**1. Abstract**

This project focuses on developing a machine learning model to analyse Flipkart product reviews and classify them into three sentiment categories: positive, negative, and neutral. The approach involves preprocessing text data, feature extraction using model building with Logistic Regression. The project provides actionable insights to improve customer experience and enhance Flipkart's product offerings.

**2. Introduction**

Customer reviews are a valuable source of feedback for e-commerce platforms like Flipkart. However, manually analysing these reviews is challenging due to the volume of data. This project automates sentiment analysis, enabling efficient and accurate classification of customer feedback. By leveraging Natural Language Processing (NLP) and machine learning, the project aims to provide real-time insights into customer sentiments.

**3. Objectives**

* To preprocess and clean review text for analysis.
* To classify reviews into positive, negative, and neutral sentiments.
* To evaluate the performance of the sentiment classification model.
* To visualize key insights from customer reviews using word clouds and graphs.

**4. Dataset Description**

* **Source**: A dataset of Flipkart reviews.
* **Features**:
  + review: Text of the customer review.
  + rating: Numeric rating given by the customer (1–5).
  + Derived feature: sentiment (Positive: rating ≥ 4, Neutral: rating = 3, Negative: rating ≤ 2).
* **Size**: Approximately 10,000 rows and 3 features.

**5. Methodology**

**5.1 Data Preprocessing**

* Lowercasing the text.
* Removing punctuation and non-alphanumeric characters.
* Tokenizing and lemmatizing text.
* Removing stop words using NLTK.

**5.2 Sentiment Assignment**

* Sentiments were assigned based on the ratings:
  + Ratings ≥ 4: Positive.
  + Ratings = 3: Neutral.
  + Ratings ≤ 2: Negative.

**5.3 Feature Extraction**

* Used **TF-IDF (Term Frequency-Inverse Document Frequency)** to convert text into numerical representations for model training.

**5.4 Model Building**

* Chose Logistic Regression for its simplicity and efficiency.
* Trained the model on 80% of the dataset and tested it on the remaining 20%.

**5.5 Evaluation Metrics**

* Accuracy, Precision, and Confusion Matrix were used to evaluate model performance.

**6. Tools and Technologies**

* **Programming Language**: Python
* **Libraries**: Pandas, NumPy, NLTK, Seaborn, Scikit-learn, Matplotlib, Word Cloud
* **Environment**: Jupyter Notebook

**7. Results**

**7.1 Model Performance**

* Accuracy: 85%
* Precision (Positive Sentiment): 87%
* Recall (Positive Sentiment): 84%

**7.2 Confusion Matrix**

* Visualized using a heatmap to analyse misclassifications.

**7.3 Word Cloud**

* A word cloud showed the most frequent terms in positive and negative reviews, highlighting customer priorities and concerns.

**8. Conclusion**

The sentiment analysis model successfully classified Flipkart reviews into positive, negative, and neutral categories. The model achieved high accuracy and provided meaningful insights into customer sentiments. Future work can focus on incorporating more sophisticated algorithms like LSTM or BERT for improved accuracy.

**9. Future Work**

* Use advanced deep learning models for sentiment analysis, such as LSTM or BERT.
* Include multilingual reviews for broader analysis.
* Implement real-time sentiment analysis for live feedback monitoring.

**10. References**

1. Python libraries: Pandas, Scikit-learn, NLTK documentation.
2. Research papers on sentiment analysis and NLP.
3. Kaggle datasets for e-commerce reviews.