

Practical No:8

Title:

Implement a movie reviews sentiment classifier

Aim:

To implement a **sentiment classifier** for movie reviews using natural language processing techniques and a machine learning model (e.g., **Logistic Regression**) to classify reviews as either **positive** or **negative**.

Pre-requisites:

1. Basic understanding of **sentiment analysis** and its applications.
 2. Knowledge of text preprocessing techniques like tokenization, removing stopwords, and feature extraction (TF-IDF, Bag of Words).
 3. Familiarity with machine learning models such as **Logistic Regression**, **Naive Bayes**, or **SVM**.
 4. Python libraries such as **scikit-learn** for machine learning and **nltk** for text processing.
-

Theory:

Sentiment Analysis:

Sentiment Analysis is the process of determining the sentiment (positive, negative, or neutral) behind a piece of text. In this case, we are analyzing **movie reviews** to predict whether they are positive or negative based on the textual content.

Approach:

1. Text Preprocessing:

- Tokenize the text, remove stopwords, perform stemming/lemmatization, and convert text into a numerical form using **Bag of Words** or **TF-IDF**.

2. Model Selection:

- Use a supervised learning algorithm such as **Logistic Regression** for binary classification (positive vs. negative).

3. Feature Extraction:

- Extract features from text data using methods like **TF-IDF** (Term Frequency-Inverse Document Frequency).

4. Evaluation:

- Evaluate the performance of the sentiment classifier using metrics like **accuracy**, **precision**, **recall**, and **F1-score**.
-

Steps for Movie Review Sentiment Classification:

1. Text Preprocessing:

- Tokenize and clean the text (remove punctuation, stopwords, etc.).
- Apply stemming or lemmatization to normalize the text.

2. Feature Extraction:

- Use **TF-IDF Vectorizer** to transform the text into feature vectors.

3. Model Training:

- Train a **Logistic Regression** model on labeled movie reviews (positive and negative).

4. Model Evaluation:

- Evaluate the trained model on test data using performance metrics like accuracy.

Conclusion:

In this assignment, we successfully implemented a movie reviews sentiment classifier using the Logistic Regression model. The process involved text preprocessing, feature extraction using TF-IDF Vectorizer, and training a classifier to predict whether a review is positive or negative. This approach provides a foundation for sentiment analysis in real-world applications, such as customer reviews and social media analysis.