IMDb Review Sentiment Analysis Using Deep Learning and Gradio

Developed By: Balkrishan Singh  
Roll No: 2302492  
B.Tech CSE, Guru Nanak Dev Engineering College, Ludhiana

# Abstract

This project presents an intelligent web-based system to predict the sentiment of movie reviews from IMDb. The application utilizes the classic IMDb Movie Reviews Dataset and implements a hybrid deep learning model combining a Convolutional Neural Network (CNN) with a Bidirectional Long Short-Term Memory (LSTM) network for classification. An interactive and user-friendly frontend is created using Gradio, allowing users to input an IMDb review URL. The system scrapes the review in real-time and provides a prediction indicating whether the sentiment is 'Positive' or 'Negative', along with a confidence score.

# Objective

To develop a deep learning-based tool that accurately classifies the sentiment of movie reviews and to provide an accessible web interface that allows for real-time analysis, demonstrating a practical application of NLP and web technologies.

# Technologies Used

• Programming Language: Python

• Libraries: TensorFlow (Keras), scikit-learn, numpy, pandas, Gradio, Requests, BeautifulSoup

• ML Model:Hybrid Conv1D + Bidirectional LSTM Network

• Frontend Styling:Gradio Web Interface

# Dataset

• Source: IMDb Movie Reviews Dataset

• Features: Text content of 50,000 movie reviews.

• Target:Sentiment (0 = Negative, 1 = Positive)

• Size: 50,000 records (25,000 negative, 25,000 positive)

# Methodology

**Data Handling:** Data was loaded and split into training and testing sets using a stratified split to maintain a balanced distribution of sentiments.

**Preprocessing:** Text reviews were cleaned, then converted into sequences of integers using the Keras Tokenizer, and finally padded to a uniform length.

**Model Architecture:** A hybrid Sequential model was built, featuring an Embedding layer, a Conv1D layer to capture local phrase patterns, and a Bidirectional LSTM layer to understand the sequence context.

**Model Training:** The model was trained using the Adam optimizer and binary cross-entropy loss. The EarlyStopping callback was used to prevent overfitting and restore the best model weights.

**Web Frontend:** A web UI was built using Gradio, containing an input for a URL and an output for the prediction.

**Integration:** A predictive function was created to scrape review content from a given IMDb URL, preprocess the text using the trained tokenizer, and feed it to the model to display the final sentiment prediction.

# Features of the Application

**Interactive UI:** A clean and simple interface built with Gradio.

**Real-Time Prediction:** Scrapes and analyzes reviews directly from a live IMDb URL.

**Confidence Score:** Displays the model's prediction confidence alongside the sentiment.

**Advanced Model:** Uses a powerful hybrid deep learning model for high accuracy on nuanced text.

**Robust Training:** The model training process is optimized with EarlyStopping to ensure generalization.

# Model Performance

Training Accuracy: ~100%  
Testing Accuracy: ~90%

# Future Scope

•Deploy the Gradio application to a permanent cloud host like Hugging Face Spaces.

•Experiment with more advanced Transformer-based models (e.g., fine-tuning DistilBERT) to further improve accuracy.

• Allow for batch predictions by enabling file uploads with multiple URLs.

•Enhance the UI to highlight specific words or phrases that most influenced the model's prediction.

# Conclusion

This project successfully demonstrates the power of combining modern deep learning techniques with web technologies to create a practical, real-world application. The hybrid CNN-LSTM model proved effective at capturing the complex patterns in text data, and the Gradio framework allowed for the rapid development of an intuitive and interactive user interface for testing the model.