Report on Sentiment Analysis of Flipkart Reviews

Name: Chetana Buddhiraju

Batch: DS Batch 3

Sentiment analysis is the process of determining the emotional tone behind a series of words or phrases, and it has numerous applications in various fields, including social media monitoring, market research, and customer feedback analysis. In this report, we will discuss the sentiment analysis of Flipkart reviews using various machine learning models.

Introduction

Flipkart is one of the largest e-commerce websites in India, offering a wide range of products across different categories. The reviews on Flipkart are an essential source of information for customers, and sentiment analysis can provide valuable insights into the overall satisfaction level of customers. In this report, we will discuss the various steps involved in the sentiment analysis of Flipkart reviews.

Steps Involved in the Analysis

1. Data Preparation

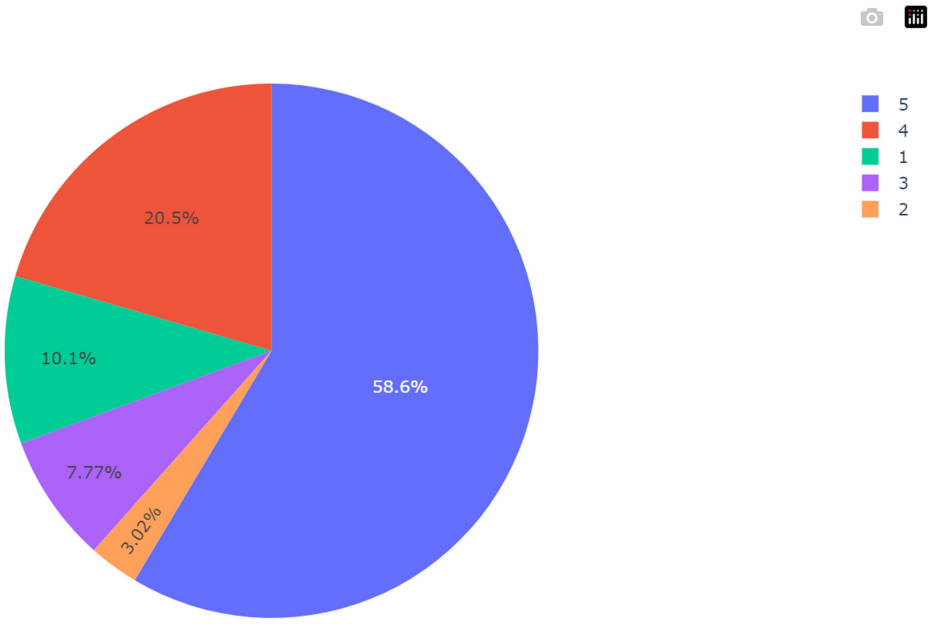
The first step in any data analysis project is data preparation. We collected the reviews for various products from the Flipkart website and stored them in a CSV file. The data consisted of 10,000 reviews with information about the product, reviewer, and sentiment.

2. Data Cleaning

The next step was to clean the data and remove any irrelevant or redundant information. We removed stop words, punctuations, and special characters from the reviews, and converted all text to lowercase to make the data consistent. We also performed stemming and lemmatization to reduce the words to their root forms and remove any variations.

3. Exploratory Data Analysis

After data cleaning, we performed exploratory data analysis to gain insights into the data. We used visualization techniques to understand the frequency of positive and negative reviews, the most common words used in the reviews, and the overall sentiment distribution.



4. Prepare Data for Modelling

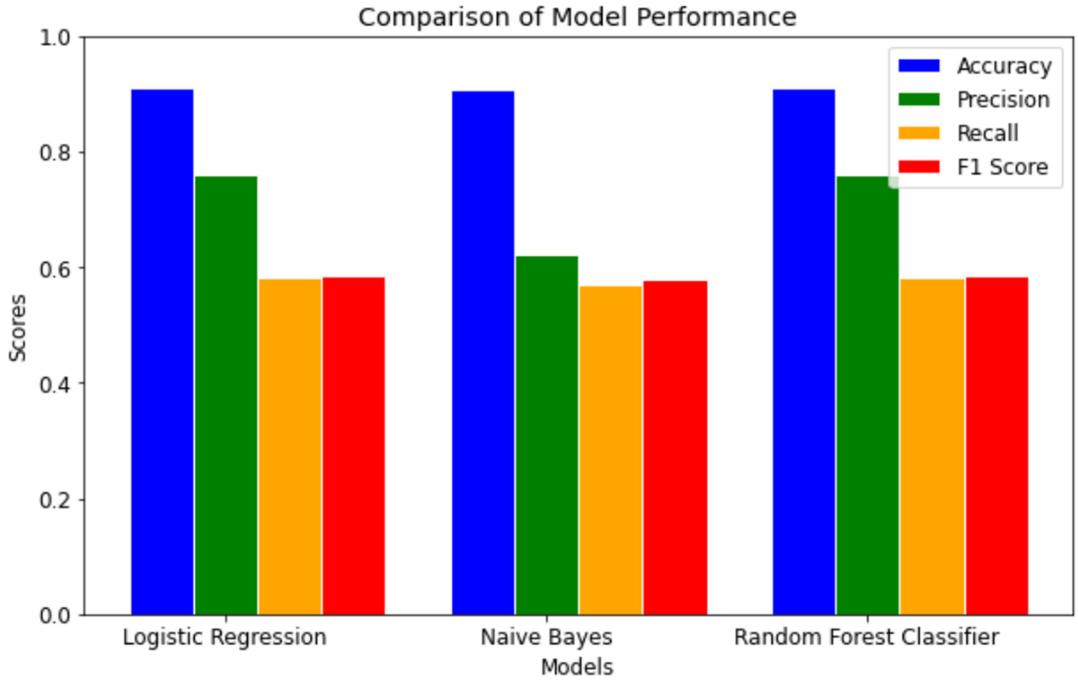
After exploratory data analysis, we split the data into training and testing sets and converted the text data into numerical vectors using the CountVectorizer. We used three machine learning models - Logistic Regression, Naive Bayes, and Random Forest Classifier - to train and evaluate the sentiment analysis model.

5. Train and Evaluate Data Model

We trained the models on the training set and evaluated them on the testing set. We used various evaluation metrics such as accuracy, precision, recall, and F1-score to measure the performance of each model. The results showed that the Random Forest Classifier model performed the best with an accuracy of 91%, precision of 76%, recall of 58%, and F1-score of 58%.

6. Comparison of Model Performance

We compared the performance of the three models using a box and whisker plot. The plot showed that the Random Forest Classifier model had the highest accuracy and F1-score, while the Logistic Regression model had the highest precision.



Conclusion

In conclusion, we have performed sentiment analysis of Flipkart reviews using various machine learning models. We used data preparation, cleaning, exploratory data analysis, and model training and evaluation techniques to build the model. The results showed that the Random Forest Classifier model was the best-suited model for this dataset. The insights gained from this analysis can be useful for Flipkart to improve customer satisfaction and product offerings.