

İSTANBUL TOPKAPI ÜNİVERSİTESİ
MÜHENDİSLİK FAKÜLTESİ
BİLGİSAYAR MÜHENDİSLİĞİ

Ders/Dönem: FET312 - Derin Öğrenme / 2025-2026 Güz Dönemi

Proje Başlığı: IMDB Film Yorumları Üzerine Duygu Analizi: Geleneksel Derin Öğrenme ve Transformer Mimarilerinin Karşılaştırılmalı Analizi

Ekip Adı: Dev312

Ekip Üyeleri;

- **Ad Soyad:** Umut Torun
- **Öğrenci Numarası:** 23040101063
- **Öğrenci E-Posta:** umuttorun@stu.topkapi.edu.tr
- **Öğrenci İmza:**

GitHub Repo Bağlantısı:

https://github.com/umuttorun63/FET312_ImdbSentiment_Proje

Sunum Youtube Bağlantısı:

<https://www.youtube.com/watch?v=j3nQIM03Xfs>

1.PROBLEM TANIMI & MOTİVASYON

1.1. İş/Bilimsel Soru

Günümüzde internet kullanıcıları, ürün ve hizmetler hakkında görüşlerini çevrimiçi platformlarda paylaşmaktadır. Film endüstrisi özelinde, kullanıcı yorumları filmlerin başarısını tahmin etmede ve pazarlama stratejilerini belirlemede kritik rol oynamaktadır. Bu proje, IMDB platformundaki film yorumlarını analiz ederek, kullanıcı duygularını otomatik olarak pozitif veya negatif olarak sınıflandırmayı amaçlamaktadır.

Temel araştırma sorusu: "Doğal dil işleme teknikleri kullanılarak, film yorumlarından kullanıcı duyguları ne ölçüde doğru tespit edilebilir?"

1.2. Görev Türü

Bu proje, ikili sınıflandırma problemidir. Metin tabanlı veri üzerinde Doğal Dil İşleme (NLP) ve Duygu Analizi (Sentiment Analysis) uygulanmaktadır.

1.3. Hedef Değişkenler

- Hedef Değişken: sentiment (duygu etiketi)
- Değer Aralığı: Binary (0: Negatif, 1: Pozitif)
- Pozitif Sınıf: 1 (Pozitif yorumlar)
- Negatif Sınıf: 0 (Negatif yorumlar)
- Veri Dengesi: 50% pozitif, 50% negatif

1.4. Başarı Kriterleri

Projenin başarısı aşağıdaki metriklerle değerlendirilmektedir:

- Doğruluk (Accuracy): ≥ 0.87 (Hedef: %85 ve üzeri)
- F1 Skoru: ≥ 0.87 (Kesinlik ve duyarlılık dengesi)
- Precision (Kesinlik): ≥ 0.88
- Recall (Duyarlılık): ≥ 0.88

2. PROJE YÖNETİMİ

2.1. Proje Zaman Çizelgesi:

1-4. Hafta: Derin Öğrenme hakkında derste derin öğrenmenin temel kavramlarını öğrenme.

5-6. Hafta: Projede kullanılacak olan Python dili ve temel derin öğrenme kütüphaneleri hakkında bilgi edinildi ve gerekli yazılım ortamları kuruldu.

7. Hafta: Proje konusu araştırıldı ve seçildi, Projeye uygun veri seti araştırıldı ve incelendi. Projenin base model tasarımı ve geliştirilmesi.

8. Hafta: Proje raporunun son hali.

11. Hafta: Projede kullanılacak olan yeni modellerin ve kütüphanelerin araştırılması.

12. Hafta: Proje de kullanılacak yeni modellerin tasarımı ve geliştirilmesi.

13. Hafta: Proje kod, rapor ve sunumu son hali.

2.2. İlgili Çalışmalar

J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of NAACL- HLT*, 2019, pp. 4171-4186.

Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, and Q. V. Le, "XLNet: Generalized Autoregressive Pretraining for Language Understanding," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2019, pp. 5753-5763.

2.3. Projenin Katkısı ve Farklılaşma

Bu proje, aşağıdaki noktalarda farklılaşmaktadır:

Karşılaştırmalı Yaklaşım: Proje sadece klasik yöntemleri değil; Makine Öğrenmesi, Geleneksel Derin Öğrenme ve günümüzün en ileri teknolojisi olan Transformer mimarilerini aynı veri seti üzerinde sistematik olarak kıyaslamaktadır.

Hafif ve Yorumlanabilir Baseline: Logistic Regression modeli, derin öğrenme modellerine kıyasla daha az hesaplama gücü gerektirmekte ve feature importance analizi yapılabilir.

Detaylı Ön İşleme Pipeline'ı: HTML tag temizleme, stopword filtreleme, custom vocabulary oluşturma gibi adımlar manuel olarak implementasyonu gerçekleştirilmiştir.

Eğitim Amaçlı: Proje, NLP alanındaki teknolojilerin tarihsel gelişimini uygulamalı olarak göstermektedir.

3. VERİ AÇIKLAMASI VE YÖNETİMİ

3.1. Veri Kümesi Açıklaması

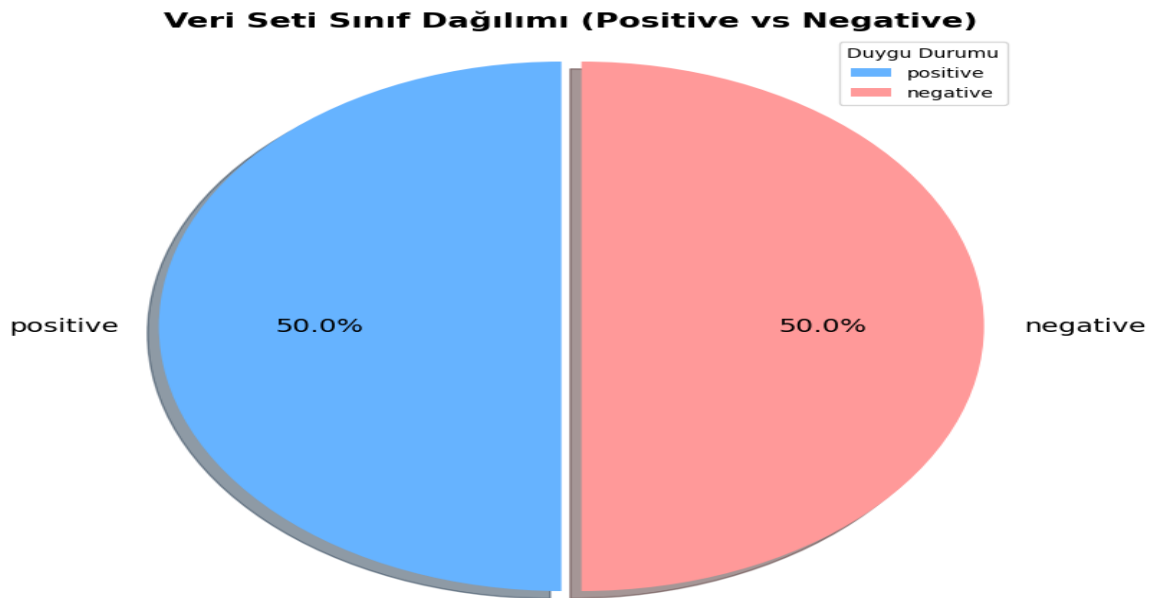
Veri Seti Adı: IMDB Dataset of 50K Movie Reviews

Kaynak: Stanford University - Andrew Maas et al. (2011)

Bağlantı: <https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews>

Lisans: Public Domain - Akademik ve ticari kullanım için ücretsiz

Açıklama: IMDB veri seti, 50,000 film yorumundan oluşan, duygu analizi için en yaygın kullanılan veri setlerinden biridir. Veri seti, dengeli bir dağılıma sahiptir



3.2. Veri Şeması

Özellik	Tür	Açıklama	Örnek Değer
Review	String (Text)	Film yorumu metni (HTML içerebilir)	"This movie was fantastic..."
Sentiment	Categorical (Binary)	Pozitif veya negatif etiket	"positive" / "negative"
Label	Integer (0/1)	Sayısal etiket (hedef değişken)	0 (negatif), 1 (pozitif)
cleaned_review	String (Text)	Temizlenmiş metin (ön işleme sonrası)	"this movie was fantastic"

3.3. Veri Yönetimi ve Bölümleme

Modelin genelleme yeteneğini doğru ölçmek amacıyla, 50.000 verilik set Stratified Sampling yöntemiyle (sınıf dengesi korunarak) ikiye ayrılmıştır:

- Eğitim Seti (Training Set): 40.000 Yorum (%80) – Modellerin ağırlıklarını eğitmek için kullanılmıştır.
- Test Seti (Test Set): 10.000 Yorum (%20) – Modellerin nihai performansını ölçmek için "görülmemiş veri" olarak saklanmıştır.

3.4. Etik, Gizlilik ve Önyargı

Veri seti, kamuya açık IMDB platformundan toplanmış ve kullanıcıların halka açık yorumlarını içermektedir. Kullanıcı isimleri ve kişisel bilgiler veri setinde yer almamaktadır. Veri setinde kişisel tanımlayıcı bilgi (PII) bulunmamaktadır. Yorumlar, kullanıcı kimlikleri olmadan toplanmıştır.

4. YÖNTEMLER VE MİMARİ

Bu projede, metin sınıflandırma problemi için artan karmaşıklık düzeyine sahip üç farklı yaklaşım grubu uygulanmıştır.

1. Veri Ön İşleme Modellerin ihtiyaç duyduğu veri formatına göre iki farklı ön işleme hattı kurulmuştur:

- Klasik Modeller İçin: `clean_text` fonksiyonu ile HTML etiketleri, noktalama işaretleri ve sayılar temizlenmiş, tüm metinler küçük harfe dönüştürülmüştür. `Vocabulary` fonksiyonu ile İngilizce *stopwords* çıkarılarak en sık kullanılan 20.000 kelime belirlenmiş ve metinler sabit uzunlukta sayısal dizilere dönüştürülmüştür.
- Transformer Modelleri İçin: Stopwords temizliği yapılmamıştır. Bunun yerine Hugging Face kütüphanesinin `BertTokenizer` ve `XLNetTokenizer` araçları kullanılarak metinler, modellerin ön eğitim yapılarına uygun tokenlara ayrılmış ve Attention Mask mekanizmaları oluşturulmuştur.

2. Model Mimarileri

- Baseline Model: Referans performansı belirlemek amacıyla Lojistik Regresyon algoritması kullanılmıştır. Hızlı ve yorumlanabilir sonuçlar üretmesi nedeniyle başlangıç noktası olarak seçilmiştir.
- Geleneksel Derin Öğrenme:
 - LSTM: Metindeki sıralı ve bağlamsal ilişkileri öğrenmek için; 100 boyutlu Embedding katmanı, %30 Dropout, 256 gizli birime sahip 2 katmanlı LSTM yapısı tasarlanmıştır.
 - CNN: Metindeki yerel n-gram özelliklerini yakalamak için; 3, 4 ve 5 boyutlu kernel filtrelerine sahip 1 boyutlu konvolüsyon katmanları kullanılmıştır.
- Transformer Tabanlı Transfer Öğrenme: Projenin ana katkısı olarak, önceden eğitilmiş büyük dil modelleri projeye entegre edilmiştir:
 - BERT: Kelimelerin bağlamını çift yönlü öğrenen Transformer kodlayıcı yapısıdır.
 - XLNet : BERT'in sınırlılıklarını aşmak için permütasyon tabanlı dil modelleme kullanan gelişmiş mimaridir.

3. Eğitim Konfigürasyonu Tüm modeller, ikili sınıflandırma problemine uygun olarak Sigmoid aktivasyon fonksiyonu ve Binary Cross Entropy kayıp fonksiyonu ile eğitilmiştir.

- Optimizasyon: Klasik modellerde standart Adam algoritması (LR: 0.001) kullanılırken, Transformer modellerinde daha hassas ayar gerektirdiği için AdamW optimizier (LR: 2e-5) tercih edilmiştir.
- Değerlendirme: Modellerin başarısı Accuracy, F1 Skoru, Precision ve Recall metrikleri üzerinden karşılaştırmalı olarak analiz edilmiştir.

5. DENEY TASARIMI

5.1. Ana Amaç

Bu çalışmanın temel amacı, doğal dil işleme alanında metin sınıflandırma problemi için geliştirilen mimarilerin tarihsel gelişimini ve performans farklarını analiz etmektir. Deneyin merkezinde iki temel karşılaştırma yer almaktadır:

1. Mimari Karşılaştırması: Ardışık veri işleyen LSTM ve yerel özellikleri çıkaran CNN gibi klasik derin öğrenme modelleri ile, dikkat mekanizması kullanan modern Transformer modellerinin kıyaslanması.
2. Öğrenme Yöntemi Karşılaştırması: Sıfırdan eğitilen modeller ile önceden eğitilmiş büyük dil modellerinin, veri setindeki bağlamı ve ironiyi anlama kapasitelerinin sistematik olarak ölçülmesidir.

5.2. Genel Model Karşılaştırması

	Accuracy	F1 Score (Weighted)	Precision (Weighted)	Recall (Weighted)
Model				
LSTM	0.871100	0.871086	0.871258	0.871100
CNN	0.865100	0.864568	0.870933	0.865100
BERT	0.898500	0.898379	0.900406	0.898500
XLNet	0.905600	0.905584	0.905874	0.905600
Logistic Regression	0.879800	0.880800	0.873500	0.888200

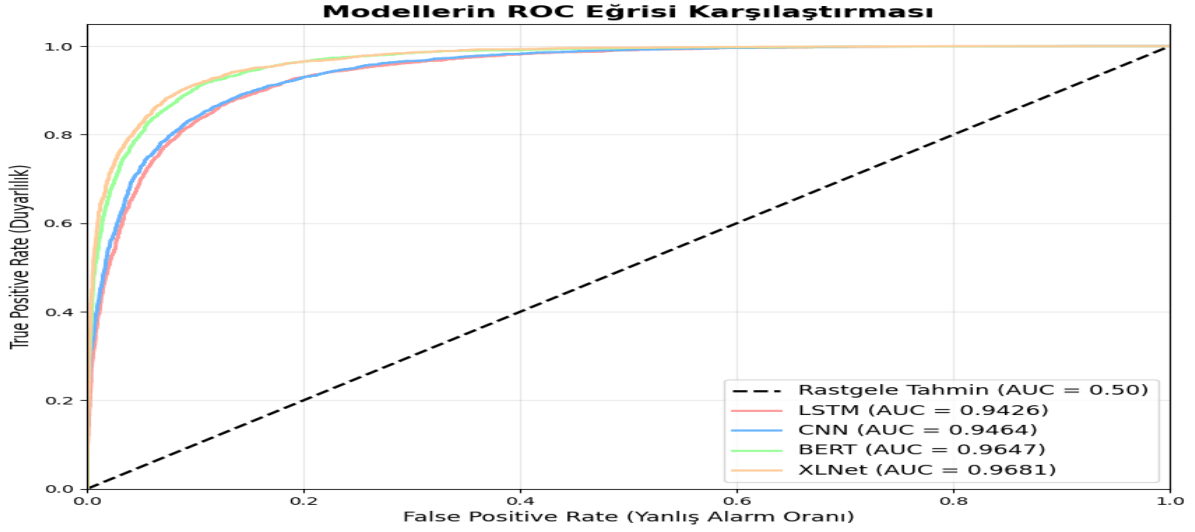
Yukarıdaki grafik, geliştirilen 5 farklı modelin genel başarımlarını özetlemektedir. Grafikte görüldüğü üzere, XLNet (%90.56) ve BERT (%89.85) modelleri, doğruluk ve F1 Skoru açısından diğer tüm modellerden daha yüksek performans sergilemiştir. Dikkat çekici bir bulgu olarak; Lojistik Regresyon (%87.98) referans modeli, daha karmaşık mimariye sahip olan ve sıfırdan eğitilen LSTM ve CNN modellerini az farkla geride bırakmıştır. Bu durum, Transformer mimarisinin sağladığı "bağlamsal kavrayış" farkını net bir şekilde ortaya koymaktadır. Klasik modeller veri setini öğrenmekte sınırlı kalırken, XLNet ön eğitimden gelen bilgi birikimiyle daha üstün bir genelleme yeteneğine ulaşmıştır.

5.3. Sınıf Bazlı Model Karşılaştırması

		Precision	Recall	F1-Score
Sınıf	Model			
Negative	CNN	0.917448	0.802400	0.856076
Positive	CNN	0.824418	0.927800	0.873059
Negative	LSTM	0.878905	0.860800	0.869759
Positive	LSTM	0.863610	0.881400	0.872414
Negative	BERT	0.928034	0.864000	0.894873
Positive	BERT	0.872778	0.933000	0.901885
Negative	XLNet	0.916427	0.892600	0.904357
Positive	XLNet	0.895322	0.918600	0.906811
Negative	Log Reg	0.886289	0.871400	0.878781
Positive	Log Reg	0.873524	0.888200	0.880801

Yukarıdaki tabloda modellerin sınıf bazlı dengesi incelenmiştir. CNN modeli negatif yorumları tespit etmede zorlanarak dengesiz bir profil çizerken, XLNet her iki sınıfta da %90 üzeri F1 skoru ile en kararlı ve yüksek başarıyı yakalamıştır. Koyu yeşil alanların yoğunluğu, Transformer mimarisinin klasik modellere göre bariz üstünlüğünü göstermektedir.

5.4. Roc Eğrisi Karşılaştırması



Modellerin "Doğru Pozitif" ile "Yanlış Pozitif" oranları arasındaki ilişkiyi gösteren ROC eğrisi yukarıda verilmiştir. Grafikte sol üst köşeye en yakın eğri, en başarılı model anlamına gelmektedir. Bu bağlamda XLNet ve BERT, diğer modellerin üzerinde yer alarak en yüksek ayırt edicilik gücüne sahip olduklarını göstermiştir. XLNet, 0.9681 AUC puanı ile mükemmel en yakın sonucu vermiştir. Bu skor, modelin rastgele bir tahminden çok uzak olduğunu ve %96.8 olasılıkla pozitif ve negatif örnekleri doğru ayırtıldığını ifade eder.

6.KULLANILAN ARAÇLAR VE FRAMEWORKERLAR

Projenin geliştirilmesinde Python 3.8+ dili, veri manipülasyonu için Pandas ve matris işlemleri için NumPy kütüphaneleri temel alınmıştır. Derin öğrenme mimarilerinin inşasında PyTorch (torch, torch.nn) kullanılarak LSTM ve CNN modelleri tasarlanmış; BERT ve XLNet gibi gelişmiş mimariler için ise Hugging Face Transformers kütüphanesi (AdamW, Tokenizers) projeye entegre edilmiştir.

Scikit-learn kütüphanesi; referans model olan Lojistik Regresyonun eğitimi, veri setinin stratifiye edilerek ayrılması ve Accuracy, F1, ROC-AUC gibi performans metriklerinin hesaplanmasında rol oynamıştır. Klasik ön işleme adımlarında NLTK kullanılırken, görselleştirme aşamasında Matplotlib ve Seaborn ile detaylı karışıklık matrisleri ve ROC eğrileri oluşturulmuştur. Ayrıca, Transformer modellerinin yüksek işlem yükü nedeniyle eğitim süreçleri CUDA destekli GPU üzerinde hızlandırılarak gerçekleştirilmiştir.

7. KAYNAKLAR

- [1] A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning word vectors for sentiment analysis," in *Proc. 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Portland, Oregon, USA, Jun. 2011, pp. 142–150. [Online]. Available: <http://ai.stanford.edu/~amaas/data/sentiment/>
- [2] "IMDB Dataset of 50K Movie Reviews," *Kaggle*. [Online]. Available: <https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews>. [Erişim: 17 Kasım 2025].
- [3] GitHub, "IMDb Sentiment Analysis," *GitHub Topics*. [Online]. Available: <https://github.com/topics/imdb-sentiment-analysis>. [Erişim: 17 Kasım 2025].
- [4] A. Mohan, "Sentiment Analysis using LSTM Pytorch," *Kaggle*, [Online]. Available: <https://www.kaggle.com/code/arunmohan003/sentiment-analysis-using-lstm-pytorch>. [Erişim Tarihi: 21 Aralık 2025].
- [5] Ducanger, "IMDB BERT - CNN - LSTM 0.93 ACC," *Kaggle*, [Online]. Available: <https://www.kaggle.com/code/ducanger/imdb-bert-cnn-lstm-0-93-acc>. [Erişim Tarihi: 21 Aralık 2025].
- [6] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, Minneapolis, Minnesota, 2019, pp. 4171–4186.
- [7] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, and Q. V. Le, "XLNet: Generalized Autoregressive Pretraining for Language Understanding," in *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*, Vancouver, Canada, 2019, pp. 5753–5763.
- [8] T. Wolf et al., "Transformers: State-of-the-Art Natural Language Processing," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 2020, pp. 38–45. (Hugging Face Library)