Sentiment Analysis of Real-time Flipkart Product Reviews

Objective
The objective of this project is to classify Flipkart customer reviews as positive or negative and identify key pain points from negative reviews to understand customer dissatisfaction.

Dataset
The dataset contains 8,518 customer reviews for the product "YONEX MAVIS 350 Nylon Shuttle" scraped by a Data Engineering team. The dataset includes review text, ratings, reviewer details, votes, and review metadata. No manual scraping was performed.

Data Preprocessing
Text cleaning involved lowercasing, removing punctuation, numbers, and stopwords. Lemmatization was applied for text normalization. Ratings were converted into sentiment labels where ratings ≥ 4 were considered positive and ≤ 2 were considered negative. Neutral reviews were removed.

Text Embedding
TF-IDF vectorization was used to convert cleaned review text into numerical features suitable for machine learning models.

Modeling and Evaluation
A Logistic Regression model was trained using TF-IDF features. The model achieved an F1-score of 0.95 and an accuracy of 92%, demonstrating strong sentiment classification performance.

Pain Point Analysis
Negative reviews were analyzed using word frequency analysis and word clouds. Common issues identified include poor product quality, durability concerns, damaged items, and dissatisfaction with value for money.

Deployment
A Streamlit/Flask web application was developed to perform real-time sentiment prediction. The application was deployed on an AWS EC2 instance for public access.

Conclusion
This project demonstrates a complete sentiment analysis pipeline, providing actionable insights into customer feedback and improving product evaluation strategies.

Author
Abhishek Jula