

Video Oyunları Kullanıcı Yorumlarının Makine Öğrenmesi İle Tahmini: Veri Bilimi Perspektifinden Bir İnceleme

Prediction of Video Game User Reviews With Machine Learning: A Review From a Data Science Perspective

Kamil Abdullah EŞİDİR

Dr. Fırat Kalkınma Ajansı,

Elazığ Yatırım Destek Ofisi

abdullahesidir@yahoo.com

<https://orcid.org/0000-0002-8106-1758>

Makale Başvuru Tarihi: 09.04.2025

Makale Kabul Tarihi: 27.04.2025

Makale Türü/Article Type: Araştırma Makalesi

ÖZET

Yapılan çalışma, video oyunları kullanıcı yorumlarını veri analizi ve makine öğrenmesi modelleri ile inceleyerek kullanıcı değerlendirme puanlarının tahmin edilmesini amaçlamaktadır. Veri seti; kullanıcı puanları, yorum metinleri, fiyat bilgileri, oyun süresi, oyun türü ve platform gibi hem nicel hem de nitel değişkenleri birlikte içermektedir. Makine öğrenmesi modellerinden Doğrusal Regresyon, Rastgele Orman, Gradyan Artırma ve Ekstra Ağaçlar kullanılarak regresyon tabanlı tahmin modelleri geliştirilmiştir. Modellerin başarımı Ortalama Mutlak Hata (MAE), Ortalama Kare Hata (MSE), Kök Ortalama Kare Hata (RMSE) ve Belirlilik Katsayısı (R^2) metrikleri ile değerlendirilmiştir. Elde edilen analiz bulguları, tüm modellerin yüksek düzeyde doğruluk sağladığını göstermekle beraber, Doğrusal Regresyon modeli en düşük hata değerleri ile en başarılı model olmuştur. Doğrusal ilişkilerin baskın olduğu veri setlerinde basit modellerin dahi üstün performans sergileyebileceği ve model seçiminde veri yapısının belirleyici rol oynadığı anlaşılmıştır. Bulgular, metin madenciliği ve makine öğrenmesi tabanlı yaklaşımların oyun sektörü bağlamında kullanıcı memnuniyetini tahmin etmede etkili olduğunu ortaya koymaktadır. Bu bağlamda çalışma, video oyun sektöründe tüketici davranışlarının veri odaklı analizi ve makine öğrenmesine dayalı karar destek sistemlerinin geliştirilmesine katkılar sağlamaktadır.

Anahtar Kelimeler: Makine Öğrenmesi, Veri Madenciliği, Kullanıcı Yorumları, Video Oyunları, Duygu Analizi.

ABSTRACT

This study aims to predict user rating scores by analyzing video game user reviews with data analysis and machine learning models. The dataset includes both quantitative and qualitative variables such as user ratings, review texts, price information, playing time, game type and platform. Regression-based prediction models were developed using machine learning models such as Linear Regression, Random Forest, Gradient Boosting and Extra Trees. The performance of the models was evaluated with the metrics of Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE) and Coefficient of Determination (R^2). The results of the analysis show that all models provide a high level of accuracy, with the Linear Regression model being the most successful model with the lowest error values. It is understood that even simple models can exhibit superior performance in datasets where linear relationships predominate and that data structure plays a decisive role in model selection. The findings reveal that text mining and machine learning-based approaches are effective in predicting user satisfaction in the context of the gaming industry. In this context, the study contributes to the development of decision support systems based on data-driven analysis of consumer behavior and machine learning in the video game industry.

Keywords: Machine Learning, Data Mining, User Reviews, Video Games, Sentiment Analysis.

1. GİRİŞ

Son yıllarda dijital teknolojilerdeki hızlı gelişim, büyük veri analitiği ve yapay zekâ temelli yaklaşımların birçok alanda yaygın şekilde kullanılmasına olanak tanımıştır. Bu dönüşüm, veri temelli karar alma süreçlerini öne çıkararak, özellikle tüketici davranışlarının daha derinlemesine analiz edilmesini mümkün kılmaktadır. Bu bağlamda, makine öğrenmesi modelleri; karmaşık, yüksek boyutlu ve çeşitli veri yapıları üzerinde tahmine dayalı modeller oluşturmak için etkin biçimde kullanılmaktadır. Makine öğrenmesi, diğer disiplinler ile yakın ilişkiler içerisinde bulunarak, çeşitli alanlarda problem çözme ve karar verme süreçlerini iyileştirmeyi hedeflemektedir (Aslan, 2025).

Veriye dayalı yöntemler, stratejik planlama, müşteri segmentasyonu, ürün geliştirme ve kullanıcı deneyimi yönetimi gibi alanlarda işletmelere önemli avantajlar sağlamaktadır. Bu avantajların başında, kullanıcı etkileşimlerinin ölçülebilir hale getirilmesi ve bu etkileşimlerin istatistiksel olarak modellenmesi gelmektedir. Özellikle oyun sektörü gibi dijital etkileşimin yoğun yaşandığı alanlarda, kullanıcı yorumları ve değerlendirme puanları; memnuniyet, kalite algısı ve tercih eğilimleri hakkında değerli bilgiler sunmaktadır. Video oyun endüstrisi, dinamik kullanıcı davranışları, sürekli yenilenen içerik yapısı ve yüksek rekabet seviyesi ile veri bilimi uygulamaları

açısından önemli bir inceleme alanı sunmaktadır. Fiyat, oyun süresi, oyun türü, grafik kalitesi ve kullanıcı yorumları gibi çok sayıda değişkenin analiz edilmesi, hem sektörel aktörler hem de akademik araştırmalar için stratejik öngörüler üretme potansiyeli taşımaktadır.

Video oyun sektörü, tüketici tercihlerinin ve davranışlarının hızlı ve sürekli değiştiği dinamik bir alandır. Oyun fiyatlandırması, içerik kalitesi, kullanıcı deneyimi ve genel kullanıcı memnuniyeti gibi değişkenler, oyun sektöründe pazar başarısını doğrudan etkileyen temel faktörlerdir. Bu nedenle, oyun sektöründe elde edilen çok boyutlu kullanıcı verilerinin analiz edilmesi ve verilere anlam kazandırılması, sektördeki firmalara rekabet avantajları sağlayabilmektedir. Kullanıcı yorumları ve değerlendirme puanları benzeri veriler ile gerçekleştirilen analizler, oyun geliştiricilerine tüketici beklentilerini daha iyi anlama ve ürünlerini bu doğrultuda şekillendirme fırsatları sunmaktadır. Geleneksel istatistiksel modeller, regresyon analizleri gibi teknikler belirli koşullar altında ve sınırlı değişkenler ile etkili olabilmektedir (Zu ve ark., 2016). Büyük veri setleri ve çok boyutlu veri yapıları söz konusu olduğunda, geleneksel modeller yetersiz kalmakta ve karmaşık ilişkileri açığa çıkaramamaktadır (Ji, 2023).

Çalışma, video oyunlarına ilişkin kullanıcı yorumlarının çok boyutlu yapısını analiz ederek, kullanıcı değerlendirme puanlarının tahmin edilebilirliğini artırmayı hedeflemektedir. Bu amaçla hem sayısal hem de metinsel verilerin bütüncül biçimde işlenmesine olanak sağlayan makine öğrenmesi modellerinden faydalanılmıştır. Böylece, oyun sektörüne yönelik veri odaklı karar destek sistemlerinin geliştirilmesine katkı sağlanması amaçlanmaktadır.

2. LİTERATÜR TARAMASI

Geleneksel istatistiksel yöntemler, sınırlı sayıda değişken içeren yapılandırılmış veri setlerinde etkin sonuçlar üretebilirken, yüksek boyutlu ve heterojen veri yapıları karşısında yetersiz kalmaktadır (Ji, 2023). Bu noktada makine öğrenmesi modelleri, geniş değişken setlerini ve değişkenler arası karmaşık ilişkileri daha etkin biçimde modelleme imkânı sunarak literatürde öne çıkmaktadır. Speer (2021) ve Wu (2023), makine öğrenmesinin doğrusal olmayan ilişkilere ve etkileşimlere duyarlılığı sayesinde, geleneksel modellere kıyasla daha yüksek doğruluk oranları sağlayabildiğini belirtmektedir.

Video oyunları, hem bireysel eğlence aracı hem de sosyal etkileşim platformu olarak öne çıkmaktadır. Bu durum, kullanıcıların oyunlara ilişkin değerlendirmelerini çok boyutlu hale getirmektedir. Mustafaoğlu (2018), e-spor ve aktif oyunlar gibi alt türlerin kullanıcı deneyimini doğrudan etkileyen yapısal farklar içerdiğini belirtmiştir. Altuntaş ve Karaarslan (2017) ise, mobil oyun tercihlerinde yıldız puanı, oyuncu sayısı ve reklam varlığının kullanıcı davranışları üzerindeki belirleyici etkilerine dikkat çekmiştir. Bu çalışmalar, oyun türü ve platform gibi değişkenlerin kullanıcı değerlendirmeleri üzerindeki etkilerini ortaya koyarak, tahmine dayalı modellere bu değişkenlerin entegrasyonunun önemini vurgulamaktadır.

Yorum verileri, kullanıcıların ürünlerle ilgili deneyimlerini, beklentilerini ve duygularını doğrudan yansıttığı için makine öğrenmesi açısından oldukça değerli bir veri kaynağıdır. Tuzcu (2020) çalışmasında, çevrimiçi kitap satış sitesine ait kullanıcı yorumları üzerinde duygu analizi yapmıştır. Python programlama dilinde Çok Katmanlı Algılayıcı (MLP) ve Rapid Miner kullanmıştır. Naive Bayes, Destek Vektör Makineleri ve Lojistik Regresyon makine öğrenmesi modelleri ile elde edilen sonuçlar karşılaştırılmış ve veri seti üzerinde en yüksek sınıflandırma başarısını MLP algoritması göstermiştir. Artan çevrimiçi kullanıcı yorumlarının analizi akademisyenler ve de ticari karar vericiler açısından kritik önem taşımaktadır.

Teknolojinin hızlı ilerlemesiyle e-ticaret platformlarında, kullanıcı katılımını artırmaya yönelik oyunlaştırma unsurlarının (hediye çekleri, rozetler, puan sistemleri, indirim kuponları) etkin kullanımı ön plana çıkmıştır. Turp Gölbaşı ve Ayhan Gökcek (2022)'in yaptıkları çalışmaya göre, bu unsurlar müşteri etkileşimini ve memnuniyetini artırmaktadır. Bu bağlamda, video oyunları kullanıcı yorumlarının makine öğrenmesiyle incelenmesi, dijital platformlardaki etkileşim dinamiklerinin anlaşılmasına teorik bir temel sunmaktadır.

Çabuk ve arkadaşları (2023) çalışmalarında, e-ticaret platformlarından toplanan ürün yorumları kullanılarak, LSTM tabanlı model eğitimiyle otomatik duygu analizi gerçekleştirmiş ve otomatik etiketlenen geniş veri seti kullanılarak dört derin öğrenme modeli ile dört geleneksel makine öğrenmesi modelinin sınıflandırma performansları karşılaştırılmıştır. Elde edilen sonuçlara göre, derin öğrenme modellerinin makine öğrenmesi modellerine kıyasla daha fazla başarı sağladığı görülmüştür. E-ticaret sektöründe artan ürün yorumlarının manuel analiz zorlukları bu şekilde kolaylıkla aşılabilmektedir.

Aydın ve Sarıca (2024) tarafından yapılan çalışmada, dijital oyun geri bildirimlerinin duygu analizi yöntemiyle incelenmesi araştırılmıştır. e-WoM verilerinin r-WoM'a dönüşüm potansiyeli oyun türleri perspektifinden değerlendirilmiştir. PlayStation 5 için en yüksek puanlı 100 oyunun kullanıcı yorumları, web kazıma teknikleri ve Python programlama dili kullanılarak toplanmış; toplamda 44.065 yorum duygu analizi yöntemi ile analiz edilmiştir. Elde edilen bulgular, yarış, rol yapma, nişancı ve benzersiz oyun türleri arasında polarite skorları açısından anlamlı farklılıklar bulunurken, aksiyon ve macera türleri arasında net bir ayırım yapılamadığını göstermektedir. Bu sonuçlar, dijital platformlardaki kullanıcı geri bildirimlerinin duygu analizine dayalı değerlendirilmesinin, oyun sektöründeki pazarlama stratejilerinin belirlenmesinde alternatif ve değerli bir yaklaşım sunduğunu ortaya koymaktadır.

Sayısal özelliklerin (fiyat, süre, puan) modellenmesi yoluyla oyun sektöründe kullanıcı deneyiminin ele alınması, literatürde araştırma boşluğunu doldurmayı hedeflemektedir. Çalışmada kullanılan dört farklı makine öğrenmesi modelinin aynı veri seti üzerinde karşılaştırılması ise yöntemsel çeşitlilik ve de uygulama açısından literatüre katkı sunmaktadır.

3. METODOLOJİ

Analizlerde dört farklı makine öğrenimi modeli kullanılmıştır. Çalışmada, veri ön işleme teknikleri, model eğitim süreçleri, özellik çıkarımı yöntemleri kullanılmıştır. Büyük veriler, çeşitlilik, hız ve hacimden dolayı çeşitli zorluklar barındırmaktadır. Makine öğrenmesi, veriye dayalı problemlerin çözümünde, uygun algoritmalar ile modelleme yapmayı hedefleyen bir çeşit hesaplama disiplini. Makine öğrenmesi, verilerden öğrenerek ve bu öğrendiklerini kullanarak tahminlerde bulunma veya karar verme üzerine kurulu bir yapay zeka alt dalıdır. Makine öğrenmesi modelleri, veri türüne ve problemin karmaşıklığına göre seçilmektedir (Eşidir, 2025a).

3.1. Yazılım ve Donanım Altyapısı

Veri analizi ve modelleme süreçleri, Python 3.12.7 sürümü kullanılarak gerçekleştirilmiştir. Python, geniş kütüphane desteği ile veri bilimi ve makine öğrenimi uygulamaları için ideal bir ortam sağlamaktadır ve aynı zamanda da ücretsizdir (Eşidir, 2025b). Gerekli analizler için NumPy, Pandas, Scikit-learn, TensorFlow ve Keras gibi önde gelen açık kaynak kütüphanelerden faydalanılmıştır. Kodlama ve görselleştirme işlemleri Jupyter Notebook 7.2.2 platformu üzerinde yürütülmüştür. Analizlerde kullanılan bilgisayar sisteminde Windows 11 Pro işletim sistemi ve 16 GB RAM'e sahip Intel Core i7 işlemci mevcuttur. Kullanılan bilgisayar altyapısı, büyük boyutlu veri setlerinin hızlı ve verimli bir biçimde işlenmesine olanak sağlamıştır.

3.2. Veri Seti ve Özellikleri

Analizlerde kullanılan veri seti, Kaggle platformunda yayınlanan ve Jahnavi Paliwal tarafından oluşturulan "Video Game Reviews and Ratings" başlıklı çalışmadan alınmıştır (Paliwal, 2021). Veri seti, Metacritic ve Steam gibi oyun değerlendirme platformlarından toplanan oyun puanlarını, kullanıcı yorumlarını, oyun türlerini, oynanma sürelerini, fiyat bilgilerini ve platform bilgilerini içermektedir. Toplamda 47.774 satır ve 16 sütundan oluşan veri seti, hem nicel (fiyat, süre, puan) hem de nitel (yorum metinleri) değişkenler sunmaktadır. Hedef değişken, kullanıcı puanlarının sayısal değerleridir. Çalışma regresyon analizi odaklı olarak kurgulanmıştır. Kayıtlar, oyun adı, platform, yayın tarihi, kullanıcı yorumları ve değerlendirme skorları gibi değişkenleri barındırarak, oyun sektöründeki tüketici davranışları ve memnuniyeti üzerine çok boyutlu analizlerin

gerçekleştirilmesine olanak sağlamaktadır. Veri seti hem nicel (fiyat, oyun süresi, kullanıcı puanları gibi) hem de nitel (yorum metinleri, grafik, müzik ve hikâye gibi) veriler sunarak, çok boyutlu araştırma ve modelleme çalışmalarında kullanılabilir.

Tablo 1. Veri Setinin Bir Bölümü

Game Title	Age Group Targeted	Price	Platform	Requires Special Device	Developer	Publisher	Release Year
Grand Theft Auto V	All Ages	41,41	PC	No	Game Freak	Innersloth	2015
The Sims 4	Adults	57,56	PC	No	Nintendo	Electronic Arts	2015
Minecraft	Teens	44,93	PC	Yes	Bungie	Capcom	2012
Bioshock Infinite	All Ages	48,29	Mobile	Yes	Game Freak	Nintendo	2015
Half-Life: Alyx	Adults	55,49	PlayStation	Yes	Game Freak	Epic Games	2022
Genre	Multiplayer	Game Length (Hours)	Graphics Quality	Soundtrack Quality	Story Quality	User Review Text	User Rating
Adventure	No	55,3	Medium	Average	Poor	Solid game, but too many bugs.	36,4
Shooter	Yes	34,6	Low	Poor	Poor	Solid game, but too many bugs.	38,3
Adventure	Yes	13,9	Low	Good	Average	Great game, but the graphics could be better.	26,8
Sports	No	41,9	Medium	Good	Excellent	Solid game, but the graphics could be better.	38,4
RPG	Yes	13,2	High	Poor	Good	Great game, but too many bugs.	30,1

3.3. Veri Ön İşleme

Veri ön işleme, veri setinin analiz ve modelleme için hazır hale getirilmesi sürecidir. Bu aşamada gerçekleştirilen işlemler, modellerin doğruluğunu ve genel performansını önemli ölçüde etkilemektedir (Eşidir, 2025c). Modellerin eğitilmesi esnasında, gereksiz veriler içeren sütunlar veri setinden çıkarılmıştır. Veri ön işleme aşamasında, gereksiz ve eksik veriler içeren sütunlar veri setinden çıkarılmış; kategorik değişkenler sayısal değerlere dönüştürülmüştür. Ardından, veri seti eğitim (%80) ve test (%20) olarak ayrılmış ve sayısal veriler, Standart Ölçekleme kullanılarak ölçeklendirilmiştir. Yapılan bu işlem, özellikler arasındaki ölçek farklılıklarını gidererek modelin daha etkili öğrenmesine olanak sağlamaktadır. Hiperparametre optimizasyonu yapılmadan, scikit-learn kütüphanesinin varsayılan parametreleri uygulanmıştır. Bu yaklaşım, modellerin ham performanslarını değerlendirmek ve modeller arasında temel karşılaştırma yapmak amacıyla tercih edilmiştir (Probst, Wright ve Boulesteix, 2019). Modelleme sürecinde random_state=42 sabitlenerek yeniden üretilebilirlik sağlanmıştır. Beş katlı çapraz doğrulama ile desteklenen yapı, modellerin genellenebilirlik düzeylerini ölçmede değerlendirme imkânı sunmuştur.

3.4. Modellerin Performans Değerlendirme Kriterleri

Makine öğrenmesi modellerinin performanslarını nesnel biçimde değerlendirebilmek amacıyla çeşitli metrikler ve analiz yöntemleri kullanılmaktadır. MAE, tahmin değerleri ile gerçek değerler arasındaki farkların mutlak değerlerinin ortalamasını verirken, MSE hataların karelerinin ortalamasını hesaplamakta ve büyük sapmaları daha fazla cezalandırmaktadır. RMSE ise MSE'nin kareköküdür ve hatanın büyüklüğünü orijinal birimlerde sunarak yorumlanabilirliği artırmaktadır. R^2 , modelin

bağımlı değişkendeki varyansı ne oranda açıkladığını ölçen determinasyon katsayısıdır (Chai & Draxler, 2014).

4. MAKİNE ÖĞRENMESİ MODELLERİ

Makine öğrenmesi, bilgisayar sistemlerinin veri analizi yoluyla öğrenmesini ve bu süreçte elde edilen bilgiler doğrultusunda tahmin yapabilme yetisi kazanmasını amaçlayan bir bilim dalıdır. Çözüm süreçlerinde istatistiksel ve matematiksel modeller kullanılarak veri kümesinin yapısına ve problem türüne uygun algoritmalar seçilmektedir (Eşidir, 2025c).

4.1 Doğrusal Regresyon

Doğrusal Regresyon, bağımlı değişkenin bir ya da daha fazla bağımsız değişken kullanılarak doğrusal bir fonksiyon ile tahmin edilmesini amaçlayan bir regresyon analizidir. Bu yöntem, verileri en iyi temsil eden bir doğrusal denklem oluşturmayı hedefler.

Bağımlı değişken ile bağımsız değişken(ler) arasında doğrusal bir ilişki varsayımı altında yapılan regresyon analizine "doğrusal regresyon (linear regression)" adı verilmektedir. Bu yöntem, bağımlı değişkenin bir ya da daha fazla bağımsız değişken aracılığıyla doğrusal bir fonksiyonla tahmin edilmesini amaçlamaktadır. Eğer yalnızca bir bağımsız değişken söz konusu ise, bu analiz "basit doğrusal regresyon (simple linear regression)" olarak adlandırılırken; birden fazla bağımsız değişkenin modele dahil edilmesi durumunda ise "çoklu doğrusal regresyon (multiple linear regression)" terimi kullanılmaktadır. Doğrusal regresyon, lojistik olmayan regresyon problemlerinin çözümünde en yaygın başvuru istatistiksel yöntemlerden bir tanesidir. Yöntemin temel amacı, veri noktalarını en iyi şekilde temsil eden bir doğrusal denklem (doğru denklemi) elde etmektir. Böyle bir modelin oluşturulabilmesi durumunda, söz konusu denklem aracılığıyla kestirimde bulunmak mümkün hale gelmektedir. Kestirim süreci, ilgilenilen bağımsız değişken değerlerinin doğrusal modelde yerine konularak bağımlı değişkenin tahmin edilmesi sürecini içermektedir (Aslan, 2025).

4.2 Rastgele Orman

Random Forest, farklı alt veri setlerinde oluşturulan birden fazla karar ağacının birleşiminden meydana gelen, topluluk öğrenme (ensemble learning) metoduna dayalı güçlü bir makine öğrenmesi modelidir. Sınıflandırma ve regresyon problemlerinde yaygın olarak kullanılmakta olup, birden fazla karar ağacını bir araya getirerek her bir ağacın bireysel zayıflıklarını dengelemekte ve daha kararlı, güçlü ve genelleştirilebilir bir model oluşturmaktadır. Bu model, bootstrap örnekleme yöntemiyle eğitim veri setinden birçok alt örneklem (subsample) oluşturur. Her bir bootstrap örneği, orijinal veri kümesinden rastgele seçilen örnekleri içermekte olup, bu örnekler kullanılarak bağımsız karar ağaçları inşa edilir. Bu süreçte, her karar ağacı veri setinin farklı bir alt kümesi üzerinde eğitilir. Ayrıca, her düğümdeki bölünme işlemi tüm özellikler yerine rastgele seçilen bir alt küme üzerinden gerçekleştirilir. Bu yöntem, modelin varyansını düşürerek aşırı öğrenmeyi (overfitting) önlemekte ve farklı karar ağaçları arasındaki korelasyonu azaltarak modelin genelleme performansını artırmaktadır. Sınıflandırma problemlerinde, nihai tahmin genellikle oylama (majority voting) yöntemiyle belirlenmektedir (Oguine ve Oguine, 2021).

4.3 Gradyan Artırma

Gradyan Artırma (Gradient Boosting), topluluk (ensemble) öğrenme yaklaşımlarından biri olup, temel olarak ardışık biçimde oluşturulan zayıf öğrenicilerin (çoğunlukla karar ağaçları) bir araya getirilmesi prensibine dayanır. Bu yöntemde her yeni model, önceki modelin yaptığı hataları düzeltmeye odaklanarak eğitilir. Her iterasyonda, modelin oluşturulma süreci gradyan inişi (gradient descent) algoritması ile optimize edilir; böylece kayıp fonksiyonunun minimize edilmesine yönelik olarak parametreler sistematik biçimde güncellenir. Gradyan artırma yaklaşımı, her bir yeni öğrenicinin, kendisinden önceki modellerin hatalarını öğrenmesi yoluyla genel model performansını artırmayı hedefler. Bu yapı sayesinde yüksek doğruluk seviyelerine ulaşmak mümkün hale gelirken, aynı zamanda aşırı öğrenme (overfitting) riski de ortaya çıkabilmektedir. Bu nedenle, hiperparametrelerin doğru biçimde ayarlanması, modelin genel geçer başarısını artırmak açısından

kritik öneme sahiptir. Buna benzer uygulamalarda, modelin doğruluk performansı artırılmakta; özellikle karmaşık veri yapılarının öğrenilmesinde ve doğru sınıflandırmalar yapılmasında önemli avantajlar sağlanmaktadır. Yöntemde problem ilk kez çözüldükten sonra tahmin edilen değerlerle gerçek değerler arasındaki farklar (residuals) hesaplanmakta ve bu farkları minimize edecek şekilde yeni modeller inşa edilmektedir. Her bir yinelemede, sanki bir önceki tahminler gerçek değerler gibi değerlendirilerek yeni tahminleyiciler üretilmektedir. En sonunda yine oylama yapılarak kestirim gerçekleştirilmektedir (Aslan, 2025).

4.4 Ekstra Ağaçlar

Ekstra Ağaçlar (Extremely Randomized Trees), topluluk öğrenmesi yaklaşımına dayanan ve karar ağaçlarının rastgeleleştirilmiş varyasyonlarından oluşan güçlü bir makine öğrenmesi modelidir. Bu model, klasik rastgele orman algoritmasına benzer biçimde birçok karar ağacından oluşmaktadır. Fakat ağaçların oluşturulmasında kullanılan rastgelelik seviyesi daha yüksektir. Her bir özelliğin optimal ayırım noktasını bulmak yerine rastgele bir ayırım noktası seçilir. Bu rastgelelik, ağaçlar arasındaki çeşitliliği artırarak aşırı uyumlamayı azaltmaya yardımcı olur. Gürültülü ve yüksek boyutlu verilerle başa çıkmada etkili bir modeldir (Özdemir, 2023). Ekstra Ağaçlar modelinde, hem eğitim verilerinin alt örneklemeleri üzerinde işlem yapılmaz hem de her düğümde en iyi bölme noktası yerine, rastgele seçilen bir eşik değeri kullanılarak karar verme işlemi gerçekleştirilir. Bu durum, modelin hesaplama süresini düşürmekte ve genellikle varyansın azaltılmasına katkı sağlamaktadır. (Bonsignori ve ark., 2021).

4.5 Makine Öğrenmesi Modellerinin Performansları

Bu bölümde, uygulanan dört makine öğrenmesi modelinin performansları karşılaştırmalı olarak analiz edilmiştir. Modeller, dört temel regresyon metriği (MAE, MSE, RMSE, R^2) kullanılarak değerlendirilmiştir. Performans ölçütleri, modelin tahmin doğruluğunu ve genellenebilirliğini objektif biçimde kıyaslamaya olanak tanımaktadır. Modellerin performans metrikleri Tablo 2’de görülmektedir.

Tablo 2. Modellerin Performans Metrikleri

Modeller	MSE	MAE	RMSE	R2
Doğrusal Regresyon	1,340	1,001	1,157	0,977
Random Forest	1,394	1,014	1,181	0,976
Gradient Boosting	1,376	1,010	1,173	0,976
Ekstra Ağaçlar	1,359	1,005	1,166	0,977

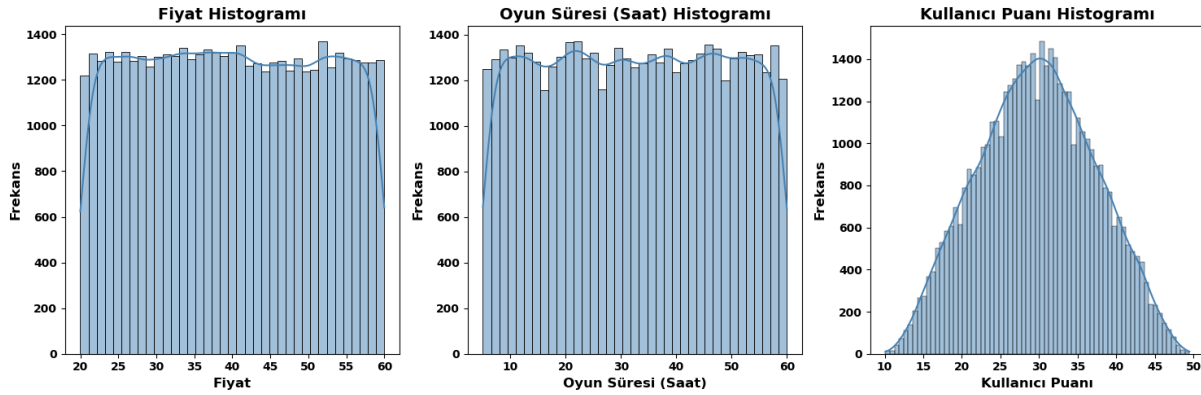
Tablo 1’de sunulan performans metriklerine göre, dört makine öğrenmesi modeli de yüksek doğruluk düzeyine ulaşmıştır. Performans metriklerine göre en başarılı modelin Doğrusal Regresyondur. Bu model, tüm hata metriklerinde (MSE = 1,340; MAE = 1,001; RMSE = 1,157) en düşük değerlere ulaşarak tahmin doğruluğu açısından üstünlük sağlamış ve R^2 değeri bakımından (0,977) diğer modellerle birlikte en yüksek açıklayıcılığa erişmiştir. Random Forest, Gradient Boosting ve Ekstra Ağaçlar gibi daha karmaşık modellerin benzer seviyede yüksek performans sergilemesine rağmen, hata düzeylerinde tutarlı bir şekilde en iyi sonuçları veren Doğrusal Regresyon modeli öne çıkmaktadır. Bu durum, ilgili veri setinde doğrusal ilişkilerin baskın olduğunu ve daha kompleks modellerin sağladığı esnekliğin anlamlı bir avantaj oluşturmadığını göstermektedir.

5. GÖRSEL İSTATİSTİKSEL ANALİZLER VE BULGULAR

Bu bölümde, video oyunlarına ilişkin kullanıcı yorumlarının görsel betimleyici analizleri sunularak, temel istatistiksel eğilimlerin anlaşılmasına katkı sağlanmıştır. Histogram, dağılım grafikleri, kelime bulutları ve duygu analizleri aracılığıyla fiyat, oyun süresi, kullanıcı puanı gibi nicel değişkenlerin yanı sıra yorum içerikleri gibi nitel veriler de değerlendirilmiştir. Amaç, veri setinde yer alan özniteliklerin genel dağılımını ve birbirleriyle olan ilişkilerini görselleştirerek, makine öğrenmesi modelleri öncesi veri yapısına dair sezgisel çıkarımlarda bulunmaktır. Söz konusu görselleştirme

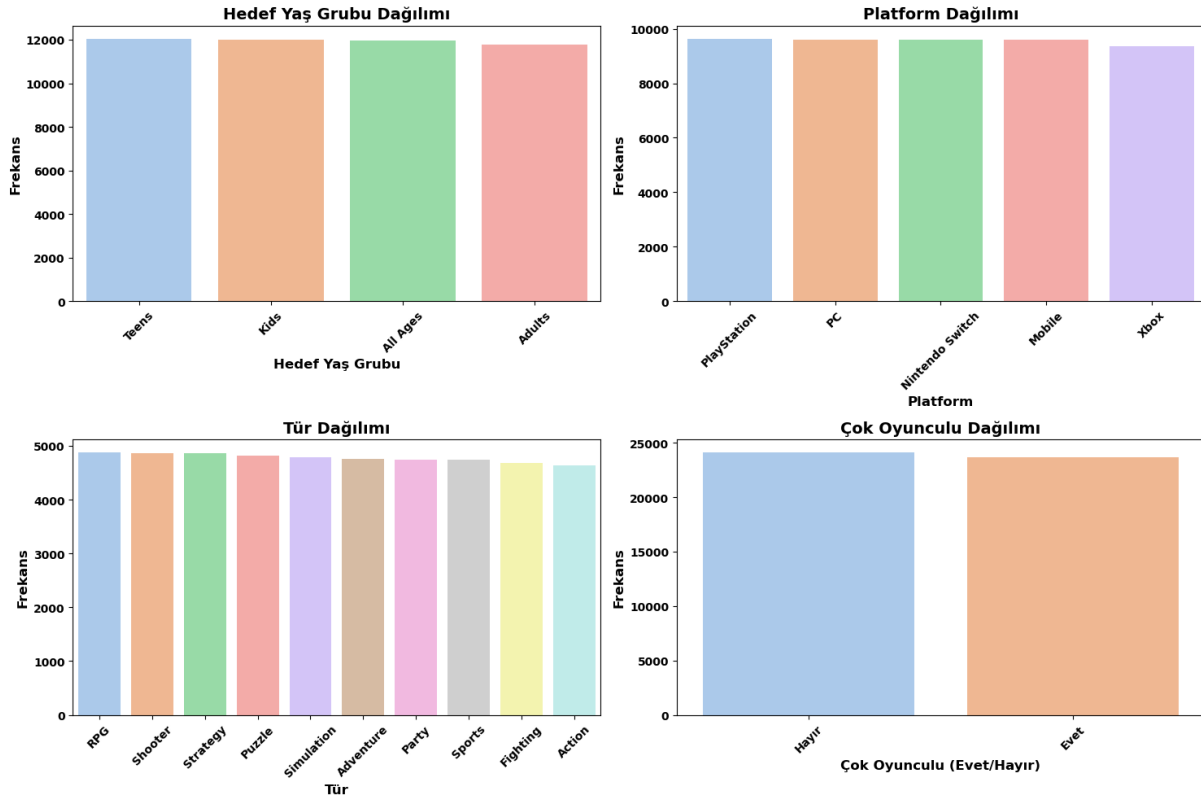
sonuçları, kullanıcı davranışlarının yorumlama biçimini, değerlendirme eğilimlerini ve oyunlara yönelik genel memnuniyet düzeyini daha iyi anlamaya imkân tanımakta; aynı zamanda sonraki modelleme aşamaları için açıklayıcı bir temel oluşturmaktadır. Şekil 1’de fiyat dağılımı, oyun süresi, kullanıcı puanı dağılımlarına ilişkin histogram grafikleri gösterilmiştir.

Şekil 1. Oyun Fiyatları, Oyun Süresi ve Kullanıcı Puanı Histogramları



Fiyat histogramı incelendiğinde, fiyatların 20 ile 60 dolar arasında düzgün (uniform) bir dağılım gösterdiği görülmektedir. Fiyat histogramında belli bir tepe noktası bulunmamaktadır, veri集中的i oyunlar fiyatlandırma açısından çeşitlilik göstermektedir. Oyun süresi (saat) histogramı ise benzer biçimde geniş bir aralıkta dağılmış olup herhangi bir yoğunlaşma noktası barındırmamaktadır. Kısa ve uzun süreli oyunlar veri setinde dengeli bir şekilde temsil edilmiştir. Kullanıcı puanı histogramı ise, çan eğrisi (normal dağılım) formu sergilemiştir. Ortalama (30 puan) çevresinde yüksek bir yoğunlaşma ve uç değerlerin az olduğu gözlemlenmektedir.

Şekil 2. Veri seti ile ilgili çeşitli dağılımlar



Şekil 2’de veri seti ile ilgili çeşitli dağılımlar görülmektedir. Hedef yaş grubu dağılımı verileri, “Çocuk”, “Genç”, “Yetişkin” ve “Aile” kategorileri ufak tefek farklar olsa da, benzer frekanslarda

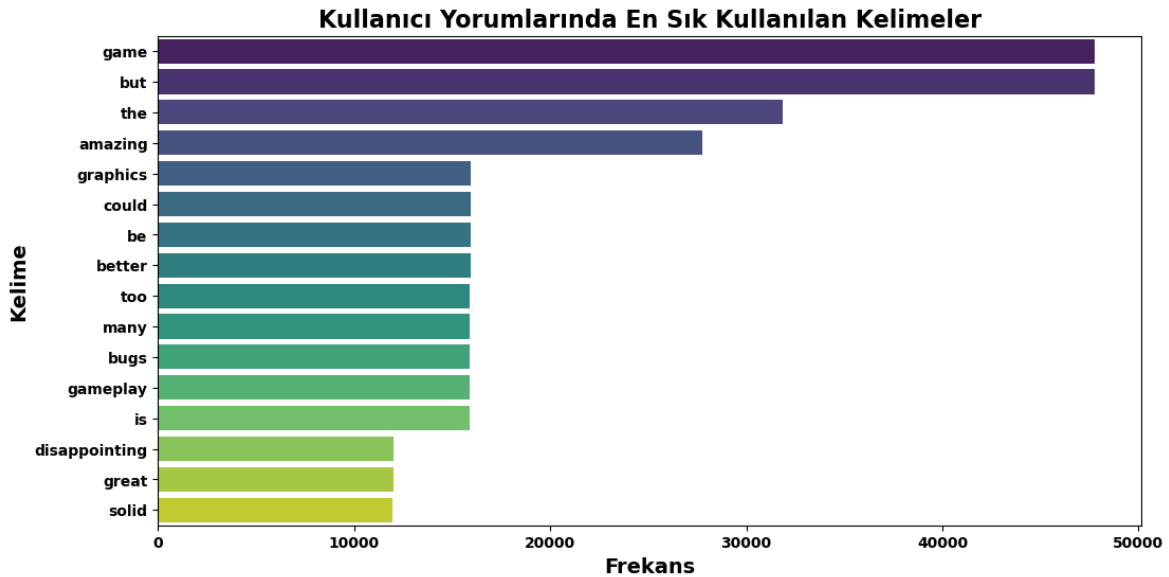
temsil edilmektedir. Platform dağılımı da (PlayStation, PC, Akıllı Telefon/Tablet, Nintendo, Xbox) yine birbirine yakın frekanslardadır. Tür Dağılımı (FPS, Strategy, RPG, Platform, Simülasyon, Yarış vb.) belirgin bir baskın tür yerine çeşitli oyun türlerinin benzer düzeyde tercih edildiğini göstermektedir. Çok Oyunculu (Evet/Hayır) dağılımı incelendiğinde de diğer dağılımlara benzer bir yapı söz konusudur.

Şekil 3. Kullanıcı Yorumları Kelime Bulutu



Şekil 3’te kullanıcı yorumları kelime bulutu görülmektedir. Kullanıcı yorumlarından oluşturmuş kelime bulutunda, teknik (“bugs”) ve görsel (“graphics”) unsurlar ile oynanış (“gameplay”) özelliklerinin en sık tartışılan başlıklar olduğu görülmektedir. “Many bugs” sözcüğü, oyunlarda meydana gelen hata ve performans ile ilgili sorunları ve hataları vurgularken, “graphics” ve “gameplay” ifadeleri de oyun ile ilgili kalite ölçütleri olarak ön plana çıkmaktadır. Kelime bulutunda “disappointing” gibi olumsuz ve “amazing”, “great” gibi olumlu sıfatlar, memnuniyet düzeylerinde çeşitli farklılıkların olduğunu belirtmektedir.

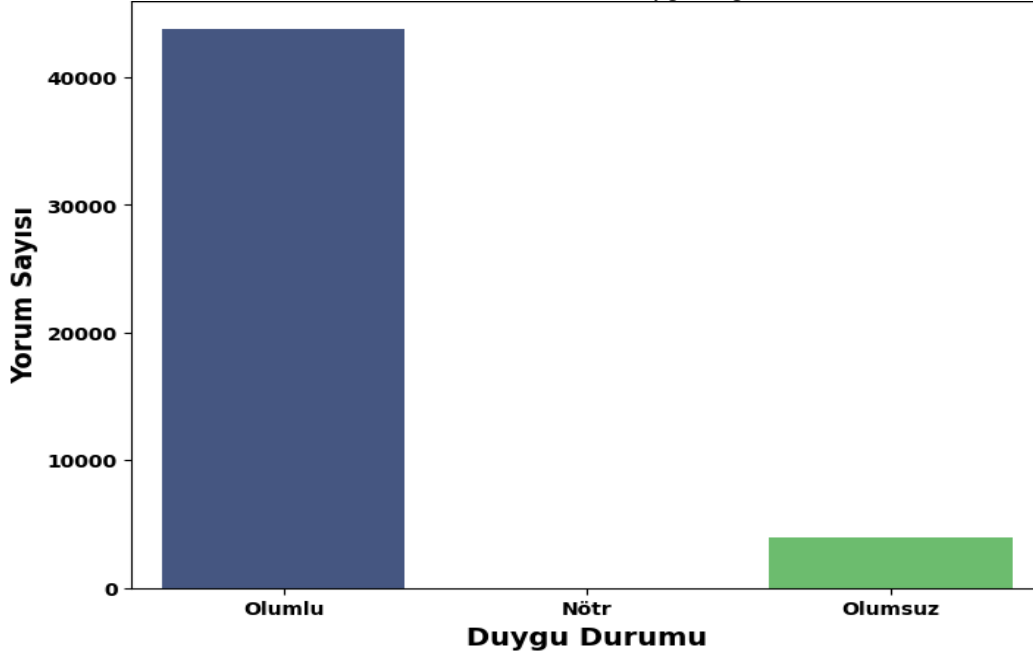
Şekil 4. Kullanıcı Yorumlarında En Sık Kullanılan Kelimeler



Şekil 4’te kullanıcı yorumlarında en fazla kullanılan kelimeler gösterilmiştir. Yorumlara ilişkin kelime frekansı verileri incelendiğinde, en sık geçen sözcüklerin oyunun temel öğelerini (“game”, “graphics”, “bugs”) ve kullanıcıların duygusal yorumlamalarını (“amazing”, “disappointing”) yansıttığı görülmektedir. Özellikle “but” gibi bağlaçların yoğun kullanımı, olumlu ve olumsuz

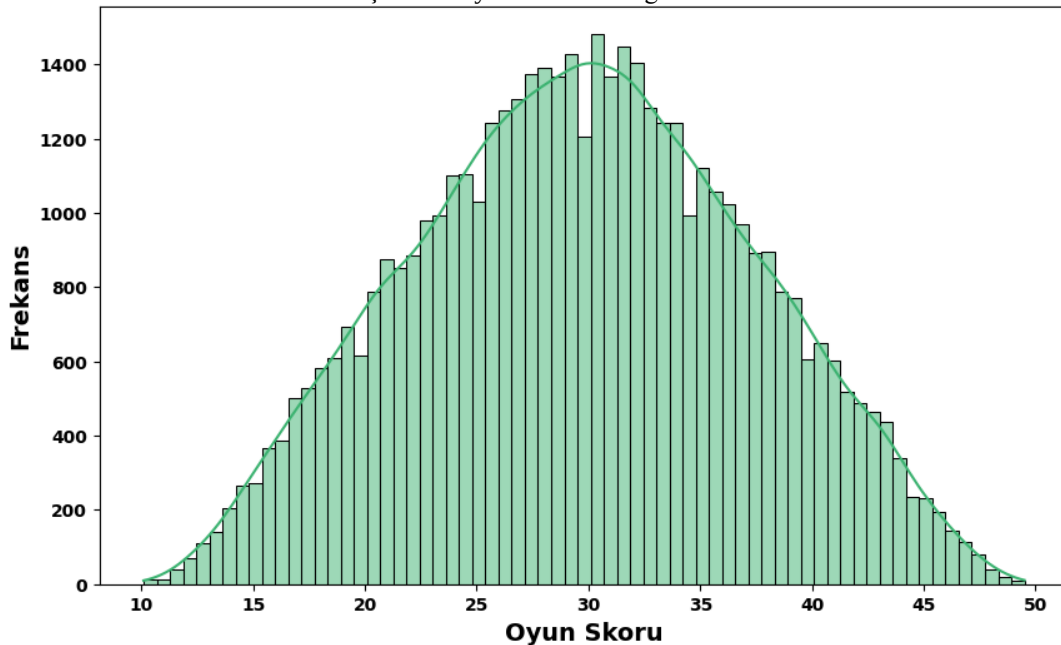
ifadelerin aynı yorum içinde sıkça bir arada yer aldığını göstermekte, bu da oyunun hem beğenilen hem de eleştirilen yönlerinin belirgin şekilde vurgulandığına işaret etmektedir. Ayrıca “could”, “be” ve “better” gibi sözcüklerin varlığı, kullanıcıların iyileştirme odaklı önerilerini dile getirdiklerini ve oyunun potansiyeline dair beklentilerinin yüksek olduğunu göstermektedir. En fazla kullanılan kelimeler, 47.774 adet ile 'game' ve 'but' olmuştur. 'the' kelimesi 31.864 adet, 'amazing', 27.733 adet, 'graphics', 'could', 'be' ve 'better' kelimeleri 15.971 adet kullanılmıştır.

Şekil 5. Kullanıcı Yorumlarındaki Duygu Dağılımı

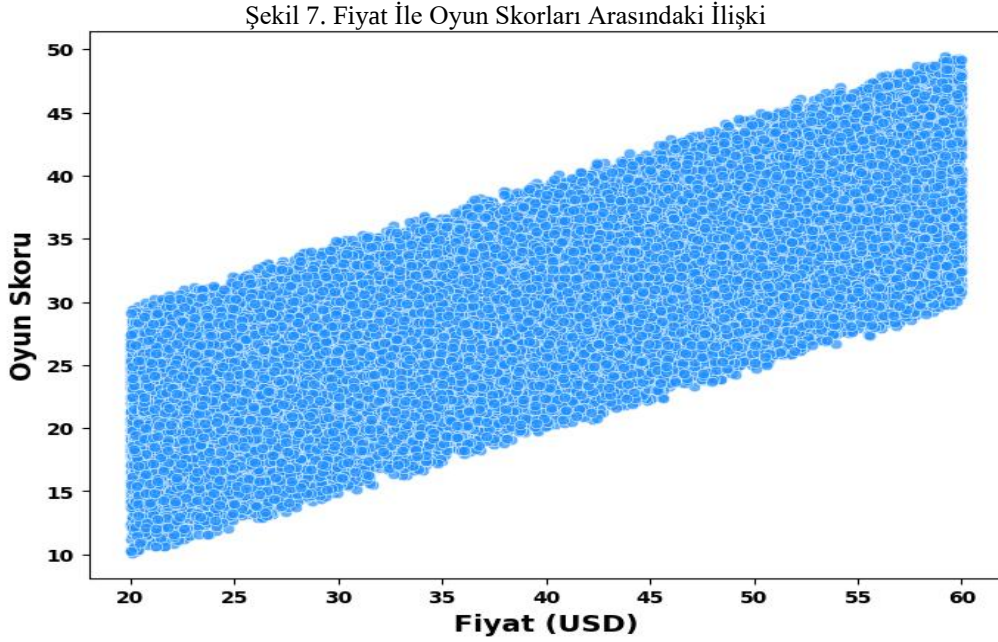


Şekil 5’te kullanıcı yorumlarındaki duygu dağılımı görülmektedir. Duygu durumunda “Olumlu” kelime sayısı 43.811 adet ve “Olumsuz” kelime sayısı ise 3.963 adettir. Grafikte de görüldüğü üzere, olumlu yorumların sayısı olumsuz yorumlara kıyasla açık ara yüksektir. Kullanıcıların genel manada video oyunlarından memnun olduğu anlaşılmaktadır. Nötr yorumların bulunmaması, kullanıcıların büyük çoğunluğunun deneyimlerini belirgin bir şekilde ifade ettiğini belirtmektedir. Olumsuz yorumlar ise ürünler de iyileştirilmesi gereken yönleri yansıtacak değerli geri bildirimler içerebilmektedir.

Şekil 6. Oyun Skorları Dağılımı



Şekil 6’da 10 ile 50 puan arasındaki oyun skorlarının dağılımı görülmekte ve dağılım 30 puan etrafında tepe yaparak simetrik bir görüntü sunmaktadır. Kullanıcı skorları orta düzeyde yoğunlaşmaktadır. Yüksek ve düşük skorların daha seyrek olduğu grafikten okunmaktadır. Histogram şekil olarak, normal dağılım göstermektedir.



Şekil 7’de oyun fiyatı ile oyun skorları arasındaki ilişkiyi göstermektedir. Grafikten, fiyat arttıkça oyun skorunun da yükseldiği gözlemlenmektedir. 20-60 USD aralığındaki oyun fiyatları ile 10-50 puan arasındaki oyun skorlarının dağılımı görülmektedir. Oyun fiyatı arttıkça oyun skorunun da yükseldiği grafikten okunmaktadır. Yüksek fiyatlı oyunlar daha fazla kaynak ayrılarak geliştirilmekte dolayısıyla bu durum oyun kalitesine de olumlu manada yansımaktadır.



Şekil 8’de oyun süresi ile oyun skoru arasındaki ilişki gösterilmiştir. Oyun süresi ile oyun skorları arasında genelde pozitif bir ilişki gözlemlenmektedir. Daha uzun süre oynanabilen oyunların içerik,

derinlik ve kalite açısından daha fazla kabul gördüğü ve oyun kullanıcılarınca daha yüksek puanlandığı anlaşılmaktadır.

SONUÇ

Yapılan çalışmada, video oyunu sektöründe kullanıcı yorumları veri analizi ve makine öğrenmesi modelleri ile incelenmiştir. Kullanıcı puanlarının tahminine yönelik etkin modeller geliştirilmiştir. Kullanıcı yorumları, fiyat, oyun süresi ve oyun türü gibi çeşitli değişkenlerden oluşan kapsamlı bir veri seti kullanılmıştır. Doğrusal Regresyon, Rastgele Orman, Gradyan Artırma ve Ekstra Ağaçlar modelleri ile tahminler yapılmıştır. Analiz sonuçları, tüm makine öğrenmesi modellerinin yüksek doğruluk oranları sergilediğini ortaya koymuştur. Kullanıcı yorumları ve oyun süresi, puan tahmininde en belirleyici rol oynamıştır. Elde edilen bulgular, tüm modellerin yüksek düzeyde performans gösterdiğini ortaya koymakla birlikte, Doğrusal Regresyon modelinin en düşük hata oranlarına ve en yüksek R^2 değerine ulaşarak en başarılı model olduğunu göstermiştir. Bu sonuç, veri setindeki ilişkilerin büyük ölçüde doğrusal olduğunu ve daha karmaşık modellere kıyasla basit doğrusal yapıların daha etkin tahmin gücü sağlayabildiğini ortaya koymaktadır. Ayrıca, bu durum model seçiminde sadece karmaşıklık düzeyine değil, veri yapısının niteliklerine de dikkat edilmesi gerektiğini göstermektedir.

Makine öğrenmesi modelleri, video oyunu kullanıcı puanlarını tahmin etmede oldukça başarılı sonuçlar vermiştir. Elde edilen tahmin sonuçları mevcut literatür ile de uyumludur. Örneğin; Olmedilla vd. (2022) tarafından yapılan çalışmada, Steam platformundan elde edilen kullanıcı verileri kullanılarak XGBoost ve Extreme Learning Machine (ELM) gibi modeller ile oyun puanı tahmin edilmiştir. Oldukça geniş bir veri tabanında yapılan analizlerde, modelin R^2 skoru 0.61 olarak elde edilmiştir. Bu bağlamda, benzer bu çalışmanın elde ettiği R^2 değerleri 0.97 seviyelerine ulaşmıştır.

Veri ön işleme kalitesi, model yapısı ve değişken seçiminin başarıda belirleyici olduğunu göstermektedir. Literatürdeki bu karşılaştırmalar, mevcut çalışmanın hem metodolojik sağlamlığını hem de katkı değerini güçlendirmektedir. Ayrıca, gelecekte farklı metin madenciliği yaklaşımlarının (örneğin BERT tabanlı duygu analizleri) regresyon modellerine girdi olarak entegre edilmesi, model başarısını artırmaya yönelik önemli bir gelişim alanı olarak önerilmektedir.

Çalışma, kullanıcı yorumlarının analiziyle oyun sektöründe hedef kitleye yönelik pazarlama stratejilerinin oluşturulması ve ürün geliştirme süreçlerinin iyileştirilmesi açısından pratik katkılar sağlamaktadır. Özellikle metin madenciliği teknikleriyle kullanıcı yorumlarındaki duygu ve eğilimlerin detaylı olarak incelenmesi, tüketici beklentilerini daha iyi anlamak adına önemli avantajlar sunmaktadır. Çalışmada kullanılan dört makine öğrenmesi modeli arasındaki karşılaştırma, verinin doğasına ve özelliklerine uygun model seçiminin önemini ortaya koymaktadır. Pratik uygulamalarda algoritma karmaşıklığından ziyade veri yapısı ile uyumlu modellerin tercih edilmesi gerektiği anlaşılmıştır.

Akademik açıdan bakıldığında, metin madenciliği ile makine öğrenmesinin bütünleşik biçimde kullanılması, hem yöntemsel çeşitlilik hem de disiplinlerarası etki bakımından özgün katkılar sunmaktadır. Çalışmanın ortaya koyduğu sonuçlar, oyun sektöründe tüketici içgörülerinin daha etkili biçimde analiz edilmesine olanak sağlayarak; veri bilimi temelli karar destek sistemlerinin geliştirilmesine yönelik katkılar sunmaktadır. Gelecek çalışmalarda, metin verilerinin işlenmesinde BERT benzeri modern derin öğrenme yaklaşımları incelenebilir.

KAYNAKÇA

- Altuntaş, B., & Karaarslan, M. H. (2017), “Kullanıcıların mobil oyun tercihinde etkili olan faktör düzeylerinin öneminin belirlenmesi”, *Uluslararası İktisadi ve İdari İncelemeler Dergisi*, (19), 277-298. <https://doi.org/10.18092/ulikidince.321570>
- Aslan, K. (2025). Yapay Zekâ, Makine Öğrenmesi ve Veri Bilimi Kursu, Sınıfta Yapılan Örnekler ve Özet Notlar, C ve Sistem Programcıları Derneği, İstanbul.

- Aydın, O. & Sarıca, Y., (2024), “Dijital Oyun Geri Bildirimlerinin Duygu Analizi ile İncelenmesi: e-WoM’den r-WoM’a Geçiş Mümkün Mü?”, *Üçüncü Sektör Sosyal Ekonomi Dergisi*, 59(4), 2667-2683.
- Bonsignori, V., Guidotti, R., Monreale, A. (2021), “Deriving a Single Interpretable Model by Merging TreeBased Classifiers”, *Discovery Science: 24th International Conference*, 11-13 Ekim, Halifax, Kanada, 347-357.
- Chai, T., & Draxler, R. R. (2014), “Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature”, *Geoscientific Model Development*, 7(3), 1247–1250. <https://doi.org/10.5194/gmd-7-1247-2014>
- Çabuk, M., Yücalar, F., & Toçoğlu, M. A. (2023), “Makine Öğrenmesi ile E-Ticaret Ürün Yorumlarının Otomatik Analizi”, *Avrupa Bilim ve Teknoloji Dergisi*, (52), 110-121.
- Eşidir, K. A. (2025a), “Makine Öğrenimi Modelleri ile Yetişkin Eğitimi Analizi: Modellerin Karşılaştırmalı Performansı”, *Elektronik Sosyal Bilimler Dergisi*, 24(2), 946-964. <https://doi.org/10.17755/esosder.1589887>
- Eşidir, K. A. (2025b), “Kanatlı Sektöründe Yapay Zekâ Uygulamaları: Etlik Piliç Yemi Fiyat Tahmini”, *Journal of History School*, 74, 269-290. <http://dx.doi.org/10.29228/joh.78735>
- Eşidir, K. A. (2025c), “TÜİK Mikro Verileri ile Çocuk İşgücü Tahmini: Makine Öğrenimi Modellerinin Performans Analizleri”, *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, 37(1), 453-466. <https://doi.org/10.35234/fumbd.1607609>
- Gür, Y. E. (2024), “Stock price forecasting using machine learning and deep learning algorithms: A case study for the aviation industry”, *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, 36(1), 25-34. <https://doi.org/10.35234/fumbd.1357613>
- Ji H. (2023), Robustness analysis on stock market prediction method, Highlights in Business”, *Economics and Management*, (2957-952X), 21:791-801.
- Mustafaoğlu, R. (2018), “E-Spor, Spor ve Fiziksel Aktivite”, *Ulusal Spor Bilimleri Dergisi*, 2(2), 84-96. <https://doi.org/10.30769/usbd.457545>
- Oguine O. C ve Oguine MB. (2021), “Comparative analysis and forecasting on the death rate of COVID-19 patients in Nigeria using random forest and multinomial Bayesian epidemiological models”, *Journal of Clinical Case Studies Reviews & Reports*, 1-7.
- Olmedilla, M., Espinosa-Leal, L., Romero-Moreno, J.C., & Li, Z. (2024), “Predicting Review Helpfulness in Video Games: A Comparative Analysis of Machine Learning Models and NLP Integration”, *IADIS International Journal on WWW/Internet*, ISSN: 1645-7641, 22(2), 1–15.
- Özdemir, A. (2023), “Makine Öğrenmesi Algoritmaları ile Aritmilerin Sınıflandırılması”, *Erciyes Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 39(3), 394-402.
- Paliwal, J. (2021). Video Game Reviews and Ratings [Data set]. Kaggle. <https://www.kaggle.com/datasets/jahnavipaliwal/video-game-reviews-and-ratings>
- Probst, P., Wright, M. N., & Boulesteix, A. L. (2019), “Hyperparameters and tuning strategies for random forest”, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(3), e1301. <https://doi.org/10.1002/widm.1301>
- Speer AB. (2021), “Empirical attrition modelling and discrimination: Balancing validity and group differences”, *Human Resource Management Journal*, 34 (1), 1-19.
- Turp Gölbaşı, B., & Ayhan Gökcek, H. (2022), “Dijital pazarlamaya yönelik olarak oyunlaştırma içeren e-ticaret siteleri üzerine bir meta analiz”, *Elektronik Sosyal Bilimler Dergisi*, 21(84), 1909-1929. <https://doi.org/10.17755/esosder.1064214>
- Tuzcu S. (2020), “Çevrimiçi Kullanıcı Yorumlarının Duygu Analizi ile Sınıflandırılması”, *ESTUDAM Bilişim*, 1(2), 1–5.
- Wu Y. (2023), “Job embeddedness review: Presentation, measurement and development”, *Advances in Economics, Management and Political Sciences*, 47 (1), 169- 174.
- Zhu X, Sawhney R, Upreti G. (2016), “Determinates of employee voluntary turnover and forecasting in departments: A case study”, *Studies in Engineering and Technology*, 3(1):64-73.