Movie Review Sentiment Analysis Report

Name: Abeer Fatima

Domain: AI/ML

Project: Movie Review Sentiment Analysis

1. Approach Used

This project aimed to classify movie reviews as either positive or negative using machine learning techniques. The following steps were followed:

Data Preprocessing

The dataset underwent various preprocessing steps to clean and standardize the text data:
• Removal of stopwords, punctuation, and special characters.
• Tokenization and stemming for text normalization.
• Text conversion into numerical format using vectorization techniques.

Model Training

Two machine learning models were implemented and trained on the processed dataset:
• Logistic Regression: Used for binary classification of reviews.
• Naive Bayes: Applied for probabilistic text classification.

Testing & Evaluation

The trained models were tested using unseen data and evaluated using performance metrics. Additionally, sample reviews were used for manual verification.

2. Challenges Faced

• Text Preprocessing Complexity: Handling noisy and unstructured text data.
• Feature Representation: Choosing the best vectorization method for improved accuracy.
• Model Comparison: Finding a balance between model efficiency and accuracy.

3. Model Performance & Improvements

Performance Metrics

• Naive Bayes showed strong performance in classifying text but struggled with complex sentences.
• Logistic Regression provided competitive accuracy with more balanced predictions.

Potential Improvements

• Implementing deep learning techniques such as LSTMs or Transformers for better sentiment analysis.
• Utilizing advanced NLP techniques like Word2Vec or GloVe for better feature representation.

Conclusion

The project successfully classified movie reviews using machine learning models. While Naive Bayes provided a strong baseline, further improvements can be made using deep learning and advanced NLP techniques to enhance accuracy and generalization.