Project: Flipkart Reviews Sentiment Analysis using Python

- **Objective: Develop a Machine Learning model to analyze Flipkart product reviews and classify them as positive or negative based on user sentiment.**
- **Dataset: Flipkart Reviews Dataset**

Project Goals:

1. Importing Necessary Libraries and Dataset:

- **Load required Python libraries:**
  - **Pandas for handling datasets.**
  - **Scikit-learn for machine learning algorithms and vectorization.**
  - **Matplotlib/Seaborn for data visualization.**
  - **WordCloud for visualizing common words in reviews.**
  - **Warnings to suppress unnecessary messages.**
- **Load the dataset using Pandas and explore its structure.**

2. Data Preprocessing:

- **Remove missing values and duplicate entries.**
- **Convert text into lowercase and remove stopwords, punctuation, and special characters.**
- **Tokenize text data using TF-IDF (Term Frequency-Inverse Document Frequency).**
- **Encode sentiment labels (Positive = 1, Negative = 0).**
- **Split dataset into training (80%) and testing (20%) sets.**

3. Exploratory Data Analysis (EDA):

- **Visualize sentiment distribution using count plots.**
- **Create a word cloud to identify common words in positive and negative reviews.**
- **Analyze correlations between review length and sentiment.**

4. Model Training and Selection:

- **Train different machine learning models:**

    - **Logistic Regression**
    - **Naïve Bayes**
    - **Random Forest Classifier**
    - **Support Vector Machine (SVM)**
- **Compare model performance using accuracy and F1-score.**

5. Model Evaluation and Prediction:

- **Evaluate the best model using:**
    - **Accuracy Score**
    - **Precision, Recall, F1-Score**
    - **Confusion Matrix**
- **Test the model on new reviews to classify sentiment as positive or negative.**

Conclusion:

- **This model helps in automatically classifying customer reviews, improving product quality insights.**
- **Future improvements can include Deep Learning models (LSTMs or Transformers) for better sentiment analysis.**