Movie Review Sentiment Analysis

1. Introduction

Sentiment analysis is a Natural Language Processing (NLP) technique used to determine the sentiment expressed in a piece of text. This project focuses on analyzing movie reviews to classify them as positive or negative using machine learning techniques.

2. Objectives

• To analyze sentiment in movie reviews.

• To apply machine learning models for sentiment classification.

• To evaluate model performance for accurate sentiment detection.

3. Dataset Description

The dataset used in this project contains:

• Movie reviews as text data.

• Labels indicating whether the review is "Positive" or "Negative".

• Preprocessed textual features for analysis.

4. Methodology

4.1 Data Preprocessing

• Removal of stopwords and punctuation.

• Tokenization and stemming.

• Text vectorization using TF-IDF.

4.2 Machine Learning Models Applied

• Logistic Regression: A baseline classifier for sentiment analysis.

• Naive Bayes: Effective for text classification tasks.

5. Results and Analysis

• Performance evaluation using accuracy, precision, recall, and F1-score.
• Comparison of the two models to determine the best-performing one.
• Visualization of sentiment classification results.

6. Conclusion

The project successfully classified movie reviews as positive or negative using Logistic Regression and Naïve Bayes. The best-performing model can be used for real-world applications such as movie recommendation systems and user feedback analysis.

7. Future Work

- Implementing real-time sentiment analysis.
- Enhancing the model using larger datasets and fine-tuned deep learning models.
- Exploring transformer-based NLP models like BERT for improved accuracy.

Prepared by: Abeer Fatima
Domain: AI/ML

DHC-3603