



Kelime Uzayları ve Derin Öğrenme Algoritmaları ile Oyun Sektöründe Geri Bildirim Etkinliğinin Değerlendirilmesi

Evaluation of Feedback Effectiveness in the Gaming Sector Using Word Spaces and Deep Learning Algorithms

Mert Yavaş ^{1*}, Mansur Alp Toçoğlu ², Emre Şatır ³

¹ İzmir Katip Çelebi Üniversitesi, Fen Bilimleri Enstitüsü, Yazılım Mühendisliği Anabilim Dalı, Tezsiz Yüksek Lisans Programı, İzmir, TÜRKİYE

² İzmir Katip Çelebi Üniversitesi Mühendislik ve Mimarlık Fakültesi Bilgisayar Mühendisliği Bölümü, İzmir, TÜRKİYE

³ İzmir Katip Çelebi Üniversitesi Mühendislik ve Mimarlık Fakültesi Bilgisayar Mühendisliği Bölümü, İzmir, TÜRKİYE

Sorumlu Yazar / Corresponding Author*: mertcyavas@gmail.com

Öz

Bu çalışmanın amacı, oyun sektörü gibi spesifik bir alandan elde edilen yorum, değerlendirme ve tavsiye gibi geri dönüşleri, daha geniş konu seçeneklerine sahip gönderilerle karşılaştırarak, bu geri dönüşlerin etkinliğini incelemektir. Açık kaynaklı platformlardan toplanan veriler, Word2Vec'in CBOW ve Skip-Gram yöntemleri ile 50, 100, 150, 200, 250, 300 vektör boyutlarında 24 farklı kelime uzayı oluşturmak için kullanılmıştır. Steam platformundan alınan ve "Tavsiye Edilen" veya "Tavsiye Edilmeyen" olarak etiketlenen 14.769 adet oyun yorumu, derin öğrenme algoritmaları olan LSTM, RNN ve GRU ile analiz edilmiştir. Çalışmada, veri kazıma yöntemiyle elde edilen etiketli veriler üzerinde çeşitli derin öğrenme algoritmaları uygulanarak tavsiye analizi gerçekleştirilmiş ve farklı kelime uzayları ile derin öğrenme yaklaşımlarının başarı skorları karşılaştırılmıştır. Deneylerden alınan skorlar, oyun sektörü gibi belirli bir alandan elde edilen yorumların genel içeriklere kıyasla daha anlamlı sonuçlar verdiğini ortaya koymaktadır. Özellikle, 250 vektör boyutlu Skip-Gram kelime uzayında GRU algoritmasının; %80,16 doğruluk, %80,21 kesinlik, %80,00 duyarlılık ve %80,02 F1 skoru ile en yüksek başarıyı gösterdiği gözlemlenmiştir. Bu bulguların, derin öğrenme algoritmalarının farklı kelime uzayları üzerindeki performansını değerlendirerek oyun sektörü geri bildirimlerinin analizine katkı sağladığı düşünülmektedir. Çalışma, oyun sektörü gibi spesifik alanlara yönelik bağlamsal analizlerin ve derin öğrenme tekniklerinin uygulamalı değerini vurgulamaktadır.

Anahtar Kelimeler: Metin madenciliği, Kelime uzayı, Derin öğrenme, Geri bildirim analizi, Veri kazıma, Doğal dil işleme

Abstract

The aim of this study is to examine the effectiveness of feedback such as comments, reviews and recommendations from a specific domain such as the gaming industry by comparing them to posts with a wider range of topics. The data collected from open-source platforms are used to create 24 different word spaces with vector sizes of 50, 100, 150, 200, 250, and 300 using Word2Vec's CBOW and Skip-Gram methods. 14,769 game reviews from the Steam platform, labeled as "Recommended" or "Not Recommended", are analyzed with deep learning algorithms LSTM, RNN and GRU. In the study, recommendation analysis is performed by applying various deep learning algorithms on the labeled data obtained by data scraping method and the success scores of different word spaces and deep learning approaches are compared. The scores obtained from the experiments reveal that reviews from a specific domain such as the gaming industry yield more meaningful results than general content. In particular, it is observed that the GRU algorithm shows the highest success with 80.16% accuracy, 80.21% precision, 80.00% recall and 80.02% F1 score in the Skip-Gram word space with 250 vector size. It is thought that these findings contribute to the analysis of gaming industry feedback by evaluating the performance of deep learning algorithms on different word spaces. The study emphasizes the practical value of contextual analysis and deep learning techniques for specific domains such as the gaming industry.

Keywords: Text mining, Word space, Deep learning, Feedback analysis, Data scraping, Natural language processing

EXTENDED ABSTRACT

Introduction

After a few decades of substantial evolution, the gaming sector is now a key component of the digital economy. The current gaming era was founded by early video games such as Tennis for Two in 1958 and the commercial debut of the Magnavox Odyssey in the 1970s [1]. Digital games are currently among the most popular digital goods consumed globally due to the widespread use of personal computers, gaming consoles, and mobile devices. User-generated reviews and ratings on platforms like Steam [2] and

Metacritic [4] are crucial in influencing judgments about what games to buy. The relevance of game reviews has grown since Steam unveiled a new feature that allows users to give direct input on titles.

This study compares comments from general domains to the gaming industry in order to evaluate the efficacy of that input. The study also investigates how different deep learning algorithms-like RNN, LSTM, and GRU-perform on various word space models made with Word2Vec's [19] CBOW and Skip-Gram

techniques. The literature review reveals a lack of specific studies focusing on sentiment analysis and recommendation analysis in game reviews. Instead, most of the research in this area seems to focus on areas such as e-commerce, social media and stock market rather than game reviews. Despite some studies on game reviews, there are noticeable gaps in both languages, highlighting the need for further research in this area.

Materials and Methods

Steam is the source of the study's data, which included 14,769 reviews with the labels "Recommended" or "Not Recommended." For comparison, posts from a wider variety of topics are included in a second dataset. Using both CBOW and Skip-Gram models, Word2Vec approaches are used to produce 24 different word spaces with vector sizes of 50, 100, 150, 200, 250, and 300. The recommendation data is analyzed using RNN, LSTM, and GRU, three deep learning models. Data cleaning, tokenization, and lemmatization using Python's NLTK module are among the preprocessing procedures. Metrics like accuracy, precision, recall, and F1-score are used to assess performance [26]. The flowchart of the method used can be seen in Figure 3.

Results and Discussion

The findings show that, in contrast to feedback from more general issues, feedback from the gaming domain yielded more consistent and relevant results. Using a 250-dimensional Skip-Gram word space and the Steam dataset, the GRU method performs best (Tables 4, 5, 6, and 7). Conversely, the larger

dataset performs worse on all criteria, suggesting that feedback relevant to a given domain—particularly in the gaming industry—usually produces more accurate insights. These results are consistent with earlier studies that showed how well GRU and LSTM models handle sequential data.

Furthermore, GRU and LSTM consistently beat RNN when evaluating the performance of deep learning models across various word spaces, especially in tasks involving longer text sequences. Previous research showing the advantages of LSTM and GRU over conventional RNNs in text analysis and recommendation systems [7,8] lends credence to these findings.

Conclusion

This study uses word space and deep learning models to provide an in-depth assessment of feedback efficacy in the gaming industry. The findings demonstrate that input from particular domains-like gaming-produces more reliable outcomes than feedback from general domains. When paired with the Skip-Gram model, the GRU algorithm proved to be the most successful method for examining game reviews. These results underscore the significance of specialized data in obtaining high-performance outcomes and add to the expanding corpus of research on sentiment analysis and recommendation systems in the gaming industry. Subsequent investigations may delve into expanding this methodology to encompass additional languages and industries, hence enhancing the models and their suitability for diverse settings.

1. Giriş

1958 yılında ilk bilgisayar oyunu "Tennis for Two" ile tohumları atılan dijital oyunlar, 1970'li yıllarda ilk ev tipi ticari oyun konsolu olan Magnavox Odyssey'in hayatımıza girmesi ile süregelen elli yıllık periyotta kültür endüstrisinin ana elemanlarından biri haline gelmiştir [1]. Kişisel bilgisayarların, internetin, konsol ve mobil cihazların küresel yaygınlaşmasıyla birlikte dijital oyunlar, kullanılan cihazların en çok tüketilen varlıklarından biri olmuştur. Başta fiziki kopyalar ile dağıtılan oyunlar, 2003 yılında Valve firmasının Steam [2] platformunu başlatması, 2013 yılında ise Sony Playstation 4 ve Microsoft Xbox One oyun konsollarının çıkışı oyunların dijital kopyalarının yaygınlaşmasında önemli rol oynadı. 2008 yılında satılan oyunların her 100 tanesinin 22 adedini dijital kopya oyunlar oluştururken, 2014 yılında dijital kopyalar fiziksel satışları geride bırakarak her 100 satıştan 69'unu oluşturmuş ve sektörde yeni bir çağı başlatmıştır [3]. Bu değişimin en önemli noktalarından biri, kullanıcılara aldıkları ürünleri herhangi bir medya kuruluşuna bağlı kalmadan değerlendirme fırsatı vermiş olmasıdır.

Küresel bazda oyun sektörü geliştikçe, bu sektöre ait veriler ve son kullanıcılar tarafından yapılan geri dönüşler daha değerli hale gelmiştir. Geliştirilen oyunlara yapılan yorumlar bir ürünün satın alınma kararlarında önemli bir rol oynamaktadır. Özellikle Metacritic [4] puanları ve Steam platformunda "Tavsiye Ediyor" veya "Tavsiye Etmiyor" etiketleri, kullanıcıların oyunları satın alıp almama kararlarını etkileyen önemli veriler haline gelmiştir. Bu verileri doğru bir şekilde analiz eden firmalar, pazar paylarını önemli ölçüde arttırabilmektedir. Nitekim, küresel oyun pazarı Newzoo Games verilerine göre, 2023 yılında 183,9 milyar dolar gelir elde etmiş ve yine aynı rapora göre sektörde yıllık bazda %6'dan fazla büyüme öngörmektedir [5].

Bu çalışmanın temel amacı, oyun sektörü gibi spesifik bir alanda yapılan geri dönüşleri (yorum, değerlendirme, tavsiye vb.) daha

geniş konu seçeneklerine sahip olan gönderilerle karşılaştırarak, bunların etkinliğini tespit etmektir. Ayrıca, bu etkinlik tespitinde kullanılan derin öğrenme (deep learning) algoritma yaklaşımları arasındaki farklar da ölçülmektedir.

Çalışmada veriler açık kaynaklı bir platformdan alınmıştır. Bu aşamada farklı platformlara ait geniş çaplı veri kümeleri toplanmıştır. Bu veriler kullanılarak, çeşitli kelime temsili yöntemleriyle dört ana başlıkta altı farklı vektör boyutunda yirmi dört farklı kelime uzayı oluşturulmuştur. İkinci aşamada, veri kazıma (data scrapping) yöntemiyle etiketli veri kümeleri elde edilmiştir. Son aşamada ise, bu etiketli veriler üzerinde derin öğrenme algoritmaları kullanılarak bir tavsiye analizi gerçekleştirilmiştir.

Bu çalışmada, metin madenciliğinde sıklıkla kullanılan derin öğrenme algoritmaları arasında yer alan Tekrarlayan Sinir Ağları (Recurrent Neural Networks - RNN) ve türevlerinden Uzun-Kısa Süreli Bellek Ağları (Long Short-Term Memory - LSTM) ile Geçitli Tekrarlayan Birimler (Gated Recurrent Unit - GRU) modelleri kullanılmıştır. RNN'ler metin, konuşma gibi birbirine bağlı bilgilerin uzun vadeli bağımlılıklarını yakalama kabiliyeti sebebiyle yaygın olarak kullanılmaktadır [6]. Bununla birlikte RNN'lerde karşılaşılan kaybolan gradyan problemi, LSTM ve GRU algoritmaları geliştirilerek aşılmıştır. Kaybolan gradyan probleminin aşılmasıyla daha uzun sıralı verilerin işlenmesi mümkün olmuştur. LSTM algoritması uzun vadeli bağımlılıkları öğrenmede üstün bir performans gösterirken GRU'lar basit yapıları ve hesaplama verimlilikleri nedeniyle giderek daha popüler hale gelmektedir [7,8]. Bu modeller, metin madenciliğinde duygu analizi, belge sınıflandırma ve makine çevirisi gibi görevlerde yaygın bir şekilde kullanılmaktadır [9,10].

2017 yılında yaptıkları çalışmalarında, Bais ve ark. [11] Steam platformunda 100 oyuna ait 5000 adet inceleme toplamışlardır. Ham veri seti üzerinde, Destek Vektör Makineleri (Support Vector Machine - SVM), Lojistik Regresyon (LR), Multinomial

Naive Bayes (MNB) ve Turney'in denetimsiz ifade etiketleme algoritması kullanarak duyguları sınıflamışlardır. Burada çalışmacıların amacı sarkastik, esprili yaklaşımları olan yorumlar kullanarak algoritmaların sınıflarını zorlamak olmuştur. Yaptıkları çalışmada çalışmacılar, SVM algoritmasının ironik yorumlarda dahi diğer algoritmalarla karşı daha başarılı bir model olduğu sonucuna varmışlardır.

İlgili çalışmalarında Toçoğlu ve ark. [12], Türkçe dilinde atılan tweetleri altı duygu kategorisinde (Korku, Mutluluk, İğrenme, Öfke, Üzüntü, Şaşkınlık) otomatik olarak etiketlemişlerdir. Araştırmacılar bu çalışmada, 195 binden fazla tweeti sözlük tabanlı bir yaklaşım ile kullanmışlardır. Yapay Sinir Ağları, Konvolüsyonel Sinir Ağları (Convolutional Neural Networks - CNN), Uzun-Kısa Süreli Bellek Ağı ile Tekrarlayan Sinir Ağları mimarilerini kullanarak testler yapmışlardır. Yapılan testlerde doğruluk oranları karşılaştırıldığında CNN mimarisi %74 ile en yüksek puanı almıştır.

2019 yılında Alpkoçak ve ark. [13] TREMO veri seti üzerinde Yapay Sinir Ağları (YSA), Destek Vektör Makineleri (DVM), Random Forest (RF) ve K-En Yakın Komşu (KEYK) algoritmaları kullanarak veri setinin sağladığı mutluluk, korku, öfke, üzüntü, tiksime ve şaşırma kategorilerinde duyguları sınıflamışlardır. Sonuçlara doğruluk metriği ile bakıldığında Yapay Sinir Ağları (YSA) 0,0045 gibi az bir farkla Destek Vektör Makineleri (DVM)'den daha başarılı olmuştur. Duygu kategorilerini bakımından incelendiğinde Yapay Sinir Ağları (YSA) ve Destek Vektör Makineleri (DVM) birbirleriyle yakın sonuçlar üretmişlerdir.

2021 yılındaki çalışmasında Tran [14], "SteamInsider" programını tanıtmaktadır. Araştırmacı çalışmasında, Steam platformundan alınan veriler için RNN tabanlı bir duygu sınıflandırıcı kullanmıştır. İncelemeler üzerinde doğal dil işleme teknikleri kullanarak inceleme özetleri çıkartmış olup müşterilerin satın alma kararlarını daha verimli bir şekilde vermesini sağladığını iddia etmektedir. Sistemin gelecek projeksiyonunda ürün, otel veya restoran yorumlarına da entegre edilmesinin amaçlandığı belirtilmektedir.

Karga ve ark [15], ilgili çalışmalarında COVID-19 pandemisinin yükseköğretim üzerindeki etkilerini otomatik olarak tespit etmek amacıyla Destek Vektör Makineleri, Lojistik Regresyon, Naive Bayes, Rastgele Orman gibi geleneksel yöntemlerle Konvolüsyonel Sinir Ağları, Tekrarlayan Sinir Ağları, Uzun-Kısa Süreli Bellek, Geçitli Tekrarlayan Birimler gibi derin öğrenme modelleri ve BERT sınıflayıcısını kullanarak modellemişlerdir. Veri seti olarak üniversite öğrencileri ile yapılan anketler sonucu oluşturulan "TURCOVIDU" kullanmışlardır. Doğruluk değerleri göz önüne alındığında, BERT sınıflayıcısının geleneksel algoritmalarla ve derin öğrenme modellerine göre 0.9598 ortalama doğruluk sonucu ile daha başarılı olduğunu gözlemlenmiştir.

Tan ve ark. [16], oyun incelemeleri üzerinden duygu analizi için makine öğrenmesi algoritmaları üzerine karşılaştırmalı bir çalışma yapmışlardır. Çalışmalarında Metacritic ve Steam platformundaki oyun incelemelerinden oluşan bir veri seti kullanmışlardır. Bu veri seti üzerinde SVC (Support Vector Classifier), MLP (Multi-layer Perceptron Classifier) , XGB (Extreme Gradient Boosting Classifier), LR (Logistic Regression) ve MNB (Multinomial Naive Bayes) algoritmalarıyla testler yapılmıştır. Aralarında en iyi performansı veren algoritma %91 doğruluk değeriyle SVC olmuştur.

Guzvinecz [17], Soul benzeri zorlu video oyunlarında kullanıcı deneyimini anlamak için Steam incelemelerini NRC Emotion Lexicon kullanarak analiz etmiştir. İlgili çalışmada, 993.932 adet

yorum kullanılmıştır. Bu yayında araştırma 4 aşamaya bölünmüştür. İlk aşamada negatif ve pozitif incelemeler arasındaki farklar ifade edilmiştir. İkinci aşamada negatif ve pozitif inceleme grubu arasındaki duygusal değerler incelenmiştir. Üçüncü aşamada her iki inceleme grubundaki kelime sıklıklarına dikkat çekilmiştir. Burada gözlemlenen şey, negatif incelemelerin genellikle yazılım/donanım/optimizasyon sorunlarını detaylandırmak için kelimeler kullandığı iken, pozitif incelemelerin oyuncu deneyimleri veya oyunlarda kullanıcıların sevdikleri şeyleri tanımlamak için kelimeler kullandıkları olmuştur. Dördüncü aşamada negatif ve pozitif incelemeler arasında oy sayısı ve kelimeler arasında bir korelasyon olup olmadığı sorusu araştırılmış ve çıkan sonuçlara göre iki kategori arasında korelasyon olmadığı sonucuna varılmıştır. Ancak negatif yorumların pozitif yorumlara göre daha fazla oy aldığı ve bunların daha uzun oldukları gözlemlenmiştir.

2023 yılındaki çalışmalarında, Çabuk ve ark. [18], e-ticaret ürün yorumları üzerinden bir duygu analizi çalışması gerçekleştirmiştir. Çalışmanın ilk aşamasında Türkiye'deki e-ticaret platformlarından 15.170 adet ürün yorumu toplanmıştır. Veri setinin etiketleme aşamasında, yorumlar pozitif ve negatif olarak farklı zamanlarda iki kere etiketlenmiş ve iki sınıflı bir veri seti hazırlanmıştır. Çalışmanın ikinci aşamasında ise, bu veri seti ile LSTM algoritması kullanılarak bir model eğitilmiş ve e-ticaret platformlarından toplanan 203.274 adet yorumun yer aldığı bir veri seti otomatik olarak etiketlemiştir. Bu otomatik olarak etiketlenmiş veri setinin performansı dört farklı derin öğrenme algoritması (Tekrarlayan Sinir Ağları, Uzun-Kısa Süreli Bellek, Geçitli Tekrarlayan Birim, Konvolüsyonel Sinir Ağları) kullanılarak manuel olarak etiketlenmiş veri seti karşılaştırmıştır. Çalışmanın son aşamasında, yüksek sınıflandırma performansı gösteren otomatik olarak etiketlenmiş veri seti kullanılarak dört geleneksel makine öğrenmesi (Lojistik Regresyon, Destek Vektör Makineleri, Rastgele Orman, Naive Bayes) ile dört farklı derin öğrenme algoritmasının sınıflandırma performansları karşılaştırmıştır. Yapılan deneysel çalışmalar sonucunda, derin öğrenme algoritmalarının makine öğrenmesi algoritmalarına kıyasla daha başarılı olduğu gözlemlenmiştir. Geliştirilen modelle otomatik olarak etiketlenen veri setinin kullanıldığı deneylerde yüksek performanslı sonuçlar elde edilmiştir.

Literatürde yapılan taramalarda, oyun sektörüyle ilgili çalışmaların sınırlı olduğu, araştırmaların genellikle e-ticaret yorumları veya genel duygu analizi gibi daha geniş kapsamlı alanlara odaklandığı görülmektedir. Oyun incelemelerini içeren kıt sayıdaki çalışmalarda ise derin öğrenme algoritmalarının etkin bir şekilde kullanılmadığı gözlenmektedir. Örneğin, Tran [14], Steam platformundaki kullanıcı yorumlarını RNN tabanlı bir duygu sınıflandırıcı kullanarak analiz etmiştir. Ancak, çalışmasında bağlamın korunmasına ve derin öğrenme algoritmalarının entegrasyonuna yönelik sınırlı bir yaklaşım sergilemiştir. Benzer şekilde, Tan ve arkadaşları [16] ise oyun yorumlarına yönelik duygu analizi çalışmalarında makine öğrenmesi algoritmalarının karşılaştırılmasına odaklanmıştır. Guzvinecz [17], oyun türlerinin duygusal etkilerini anlamak amacıyla Steam platformundan alınan kullanıcı yorumlarını analiz etmiş, ancak yöntemlerinde derin öğrenme algoritmalarını aktif bir şekilde kullanmamıştır. Bu çalışma, önceki çalışmalardan farklı olarak, belirli alanlarda bağlamın korunduğu durumların daha anlamlı sonuçlar üreterek sektörün daha hedefli analizler yapmasını sağlamayı hedeflemektedir.

Bu amaç doğrultusunda, Word2Vec ile oluşturulan kelime uzayları ve derin öğrenme algoritmaları (LSTM, GRU, RNN) birleştirilerek oyun sektörüne özgü verilerin anlamsal ilişkileri

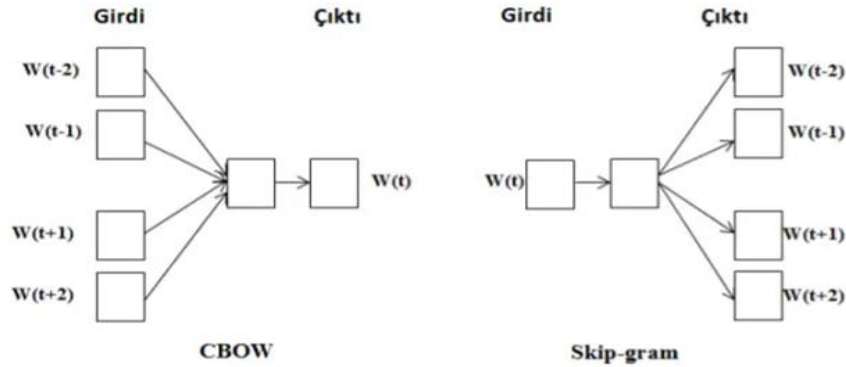
test edilmiştir. Word2Vec modeli, geniş bir veri seti ile eğitilmiş ve ardından daha küçük, etiketli bir veri seti üzerinde derin öğrenme algoritmalarının performansı değerlendirilmiştir. Bu yöntem, geniş ölçekli bağlamsal analizle hedefli analizlerin birleştirilmesini sağlamıştır.

Sonuçlar, oyun sektöründen alınan yorumların, genel veri setlerine kıyasla daha tutarlı ve anlamlı olduğunu ortaya koymaktadır. Bu bulguların, oyun sektörü geri bildirimlerinin Word2Vec ile güçlendirilmiş derin öğrenme algoritmaları ile analizine dair literatüre katkıda bulunduğu düşünülmektedir.

Bu makale şu şekilde düzenlenmiştir: Giriş bölümünden sonra, Bölüm 2'de çalışmada kullanılan teknolojiler hakkında temel ön bilgiler verilmektedir. Bu bölümde ayrıca, kullanılan veri seti ve çalışmanın yöntemi açıklanmıştır. Bölüm 3'te deneyler sonucunda alınan sonuçlar verilmektedir. Bölüm 4'te makale, çalışmayla ilgili son sözlerle ve önerilerle sonlandırılmaktadır.

2. Materyal ve Metot

Çalışmanın bu bölümü ön teknik bilgiler ve yöntem olarak ikiye ayrılmıştır. Ön teknik bilgiler kısmında kullanılan daha kolay anlaşılabilirliği için kısa bilgiler verilmiştir. Yöntem kısmında ise modellemenin veri setlerinden ve teknik detaylardan bahsedilmiştir.



Şekil 1. CBOW – Skip-Gram modelleri gösterimi [19].

Figure 1. CBOW - Skip-Gram models representation.

2.1.2. Derin Öğrenme Algoritmaları

Derin öğrenme, bir makine öğrenmesi (machine learning) alt dalıdır ve yapay sinir ağlarından oluşan yöntemler bütünüdür [20]. Verilerin doğrusal veya doğrusal olmayan dönüşümlerle birden çok işlem katmanında otomatik olarak işlendiği bir tekniktir. Derin öğrenme temelde insan beyninin çalışma prensibini taklit etmeye çalışmaktadır. Derin öğrenme mimarileri katman katman oluşmuş olup her bir katman bir üst katmana özellik aktararak en basit özelliğe kadar bir öğrenme süreci oluşturur [21].

Derin öğrenme algoritmaları farklı matematiksel yaklaşımlarla çalışan bir dizi tekniğe sahiptir. Bu çalışmada, ilgili yaklaşımlardan sıralı verilerin işlenmesinde yaygın olarak kullanılan Tekrarlayan Sinir Ağı (Recurrent Neural Network – RNN), Uzun–Kısa Süreli Bellek Ağı (Long Short Term Memory – LSTM) ve Geçitli Tekrarlayan Birim (Gated Recurrent Unit – GRU) kullanmıştır.

Tekrarlayan Sinir Ağları (RNN) [22], sıralı veri işlemede kullanılan bir yapay sinir ağıdır. Jeffrey Elman 1990 yılında yayımladığı çalışmasıyla RNN'lerin dil modelleme ve dil işleme

2.1. Ön Teknik Bilgiler

Yazının bu bölümünde, çalışmanın anlaşılmasını kolaylaştıracak bazı teknik bilgiler verilmektedir.

2.1.1. Word2Vec ile Kelime Uzaı

Word2Vec; tahmin tabanlı (prediction-based) bir kelime temsil yöntemi olup, 2013 yılında Google araştırmacısı Thomas Mikolov ve ekip arkadaşları tarafından geliştirilmiştir [19]. Yöntemin ortaya çıkarılma amacı, temelinde yapay sinir ağı (artificial neural network - ANN) ile iki farklı metot kullanarak kelimelerin eğitilmesidir (training). Word2Vec'in kullandığı iki metot CBOW(Continuous Bag of Words) ve Skip-Gram'dır. Bu yöntemde, kelimeler n boyutlu bir uzayda vektör olarak temsil edilmekte olup, teknikteki mantık, kelimelerin birbirleri arası mesafelerin hesaplanarak, aralarındaki anlamsal ilişkilerin tespit edilmesi şeklindedir.

CBOW modelinde, bir kelimenin çevresindeki kelimeler kullanılarak hedef kelime tahmin edilmeye çalışılmaktadır. Skip-gram modeli ise, CBOW modelinin tam tersi bir işleyişe sahip olup, hedef kelime kullanılarak bağlam ifadeleri tahmin edilmeye çalışılmaktadır. Bu yöntemlerin işleyiş biçimi Şekil 1'de verilmiştir.

uygulamaları için kullanılabilirliğini göstermiştir [23]. Bu teknikte, önceki çıktılar bir sonraki girdi olarak kullanılarak döngüsel bir yapı oluşturulmaktadır. Bu döngüsel yapı, ardışık gelen verinin birbirleriyle ilişkilendirilerek kullanılmasıyla, model eğitilirken veri içerisindeki tüm ilişkilerin algoritma tarafından hatırlanabilmesini sağlar. Bu özellik, RNN'leri özellikle zaman serileri, dil modeli, el yazısı tespiti ve konuşma tanıma gibi sıralı veri içeren uygulamalarda etkili kılmaktadır.

Uzun–Kısa Süreli Bellek Ağları (LSTM), RNN tabanlı bir derin sinir ağı mimarisidir. 1997 yılında Sepp Hochreiter ve Jürgen Schmidhuber tarafından, RNN mimarisindeki eksiklikleri gidermek amacıyla oluşturulmuştur [8]. RNN mimarisinde, kelimeler arası mesafeler arttıkça, geçmişten gelen ve sistemin döngüye soktuğu bilgileri kullanmak zor hale gelmektedir. LSTM, RNN'den farklı olarak üç kapı içermektedir. Bunlar Giriş Kapısı (Input Gate), Unutma Kapısı (Forget Gate) ve Çıkış Kapısıdır (Output Gate). Hücre, hangi bilgilerin saklanacağına ve ünitelerin bilgilere ne zaman erişeceğine kapıların açık ve kapalı olma durumlarına göre karar vermektedir. LSTM modelleri, konuşma, metin işleme, resimleri tarayarak otomatik başlık oluşturma, robotik kalp cerrahisi ve birçok alanda kullanılmaktadır.

Geçitli Tekrarlayan Birimler (Gated Recurrent Unit - GRU), yeni nesil RNN'dir [23]. LSTM ile aralarında oldukça benzer bir yapı vardır. LSTM üç kapıya sahip iken, GRU ise Güncelleme kapısı (Update Gate) ve Sıfırlama Kapısı (Reset Gate) olmak üzere iki kapıya sahiptir. GRU mimarisinde LSTM'deki iki kapı güncelleme kapısıyla değiştirilmektedir. Bu daha hafif bir yapı ve daha basit bir mimari sağlamaktadır.

2.1.3. Ölçüm metrikleri

Makine öğrenmesi modellerinde performansın değerlendirilmesi temel bir husustur. Bu performansları ölçmek için Doğruluk (Accuracy), Kesinlik (Precision), Duyarlılık/Hassasiyet (Recall), F1 Skoru (F1-Score) gibi temel metrikler kullanılmaktadır.[24]

Doğruluk, makine öğrenmesi modelinin doğru tahmin ettiği örneklerin tüm veri noktalarına oranıdır. Doğruluk, modelin genel performansını yansıtan bir ölçüttür, ancak dengesiz veri setlerinde yanıltıcı olabilir. Genel haliyle, bir modelin performansı hakkında ilk bakışta bir fikir edinilmesini sağlamaktadır. Doğruluk formülü aşağıdaki şekildedir:

Formüllerde kullanılan kısaltmalar;

D = Doğru, Y = Yanlış, P = Pozitif, N = Negatif

$$\text{Doğruluk} = \frac{D.P + D.N}{D.P + Y.P + D.N + Y.N} \quad (1)$$

Kesinlik, bir sınıflandırma modelinin pozitif sınıfı tahmin ettiğindeki yüzdelik başarısını ifade etmektedir. Matematiksel olarak ifade edilecek olursa, pozitif doğru tahminlerin tüm pozitif tahminlere oranı olarak söylenebilir. Yani bu formül, modelin pozitif olarak sınıflandırdığı örneklerin ne kadarının gerçekten pozitif olduğunu gösterir. Yüksek bir kesinlik skoru, yanlış pozitiflerin az olduğu anlamına gelmektedir. Bu bağlamda kesinlik formülü aşağıdaki gibi yazılabilir:

$$\text{Kesinlik} = \frac{D.P}{D.P + Y.N} \quad (2)$$

Duyarlılık ise, gerçek pozitif değerlerin ne kadarının pozitif olarak tahmin edildiğini ölçmektedir. "Pozitif tahmine dayalı değer" olarak da adlandırılabilir. Duyarlılık, modelin pozitif olan örnekleri ne kadar iyi bulduğunu gösterir. Yüksek bir duyarlılık skoru, yanlış negatiflerin az olduğunu ve modelin pozitif sınıfları kaçırmadığını ifade eder. Formülü aşağıda görülebilir.

$$\text{Duyarlılık} = \frac{D.P}{D.P + Y.N} \quad (3)$$

F1 Skoru, Kesinlik ve Duyarlılık skorlarının harmonik ortalamasıdır. Bu iki değer arasındaki dengeyi sağlamak için kullanılmaktadır. Yüksek kesinlik ve yüksek duyarlılık aranıyorsa F1 Skoru bu dengeyi sağlamak için kullanışlı hale gelmektedir. Özellikle dengesiz veri setlerinde oldukça işlevsel olan bu ölçüt, modelin dengeli bir şekilde hem kesinlik hem de duyarlılık performansını ölçer. Formül aşağıda verilmiştir.

$$F1 \text{ Skoru} = 2 * \frac{\text{Kesinlik} * \text{Duyarlılık}}{\text{Kesinlik} + \text{Duyarlılık}} \quad (4)$$

2.2. Yöntem

Bu çalışmada, farklı kelime uzaylarını karşılaştırmak amacıyla iki farklı veri seti kullanılmıştır. Bu veri setlerine, açık kaynaklı veri sağlayan Kaggle platformundan ulaşılmıştır [25]. İlk veri seti, popüler bir dijital oyun platformu olan Steam üzerindeki İngilizce dilindeki oyun yorumlarını içermektedir ve toplamda 6.417.106 satırdan oluşmaktadır. Bu veri seti, çeşitli kategorideki yüzlerce oyuna yapılan kullanıcı yorumlarını kapsamaktadır.

İkinci veri seti ise, farklı konu başlıkları altında toplanmış yine İngilizce dilindeki 6,5 milyon X platformu gönderisi içermektedir.

Bu veri seti daha geniş bir dilsel çeşitlilik ve konusal zenginlik sunarken, kullanıcıların düşüncelerini, sorunlarını ve genel tepkilerini içermektedir. Başarı oranlarının eşit faktörlerden etkilenmesi amacıyla, birbirine yakın büyüklükte kelime uzayları tercih edilmiştir. Kelime uzaylarının büyüklüğü, veri setlerinin boyutları ve içerdiği kelime çeşitleri sayesinde kurulan modellerin performansları daha kapsamlı bir şekilde test edilebilmiştir.

Eğitim veri seti ise, Steam platformundan Python programlama dilinin BeautifulSoup kütüphanesi kullanılarak veri kazıma yöntemi ile elde edilmiştir. Bu süreçte 169 farklı oyun kullanılarak 14.769 satır İngilizce yoruma ulaşılmıştır. Steam platformunda her bir oyunun benzersiz bir kimlik numarası bulunmaktadır. Bu kimlik numaraları sayesinde, Steam platformundaki oyunların topluluk bölümündeki incelemelerine, web adresindeki kimlik numarası değiştirilerek kolayca erişilebilmektedir. Bu yöntem kullanılarak BeautifulSoup ile web adresinin kaynak kodlarına ulaşılıp metin kutuları içerisindeki ifadelerle ulaşılmıştır. Ayrıca veri kazıma işlemi sayesinde, istenilen formatta etiketlenmiş verilere kolaylıkla erişilebilmiştir. Siteden alınan veriler hazır olarak kategorize edildiği için herhangi bir manuel etiketleme işlemi yapılmasına gerek kalmamıştır. Tablo 1'de eğitim veri setinin etiket dağılımına yer verilmektedir.

Tablo 1. Eğitim veri seti etiket dağılımı.

Table 1. Label distribution of the training data set.

Sınıf	Yorum Adedi
Tavsiye Edilen (Recommend)	7007
Tavsiye Edilmeyen (Not Recommend)	7762
Toplam	14769

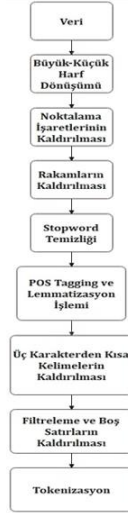
Hazırlanan veri setleri, derin öğrenme algoritmaları tarafından eğitilmeden önce ön işlem (preprocessing) aşamalarından geçirilmiştir.

Metin temizleme (text cleaning) ve metin normalleştirme (text normalization) işlemleri ile ilk önce ham veri seti işlenmeye başlanmıştır. Veri seti, ham veride bulunan ve çalışmaya dahil edilmemesi gereken noktalama işaretleri, sayılar, internet adresleri, emojiler ve benzeri karakterlerden temizlenmiştir. Ayrıca etkisiz kelimeler (stop words) Python'un NLTK(Natural Language Toolkit) kütüphanesi [26] kullanılarak veri setinden çıkarılmıştır. Son aşama olarak yazım yanlış kaynaklı anlamı olmayan tek karakter uzunluğundaki kelimeler de veri setinden temizlenmiştir.

Makine öğrenmesinde, lemmatization (kök bulma) işlemi ile metin verisi daha etkili bir şekilde işlenmektedir. Bu işlem, kelimeleri morfolojik olarak analiz ederek, bunların cümle içerisindeki görevlerini tespit ederek doğru köke ulaşmayı hedefler. Böylece, modele performans artışı ve kelime anlamlarının korunması gibi faydalar sağlamaktadır. Çalışmada NLTK kütüphanesinin "POS (Part of Speech) Tagging" veri setindeki kelimeler cümle yapısına göre isim, fiil, sıfat, zarf olarak etiketlenmiştir. NLTK kütüphanesinin "WordNetLemmatizer" işlevinin bu etiketleri takip ederek daha doğru bir kök bulma işlemi sunması sağlanmıştır. Temizleme ve kök bulma işlemlerinin sona ermesi ile üç karakterden kısa kelimeler de tespit edilerek veri setinden çıkarılmıştır. Ayrıca silinen verilerden kaynaklı boş satırlar oluşmuştur. Bu satırlara NumPy [27] kütüphanesi yardımıyla "NaN" değeri atanmıştır ve bunlar veri setinden temizlenmiştir. Gerekli temizlik adımları

gerçekleştirildikten sonra tokenizasyon işlemi uygulanmıştır. Şekil 2’de metin ön işleme sürecinin görsel özeti verilmiştir.

Hazırlanan bu veriler kullanarak Word2Vec metodu ile bu metoda ait yöntemler olan CBOW ve Skip-Gram yöntemleri kullanarak dört ana başlıkta altı farklı vektör boyutunda yirmi dört farklı kelime uzayı oluşturulmuştur. Elde edilen etiketli



Şekil 2. Metin ön işleme sürecinin görsel özeti.

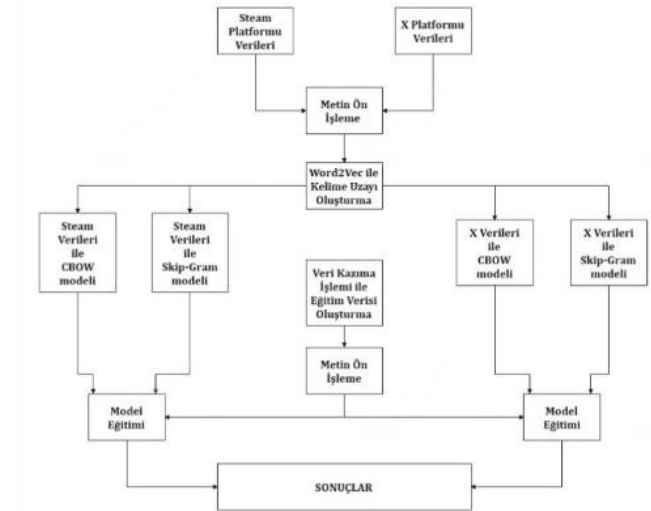
Figure 2. Visual summary of the text preprocessing process.

3. Bulgular ve Tartışma

Bu çalışmada, CBOW ve Skip-Gram modelleri kullanarak derin öğrenme algoritmalarının farklı kelime uzaylarındaki performansları ölçülmüştür. Çalışmada Word2Vec yönteminin tercih edilmesinin temel nedeni, metin analizlerinde kelimeler arasındaki bağlamsal ilişkileri iyi bir şekilde temsil etmesidir. Skip-Gram yöntemi, özellikle nadir kelimeler için daha güçlü sonuçlar verirken, CBOW yöntemi hızlı eğitim süresiyle öne çıkmaktadır. Bu özellikler, oyun sektörüne özgü ve bağlam odaklı analizlerde Word2Vec'i etkili bir seçenek haline getirmiştir [19]. Modeller tasarlanırken, öncelikle bazı parametrelere ayarlamaları yapılmıştır. Word2Vec parametreleri düzenlenirken en iyi performans veren değerler göz önüne alınmıştır. Bu doğrultuda asgari kelime frekansı değeri 5 olarak belirlenmiştir. Bu değer ile, en az beş kez tekrar eden kelimelerin seçilmesi sağlanmıştır. Böylece, nadir kelimeler modelden çıkarılarak hesaplama maliyetleri azaltılmış ve modelin daha yaygın olan kelimeler üzerine yoğunlaşması sağlanmıştır. Vektör boyutu 50, 100, 150, 200, 250, 300 değerleri sırasıyla test edilmiştir. Bu, kelime vektörlerinin vektör boyutu parametresinde belirlediğimiz değer boyutunda bir uzayda temsil edileceği anlamına gelmektedir. Kelime anlamlarını daha iyi yakalamak ve modelin genel performansını artırmak için karşılaştırmalı skorlar elde edilmiştir. Karşılaştırılan skorlar arasında en başarılı vektör boyutunun hangisi olduğu anlamaya çalışılmıştır. Pencere boyutu ise 5 olarak ayarlanmıştır. Bu değer, hedef kelimenin sağ ve sol tarafındaki 5 komşusunun model eğitimi sırasında hesaba katılacağını belirtmektedir. Bu şekilde, kelimenin bağlamını daha iyi yakalayarak daha anlamlı vektörler oluşturulması hedeflenmiştir. Açıklanan bu parametre değerleri Tablo 2’de incelenebilir.

Çalışmada, modelin eğitim aşamasında veri kazıma yöntemiyle elde edilen 14.769 adet oyun yorumu kullanılmıştır. Veri, yapı

veriler kullanılarak, oluşturulan kelime uzayları üzerinde derin öğrenme algoritmaları uygulanarak bir tavsiye analizi yapılmıştır. Bu analizde RNN, RNN ilkeleri ile modellenen LSTM’ler ve LSTM modeline çok yakın bir mantık anlayışını takip eden GRU modelleri kullanılmıştır. Şekil 3’de çalışmayı daha anlaşılır kılmak amacıyla kullanılan yöntemin akış şemasına yer verilmiştir.



Şekil 3. Kullanılan yöntemin akış şeması.

Figure 3. Flowchart of the method used.

itibariyle “Tavsiye Edilen (Recommend)” ve “Tavsiye Edilmeyen (Not Recommend)” olarak ayrılmış halde kazanmıştır. Bu kategorik veriler, “Recommend = 1” ve “Not Recommend = 0” olacak şekilde sayısallaştırılmıştır. Word2Vec modelini inşa ederken yapıldığı gibi, eğitim veri seti de metin ön işleme işlemlerine tabi tutulmuştur. Bu veri seti kullanılarak, modeller üç farklı derin öğrenme algoritması olan RNN (Recurrent Neural Networks), LSTM (Long Short-Term Memory) ve GRU (Gated Recurrent Unit) ile eğitilmiştir.

Tablo 2. Word2Vec parametreleri.

Table 2. Word2Vec parameters.

Parametre Adı	Değeri
Asgari Sayı (Minimum Count)	5
Vektör Boyutu (Vector Size)	50, 100, 150, 200, 250, 300
Pencere Boyutu (Window)	5

Steam ve X platformlarından alınan veriler işlenerek, Word2Vec metoduyla altı farklı vektör boyutunda yirmi dört adet kelime uzayı elde edilmiştir.

Çalışmada model, Keras [28] ve Scikit-learn [29], kütüphaneleri kullanılarak Python ortamında eğitilmiştir. Derin öğrenme algoritmalarında kullanılan parametreler, en iyi performansı sağlayan değerlere göre belirlenmiştir. Bununla birlikte, gömme boyutlarının performansa etkisini değerlendirmek amacıyla altı farklı gömme boyutu test edilmiştir. Ayrıca, gömme dolgu boyutu (padding size) en uzun yoruma uygun olacak şekilde “max size” değeri olarak belirlenmiş ve bu sayede tüm yorumların eşit uzunlukta işlenmesi sağlanmıştır. Model eğitilirken kullanılan parametreler Tablo 3’te verilmiştir.

Tablo 3. Derin öğrenme algoritma eğitim parametreleri.**Table 3.** Deep learning algorithm training parameters.

Parametre Adı	Değeri
Gömme Boyutu (Embedding Dimension Size)	50, 100, 150, 200, 250, 300
Gömme Dolgu Boyutu (Embedding Padding Size)	Max Size
Nöron Boyutu (Neuron Size)	128
Bırakma/Düşürme Oranı (Droupout Rate)	0,1
Dönem/Epok Sayısı (Number of Epochs)	5
Yığın Boyutu (Batch Size)	128
Optimize Edici (Optimizer)	Adamax
Aktivasyon Fonksiyonu (Activation Function)	ReLU
Rastgelelik Tohumu (Random Seed)	42
Kayıp Fonksiyonu (Loss Function)	Binary Crossentropy
K-Katlı Çapraz Doğrulama Bölmeleri (K-Fold Cross-Validation Splits)	10

Yapılan deneyler sonucunda ortaya çıkan sonuçlar; Tablo 4, Tablo 5, Tablo 6 ve Tablo 7’de verilmiştir (Elde edilen skorlar ısı haritası şeklinde verilmiştir ve en başarılı skor kalın olarak işaretlenmiştir)

Tablo 4. Doğruluk sonuçları ve ısı haritası.**Table 4.** Accuracy scores and heat map.

ACCURACY SCORE					
Platform	WORD2VEC	VEKTÖR	RNN	LSTM	GRU
STEAM	CBOW	50	74,42%	79,01%	79,19%
		100	75,52%	76,56%	79,54%
		150	73,68%	79,08%	79,14%
		200	74,61%	79,46%	79,20%
		250	74,37%	78,16%	79,20%
		300	73,37%	79,46%	79,45%
	SKIPGRAM	50	69,64%	78,06%	78,72%
		100	72,01%	79,16%	79,82%
		150	70,47%	78,86%	79,77%
		200	71,75%	79,36%	79,59%
		250	71,75%	79,55%	80,16%
		300	70,01%	79,13%	79,22%
X	CBOW	50	66,53%	75,27%	75,88%
		100	67,88%	75,86%	76,43%
		150	68,59%	76,44%	76,49%
		200	65,53%	76,28%	76,58%
		250	68,28%	74,56%	77,92%
		300	70,30%	75,82%	76,57%
	SKIPGRAM	50	66,23%	72,81%	74,05%
		100	66,66%	74,50%	76,80%
		150	66,10%	74,83%	75,20%
		200	65,52%	75,27%	76,13%
		250	69,74%	75,45%	76,44%
		300	66,23%	76,36%	76,00%

Tablo 5. Kesinlik sonuçları ve ısı haritası.**Table 5.** Precision scores and heat map.

PRECISION SCORE					
Platform	WORD2VEC	VEKTÖR	RNN	LSTM	GRU
STEAM	CBOW	50	74,56%	79,12%	79,26%
		100	75,53%	77,44%	79,54%
		150	73,64%	79,16%	79,30%
		200	74,56%	77,69%	79,16%
		250	74,48%	78,40%	79,27%
		300	73,75%	79,96%	79,42%
	SKIPGRAM	50	70,22%	78,44%	79,01%
		100	72,91%	79,20%	79,97%
		150	70,19%	79,18%	79,85%
		200	73,15%	79,52%	79,63%
		250	73,15%	79,68%	80,21%
		300	72,75%	79,50%	79,72%
X	CBOW	50	66,76%	75,44%	75,92%
		100	67,88%	76,01%	76,58%
		150	68,88%	76,63%	76,57%
		200	65,85%	76,50%	76,63%
		250	68,44%	75,71%	77,02%
		300	70,83%	76,15%	76,93%
	SKIPGRAM	50	67,52%	73,05%	74,38%
		100	67,58%	74,91%	76,20%
		150	66,81%	75,05%	75,70%
		200	67,24%	75,44%	76,61%
		250	69,85%	75,76%	76,62%
		300	68,16%	75,72%	76,67%

Tablo 6. Duyarlılık sonuçları ve ısı haritası.**Table 6.** Recall scores and heat map.

RECALL SCORE					
Platform	WORD2VEC	VEKTÖR	RNN	LSTM	GRU
STEAM	CBOW	50	74,41%	78,94%	79,08%
		100	75,39%	76,37%	79,45%
		150	73,44%	79,01%	79,24%
		200	74,46%	77,04%	79,16%
		250	74,08%	78,10%	79,01%
		300	73,53%	79,47%	79,37%
	SKIPGRAM	50	69,17%	78,18%	78,76%
		100	71,77%	79,20%	79,58%
		150	69,85%	78,91%	79,70%
		200	71,07%	79,38%	79,51%
		250	71,42%	79,60%	80,00%
		300	69,49%	79,17%	79,00%
X	CBOW	50	66,01%	75,16%	75,76%
		100	67,44%	75,72%	76,26%
		150	68,76%	76,31%	76,43%
		200	65,03%	76,14%	76,45%
		250	67,84%	75,29%	76,82%
		300	70,41%	75,82%	76,73%
	SKIPGRAM	50	65,64%	72,53%	73,83%
		100	66,52%	74,50%	75,70%
		150	65,78%	74,69%	75,13%
		200	65,21%	75,07%	76,10%
		250	69,41%	75,39%	76,33%
		300	67,08%	76,37%	76,25%

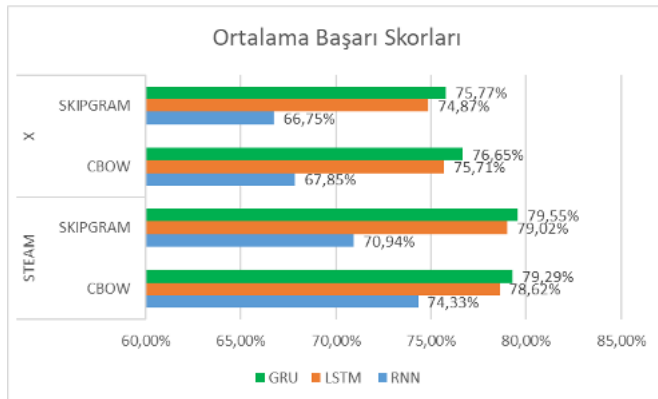
Tablo 7. F1 skoru sonuçları ve ısı haritası**Table 7.** F1 scores and heat map

F1-SCORE					
Platform	WORD2VEC	VEKTÖR	RNN	LSTM	GRU
STEAM	CBOW	50	74,29%	78,84%	79,06%
		100	75,36%	76,01%	79,44%
		150	73,45%	78,95%	79,08%
		200	74,47%	76,84%	79,12%
		250	74,10%	77,94%	79,04%
		300	73,52%	79,30%	79,36%
	SKIPGRAM	50	68,45%	77,96%	78,61%
		100	71,27%	79,09%	79,64%
		150	67,59%	78,84%	79,63%
		200	70,33%	79,29%	79,47%
		250	69,90%	79,47%	80,02%
		300	67,78%	79,01%	78,97%
X	CBOW	50	65,29%	75,09%	75,68%
		100	67,32%	75,68%	76,24%
		150	68,72%	76,26%	76,37%
		200	64,32%	76,08%	76,43%
		250	67,65%	75,15%	76,85%
		300	70,30%	75,82%	76,57%
	SKIPGRAM	50	63,22%	72,48%	73,74%
		100	65,39%	74,31%	75,62%
		150	64,10%	74,61%	74,95%
		200	63,70%	75,05%	75,93%
		250	69,26%	75,26%	76,27%
		300	66,23%	76,36%	76,00%

Ortaya çıkan sonuçlar göz önüne alındığında, Steam platformundan alınan verilerin X platformundan alınan verilere göre daha başarılı sonuçlar verdiği gözlemlenmektedir. Steam platformundan alınan verilerde, 250 vektör boyutunda Skip-Gram modeline GRU algoritması uygulandığında Doğruluk, Kesinlik ve Duyarlılık ve F1 Skorları puanları bakımından (%80,16, %80,21, %80,00 ve %80,02) ile en başarılı model olarak görülmektedir. Modellerin farklı vektör boyutları ve derin öğrenme algoritmalarından alınan skorların ısı haritaları incelendiğinde daha başarılı olan modellerin Steam veri setine ait olduğu gözlenmektedir. Ayrıca, GRU algoritması LSTM algoritmasına çok yakın bir mantık anlayışını takip ettiği için genel olarak başarı puanları ve ısı haritası renk yoğunluğu LSTM ve GRU algoritması sütunlarında toplandığı görülmektedir. Buna karşılık, RNN algoritması beklendiği gibi uzun cümlelerde başarısız olmuş ve performans skorlarında en düşük performansı X platformunun Skip-Gram modellemesinde göstermiştir.

Tablolar daha detaylı incelendiğinde, aşağıdaki bulgulara ulaşılabilmektedir.

- Doğruluk Skorları: Steam-SkipGram-250 modeline uygulanan GRU algoritması %80,16 ile en yüksek doğruluk oranına sahip olurken, X-SkipGram-200 modeline uygulanan RNN algoritması %65,52 ile en düşük doğruluk oranına sahip olmuştur.
- Kesinlik Skorları: Steam-SkipGram-250 modeline uygulanan GRU algoritması %80,21 ile en yüksek hassasiyet oranına sahipken, X-CBOW-200 modeline uygulanan RNN algoritması %65,85 ile en düşük hassasiyet oranına sahip olmuştur.
- Duyarlılık Skorları: Steam-SkipGram-250 modeline uygulanan GRU algoritması %80,00 ile en yüksek duyarlılık oranına sahipken, X-CBOW-200 modeline uygulanan RNN algoritması %65,03 ile en düşük duyarlılık oranına sahip olmuştur.
- F1 Skorları: Steam-SkipGram-250 modeline uygulanan GRU algoritması %80,02 ile en yüksek F1 skoruna sahipken, X-SkipGram-50 modeline uygulanan RNN algoritması %63,22 ile en düşük F1 skoruna sahip olmuştur. Model performanslarının görsel olarak daha iyi incelenebilmesi için çubuk grafikleri oluşturulmuş ve bu grafik Şekil 4'te verilmiştir. Ortalama başarı skorları grafiğine göre Steam-SkipGram modeli farklı vektör boyutları da göz önüne alındığında %79,55 ile daha başarılı görülmektedir. Grafikteki RNN başarı ortalaması da X platformundaki verilerden STEAM platformundaki verilere göre %7,58 artış ile kayda değer bir gelişim göstermiştir.

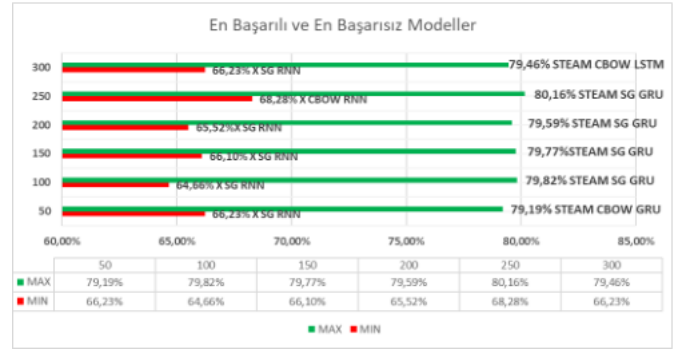


Şekil 4. Derin öğrenme algoritmalarının ortalama doğruluk skorları.

Figure 4. Average accuracy scores of deep learning algorithms.

Şekil 5'te farklı vektör boyutlarındaki başarılı ve başarısız modeller ele alınmıştır. Burada da görüldüğü üzere Steam-

SkipGram algoritmaları diğer algoritmalarla göre daha iyi sonuçlar verdiği gözlemlenmektedir.



Şekil 5. Doğruluk sonuçlarına göre en başarılı ve en başarısız modeller.

Figure 5. Most successful and least successful models based on accuracy scores.

4. Sonuç

Bu çalışmada, oyun sektörü gibi spesifik bir alandan elde edilen geri dönüşler ile daha geniş konu seçeneklerine sahip gönderiler karşılaştırılarak, bu geri dönüşlerin etkinliği incelenmiştir. Açık kaynaklı platformlardan toplanan geniş çaplı veri kümeleri kullanılarak Word2Vec'in CBOW ve Skip-Gram yöntemleri ile dört ana kategoride altı farklı vektör boyutu ile 24 farklı kelime uzayı oluşturulmuştur. Veri kazıma yöntemiyle etiketlenen 14.769 oyun yorumu üzerinde LSTM, RNN ve GRU gibi derin öğrenme algoritmaları uygulanarak bir tavsiye analizi gerçekleştirilmiştir. Bu süreçte, farklı kelime uzayları ve derin öğrenme yaklaşımları ile elde edilen başarı skorları karşılaştırılmış ve algoritmaların etkinlikleri değerlendirilmiştir.

Yapılan çalışmanın sonuçları değerlendirildiğinde, Steam platformundan alınan yorumların CBOW ve Skip-Gram yöntemleri ile oluşturulan kelime uzaylarında, özellikle LSTM ve GRU algoritmaları ile yüksek performans gösterdiği, X platformundan alınan gönderilerin ise aynı modellerde daha düşük performans sergilediği belirlenmiştir. Bu durum, spesifik bir alandan alınan verilerin, genel konu başlıklarına göre daha tutarlı ve anlamlı sonuçlar ürettiğini ortaya koymaktadır.

Oyun sektörü gibi belirli bir alandan alınan veri kümeleri, analiz edilen içeriklerle ortak bir bağlam ve dil sunmaktadır. Bu tür yorumlar, genellikle oynanış, grafikler ve mekanikler gibi benzer konulara odaklandığından, algoritmaların metinler arasındaki ilişkileri daha iyi anlamasına ve daha tutarlı sonuçlar üretmesine olanak tanımaktadır. Buna karşılık, genel konulara odaklanan veri kümeleri çok farklı bağlam ve anlamlar içerebilmektedir. Bu çeşitlilik, algoritmaların tutarlılığını azaltarak analiz sonuçlarının güvenilirliğini olumsuz etkileyebilmektedir.

Oyun sektörü giderek gelişmekte olup kullanıcı deneyimlerinin anlık ve etkili bir şekilde analiz edilmesi her geçen gün daha değerli hale gelmektedir. Ülkemizde de oyun sektörü gün geçtikçe büyümekte ve bu gibi geri dönüşlere ihtiyaç duyulmaktadır. Bu çalışmanın sonuçları doğrultusunda, çeşitli platformlarda üretilen yerli oyunların incelemeleri kullanılarak bir veri seti oluşturulabileceği ve yerli oyun geliştiricilerine, oyunlarının pozitif ve negatif yönlerini özetleyen yararlı bilgilerin sağlanabileceği düşünülmektedir. Bunun yanı sıra, elde edilen geri dönüşlerin analiz edilmesiyle geliştiricilerin yalnızca oyunlarını iyileştirmekle kalmayıp, oyuncuların beklentilerine global ölçekte daha hızlı adapte olarak sektörde rekabet avantajı sağlamalarına da katkıda bulunabileceği düşünülmektedir.

Etik kurul onayı ve çıkar çatışması beyanı

Hazırlanan makalede etik kurul izni alınmasına gerek yoktur. Hazırlanan makalede herhangi bir kişi/kurum ile çıkar çatışması bulunmamaktadır.

Kaynaklar

- [1] Alpkoçak, A., Tocoglu, M. A., Çelikten, A., ve Aygün, İ., 2019. Türkçe Metinlerde Duygu Analizi için Farklı Makine Öğrenmesi Yöntemlerinin Karşılaştırılması. Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen Ve Mühendislik Dergisi, Cilt.21(63), s.719-725. DOI: 10.21205/deufmd.2019216303.
- [2] Bais, R., Odek, P., ve Ou, S., 2017. Sentiment Classification on Steam Reviews. Stanford University. <https://cs229.stanford.edu/proj2017/final-reports/5244171.pdf>.
- [3] Bird, S., Loper, E., ve Klein, E., 2009. Natural Language Processing with Python. O'Reilly Media Inc.
- [4] Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., ve Bengio, Y., 2014. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. arXiv preprint arXiv:1406.1078. DOI: 10.48550/arXiv.1406.1078.
- [5] Chollet, F., 2015. Keras. <https://github.com/fchollet/keras>.
- [6] Chung, J., Gulcehre, C., Cho, K., ve Bengio, Y., 2014. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. arXiv:1412.3555.
- [7] Çabuk, M., Yücalar, F., ve Toçoğlu, M. A., 2023. Makine Öğrenmesi ile E-Ticaret Ürün Yorumlarının Otomatik Analizi. Avrupa Bilim ve Teknoloji Dergisi, (52), s.110-121.
- [8] Elman, J. L., 1990. Finding structure in time. Cognitive Science, Cilt.14(2), s.179-211. DOI: 10.1207/s15516709cog1402_1.
- [9] Fandom Inc., t.y. Metacritic. <https://www.metacritic.com> (Erişim Tarihi: 25.03.2024).
- [10] Goodfellow, I., Bengio, Y., ve Courville, A., 2016. Deep Learning. MIT Press.
- [11] Guzsvinecz, T., 2023. Sentiment Analysis of Souls-like Role-Playing Video Game Reviews. Acta Polytechnica Hungarica, Cilt.20(5), s.149-168.
- [12] Güvenli İnternet Merkezi, 2019. Dijital oyunlar raporu 2019. Bilgi Teknolojileri ve İletişim Kurumu.
- [13] Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., vd., 2020. Array programming with NumPy. Nature, Cilt.585(7825), s.357-362. DOI: 10.1038/s41586-020-2649-2.
- [14] Hochreiter, S., ve Schmidhuber, J., 1997. Long short-term memory. Neural Computation, Cilt.9(8), s.1735-1780. DOI: 10.1162/neco.1997.9.8.1735.
- [15] Kaggle, t.y. <https://www.kaggle.com> (Erişim Tarihi: 12.04.2024).
- [16] Karga, K., Toçoğlu, M. A., ve Onan, A., 2022. Deep Learning-Based Sentiment Analysis on Education During the COVID-19 Pandemic. Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen Ve Mühendislik Dergisi, Cilt.24(72), s.855-868. DOI: 10.21205/deufmd.2022247215.
- [17] Kent, S. L., 2001. The ultimate history of video games: From Pong to Pokémon and beyond—The story behind the craze that touched our lives and changed the world. Three Rivers Press.
- [18] Microsoft, t.y. Karışıklık Matrisi. Microsoft Learn. <https://learn.microsoft.com/tr-tr/dynamics365/finance/finance-insights/confusion-matrix> (Erişim Tarihi: 19.09.2025).
- [19] Mikolov, T., Chen, K., Corrado, G., ve Dean, J., 2013. Efficient Estimation of Word Representations in Vector Space. arXiv preprint arXiv:1301.3781.
- [20] Newzoo, 2023. Global games market report 2023. Newzoo.
- [21] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., vd., 2011. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, Cilt.12, s.2825-2830.
- [22] Rumelhart, D. E., Hinton, G. E., ve Williams, R. J., 1986. Learning representations by back-propagating errors. Nature, Cilt.323(6088), s.533-536.
- [23] Şeker, A., Diri, B., ve Balık, H. H., 2017. Derin öğrenme yöntemleri ve uygulamaları hakkında bir inceleme. Gazi Mühendislik Bilimleri Dergisi, Cilt.3(3), s.47-64.
- [24] Tan, J. Y., Chow, A. S. K., ve Tan, C. W., 2022. A comparative study of machine learning algorithms for sentiment analysis of game reviews. The Journal of The Institution of Engineers Malaysia.
- [25] Toçoğlu, M. A., Öztürkmenoğlu, O., ve Alpkoçak, A., 2019. Emotion Analysis from Turkish Tweets using Deep Neural Networks. IEEE Access, Cilt.7, s.183061-183069.
- [26] Tran, N. K. H., 2021. Development of a machine learning system for aspect-based sentiment analysis and text summarization of video game reviews on Steam. Hamburg University of Applied Sciences.
- [27] Valve Corporation, 2003. Steam. <https://store.steampowered.com> (Erişim Tarihi: 15.03.2024).
- [28] Young, T., Hazarika, D., Poria, S., ve Cambria, E., 2018. Recent Trends in Deep Learning Based Natural Language Processing. IEEE Computational Intelligence Magazine, Cilt.13(3), s.55-75.
- [29] Zhou, J., ve Xu, J., 2015. A Comprehensive Review on Long Short-Term Memory Networks. IEEE Transactions on Neural Networks and Learning Systems, Cilt.26(10), s.2225-2239.