Flipkart Reviews Sentiment Analysis using Python

Project Report

Soumyadeep Poddar

Introduction:

This project focuses on sentiment analysis of Flipkart reviews using Python. The dataset was pre-processed to clean and tokenize text data. A Decision Tree Classifier was trained on TF-IDF features, and its performance was evaluated on a training set, displaying accuracy and a confusion matrix. The project aims to understand sentiment patterns and classify reviews as positive or negative.

Objective:

The objective of this project is to perform sentiment analysis on Flipkart reviews using machine learning. By preprocessing textual data, training a Decision Tree Classifier on TF-IDF features, and evaluating its performance, the project aims to develop a model that can classify customer reviews into positive and negative sentiments. The ultimate goal is to gain insights into customer satisfaction and sentiment patterns on the Flipkart platform.

Methodology:

1. Data Collection & Loading:
    - The dataset is taken from dataset.
    - The dataset consists of two features - customer ratings and reviews.
2. Data Preprocessing:
    - Classifying the rating column into binary.
    - Removing punctuations, stopwords and converting each word to lowercase.
3. Data Analysis & Visualisation:
    - Creating a WordCloud, the resulting word cloud visually represents the most frequent words in the positive reviews. Larger and bolder words in the cloud indicate higher frequency.
4. Text Representation:
    - Converting text into vectors using TF-IDF.
5. Model Development:
    - Performing train-test split.
    - For prediction purpose, using Decision Tree Classifier.
    - Calculating Accuracy and the Confusion Matrix.

Conclusion:

In conclusion, this project successfully implemented sentiment analysis on Flipkart reviews using a Decision Tree Classifier trained on TF-IDF features. The preprocessing of textual data enhanced the model's ability to capture meaningful patterns. The achieved accuracy on the training set was evaluated using a confusion matrix, providing insights into the model's performance. This project contributes to understanding sentiment trends in Flipkart reviews, laying the groundwork for further analysis and improvements in customer satisfaction assessment.

Limitations:

A limitation of this project lies in its reliance on a Decision Tree Classifier, which may not capture complex relationships in text data as effectively as more advanced models. Additionally, the evaluation is based solely on the training set, and the absence of testing on an independent dataset may impact the model's generalization performance. Further improvements could involve exploring alternative models and incorporating a separate test set for a more robust assessment of the model's predictive capabilities.