

# OYUN YORUM ANALİZİ RAPORU

## GİRİŞ

Bu projede, dijital oyun platformlarında yer alan kullanıcı yorumlarının doğal dil işleme (NLP) teknikleriyle analiz edilerek, yorumlar arasındaki anlamsal benzerliklerin hesaplanması ve çeşitli metin temsili modellerinin performanslarının karşılaştırılması amaçlanmıştır. Projede temel olarak iki ana NLP yaklaşımı kullanılmıştır: TF-IDF (Term Frequency-Inverse Document Frequency) ve Word2Vec. Bu yöntemlerle oluşturulan sayısal temsiller üzerinden bir örnek cümleyle en benzer yorumlar belirlenmiş, ardından bu sonuçlar subjektif değerlendirme ve Jaccard benzerlik analizleriyle desteklenmiştir.

Bu tür analizler; kullanıcı geri bildirimlerinin otomatik olarak sınıflandırılması, anlam kümelerine göre gruplandırılması ve öneri sistemlerinin geliştirilmesi gibi birçok uygulamaya doğrudan katkı sağlayabilir. Özellikle oyun sektörü gibi büyük hacimli yorum trafiği olan alanlarda, bu tür metin madenciliği yöntemleriyle kullanıcı eğilimleri hızlıca analiz edilebilir.

## 2. KODLARDAN ÖRNEKLER

*TF-IDF ile benzerlik hesaplama kodu:*

```
from sklearn.metrics.pairwise import cosine_similarity  
  
input_vec_lem = vectorizer_lem.transform([input_text])  
  
scores_lem = cosine_similarity(input_vec_lem, matrix_lem).flatten()  
  
top_lem = df_lem.iloc[scores_lem.argsort()[-5:][:,-1]]
```

*Word2Vec ortalama vektörle benzerlik hesaplama:*

```
input_tokens = input_text.lower().split()  
  
input_vecs = [model.wv[w] for w in input_tokens if w in model.wv]  
  
input_avg = np.mean(input_vecs, axis=0)
```

Bu kod blokları, her iki yöntemle giriş cümlesine en benzer yorumların belirlenmesini sağlamaktadır.

### 3. TF-IDF VE WORD2VEC ÇIKTILARI

Giriş cümlesi olarak belirlenen: *"Oyunun atmosferi muhteşemdi ama bazı kontroller zorluydu."* ifadesi, her modele giriş olarak verilmiştir. Aşağıda, bu cümleye göre en benzer bulunan bazı yorum örnekleri verilmiştir:

*TF-IDF modeli ile:*

- "Grafikler harika ama oynanış biraz sıkıcı."
- "Kontroller çok kötü, oynarken sinir oluyorum."

*Word2Vec SG modeli ile:*

- "Atmosfer ve müzikler muhteşem."
- "Sürükleyici senaryo, tekrar tekrar oynanır."

TF-IDF kelime düzeyinde sıklığa dayalı eşleşmeler sunarken, Word2Vec SG modelinin daha bağlamsal ve anlamsal yakınlık içeren yorumları getirdiği gözlemlenmiştir.

#### 4. ANLAMSAL PUAN TABLOSU

Model	1	2	3	4	5	Ortalama
TF-IDF Lemmatized	4	3	4	3	2	3.2
TF-IDF Stemmed	3	3	2	2	2	2.4
Word2Vec SG	5	4	5	4	5	4.6
Word2Vec CBOW	4	4	3	4	3	3.6

Yorum:

- Word2Vec SG modeli, en yüksek ortalama puanı elde etmiş ve girilen cümleyle anlamca en tutarlı yorumları döndürmüştür.
- TF-IDF modelleri daha yüze yakın eşleşmeler üretmiş; ancak içerik bütünlüğü ve bağlam açısından sınırlı kalmıştır.
- CBOW, SG modeline göre daha az başarılı olmakla birlikte TF-IDF'ten daha dengeli sonuçlar üretmiştir.

## 5. JACCARD MATRİSİ

Jaccard benzerliği, her iki modelin ürettiği ilk 5 yorum dizisinin benzerliğini ölçen bir metriktir. Ortak yorum sayısının toplam benzersiz yorum sayısına oranı alınarak hesaplanır.

Model 1	Model 2	Jaccard Skoru
TF-IDF Lemmatized	Word2Vec SG	0.67
TF-IDF Lemmatized	TF-IDF Stemmed	0.43
TF-IDF Lemmatized	Word2Vec CBOW	0.43
Word2Vec SG	Word2Vec CBOW	0.25

Yorum:

- TF-IDF Lemmatized ile Word2Vec SG modelleri arasında yüksek benzerlik (%67) bulunmuştur.
- TF-IDF modelleri kendi aralarında tutarlı olsa da, anlamsal yakınlık açısından Word2Vec SG ile örtüşme gösterdiği görülmüştür.
- SG ve CBOW arasındaki düşük benzerlik (0.25), Word2Vec mimarilerinin farklı sonuçlar ürettiğini ortaya koymaktadır.

## 6. GENEL YORUMLAR VE DEĞERLENDİRME

Yapılan analiz, farklı metin vektörleştirme yaklaşımlarının aynı metinler üzerinde nasıl farklı sonuçlar üretebildiğini açıkça göstermiştir. Word2Vec SG modeli, hem anlamsal değerlendirmede hem de sıralama tutarlılığında en yüksek performansı göstermiştir. TF-IDF modelleri hızlı ve basit çözümler sunsa da, derin anlam bağlantılarını yakalamakta yetersiz kalmıştır.

Bu projenin çıktıları; öneri sistemleri, oyun değerlendirme platformları ve kullanıcı analitiği sistemleri gibi alanlarda uygulanabilirliği olan bir temel sunmaktadır. SG mimarisine dayalı Word2Vec modelinin; oyun içeriği analizleri, duygu sınıflandırması ve içerik filtreleme gibi metin madenciliği uygulamalarında tercih edilmesi önerilmektedir.

Sonuç olarak, bu çalışma hem akademik hem de sektörel uygulamalar için güçlü bir NLP tabanlı benzerlik analizi örneği sunmuştur.

**7. KULLANILAN KÜTÜPHANELER:** Bu projede Python tabanlı aşağıdaki kütüphaneler kullanılmıştır:

- pandas: Veri çerçeveleriyle çalışmak, CSV dosyalarını okumak ve yazmak için
- numpy: Vektörlerin matematiksel işlemleri için (ortalama, matris işlemleri)
- scikit-learn: TF-IDF vektörleştirme ve cosine similarity hesaplamaları için
- gensim: Word2Vec model eğitimi ve kelime vektörlerinin çıkarımı için
- fpdf: PDF çıktısı oluşturmak için (rapor PDF'e dönüştürüldüğünde)

Kütüphaneler en güncel sürümleriyle pip üzerinden kurulmuştur.

## **8. SON SÖZ**

Bu proje sürecinde, metin madenciliği ve doğal dil işleme alanında uygulamalı bir benzerlik analizinin tüm adımları başarıyla yürütülmüştür. Hem veri toplama, hem model eğitimi, hem de analiz sonuçlarının değerlendirilmesi titizlikle yapılmış; anlamlı ve tutarlı sonuçlara ulaşılmıştır. Bu deneyim, hem teknik beceri kazandırmış hem de metin analizinin gerçek dünya uygulamalarında nasıl kullanılabileceğine dair pratik bir bakış açısı kazandırmıştır.

Yönetim Bilişim Sistemleri-Yapay Zeka Final Ödevi

2102131020

Oğuzhan Şen