# Performance Analysis of different Neural Network Models for Sentiment Analysis On Movie Reviews

**Team members:** Insha Javed (21K-3279)
Muhammad Tahir (21K-4503)
Muhammad Samamah (21K-3205)

# Introduction to Sentiment Analysis

### Introduction to Sentiment Analysis

Understanding Sentiment Analysis: A process that determines the sentiment expressed in textual data, categorizing it as positive or negative.

Importance in Movie Reviews: Analyzing sentiments in movie reviews helps gauge audience reactions and preferences, influencing marketing and production decisions.

Neural Network Models in Focus: Evaluation of various neural network architectures, including CNN, LSTM, and hybrid models, to assess their effectiveness in sentiment classification tasks.

Dataset Utilization: The analysis utilizes the Large Movie Review Dataset v1.0, comprising 50,000 labeled reviews, ensuring a balanced representation of sentiments with 25,000 positive and 25,000 negative reviews.

# Dataset Description

### Overview of the Dataset

The dataset consists of 50,000 labeled movie reviews, specifically designed for sentiment analysis tasks.

### Data Split

Reviews are divided into:

Training Set: 25,000 reviews

Testing Set: 25,000 reviews

### Sentiment Distribution

The dataset is balanced with:

Positive Reviews: 25,000 (scores of 7 or higher out of 10)

Negative Reviews: 25,000 (scores of 4 or lower out of 10)

### Unlabeled Data for Unsupervised Learning

In addition to the labeled reviews, there are 50,000 unlabeled reviews available for exploring unsupervised learning techniques.

# Project Objectives

### Evaluate Neural Network Models

Assess the performance of CNN, LSTM, CNN-LSTM, and LDA in sentiment analysis tasks.

### Performance Metrics Analysis

Measure accuracy, precision, and recall to determine the effectiveness of each model.

### Identify Optimal Model

Highlight the most effective model for binary sentiment analysis and discuss the strengths and limitations of each approach.

# Methodology Overview

## 01

### Data Preprocessing

Begin with the preparation of text data through tokenization, stop-word removal, and stemming to ensure the data is clean and suitable for model training.

## 02

### Model Training

Train various neural network models, including CNN, LSTM, CNN-LSTM, and LDA, using a labeled dataset consisting of 50,000 movie reviews, split evenly into 25,000 for training and 25,000 for testing.

## 03

### Performance Evaluation

Calculate performance metrics such as accuracy, precision, and recall using the testing set to assess the effectiveness of each model in sentiment classification tasks.

## 04

### Unsupervised Learning Exploration

Analyze an additional 50,000 unlabeled movie reviews to evaluate the capabilities of the models in handling unsupervised tasks, providing insights into their performance beyond labeled data.

# Model Training Process

## 01

### Data Preprocessing

Begin by preparing the text data through tokenization, stop-word removal, and stemming to ensure it is suitable for training the neural network models.

## 02

### Model Training

Train the selected neural network models—CNN, LSTM, CNN-LSTM, and LDA—using the labeled training dataset, which consists of 25,000 positive and 25,000 negative movie reviews.

## 03

### Performance Evaluation

After training, evaluate the models using the testing set of 25,000 reviews, calculating performance metrics such as accuracy, precision, and recall to determine the effectiveness of each model.

# Performance Metrics

### Performance Metrics

Model Accuracy

CNN: 0.86

LSTM: 0.84

CNN-LSTM: 0.87

LDA: 0.51

### Evaluation Criteria

Metrics used: Accuracy, Precision, Recall

Focus on binary sentiment classification effectiveness

### Model Comparison

Hybrid CNN-LSTM outperforms standalone models

LDA shows significantly lower performance, indicating lesser suitability

### Insights and Implications

High effectiveness of CNN and CNN-LSTM models in sentiment analysis

Importance of model selection based on performance metrics for optimal results

# Model Comparison

## Hybrid CNN-LSTM Model

- Achieved the highest accuracy score of 0.87
- Demonstrating superior performance in sentiment analysis tasks by effectively integrating both convolutional and recurrent neural network techniques.

## Linear Discriminant Analysis (LDA)

- Scored significantly lower at 0.51
- Indicating its lesser suitability for sentiment analysis compared to the neural network models, highlighting the limitations of traditional statistical methods in this context.

# Results and Analysis

## 87%

### Accuracy

The hybrid CNN-LSTM model achieved the highest accuracy of 87%, demonstrating its effectiveness in sentiment analysis tasks.

## 86%

### Accuracy

The CNN model followed closely with an accuracy of 86%, indicating strong performance in classifying sentiments in movie reviews.

## 84%

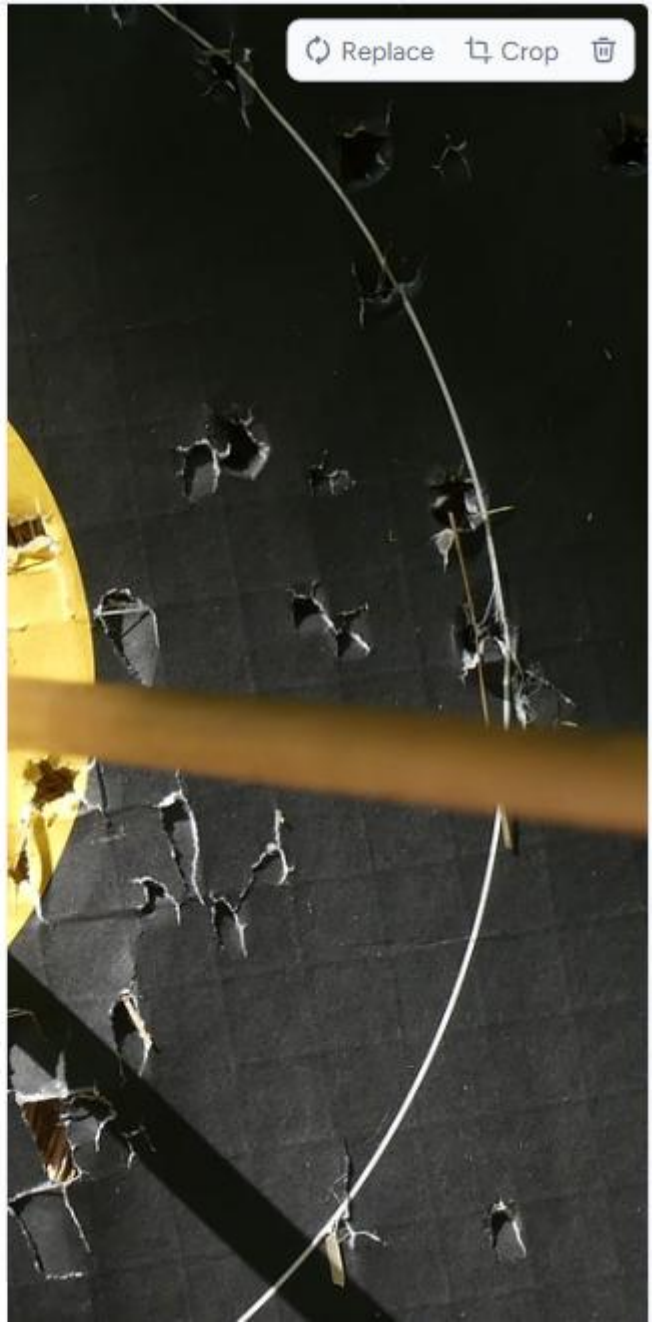
### Accuracy

The LSTM model recorded an accuracy of 84%, showcasing its capability in handling sequential data for sentiment analysis.

## 51%

### Accuracy

In contrast, the Linear Discriminant Analysis (LDA) model scored only 51%, highlighting its lesser suitability for this specific task compared to the neural network models.



# Strengths and Limitations

## S Strengths

- The hybrid CNN-LSTM model achieved the highest accuracy of 0.87, demonstrating superior performance in sentiment analysis tasks compared to other models.

## W Weaknesses

- Linear Discriminant Analysis (LDA) scored significantly lower at 0.51, indicating its limited effectiveness for sentiment classification in this context.

## O Opportunities

- The availability of 50,000 unlabeled reviews presents an opportunity for further exploration of unsupervised learning techniques to enhance model performance.

## T Threats

- The diverse and challenging nature of movie review data may introduce variability that could affect model accuracy and generalizability in real-world applications.

### Exploration of Unlabeled Data

Utilization of 50,000 unlabeled movie reviews to assess model capabilities in unsupervised tasks.

### Model Performance Comparison

Evaluation of various neural network models, with the hybrid CNN-LSTM model demonstrating superior effectiveness in sentiment analysis.

### Key Findings

Significant performance disparity noted, with LDA scoring only 0.51, while CNN and LSTM models achieved scores of 0.86 and 0.84, respectively.

# Unsupervised Learning Insights

# Conclusion and Future Work

### Model Performance Insights

The hybrid CNN-LSTM model demonstrated the highest effectiveness with an accuracy score of 0.87, outperforming standalone models.

### LDA Limitations

Linear Discriminant Analysis (LDA) scored significantly lower at 0.51, indicating its lesser suitability for sentiment analysis tasks compared to other models.

### Unsupervised Learning Potential

The exploration of unsupervised learning techniques using the unlabeled dataset presents opportunities for enhancing model capabilities in sentiment analysis.

### Future Research Directions

Future work could focus on integrating additional neural network architectures and exploring advanced preprocessing techniques to further improve sentiment classification accuracy.