Sentiment Analysis on IMDB Movie Reviews

This project aims to perform sentiment analysis on IMDB movie reviews using both Machine Learning (ML) and Deep Learning (DL) techniques. The dataset consists of labeled reviews (positive/negative). The goal is to classify reviews into positive or negative sentiments.

Dataset: IMDB Dataset of 50,000 movie reviews, equally balanced between positive and negative sentiments. The dataset was split into training (80%) and testing (20%) sets for model evaluation.

Methodology: 1. Data Preprocessing: Cleaned dataset, converted sentiment labels to numeric (positive=1, negative=0). 2. Machine Learning Models: - Logistic Regression (TF-IDF features) - Naive Bayes (TF-IDF features) 3. Deep Learning Model: - LSTM (Long Short-Term Memory) with Embedding layer for sequence modeling.

Results: - Logistic Regression: ~86% Accuracy - Naive Bayes: ~83% Accuracy - LSTM Model: ~87% Accuracy The LSTM model slightly outperformed traditional ML models due to its ability to capture sequential dependencies in text.

Model	Accuracy
Logistic Regression	86%
Naive Bayes	83%
LSTM	87%

Conclusion: This project demonstrated the use of ML and DL approaches for sentiment analysis. Logistic Regression and Naive Bayes proved effective baselines, while LSTM achieved the best performance. This shows the importance of deep learning models for handling sequential data like text. Future work can include experimenting with pre-trained embeddings (e.g., GloVe, Word2Vec) or transformer models for improved performance.