

İSTANBUL TOPKAPI ÜNİVERSİTESİ

MÜHENDİSLİK FAKÜLTESİ

BİLGİSAYAR MÜHENDİSLİĞİ

Proje Başlığı: IMDB Film Yorumları Üzerine Duygu Analizi: Makine Öğrenmesi ve
Derin Öğrenme Yaklaşımlarının Karşılaştırılmalı Analizi

Ad Soyad: Umut Torun

Öğrenci Numara: 23040101063

Öğrenci E-Posta: umuttorun@stu.topkapi.edu.tr

GitHub: https://github.com/umuttorun63/FET312_ImdbSentiment_Proje

Problem Tanımı

Günümüzde internet kullanıcıları, ürün ve hizmetler hakkında görüşlerini çevrimiçi platformlarda paylaşmaktadır. Film endüstrisi özelinde, kullanıcı yorumları filmlerin başarısını tahmin etmede ve pazarlama stratejilerini belirlemeye 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?"

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

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

Başarı Kriterleri

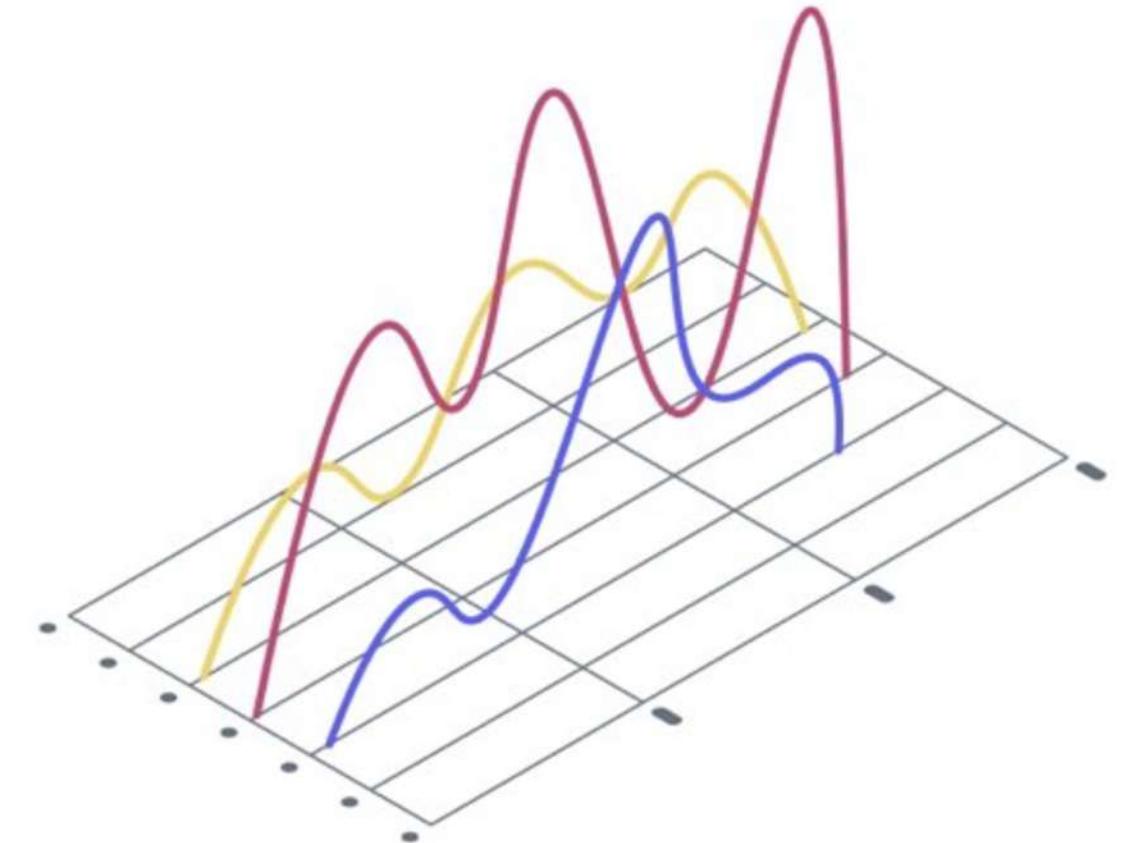
Projenin başarısı aşağıdaki metriklerle değerlendirilmektedir

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



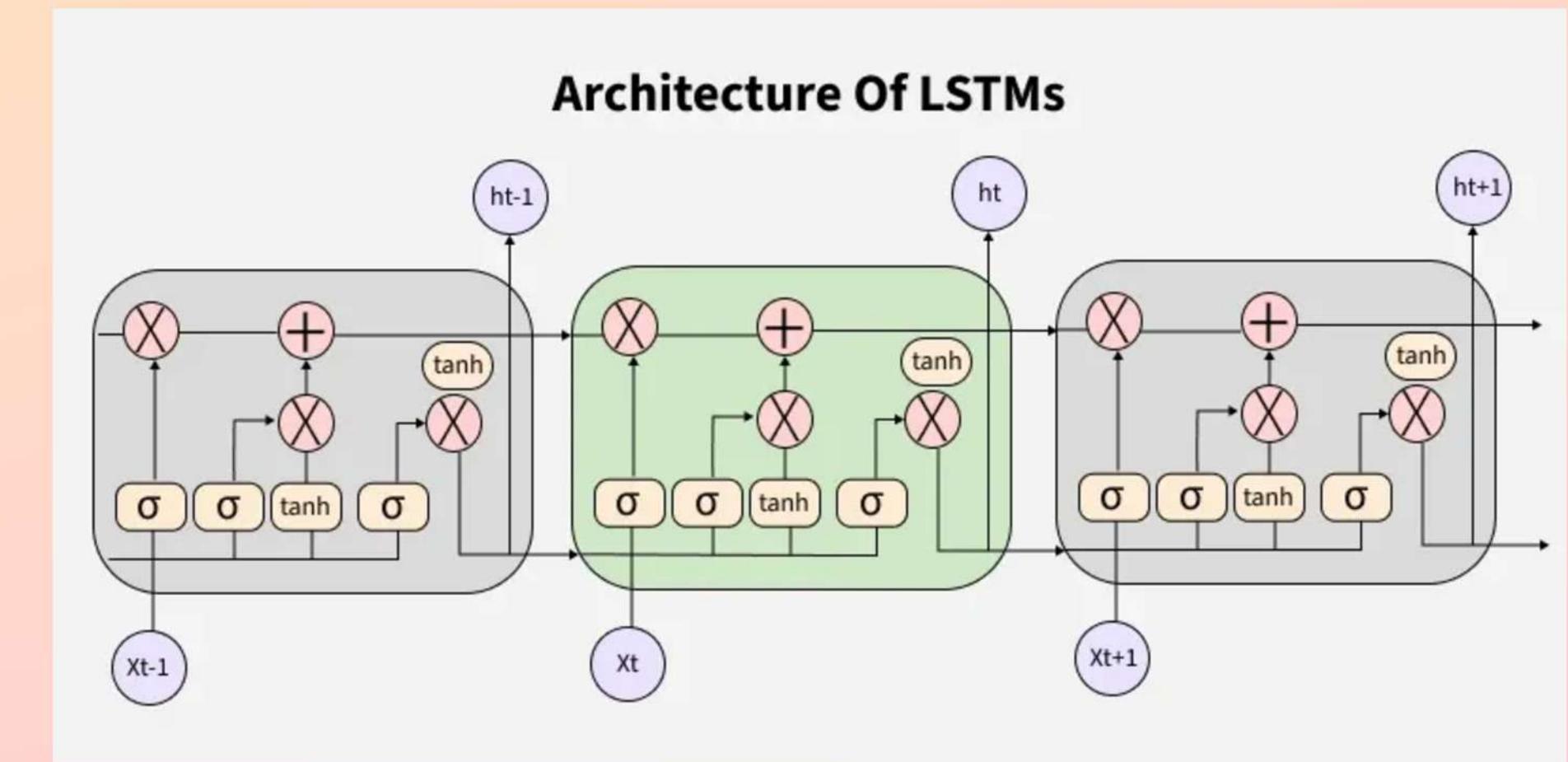
LR (Lojistik Regresyon)

- Hızlı eğitilebilir model
- Düşük donanım (CPU) kaynağı ile çalışabilir
- Daha az veri setleri için uygun
- Sonuçları yorumlanabilir ve şeffaf
- Seyrek verilerde yüksek performans gösterir



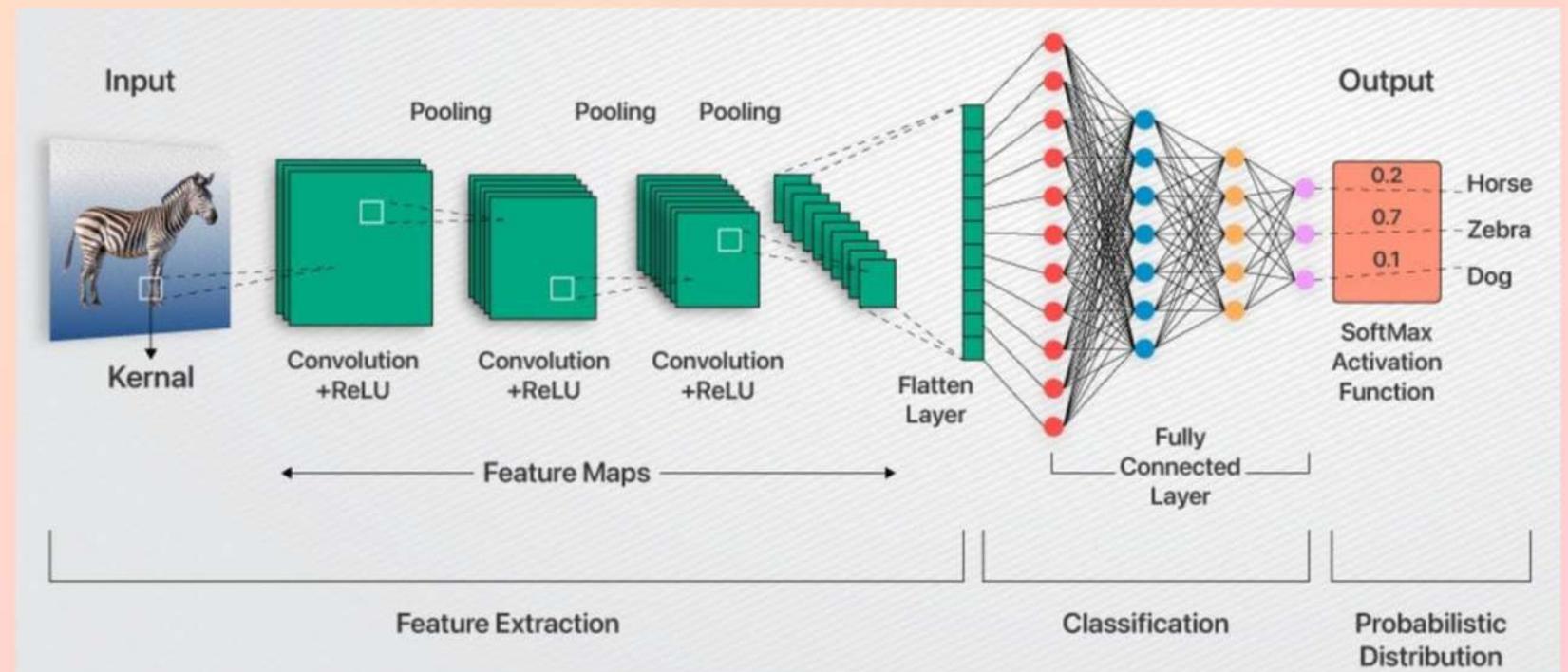
LSTM (Long Short-Term Memory)

- Yavaş eğitilebilir model
- Önceki kelimededen sonraki kelimeyi tahmin etmede başarılıır
- Karmaşık dil yapılarını ve zıtlıklarını çözebilir
- Kelime sırasını ve bağlamı öğrenir
- Büyük veri setlerinde performansı artar



CNN (Convolutional Neural Network)

- LSTM modeline kıyasla daha hızlı eğitilir
- Önemli kelime gruplarını tespit etmede başarılıdır
- Pararel işlem (Hesaplama) yeteneği yüksektir
- Kelimenin cümledeki yerinden bağımsız öğrenir
- Otomatik öznitelik çıkarımı yapar



Model Karşılaştırılması



LR: %87.98



LSTM: %87.98

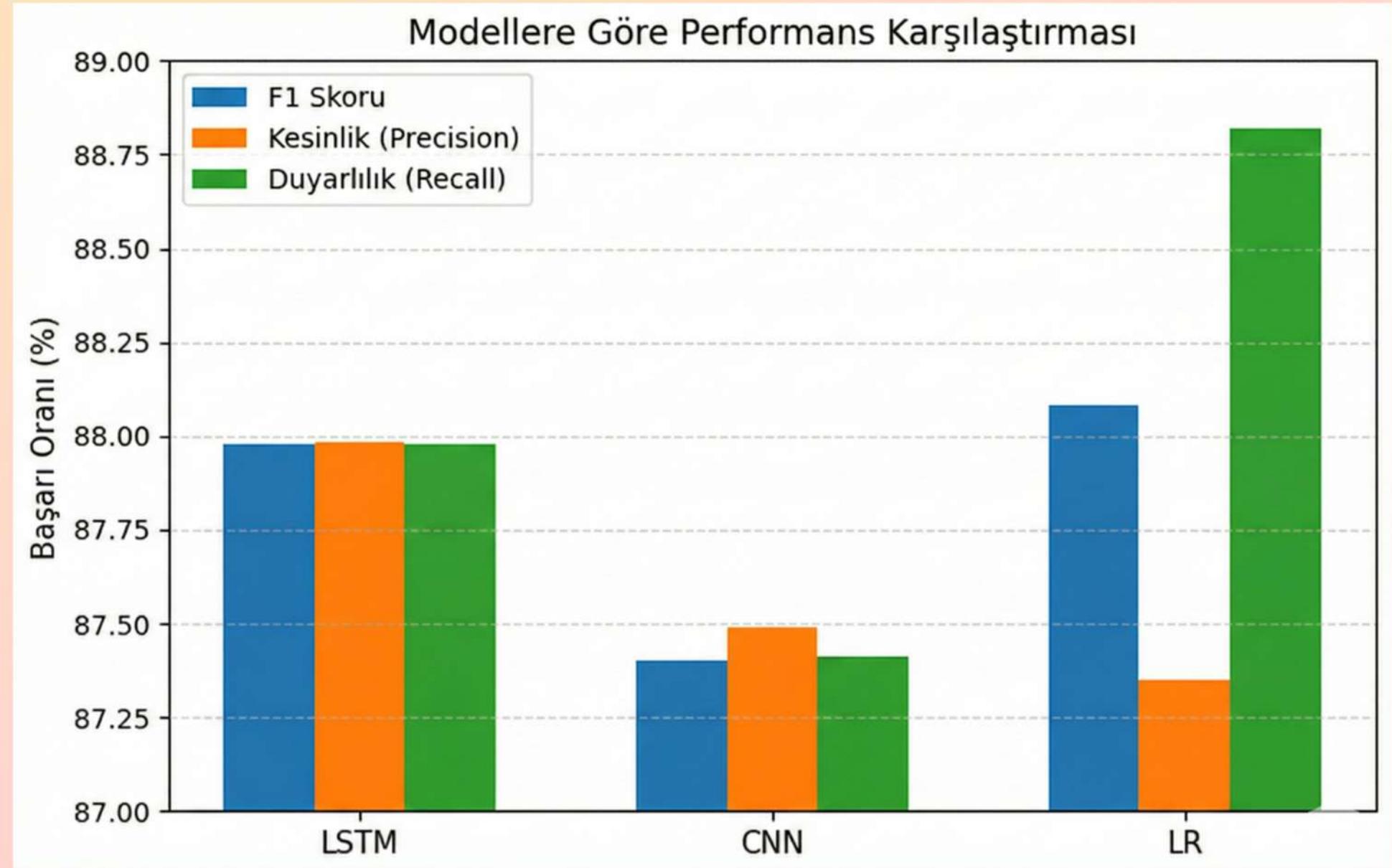


CNN: %87.41

MODEL	ACCURACY	F1 SCORE	PRECISION	RECALL
LSTM	%87.9800	%87.9799	%87.9814	%87.9800
CNN	%87.4100	%87.4035	%87.4873	%87.4100
LR	%87.9800	%88.0800	%87.3500	%88.8200

Model Karşılaştırılması

(Sütun Grafiği)

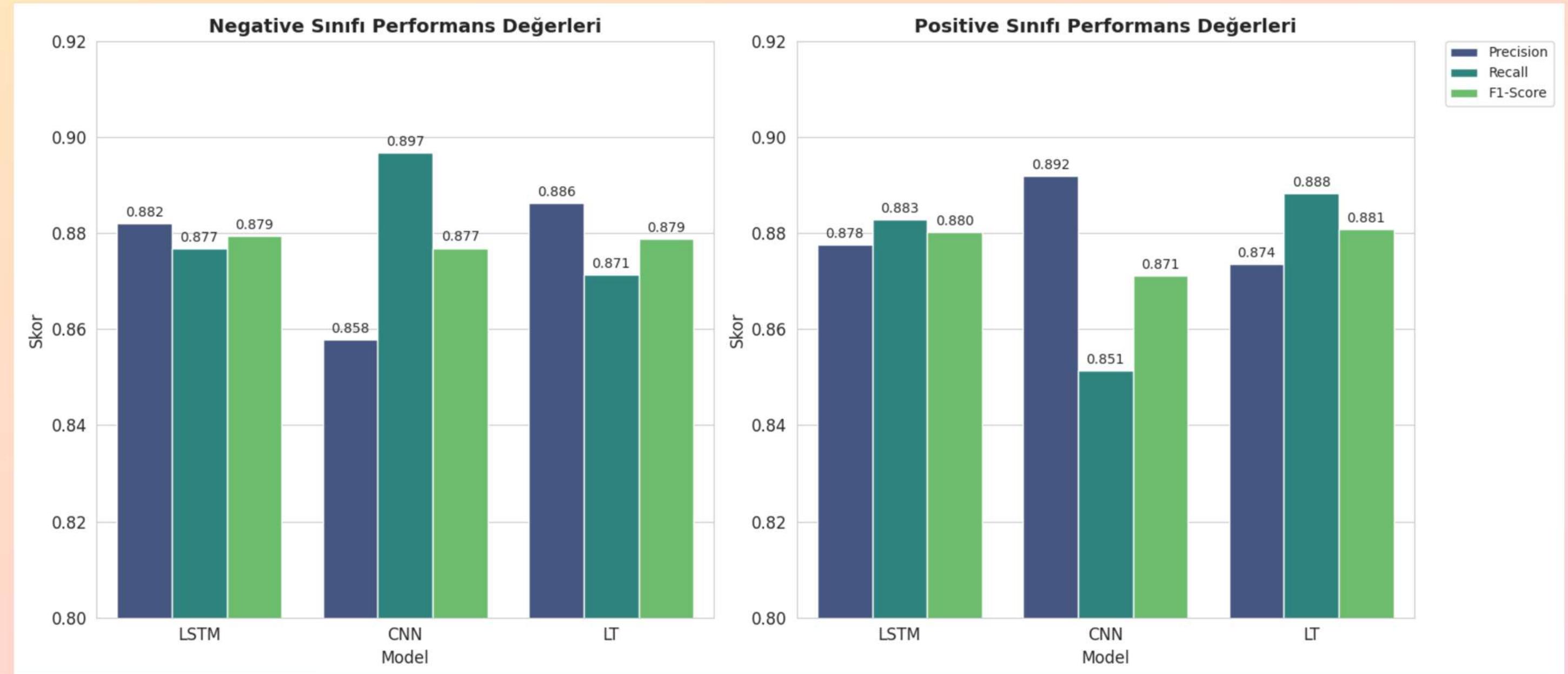


Sınıf Bazlı Model Karşılaştırılması

SINIF	MODEL	PRECISION	RECALL	F1-SCORE
NEGATIVE	LSTM	%88.2093	%87.6800	%87.9438
	CNN	%85.7853	%89.6800	%87.6894
	LG	%88.6289	%87.1400	%87.8781
POZİTİVE	LSTM	%87.7535	%88.2800	%88.0160
	CNN	%89.1892	%85.1400	%87.1176
	LG	%87.3524	%88.8200	%88.0801

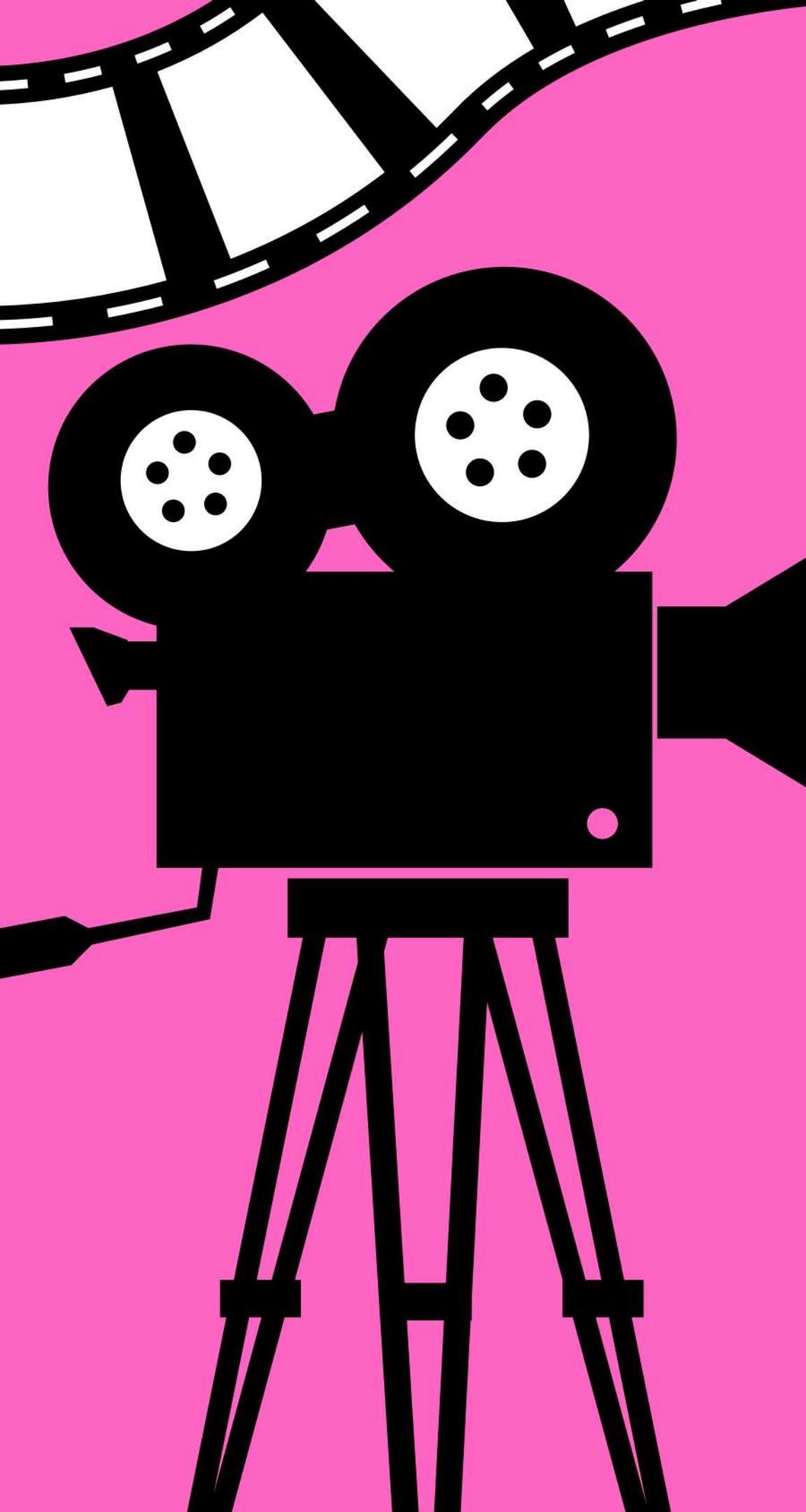
Sınıf Bazlı Model Karşılaştırması

(Sütun Grafiği)



Sonuç ve Değerlendirme

Bu proje, metin sınıflandırımda geleneksel istatistiksel yöntemler ile modern derin öğrenme mimarilerini LSTM ve CNN doğrudan kıyaslayarak, doğru veri temizliği yapıldığında basit bir algoritmanın bile karmaşık sinir ağları kadar başarılı olabileceğini somut verilerle ortaya koyması açısından değerlidir. yüksek doğruluk için her zaman yüksek işlem gücüne ihtiyaç duyulmadığını kanıtlamaktadır. Lojistik Regresyonun derin öğrenme modelleriyle yarışan performansı, kaynak kısıtı olan sistemlerde verimli yapay zeka çözümleri geliştirilebileceğini göstermektedir.



Teşekkürler!