöntemleri kullanıldı. En yüksek sınıflandırma başarısına sahip sınıflandırıcı %90 doğru sınıflandırma oranı ile çok katmanlı algılayıcı oldu.



3. METOT

3.1. Veri Kaynağı ve Değişken Özellikleri

Veriler Google Play Store Türkiye mağazasından elde edildi. “Puan” değişkeni oyunların aldığı puanları göstermektedir ve 0 ile 5 arası değer almaktadır. “Yorum” değişkeni oyunların aldığı yorum sayısını belirtir. “Ücretli” değişkeni oyunun ücretli olup olmadığını gösteren ikili bir değişkendir. “Fiyat” değişkeni oyunun ücret bilgisini taşımaktadır. “Editörün Seçimi” değişkeni oyunun editörün seçimi olup olmadığını belirten ikili bir değişkendir. “Boyut” değişkeni oyunun boyut bilgisini kilobyte cinsinden belirtmektedir. “Ekran Görüntüsü” değişkeni yükleme ekranında oyuna ait kaç tane görsel bulunduğunu belirtmektedir. “PEGI Skoru” değişkeni oyunun hangi yaş grubuna uygun olduğunu belirleyen PEGI standardının göstergesidir. “Baskın Kırmızı” oyuna ait logoda bulunan baskın rengin RGB kodlarından “Red” (kırmızı) olanının derecesini göstermektedir. Aynı şekilde “Baskın Yeşil” ve “Baskın Mavi” değişkenleri de sırasıyla “Green” (yeşil) ve “Blue” (mavi) derecelerini belirtmektedir. “İsim Uzunluğu” değişkeninde oyun isminin uzunluğu bilgisi mevcuttur. “İndirilme” değişkeni oyunun indirme sayısını belirten kesikli bir değişkendir. Mağazada oyunlar indirilme sayısı tam olarak belirtilmedi. Örneğin “10.000+”, “100000+” vb. şekilde göstergelerle sayılar belirtildi. “Reklam” değişkeni oyunun reklam içerip içermediğini, “Uygulama İçi Satın Alma” değişkeni de uygulama içi ürün satışının olup olmadığını belirtmektedir. “İndirilme/Yorum” değişkeni indirilme başına kaç yorum alındığını ifade eden bir değişkendir. “En Pahalı” ve “En Ucuz” değişkenleri uygulama içi satılan ürünlerin alt ve üst sınırlardaki fiyatlarını temsil eden değişkenlerdir. “Tür” değişkeni oyunların hangi türe ait olduğunu belirtmektedir.

3.2. Verilerin Elde Edilmesi

Veriyi elde etmek için Python programlama diline ait Selenium kütüphanesi kullanıldı (Web 8, 2018). Bu çalışma kapsamında, yaklaşık 20.000 ayrı Google Play Store bağlantısına gidilerek web adresindeki bilgilere erişildi. Bu linkleri elde etmek için oyun kategorilerine ait 100 adet anahtar kelime kullanıldı. Veriler anahtar kelimelere göre eşsiz olacak şekilde tek tek gezildi. Oyunlara ait ilgili internet adresindeki ayırt edici bilgiler kaydedildi.

3.3. Verilerin Analize Hazırlanması

Verilerin analize hazırlanması için bilimsel çalışma kütüphaneleri olan Numpy ve Pandas kütüphaneleri seçildi. Numpy (Web 5, 2021) ve Pandas (Web 6, 2021) kullanılarak veriler ön işlemeden geçirildi. “İndirilme” değişkenlerinde mevcut olan “+” ibareleri kaldırıldı. Veriler sayısallaştırılarak matematiksel işleme uygun hale getirildi.

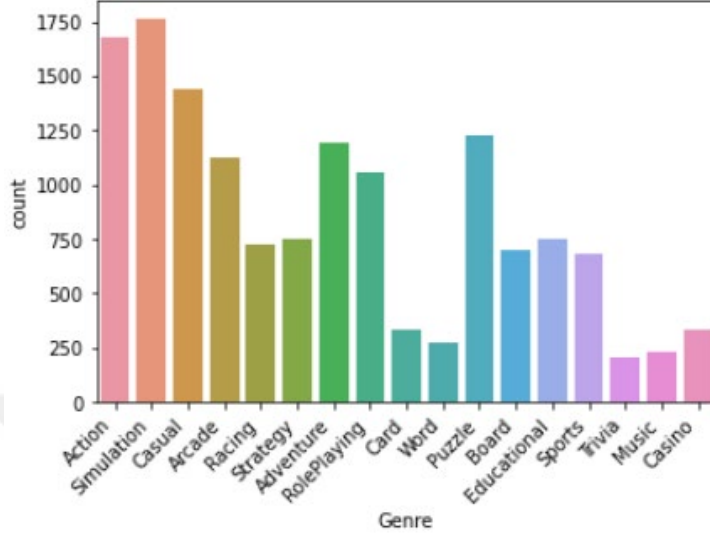
3.4. Eksik Verilerin Çıkarılması

Çalışma kapsamında 19.660 veriye erişildi. Ancak bu verilerin hepsi tüm değişkenlere ait bilgileri tam olarak içermemektedir. Veriler analize hazırlanırken “Puan” ve “İndirilme” değeri belirli olmayan veriler veri setinden çıkarıldı. Bununla birlikte “Boyut” değişkenindeki eksik bilgiler kategori ortalamaları ile dolduruldu. Bu yolla analize hazır 14.486 veri elde edildi.

4. BULGULAR

4.1. Frekanslar

Yapılan analizler sonucunda aşağıdaki bulgulara ulaşılmıştır.



Şekil 4.1 Oyun Türlerine Göre Frekansları Gösteren Sütun Grafik

Şekil 4.1.'de oyun türlerine göre frekanslar gösterilmektedir. Şekil 4.1.'nin detaylarına Tablo 4.1.'de yer verildi.

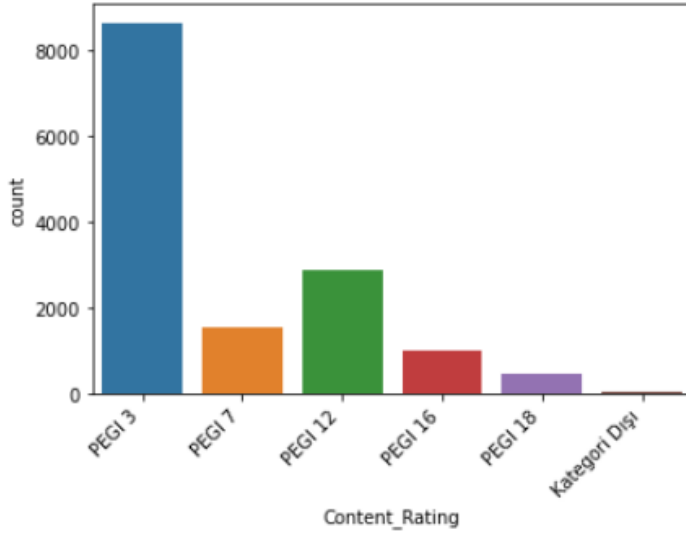
Tablo 4.1. Oyun Türlerine Göre Frekanslar ve Oranlar

Tür	Frekanslar(#)	Oranlar(%)
Simülasyon	1763	12.17
Aksiyon	1678	11.58
Klasik	1441	9.94
Bulmaca	1226	8.46
Macera	1195	8.24
Arcade	1129	7.79
Rol Yapma	1057	7.29
Eğitici	753	5.19
Strateji	749	5.17
Yarış	725	5.00
Masa	702	4.84
Spor	686	4.73
Kağıt	334	2.30

Tablo 4.1. Devam

Kumarhane	333	2.29
Kelime	273	1.88
Müzik	235	1.62
Eğlencelik Bilgi	207	1.42

Tablo 4.1.'ye göre mağazadaki tüm oyunların %12.17'sini Simülasyon oyunları oluşturmaktadır. %11.58 ile aksiyon oyunları ikinci sıradadır. Üçüncü sırada ise %9.94 ile klasik oyunlar bulunmaktadır. En az frekansa eğlencelik bilgi oyunları sahiptir (%1.42). İkinci en az frekansa sahip oyun türü müzik oyunları iken, üçüncü sırada kelime oyunları vardır.



Şekil 4.2 Oyunların İçerik Derecelendirmeye Göre Frekanslarını Gösteren Sütun Grafik

Şekil 4.2.'de oyunların içerik derecelendirmeye göre frekansları gösterilmektedir. Grafiğin detaylarına Tablo 4.2.'de yer verildi.

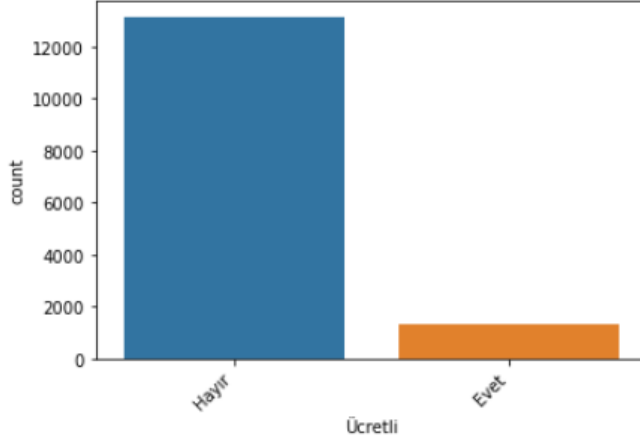
Tablo 4.2. İçerik Derecelendirmeye Göre Oyunların Frekansları ve Yüzdeleri

İçerik Derecelendirme	Frekanslar(#)	Oranlar(%)
PEGI 3	8656	59.75
PEGI 7	1516	10.46
PEGI 12	2860	19.74
PEGI 16	981	6.77
PEGI 18	457	3.15

Tablo 2.2. Devam

Kategori Dışı	16	0.11
---------------	----	------

Tablo 4.2.'ye göre mağazadaki tüm oyunların yaklaşık %60'ını 3 yaş ve üzeri oyunlar oluşturmaktadır. %19'unu 12 yaş ve üzeri %10'unu ise 7 yaş ve üzeri oyunlar oluşturmaktadır.



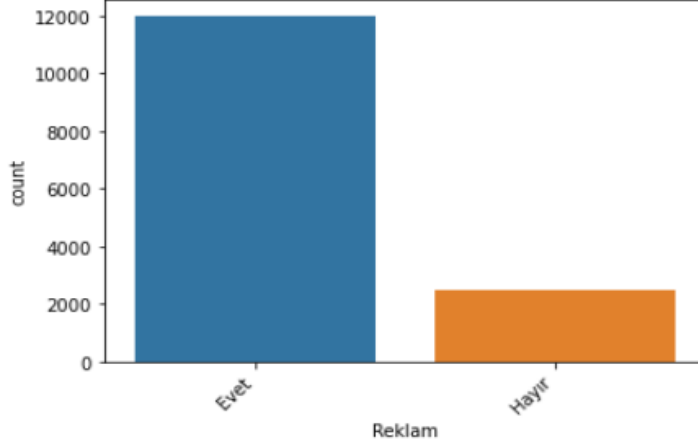
Şekil 4.3 Oyunların Ücretli Olmasına Göre Frekanslarını Gösteren Sütun Grafik

Şekil 4.3.'deki grafiğin detaylarına Tablo 4.3.'de yer verildi.

Tablo 4.3. Ücretli ve Ücretsiz Oyunların Frekansları ve Yüzdeleri

	Frekanslar(#)	Oranlar(%)
Ücretsiz	13145	90.74
Ücretli	1341	9.26

Tablo 4.3.'ye bakılacak olursa, mağazadaki tüm oyunların %90.74'ünü ücretsiz oyunlar oluşturmaktadır. Bu oyunlar genellikle freemium gelir modeline uygun şekilde geliştirilmiştir ve ücretsiz oyun içerisinde ücretli ve ayrıcalıklı hizmetler sunarlar.



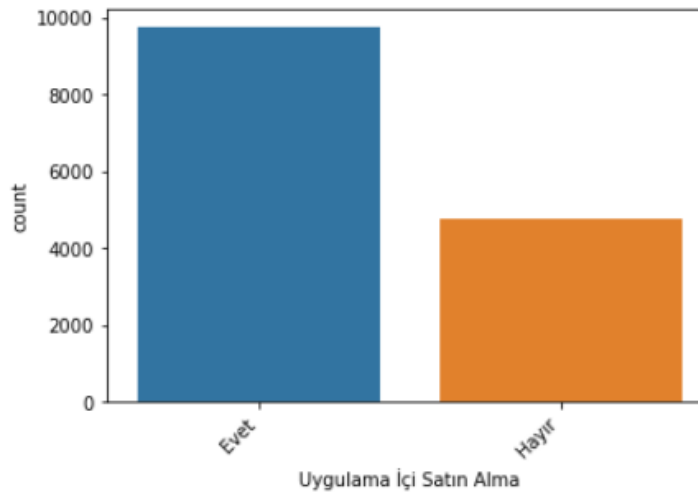
Şekil 4.4. Reklam İçerme Durumunun Frekanslarını Gösteren Sütun Grafik.

Şekil 4.4.’deki grafiğin detaylarına Tablo 4.4.’de yer verilmiştir. Şekil 4.4.’de görüldüğü gibi reklam içeren oyunların sayısı içermeyen oyunların sayısından oldukça fazladır.

Tablo 4.4. Mağazadaki Oyunların Ne Kadarının Reklam İçerdiğini Gösteren Frekanslar ve Yüzdelere

	Frekanslar(#)	Oranlar(%)
Reklam içermiyor	2489	17.18
Reklam içeriyor	11997	82.82

Tablo 4.4.’de verilen bilgilere göre mağazadaki tüm oyunların %82.82’si reklam içermektedir. Oyunların çoğunun ücretsiz olduğu göz önünde bulundurulursa bu sonuç şaşırtıcı değildir.



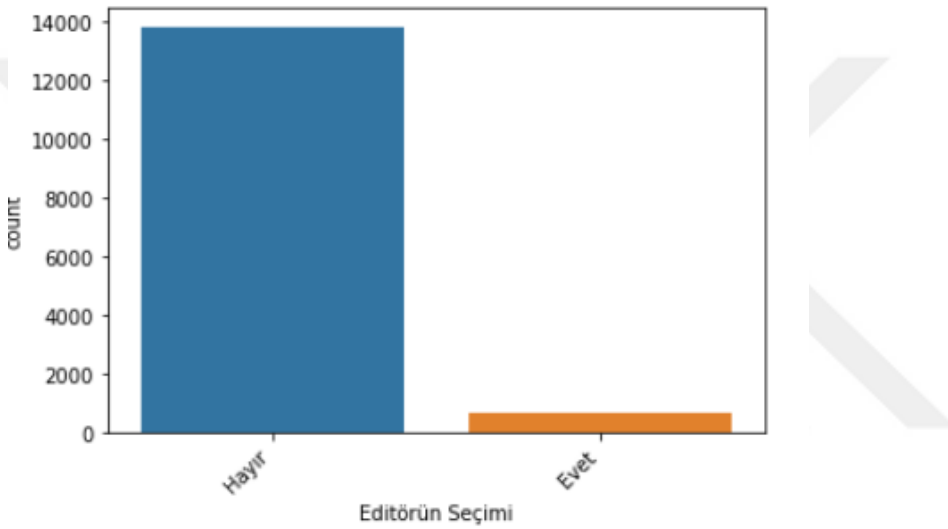
Şekil 4.5. Uygulama İçi Satın Alma İçeren Oyunların Frekanslarını Gösteren Sütun Grafik

Şekil 4.5.'ye göre uygulama içi satın alma içeren oyunların sayısı oldukça fazladır. Grafiğin detaylarına Tablo 4.5.'de yer verildi.

Tablo 4.5. Uygulama İçi Satın Alma İçeren Oyunların Frekansları ve Yüzdeleri

	Frekanslar(#)	Oranlar(%)
Uygulama içi satın alma içermiyor	4738	32.71
Uygulama içi satın alma içeriyor	9748	67.29

Tablo 4.5.'ye göre mağazadaki oyunların %67.29'unun uygulama içi satın alma içerdiği görülmektedir.



Şekil 4.6. Editörün Seçimi Olma Durumunun Frekanslarını Gösteren Sütun Grafik

Şekil 4.6.'ya bakılacak olursa editörün seçimi olan oyunların sayısı oldukça azdır. Grafiğin detaylarına Tablo 4.6.'da yer verildi.

Tablo 4.6. Editörün Seçimi Olan Uygulamaların Sıklıkları ve Oranları

	Frekanslar(#)	Oranlar(%)
Editörün Seçimi Değil	13838	95.53
Editörün Seçimi	648	4.47

Tablo 4.6.'da görüldüğü gibi uygulamaların yalnızca %4.47'si editörler tarafından önerilmektedir.

4.2. Tanımlayıcı İstatistikler

Tablo 4.7.’de sayısal değişkenlere ait tanımlayıcı istatistiklere yer verildi. Bu sayısal değişkenlerden “En Ucuz” ve “En Pahalı” değişkenleri sonradan türetilmiş değişkenlerdir ve oyun içi satılan ürünlerin en düşük ve en yüksek fiyat aralıklarından elde edilmiş sınır değerleri taşıyan değişkenlerdir.

Tablo 4.7. Sayısal Değişkenlere Ait Tanımlayıcı İstatistikler

	Sayı	Ortalama	Std.	Min	%25	%50	%75	Maks.
Puan	14486	4.09	0.41	1.30	3.90	4.10	4.40	5.00
Yorum	14486	137256.57	1249790.01	5.00	1275.00	7400.50	43200.50	90670981.00
Fiyat	14486	2.29	26.22	0	0	0	0	1329.99
Ekran Görüntüsü	14486	15.21	6.33	2.00	10.00	15.00	21.00	32.00
Boyut	14486	57427.60	47503.22	59	31000	50000	73000	1400000
İsim Uzunluğu	14486	30.30	13	2	19	29	43	50
İndirilme	14486	4902552.33	24147867	10.00	100000	1000000	1000000	1000000000
En Ucuz	14486	7.99	26.99	0	0	3.91	8.59	1329.99
En Pahalı	14486	229.53	345.68	0	0	35.99	339.99	1529.99

Tablo 4.7. incelendiğinde mağazadaki en düşük derecelendirme puanının 1.30 olduğu görülmektedir. En yüksek derecelendirme puanı ise 5’tir. Veri setinin çoğunluğu ücretsiz oyunlardan oluştuğu için fiyat değerinin 3. Kartil değeri dahi 0’dır. En düşük indirme sayısı 10 iken en yüksek indirme sayısı 1 milyardır. Uygulama içi satılan en ucuz ürünler en az 0 lira iken en fazla 1329 liradır. Oyun içi satılan en pahalı ürünlerin en az 0 lira iken en fazla 1529 liradır. Elde edilen frekanslarla araştırmanın 4. sorusu olan “Mağazadaki oyunların, editörün seçimi olma, ücretli olma, reklam içermesi, uygulama içi satın alma içermesi vb. frekansları nelerdir?” sorusu cevaplanmıştır.

4.3. Kruskal Wallis Testleri

Tablo 4.8. Farklı Oyun Türlerine Göre Yorum Sayılarının Karşılaştırılmasına İlişkin Kruskal Wallis Testi Sonucu

Tür	N	Ortc.	H	p
Aksiyon	1678	8118.5	461.927	.000
Arcade	1129	7315		
Bulmaca	1226	6302.5		
Eğitici	753	2519		
Eğlencelik Bilgi	207	5632		
Kağıt	334	12399.5		
Kelime	273	13899		
Klasik	1441	10051		
Kumarhane	333	17070		
Macera	1195	4485		
Masa	702	3739		
Müzik	235	9256		
Rol Yapma	1057	9564		
Simülasyon	1763	6278		
Spor	686	8300.5		
Strateji	749	13136		
Yarış	725	11037		

Tablo 4.8.'deki Kruskal Wallis testi sonucuna göre, oyun türlerine göre yorum sayıları arasında anlamlı bir fark vardır($H(16)= 461.927$, $p=.000$). Eğitici oyunlarla tüm oyun türleri arasında anlamlı bir farklılık vardır. Masa oyunlarının macera oyunları ve eğlencelik bilgi oyunları dışındaki tüm oyun türleriyle arasında anlamlı bir fark vardır. Macera oyunlarının eğlencelik bilgi oyunları dışında tüm oyunlarla arasında anlamlı bir fark vardır. Eğlencelik bilgi oyunlarının yalnızca strateji, kelime ve kumarhane oyunları ile arasında anlamlı bir farklılık vardır. Simülasyon oyunlarının klasik, rol yapma, kağıt, yarış, strateji, kelime ve kumarhane oyunları arasında anlamlı bir farklılık vardır. Bulmaca oyunlarının yarış, strateji, kelime ve kumarhane oyunları ile arasında anlamlı bir fark vardır. Arcade oyun türünün yarış, strateji, kelime ve kumarhane oyunları ile arasında anlamlı bir fark vardır. Aksiyon

oyunlarının strateji oyunları ve kumarhane oyunları arasında anlamlı bir fark vardır. Bu sonuca göre yorum sayıları yani alınan reaksiyon önemseniyorsa kelime türünden oyun geliştirilmesi önerilebilir. Çünkü medyan değeri en yüksek oyun kategorisi kelime türüdür. Bununla birlikte eğitici oyunlar en az yorumu almaktadır. Bu durum göz önünde bulundurulduğunda alınan reaksiyona önem veren bir geliştiricinin eğitici oyun türünde yayın yapması önerilmez.

Tablo 4.9. Farklı Oyun Türlerine Göre Boyutların Karşılaştırılmasına İlişkin Kruskal Wallis Testi Sonucu

Tür	N	Ortc.	H	p
Aksiyon	1678	61000	1,092.881	.000
Arcade	1129	47000		
Bulmaca	1226	41926,84		
Eğitici	753	44260,55		
Eğlencelik Bilgi	207	29000		
Kağıt	334	50377,43		
Kelime	273	42963,12		
Klasik	1441	46739,39		
Kumarhane	333	69000		
Macera	1195	54479,97		
Masa	702	33000		
Müzik	235	50000		
Rol Yapma	1057	64000		
Simülasyon	1763	61000		
Spor	686	49905,46		
Strateji	749	65437,55		
Yarış	725	56292,6		

Tablo 4.9.'daki Kruskal Wallis testi sonucuna göre, oyun türlerine göre boyutlar arasında anlamlı bir fark vardır($H(16)= 1,092.881$, $p=.000$). Masa Oyunlarının eğlencelik bilgi oyunları dışındaki tüm oyunlarla arasında anlamlı bir farklılık vardır. Eğlencelik bilgi oyunları için de aynı durum geçerlidir. Bulmaca oyunları ile spor, macera, yarış, aksiyon, simülasyon, strateji, rol yapma oyunları ve kumarhane oyunları arasında anlamlı bir fark vardır. Eğitici oyunların spor, macera, yarış, aksiyon, simülasyon, strateji, rol yapma oyunu ve kumarhane oyunları arasında

anlamli bir fark bulunmaktadir. Kelime oyunlarinin yariş, aksiyon, simülasyon, strateji, rol yapma oyunu ve kumarhane oyunları arasında anlamli bir farklılık vardır. Arcade oyunların macera, yariş, aksiyon, simülasyon, strateji, rol yapma ve kumarhane oyunları arasında anlamli bir farklılık vardır. Klasik oyunların macera, yariş, aksiyon, simülasyon, strateji, rol yapma ve kumarhane oyunları arasında anlamli bir farklılık vardır. Müzik oyunlarının ise aksiyon, strateji, simülasyon, rol yapma oyunu ve kumarhane oyunları ile arasında anlamli bir farklılık vardır. Bu grupların her birinde p değeri 0.05'ten küçüktür. Sonuçlara göre eğlencelik bilgi oyunlarının boyut medyan değeri en düşüktür. Yani bir geliştirici düşük boyutta bir oyun geliştirerek piyasaya çıkmayı hedefliyorsa eğlencelik bilgi oyunları kategorisini tercih etmelidir.

Tablo 4.10. Farklı Oyun Türlerine Göre İndirilme Sayılarının Karşılaştırılmasına İlişkin Kruskal Wallis Testi Sonucu

Tür	N	Ortc.	H	p
Aksiyon	1678	1000000	327.316	.000
Arcade	1129	1000000		
Bulmaca	1226	500000		
Eğitici	753	500000		
Eğlencelik Bilgi	207	500000		
Kağıt	334	500000		
Kelime	273	1000000		
Klasik	1441	1000000		
Kumarhane	333	1000000		
Macera	1195	500000		
Masa	702	100000		
Müzik	235	1000000		
Rol Yapma	1057	500000		
Simülasyon	1763	1000000		
Spor	686	1000000		
Strateji	749	500000		
Yariş	725	1000000		

Tablo 4.10.'daki Kruskal Wallis testi sonucuna göre, oyun türlerine göre indirilme sayıları arasında anlamli bir fark vardır ($H(16) = 327.316$, $p = .000$). Masa

oyunlarının eğitici, strateji, bulmaca, rol yapma, simülasyon, kumarhane, arcade, aksiyon, spor, kelime, müzik, klasik ve yarış oyunları ile arasında anlamlı bir farklılık vardır. Macera oyunlarının rol yapma, simülasyon, kumarhane, arcade, aksiyon, spor, kelime, müzik, klasik ve yarış oyunları ile arasında anlamlı bir farklılık vardır. Eğlencelik bilgi oyunlarının sadece klasik ve yarış oyunları ile arasında anlamlı bir farklılık vardır. Kağıt oyunlarının müzik, klasik ve yarış oyunları ile arasında anlamlı bir farklılık vardır. Eğitici oyunların aksiyon, müzik, klasik ve yarış oyunları ile arasında anlamlı bir farklılık vardır. Strateji, bulmaca ve rol yapma oyunlarının klasik ve yarış oyunları arasında anlamlı bir farklılık vardır. Ne var ki pek çok oyunun indirilme medyan değerleri benzerdir. Bu sebeple indirme sayılarını önemseyen bir geliştiricinin önünde pek çok seçenek vardır.

Tablo 4.11. Farklı Oyun Türlerine Göre Derecelendirme Puanlarının Karşılaştırılmasına İlişkin Kruskal Wallis Testi Sonucu

Tür	N	Ortc.	H	p
Aksiyon	1678	4.1	1,122.58	.000
Arcade	1129	4.1		
Bulmaca	1226	4.3		
Eğitici	753	4.1		
Eğlencelik Bilgi	207	4.2		
Kağıt	334	4.4		
Kelime	273	4.4		
Klasik	1441	4.1		
Kumarhane	333	4.5		
Macera	1195	4.1		
Masa	702	4.2		
Müzik	235	4.1		
Rol Yapma	1057	4.1		
Simülasyon	1763	4.0		
Spor	686	4.1		
Strateji	749	4.2		
Yarış	725	4.1		

Kruskal Wallis testi sonucuna (Tablo 4.11.) göre, oyun türlerine göre derecelendirme puanları arasında anlamlı bir fark vardır ($H(16) = 1,122.58$, $p = .000$).

Simülasyon oyunları ile tüm oyunlar arasında anlamlı bir fark vardır. Yarış oyunlarının rol yapma, masa, strateji, eğlencelik bilgi, bulmaca, kağıt, kelime ve kumarhane oyunları ile aralarında anlamlı bir fark vardır. Aksiyon oyunlarının, rol yapma, masa, strateji, eğlencelik bilgi, bulmaca, kağıt, kelime ve kumarhane oyunları ile aralarında anlamlı bir fark vardır. Spor oyunlarının strateji, bulmaca, kağıt, kelime ve kumarhane oyunları ile aralarında anlamlı bir fark vardır. Macera oyunlarının, strateji, bulmaca, kağıt, kelime ve kumarhane oyunları ile aralarında anlamlı bir fark vardır. Eğitici oyunların, bulmaca, kağıt, kelime ve kumarhane oyunları aralarında anlamlı bir farklılık vardır. Klasik oyunların ise bulmaca, kağıt, kelime ve kumarhane oyunları ile aralarında anlamlı bir fark vardır. Arcade oyunlara bakılacak olursa, bu oyun türünün de bulmaca kağıt kelime ve kumarhane oyunları arasında anlamlı bir fark vardır. Tüm gruplarda p değeri 0.05'ten küçüktür. Geliştiriciler açısından düşünülecek olursa, yüksek puanlara sahip bir oyun üretmek istenildiği takdirde kumarhane oyunları seçilebilir.

Elde edilen Kruskall Wallis sonuçları araştırmaya ait 5. soru olan “oyun türlerine ait öznelilikler arasında anlamlı bir farklılık var mıdır?” sorusunu cevaplamaktadır.

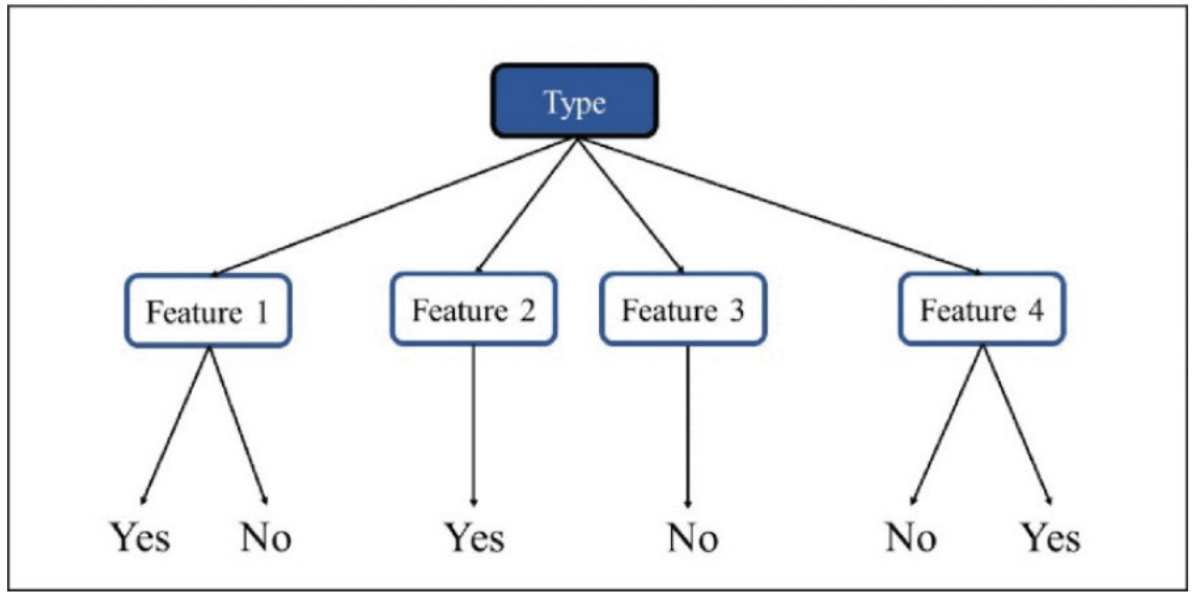
4.4. İkili Sınıflandırma

4.4.1. Tahmin Modelleri

İkili sınıflandırmada tahmin modelleri için aşağıdaki modeller kullanıldı (Nabipour et al., 2020).

A. Decision Tree (Karar Ağacı)

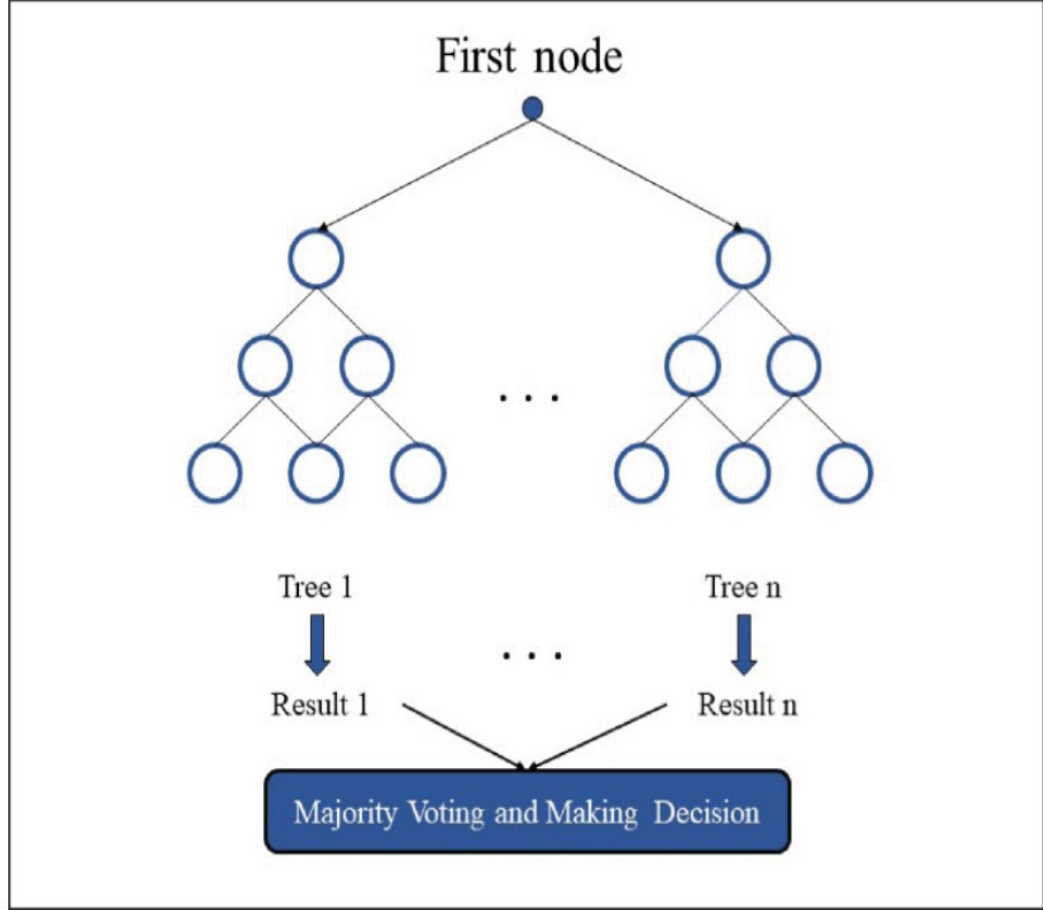
Karar Ağacı hem regresyon hem de sınıflandırma problemleri için kullanılan ortak bir denetimli öğrenme yaklaşımıdır. Tekniğin amacı, veri seti ve ilgili özelliklerden şekillenen kolay karar kurallarını kullanarak bir hedef tahmin etmektir.



Şekil 4.7. Karar Ağacının Şematik Gösterimi (Nabipour et al., 2020).

B. Random Forest (Rastgele Orman)

Çok sayıda karar ağacı rastgele bir orman modeli oluşturur. Model temelde orman olarak adlandırılan ağaçların tahmin sonucunun ortalamasını alır. Ayrıca algoritma, ağaçlar oluştururken eğitim verilerini rastgele seçme, düğümleri bölerken değişkenlerin bazı alt kümelerini rastgele seçme ve her bir temel karar ağacındaki her düğümü bölmek için tüm değişkenlerin yalnızca bir alt kümesini kabul etme olmak üzere üç rastgele fikir içerir.



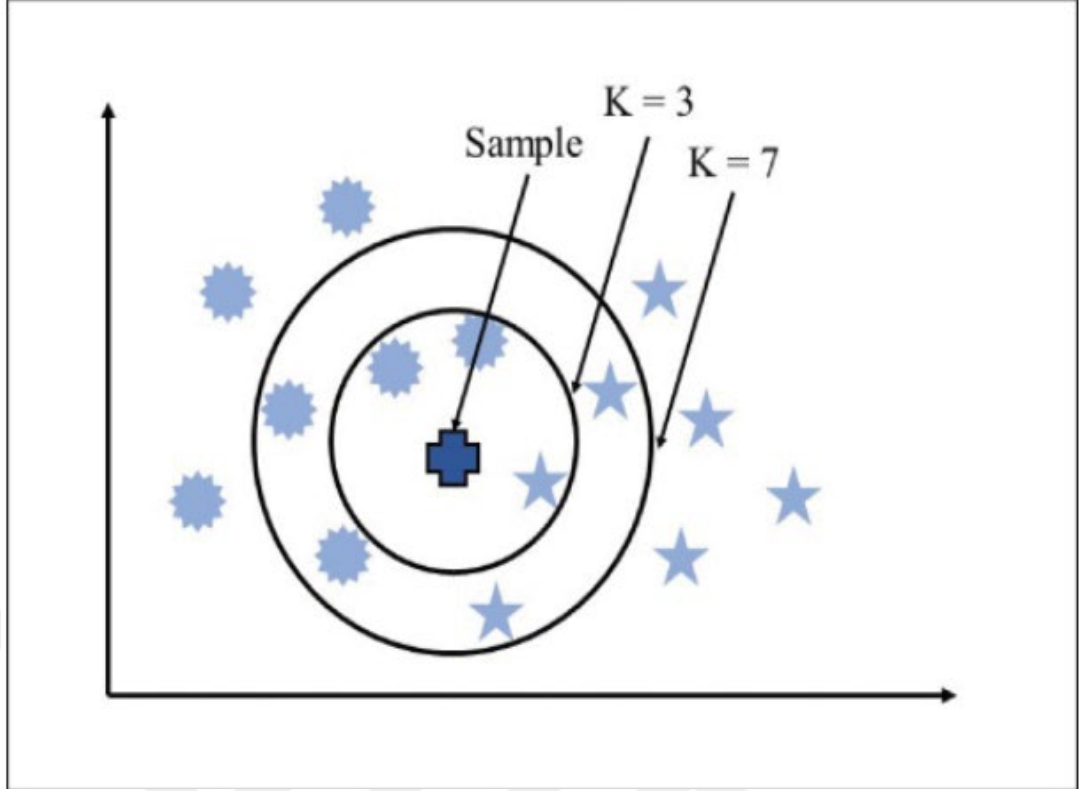
Şekil 4.8. Rastgele Orman Algoritmasının Şematik Gösterimi (Nabipour et al., (2020).

C. Naïve Bayes

Naive Bayes sınıflandırıcısı, sınıf değişkeninin değeri verilen özellikler arasında güçlü bağımsızlık varsayımları olan Bayes teoremine dayalı olasılıksal sınıflandırıcıların bir üyesidir. Bu yöntem, bir dizi denetimli öğrenme algoritmasıdır.

D. KNN (K En Yakın Komşu Yöntemi)

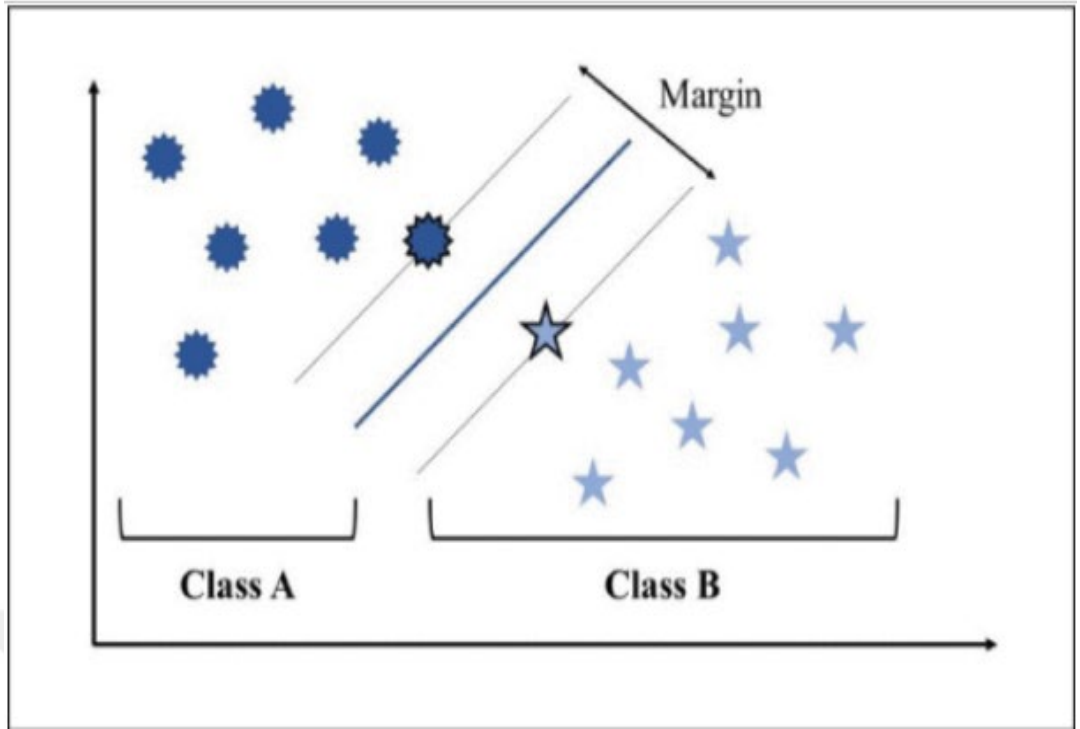
KNN tarafından temel veri dağıtımı için herhangi bir varsayım olmadığı için, genellikle KNN için tembel öğrenme ve parametrik olmayan algoritma olmak üzere iki özellik önerilmektedir. Yöntem, hedefleri bulmak için bazı adımları takip eder: Veri setini eğitim ve test verilerine bölme, K değerini seçme, hangi uzaklık fonksiyonunun kullanılması gerektiğini belirleme, test verisinden bir örnek seçme (yeni bir örnek olarak) ve n'ye olan uzaklığı hesaplama. Bununla birlikte eğitim örnekleri, elde edilen mesafelerin sıralanması ve k-en yakın veri örneklerinin alınması ve son olarak, test sınıfının k komşusunun çoğunluk oyu ile örneğe atanması şeklinde devam eder.



Şekil 4.9. KNN Algoritmasının Şematik Gösterimi (Nabipour et al., 2020).

E. SVC (Destek Vektör Makineleri)

Destek Vektör Makineleri (SVM'ler), sınıflandırma ve regresyon problemleri için kullanılabilecek bir dizi denetimli öğrenme yaklaşımıdır. Sınıflandırıcı sürümü SVC olarak adlandırılır. Yöntemin amacı, vektörlerle iki sınıf arasında bir karar sınırı bulmaktır. Sınır, veri kümesindeki herhangi bir noktadan uzakta olmalıdır ve destek vektörleri, kenar boşluğu adlı bir boşluk ile gözlem koordinatlarının işaretidir. SVM, bir çizgi veya hiperdüzlem kullanarak iki sınıfı en iyi şekilde ayıran bir sınırdır.



Şekil 4.10. SVC Algoritmasının Şematik Gösterimi (Nabipour et al., 2020).

4.4.2. İkili Sınıflandırma Sonuçları

İndirme sayıları kesikli olarak bulunmaktadır. Bu sebeple regresyon analizine uygun değildir. Çalışma kapsamında indirme oranları ikili gruplara bölündü. Bu grupları sınır değeri (ör. 100 bin) geçip geçmeme durumuna göre oluşturuldu. Birinci grup 10 bin indirmeden büyük olup olmadığı bilgisini taşıyan gruptu. Aynı şekilde ikinci grup 100 bin indirmeyi geçip geçmediğini gösteren gruptu. Üçüncü grupta indirilme sayısının 500 bini geçip geçmediği sorgulandı. Son grupta ise indirilme sayısının 1 milyonu geçip geçmediği bilgisi vardı. Bu bilgiler ışığında elde bulunan oyunun indirme barajını tahmin etmek için bahsedilen tahminciler kullanıldı. Her grup için yukarıda bahsedilen tahminciler kullanıldı. Bu tahmincilerin seçilme sebeplerine gelinecek olursa, karar ağacı algoritmasının kullanılma sebebi yorumlanmasının kolay olması veya farklı çıktılara sahip problemleri çözebilmesiydi. Rastgele orman algoritması ise rassallığı içerisinde barındırdığı için tercih edildi. Bu tekniklere ek olarak Naive Bayes, SVC ve KNN algoritmaları da kullanıldı. Bu algoritmalar literatürde sıklıkla yer alan algoritmalar. Bütün algoritmalar için aynı değişkenler kullanıldı.

Tablo 4.12. İkili Sınıflandırma İçin Seçilmiş Karar Değişkenleri

Değişkenler	10,000	100,000	500,000	1,000,000
Puan	X	X	X	X
Yorum				
Ücretli	X	X	X	X
Fiyat				
Editörün Seçimi	X	X	X	X
Boyut	X	X	X	X
Ekran Görüntüsü	X	X	X	X
PEGI skoru				
Baskın Kırmızı				
Baskın Yeşil				
Baskın Mavi				
İsim Uzunluğu	X	X	X	X
İndirilme				
Reklam	X	X	X	X
Uygulama İçi Sat.	X	X	X	X
İndirilme/Yorum				
En Ucuz	X	X	X	X
En Pahalı	X	X	X	X
Not: X sınıflandırma modeline dahil edilen değişkenleri göstermektedir.				

Tablo 4.12.'de tahmin modeline dahil edilen değişkenler gösterilmektedir. Yorum sayısı değişkeni doğası gereği indirilme sayısı değişkeni ile ilişkili olduğu için modele dahil edilmedi. Tüm tahmin modellerinde aynı değişkenler modele dahil edildi. Bu değişkenlere karar verilirken, hedef değişkeni ve tahmin değişkenleri arasında oluşturulan korelasyon tablosundaki değerlerden yararlanıldı. Ürünlerin içerik derecelendirmesi, fiyatı ve logolarına ait baskın renk kodları tahmin başarısında anlamlı bir değişiklik göstermediği için modele dahil edilmedi.

Tablo 4.13. İkili Sınıflandırma Doğruluk Skorları

Algoritma	İndirilme >10,000 (acc.)	İndirilme >100,000 (acc.)	İndirilme >500,000 (acc.)	İndirilme >1,000,000 (acc.)
Random Forest	0.95	0.85	0.71	0.68
Naive Bayes	0.95	0.81	0.62	0.56
Decision Tree	0.93	0.77	0.66	0.61
KNeighbors	0.95	0.77	0.61	0.58
SVC	0.95	0.81	0.62	0.54

Accuracy (doğruluk) skoru sınıflandırma problemlerinde yapılan doğru tahmin oranını bulmak için kullanılan bir metriktir. Tablo 4.13.'de görüldüğü gibi, 10 bin ve üzeri indirmeyi tahmin eden ikili sınıflandırmada en başarılı modeller %95 başarı ile Random Forest, Naive Bayes, K en yakın komşu ve SVC oldu. 100 bin ve üzeri indirmeyi tahmin eden ikili sınıflandırmada ise en başarılı modeller %85 ile Random Forest,%81 ile Naive Bayes ve %81 SVC algoritmaları oldu. 500 bin ve üzeri indirmeyi tahmin eden ikili sınıflandırmada en başarılı modeller %71 başarı ile Random Forest ve %66 ile Decision Tree sınıflayıcıları oldu. 1 milyon ve üzeri indirmeyi tahmin eden sınıflamaya göre ise en başarılı algoritmalar Random Forest ve Decision Tree oldu. Tablo 4.13.'ye göre indirme sayısı arttıkça accuracy (doğruluk) skoru düşmektedir. Böylelikle araştırmaya ait ilk soru olan “Oyun mağazaya yüklenmeden önce indirilme sayısı tahmin edilebilir mi?” sorusu da cevaplanmış oldu. İndirme sınırları yüksek doğrulukla tahmin edildi.

4.5. Regresyon Analizi

“Puan” ve “Yorum” hedef değişkenleri kullanılarak regresyon analizi yapılmıştır. Tablo 4.10.'da regresyon analizine hangi değişkenlerin dahil edildiği gösterilmektedir. Regresyonda araştırmacı, bağımsız değişkenin değerine dayanarak bağımlı değişkenin değerini tahmin etmeye çalışır. Çoklu doğrusal regresyonda, bağımlı değişkenin değerini tahmin etmek için iki veya daha fazla bağımsız değişken aynı anda ağırlıklandırılır. Çoklu regresyon analizi bağımlı ve bağımsız değişkenlerden oluşur. Bağımlı değişken sayısı tektir fakat bağımsız değişken sayısı

birden fazla olabilir. Eğer tek bağımsız değişken var ise basit doğrusal regresyon, iki ya da daha fazla bağımsız değişken var ise çoklu doğrusal regresyon adı verilmektedir (Tabar vd. 2021). Regresyon analizinde değişkenler arasındaki ilişkiyi fonksiyonel olarak anlamlandırmak ve bu ilişkiyi bir modelle açıklamak hedeflenmektedir (Chatterjee and Hadi, 2015). Değişkenlerden birini veya değişkenin kategorisini önceden belirlenen düzeylerde tutarak diğer değişkenin bu düzeye göre nasıl değişim gösterdiğini inceleme kuralına dayanmaktadır (Tabar et al., 2021). “Ne neyin fonksiyonudur?” sorusuna cevap arayan regresyon çözümlemesinde bilinenler yardımıyla bilinmeyenlerin tahmin edilmesi amaçlanır (Serper, 2017). Regresyon çözümlemesinin birçok şartı vardır. Bunlardan birisi verilerin normal dağılıma uymasıdır (Serper, 2017). Ne var ki makine öğrenmesi metotlarında bu varsayımlar ihlal edilebilir. Bunun nedeni, istatistiklerin model çıkarımına, makine öğreniminin ise doğru tahminlere vurgu yapmasıdır. Makine öğreniminde genellikle çoğu zaman cevabın nasıl alındığı değil sadece sonucun örneklemin hem içinde hem de dışında sıkı uyum içerisinde olması önemsenir. (Breiman, 2001).

Tablo 4.14. Tahmin Edici Regresyon Modelleri İçin Seçilen Karar Değişkenleri

Değişkenler	Puan	Yorum
sabit	3.841***	-191629.694***
Puan		57600.664***
Yorum	0.000***	
Ücretli	0.085***	-62069.044***
Fiyat		
Editörün Seçimi	0.066***	137851.578***
Boyut	0.000***	0.001***
Ekran Görüntüsü	0.008***	1969.864***
PEGI Skoru		
Baskın Kırmızı		
Baskın Yeşil		
Baskın Mavi		
İsim Uzunluğu		-2685.682***
İndirilme	-0.000***	

Tablo 4.14. Devam

Reklam	0.037***	11981.465**
Uygulama içi satın alma	0.141***	40425.546***
İndirilme/Yorum	-0.000***	
En Pahalı	0.000***	96.635***
En Ucuz		-1947.802***
Not: * p<0.10, ** p<0.05, *** p<0.01		

Tablo 4.14.’deki yıldızlar anlamlı p değerlerini göstermektedir. Üç yıldız p değerinin 0.01’den küçük olduğunu, iki yıldız 0.05’ten küçük olduğunu tek yıldız ise 0.10’dan küçük olduğunu ifade eder. Hem “Puan” hem de “Yorum” tahmin modellerinde, modele dahil edilen tahmin değişkenlerinin hedef değişken üzerinde istatistiksel olarak anlamlı etkisi olduğu görülmektedir. Aralarında doğal bir ilişki olmasından dolayı “İndirilme” değişkeni ve ilişkili diğer değişkenler “Yorum” değişkeninin hedef değişken olduğu modele dahil edilmedi. İki tahmin modeli için de farklı tahmin değişkenleri modellere dahil edildi. Değişkenler seçilirken hedef değişkenlerin tahmin değişkenleri ile ilişkisini gösteren korelasyon tablolarından ve p değerlerinden faydalanıldı. Düşük ilişki gösteren değişkenler modelden çıkarılmıştır. Örneğin renk kodlarını barındıran “Baskın Kırmızı”, “Baskın Yeşil” ve “Baskın Mavi” değişkenlerinin tahmin modelinde herhangi bir etkisi bulunamadığı için modele dahil edilmedi.

Tablo 4.15. Lineer Regresyon Modelleri İçin Performans Metrikleri

	MAE	MSE	RMSE	R²
Puan	0.28	0.14	0.37	13.38
Yorum	93000.40	28326384621.25	168304.44	17.27

Tablo 4.15.’de lineer regresyon modeline ait performans metrikleri incelenmektedir. Ortaya çıkan regresyon modelinin başarısını ölçmek için ortalama mutlak hata (MAE), ortalama kare hata (MSE), hata kareler ortalamasının kare kökü (RMSE) ve düzeltilmiş belirleme katsayısı (R²) kriterleri kullanıldı. Bu kriterlere ait denklemler aşağıda verildi (Kayakuş, 2021).

$$R^2 = 1 - \frac{\sum (y_i - x_i)^2}{\sum ((Y_i - Y_{ort})^2)} \quad (4,1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2}$$

(4,2)

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2 \quad (4,3)$$

$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n} \quad (4,4)$$

Burada i veri sayısını, y gerçek değeri, x tahmin değerini göstermektedir. Bu kriterlere göre yüksek R^2 ve en düşük RMSE, MSE ve MAE değerleri en başarılı modeli belirlemektedir (Kayakuş, 2021). Tablo 4'e göre "Ratings" hedef değişkenini tahmin eden modele göre % 13.38'lik bir varyans açıklanmaktadır. Bu sayı Reviews değişkeni için %17.27'dir. R^2 değerleri sosyal bilimler için anlamlı düzeyde kabul edilmektedir. Cohen'e (1988) göre R^2 için önerilen değerler şu şekildedir: 0.02 (küçük), 0.13 (orta) ve 0.26 (büyük). Çalışmadaki iki değer de orta düzeyde anlamlılık taşımaktadır. Bu sonuca göre modeldeki değişkenlerde yorum sayısı ve değerlendirilme puanlarının anlamlı bir şekilde tahmin edilebileceği yorumu yapılabilir. Bu sonuçlar araştırmaya ait 2. ve 3. sorular olan "Yayınlanmadan önce bir oyunun ne kadar reaksiyon alacağı yani yorum sayısı tahmin edilebilir mi?" ve "Değerlendirilme puanı oyun mağazaya yüklenmeden önce tahmin edilebilir mi?" sorulara cevap niteliğindedir.

5. Sonuç ve Tartışma

Mobil oyun sektörü giderek büyümektedir ve Google Play en çok uygulama indirilen mağazalardan biridir. Bu çalışmada Google Play Store Oyunlar kategorisine dair önemli bulgulara yer verildi. Çalışma bulgularına göre mağazada en çok simülasyon oyunları bulunmaktadır. Mağazadaki tüm oyunların %60'ını 3 yaş ve üzeri için uygun olan oyunlar oluşturmaktadır. Mağazadaki oyunların %91'i ücretsiz oyunlardan oluşmaktadır. Oyunların %83'ü reklam içermektedir. Oyunların %67'si uygulama içi satın alma içermektedir. Bu sonuçlara göre mağazada freemium stratejisinin hakim olduğu yorumu yapılabilir. Öyle ki uygulamaların yalnızca %9'u ücretlidir.

Çalışma kapsamında ikili sınıflandırma kullanılarak indirme sayıları tahmin edilmeye çalışıldı. İndirme sayılarının 10 bin , 100 bin, 500 bin ve 1 milyonu geçip geçmediğini tahmin eden modeller kuruldu. İndirme sayısının 10 bini geçip geçmediğini tahmin eden grupta en başarılı tahmin algoritmaları %95 ile Random Forest, Naive Bayes, Kneighbors ve SVC oldu. İndirme sayısının 100 bini geçip geçmediği sorusunu cevaplayan grupta en başarılı algoritmalar Random Forest (%85), Naive Bayes (%81) ve SVC (%81) oldu. İndirme sayısının 500 bini geçip geçmediğini tahmin eden grupta en başarılı algoritmalar %71 ile Random Forest ve %66 ile Decision Tree Classifier oldu. 1 milyonu geçip geçmeme durumunu anlayan grupta ise yine en başarılı tahminci %68 ile Random Forest oldu. Bu sonuçlardan ilham alınarak uygulamaların mağazaya yüklenmeden önce hangi indirme barajını geçeceğini tahmin eden bir uygulama geliştirilebilir. Geliştirilecek olan bu uygulama bir internet sitesi aracılığıyla yayınlanıp geliştiricilerin hizmetine sunulabilir. Bu tahmin sonuçlarına göre yayınlanacak olan oyuna yapılacak reklam ve yatırımların akıbeti şekillendirilebilir. Değişkenlerden biri farklılık gösterdiğinde oyunun indirme sayısının nasıl etkileneceğini belirleyen bir çalışma bu alan için oldukça faydalı olacaktır.

Regresyon modelinde ise yorum ve değerlendirme puanı değişkenleri analiz edildi. Bu çalışmanın sonuçlarına göre yorum sayılarını tahmin eden model %17.27'lik bir varyans açıklamaktaydı. Bu sayı değerlendirme puanı için %13.38'dir. Sonuçlar değerlendirildiğinde, yorum sayısı ve değerlendirme puanlarının tahmin edilebileceği görülmektedir. Regresyon modeli incelendiğinde, “Puan” ve “Yorum” hedef değişkenlerini tahmin edildiğinde değişkenlerin önemi

tabloda görülmektedir. Bu sonuçlara göre oyunun alacağı reaksiyon yani yorum sayısı tahmin edilebilir. Yaratacağı sosyal kanıt etkisini anlamak açısından yorum sayılarıyla birlikte puanların anlaşılması büyük önem arz eder. Tablo 4.14. incelendiğinde yıldızlı değerler ile belirtilen değişkenler görülmektedir. Bu değişkenler hedef değişkenleri açıkladıkları için değerlerdeki değişimler hedef değişkenleri etkileyecektir. İndime sayısını tahmin eden uygulamaya ek olarak bu değişkenleri tahmin eden bir uygulama da geliştiricileri için üretilebilir.

Çalışmaya oyunlara ait logoların baskın renklerinin RGB kodları da dahil edildi. Ancak tahmin modellerinde bu değişkenler etkisiz kaldıkları için kullanılmadılar. Değerlendirme puanlarına bakıldığında örneğin yorum sayısının pozitif bir etkiye sahip olduğu görüldü. Bu sebeple yorum sayısını arttırmanın yani alınan reaksiyonu arttırmanın puanlara bir etkisi olacağı belirlendi. Ekran görüntüsü ve ücretli olma durumu da pozitif bir etkiye sahiptir. Bu sebeple olabildiğince çok ekran görüntüsü kullanmaya çalışmak olumlu bir etki yaratabilir. Bununla birlikte editörün seçimi olmanın da puan üzerinde anlamlı bir etkisi vardır. Oyunların puanlarının yüksek olması isteniyorsa, editörün seçimi olmaya çalışmak önemlidir. PEGI skorunun modelde anlamlı bir etkisi bulunmamaktadır. Ücretli olmanın ise anlamlı bir etkisi vardır. Yani ücret ödemek oyun puanını yükselten bir etmendir. Reklam içerme ve uygulama içi satın alma içermeye de anlamlı bir etki bulunmaktadır. Buradan hareketle puan değişkeninin etkilerini değiştirmek için bahsedilen değişkenlerdeki farklılıkların önemli olduğu sonucu çıkabilir.

Yorum hedef değişkeni için önemli olan tahmin değişkenleri “Puan, ücretli, editörün seçimi, boyut, ekran görüntüsü, isim uzunluğu, reklam, uygulama içi satın alma, en pahalı ve en ucuz” değişkenleridir. Bu değişkenlerde yaratılacak farklılıklar yorum sayısında farklılıklar oluşturacaktır.

Google Play Store’daki verilerin kolay bir şekilde elde edilememesi bu alanda yapılan çalışmaların kısıtlı olmasına sebebiyet verdi. Google Play Store üzerine yapılan makine öğrenmesi çalışmalarında Lengkong ve Marinka (2020) ya da Bashir ve arkadaşları (2019), lineer regresyona yer vermedi. Bu alanda lineer regresyona yer verilmemesi sebebiyle çalışmanın doğrudan kıyaslanabileceği bir çalışma örneğine henüz rastlanmadı. Ancak Google Play Store konusu dışında örneğin satış tahmini alanında bu çalışmaya benzer şekilde ikili sınıflandırma modeli kullanıldı (Dılkı, 2020). İkili sınıflandırma “satıldı”, ”satılmadı” şeklindeki kategorilerden oluşmaktaydı. Çalışmada 0.71 doğruluk ile KNN algoritması en

yüksek doğruluğu sağlarken, bu çalışmada KNN algoritması 0.95 doğruluk ile sonuç verdi. Ne var ki bu sonuç 10,000 ve üzeri indirmeyi tahmin ederken geçerli oldu. Bu çalışmada da indirilme sayısı arttıkça KNN tahmincisi 0.58'e kadar geriledi.

Uygulama geliştiricileri açısından bu çalışma ve bu çalışmadan ortaya çıkan diğer çalışmalar incelendiğinde şu sonuçlara varılabilir: Ücretsiz oyunlar ücretli oyunlardan daha çok indirme sayısına sahiptir. Ancak ücretli oyunlar şaşırtıcı bir şekilde indirilme oranıyla yorum sayısı kıyaslandığında daha yüksek bir yorum oranına sahiptir (Arıcı ve Huseynov 2021b). Bu sonuca göre insanların ücretini ödediği oyuna reaksiyon verme durumlarının daha yüksek olduğu yorumu yapılabilir. Reklam içeren oyunların indirme sayıları ile içermeyenlerin indirme sayıları kıyaslandığında, reklam içeren uygulamaların daha çok indirildiği görüldü (Arıcı ve Huseynov, 2022). Bu sonuca göre uygulama içi reklam vermenin indirilme üzerinde olumsuz bir etkisi olduğu söylenemez. Bu tezden çıkan sonuçlara göre uygulama geliştiricilere yapılabilecek öneriler şunlardır:

1. Yorum ve Puan değişkenleri için etki eden değişkenler önemsenerek uygulama geliştirilmelidir.
2. Uygulama içi satın alma ve reklam öğeleri kullanıcı davranışında olumsuz bir yaklaşım yaratmadığı için tercih edilebilir.
3. Editörün seçimi olmak önemlidir ve bunun için çalışılmalıdır.
4. İndirme sayısını etkileyen değişkenler incelenerek bu alanda çalışmalar arttırılabilir.
5. Ücretli oyunlar daha çok yorum ve puan almaktadır. Ücretli oyun geliştirmek tercih edilebilir.

5.1. Kısıtlamalar ve Öneriler

Bu çalışmaya dair birtakım kısıtlar da ortaya çıktı. Bunlardan en önemlisi ve zorlayıcısı indirme sayılarının kesikli değerlerden oluşmasıdır. Eğer indirme sayısı tam olarak bilinebilseydi regresyon modeliyle uyumlu olabilir ve ikili sınıflandırma yerine sayısal tahmin yapılabilirdi. Bir diğer kısıt oyunların ilk çıkış tarihlerinin bilinmemesiydi. Eğer tarihler bilinseydi indirme sayısındaki zaman faktörü de incelenebilirdi. Çalışmadaki kısıtlardan biri de arama sırasında bazen çok düşük indirmeli oyunlar geldiği için değerlendirme puanları, indirme sayısı ya da yorum

sayısı oluşmayan oyunlar da veri setinde bulunmasıydı. Bu veriler olduğu gibi veri setinden çıkarıldı. Aslında daha çok oyun verisi elde edilmişti.

Yorum değişkeni indirme sayısı tahmin edilirken kullanılmadı çünkü çok indirilen bir oyunun bu sebeple çok yorum alması gibi bir olgu söz konusuydu. Ancak bu durum mağazaya ilk defa yüklenen oyunlar için geçerli olabilir. Örneğin mevcut oyunda yapılan bir değişikliğin indirme sayısında ne kadar değişikliğe sebep olacağını tahmin etmeye çalışan bir tahmincide yorum sayısı da kullanılabilir.



KAYNAKÇA

Aktaş Potur, E., Erginel, N., (2021). "Kalp Yetmezliği Hastalarının Sağ Kalımlarının Sınıflandırma Algoritmaları ile Tahmin Edilmesi", Avrupa Bilim ve Teknoloji Dergisi, Ejosat Özel Sayı 2021 (ARACONF), pp.112-118.

Arı, B., Arı, A. Şengür, A., (2020). "Suicide Prediction from Hemogram with Machine Learning", Avrupa Bilim ve Teknoloji Dergisi, Ejosat Özel Sayı.

Arıcı, N., Huseynov, F., (2021a). "A Descriptive Study Understanding The Current Status of The Games In Google Play Store", In International European Conference on Interdisciplinary Scientific Research-IV, Warsaw, Poland, 08-09.

Arıcı, N., Huseynov, F., (2021b). "An Explanatory Study Comparing The Paid And Free Games In The Google Play Store", In International European Conference on Interdisciplinary Scientific Research-IV, Warsaw, Poland, 08-09.

Arıcı, N., Huseynov, F., (2021c). "The Effect of Being The "Editors' Choice" On Google Play Store Games: Does It Really Make Any Difference?", In III. International Ankara Multidisciplinary Studies Congress, Ankara, Turkey, 5-7.

Arıcı, N., Huseynov, F., (2022). "The Effects Of In-App Advertisements On User Behavior In Google Play Store Games", In 5TH International Aegean Conferences (On Social Sciences & Humanities), İzmir, Turkey, 25-26.

Breiman, L., (2001). "Statistical modeling: The two cultures (with comments and a rejoinder by the author)", Statistical science, 16(3), pp. 199-231.

Çelik, E., Dal, D., Aydın, T., (2021). "Duygu Analizi İçin Veri Madenciliği Sınıflandırma Algoritmalarının Karşılaştırılması", Avrupa Bilim ve Teknoloji Dergisi, (27), pp. 880-889.

Celik, U., (2019). "Veri Madenciliği Yöntemleri İle İş Yaşam Dengesinde Yıpranma Durumu Tahmini", Yönetim ve Ekonomi Araştırmaları Dergisi, 17 (1), pp. 63-76.

Chatterjee, S., Hadi, A.S., (2015). "Regression analysis by example", John Wiley & Sons.

Cialdini, R. B., Cialdini, R. B., (2007). "Influence: The psychology of persuasion", Vol. 55, p. 339, New York: Collins.

Cialdini, R. B., Trost, M. R., (1998). "Social influence: Social norms, conformity and compliance".

Cohen, J., (1988). "Statistical Power Analysis for the Behavioral Sciences", 2nd Ed. Hillsdale, NJ: Laurence Erlbaum Associates,

Demir, H., Erdoğmuş, P., Kekeçoğlu, M., (2018). "Destek Vektör Makineleri, YSA, K-Means ve KNN Kullanarak Arı Türlerinin Sınıflandırılması", Düzce Üniversitesi Bilim ve Teknoloji Dergisi, 6 (1), pp. 47-67.

Demirci, M., (2019). "Destek Vektör Makineleri ve M5 Karar Ağacı Yöntemleri Kullanılarak Yağış Akış İlişkisinin Tahmini", Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi, c. 10, sayı. 3, pp. 1113-1124.

Dılkı, G., (2020). "Makine Öğrenmesi Algoritmalarının Sınıflama Problemleri Üzerinden Karşılaştırılması: Satış Tahmini", PressAcademia Procedia, 12 (1), pp. 82-83.

Göçgün, Ö. F., Onan, A., (2021). "Amazon Ürün Değerlendirmeleri Üzerinde Derin Öğrenme/Makine Öğrenmesi Tabanlı Duygu Analizi Yapılması", Avrupa Bilim ve Teknoloji Dergisi, Ejosat Özel Sayı 2021 (ARACONF), pp. 445-448.

Göktaş, M. E., Yağanoğlu, M., (2020). "Veri Bilimi Uygulamalarının Hastalık Teşhisinde Kullanılması: Kalp Krizi Örneği", Journal of Information Systems and Management Research, 2 (2), pp. 26-32.

Kayakuş, M., (2021). "Yazılım Çaba Tahmininde Yapay Sinir Ağları İçin Optimum Yapının Belirlenmesi", Avrupa Bilim ve Teknoloji Dergisi, Ejosat 2021 Ocak, pp. 43-48.

Kop, B. Ş., Bayındır, L., (2021). "Bebek Ağlamalarının Makine Öğrenmesi Yöntemleriyle Sınıflandırılması", Avrupa Bilim ve Teknoloji Dergisi, (27), pp. 784-791.

Latif, R. M. A., Abdullah, M. T., Shah, S. U. A., Farhan, M., Ijaz, F., Karim, A., (2019). "Data Scraping from Google Play Store and Visualization of its Content for Analytics. In 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies, pp. 1-8.

Nabipour, M., Nayyeri, P., Jabani, H., Shahab, S., Mosavi, A., (2020). "Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data; a comparative analysis", IEEE Access, 8, 150199-150212.

Qiang, Z., Dai, F., Lin, H., Dong, Y., (2019). "Research on the Course System of Data Science and Engineering Major", In 2019 IEEE International Conference on Computer Science and Educational Informatization, pp. 90-93.

Reddy, P. B. P., Nallabolu, R., (2020). "Machine learning based Descriptive Statistical Analysis on Google Play Store Mobile Applications", In 2020 Second International Conference on Inventive Research in Computing Applications, pp. 647-655.

Roethke, K., Klumpe, J., Adam, M., Benlian, A., (2020). "Social influence tactics in e-commerce onboarding: The role of social proof and reciprocity in affecting user registrations", Decision Support Systems, 131, 113268.

Sel, A., (2020). "Pandemi Sürecinde Altın Fiyatları ile Kripto Para İlişkisinin Makine

Öğrenme Metotları ile İncelenmesi”, İstatistik ve Uygulamalı Bilimler Dergisi, 1 (2), pp. 85-98.

Şenel, F. A., (2020). “Makine Öğrenmesi Algoritmaları Kullanılarak Kayısı İç Çekirdeklerinin Sınıflandırılması”, Bitlis Eren Üniversitesi Fen Bilimleri Dergisi, 9 (2), pp.807-815.

Serdarer Kuzu, B., Giray Yakut, S., (2020). “Destek Vektör Makineleri Yardımıyla İmalat Sanayisinde Mali Başarısızlık Tahminlerinin Teknoloji Yoğunluğuna Göre İncelenmesi”, Osmaniye Korkut Ata Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 4 (2), pp. 36-54.

Serper, Ö., (2017). “Uygulamalı İstatistik”, Ezgi Kitabevi.

Tabar, M. E., Başara, A. C., Şişman, Y., (2021). “Çoklu Regresyon ve Yapay Sinir Ağları ile Tokat İlinde Konut Değerleme Çalışması”, Türkiye Arazi Yönetimi Dergisi, 3 (1), pp. 1-7.

Uyar, T., Karacan Uyar, K., Yağlı, E., (2021). “Gözetimli Makine Öğrenmesiyle Noktalama ve Etkisiz Kelime Sıklıkları Kullanarak Yazar Tanıma”, Bilişim Teknolojileri Dergisi, 14 (2), pp. 183-190.

Web 1, (2013). <https://play.google.com/console/signup>, (Erişim Tarihi: 11/12/2013).

Web 2, (2012). March 6). Introducing Google Play: All your entertainment, (Erişim Tarihi: 12/05/2012).

Web 3, (2022). <https://www.statista.com/statistics/266210/number-of-available-applications-in-the-google-play-store/>, (Erişim Tarihi: 11/02/2022).

Web 4, (2022). <https://www.statista.com/statistics/742304/annual-new-developer-google-play/>, (Erişim Tarihi: 05/01/2022).

Web 5, (2021). <https://numpy.org/Doc/Stable/Reference/>. Retrieved July 11, 2021, from <https://numpy.org/doc/stable/reference/>, (Erişim Tarihi: 25/04/2021).

Web 6, (2021). <https://zenodo.org/record/5060318>, (Erişim Tarihi: 11/03/2021).

Web 7, (2012). <https://googleblog.blogspot.com/2012/03/introducing-google-play-all-your.html>, (Erişim Tarihi: 18/09/2012).

Web 8, (2018). <https://selenium-python.readthedocs.io/>, (Erişim Tarihi: 10/03/2021).

ÖZGEÇMİŞ

2011 yılında MTSO Anadolu Lisesinden mezun olduktan sonra 2016 yılında Marmara Üniversitesi İstatistik bölümünden mezun olmuştur. Lisans eğitiminden sonra Gebze Teknik Üniversitesi Sosyal Bilimler Enstitüsü İşletme Anabilim dalında yüksek lisans eğitimine başlamıştır.

