Sentiment Analysis on Movie Reviews

Introduction

This project focuses on performing sentiment analysis on movie reviews using Natural Language Processing (NLP). We use the NLTK movie reviews dataset to classify reviews as positive or negative.

Objective

The main objective is to build a machine learning model that can classify movie reviews into positive or negative categories with good accuracy.

Methodology

We use Python and the NLTK library. The dataset is the built-in 'movie_reviews' corpus from NLTK. Text preprocessing is applied, and Naive Bayes is used as the classifier. Accuracy and most informative features are evaluated.

Code and Implementation Details

The implementation includes loading the dataset, extracting features, splitting into training and testing sets, and training a Naive Bayes classifier. The code is provided in sentiment_analysis.py.

Results and Observations

The model achieves around 70-80% accuracy depending on the random shuffle. The classifier highlights the most informative words that indicate sentiment polarity.

Conclusion

This project demonstrates how NLP and machine learning can be combined to analyze text sentiment. It can be extended to larger datasets and advanced models for higher accuracy.