E-Commerce Sentiment Analysis Report

Samsung Galaxy S24 & iPhone 15 (Flipkart Reviews)

Aakash Jaiswal

November 25, 2025

Contents

1 Project Overview

This project performs an end-to-end sentiment analysis on Flipkart customer reviews for Samsung Galaxy S24 and iPhone 15.

The pipeline includes data loading, preprocessing, EDA, sentiment labelling, dataset balancing, model training, FastAPI deployment, and insight reporting.

2 Dataset Summary

Raw reviews were manually collected due to scraping restrictions. Both positive and negative reviews were included, and additional negative reviews were manually added.

Sentiment labelling rules:

- Rating $\geq 4 \rightarrow$ Positive
- Rating $\leq 2 \rightarrow$ Negative
- Rating $= 3 \rightarrow$ Neutral (used only in EDA)

3 Major Issues Faced & Resolutions

Issue 1: Flipkart Scraping Blocked

Flipkart blocks automated extraction.

Resolution: Manual CSV extraction.

Issue 2: Strong Sentiment Imbalance

Dataset had far more positive reviews.

Resolution: Added negative reviews manually and used SMOTE + undersampling.

Issue 3: Inconsistent Date Formats

Many dates were invalid or unparseable.

Resolution: Regex cleaning and standardization.

Issue 4: Model Predicting Only Positive

Due to imbalance.

Resolution: Retrained on balanced dataset.

Issue 5: Docker Not Supported on HP ProBook 4530s

Older hardware lacks virtualization support.

Resolution: API run locally using FastAPI.

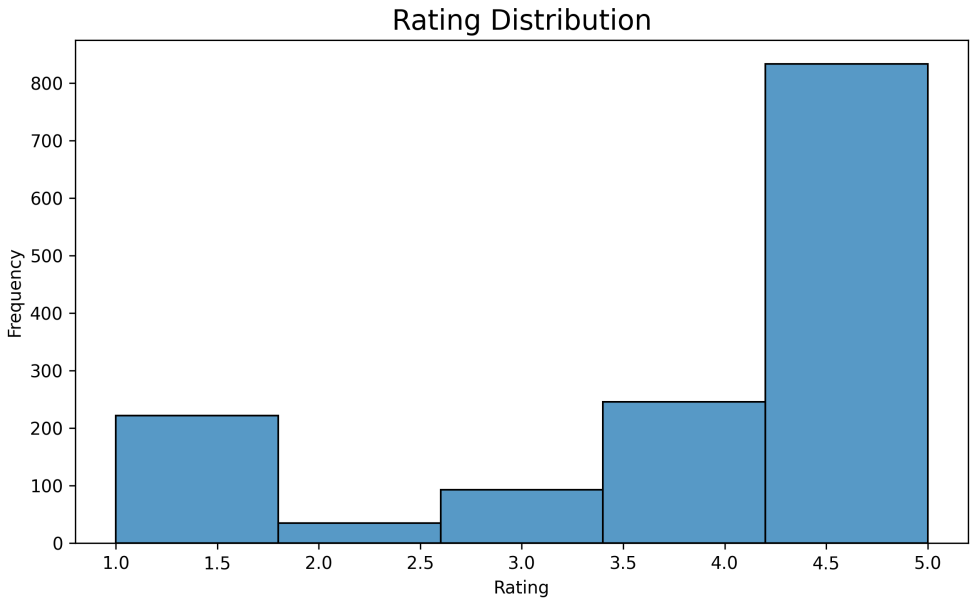
Issue 6: FastAPI Served Old Model

Old model.pkl was loaded.

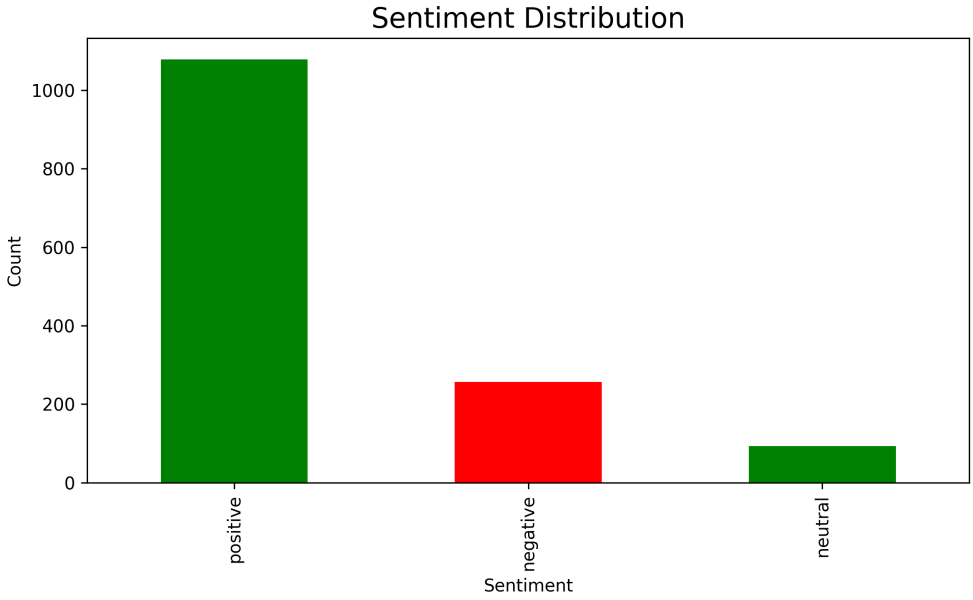
Resolution: Replaced with new balanced model.

4 Key Visual Insights

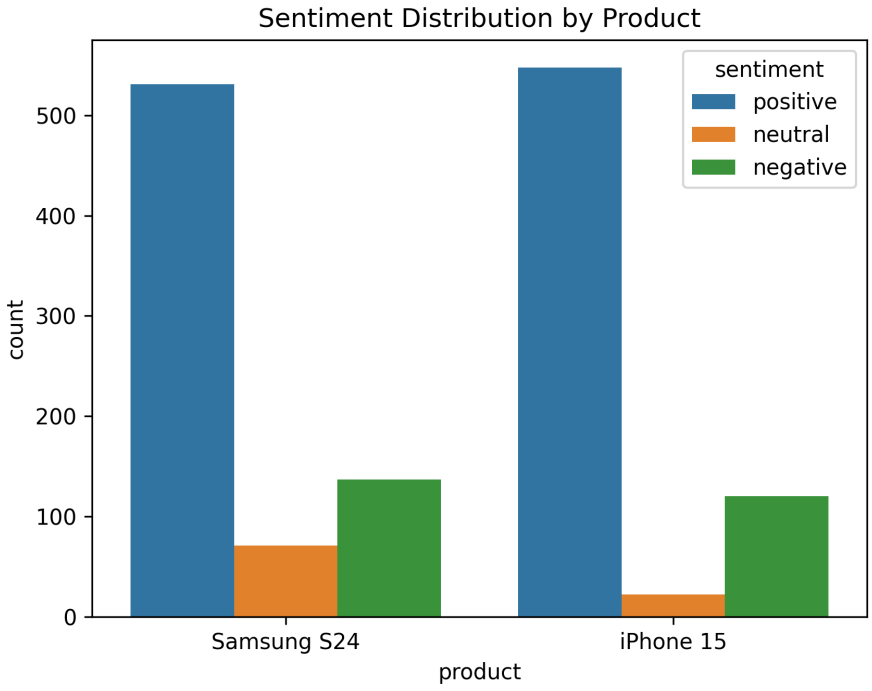
4.1 Rating Distribution



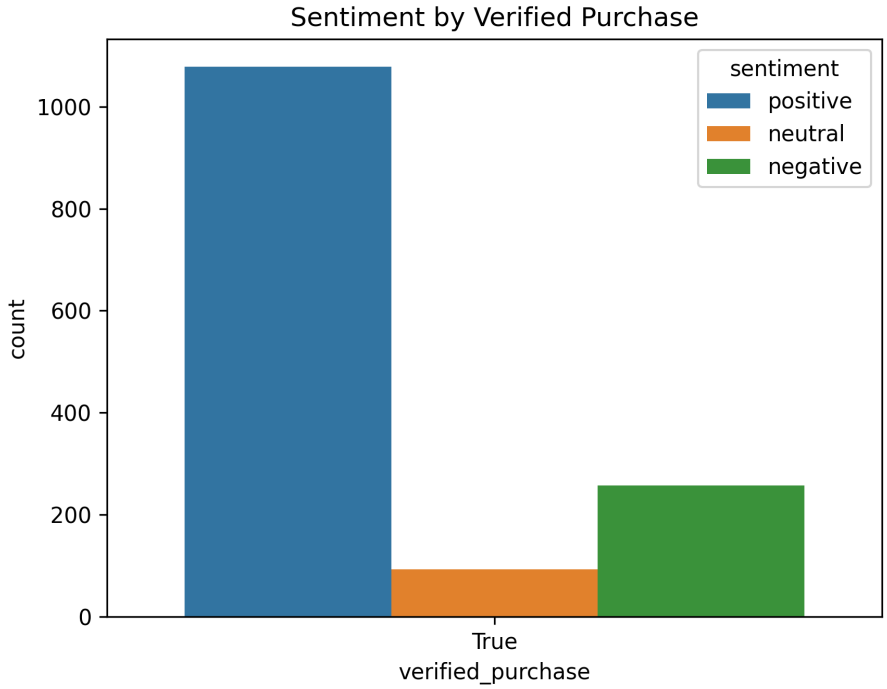
4.2 Sentiment Distribution



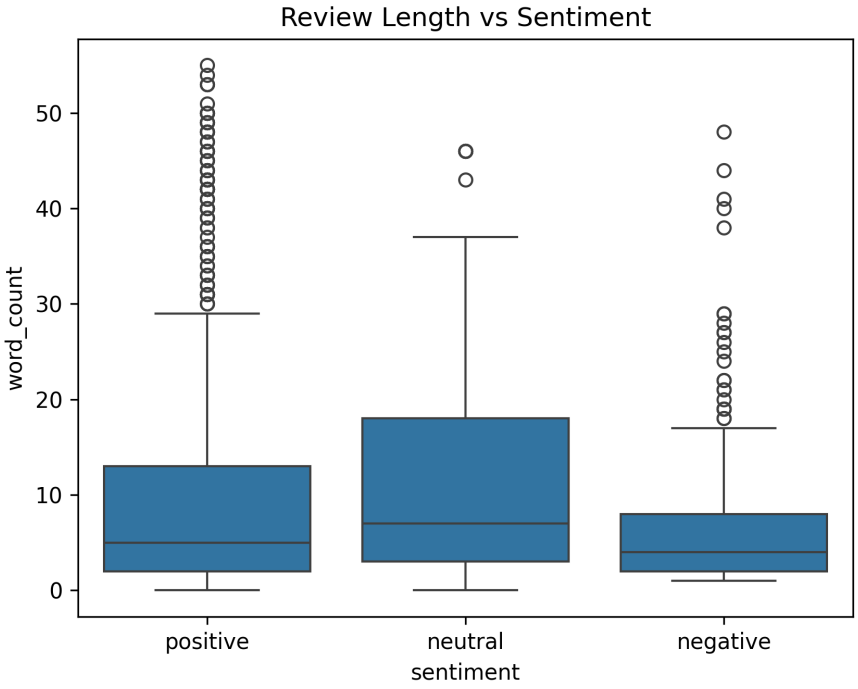
4.3 Sentiment by Product



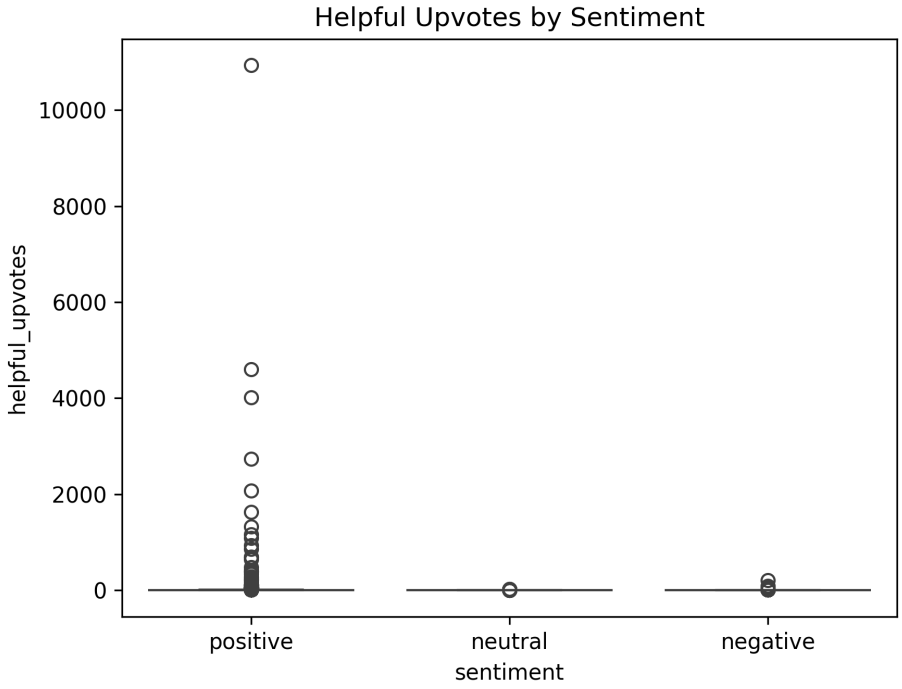
4.4 Sentiment by Verified Purchase



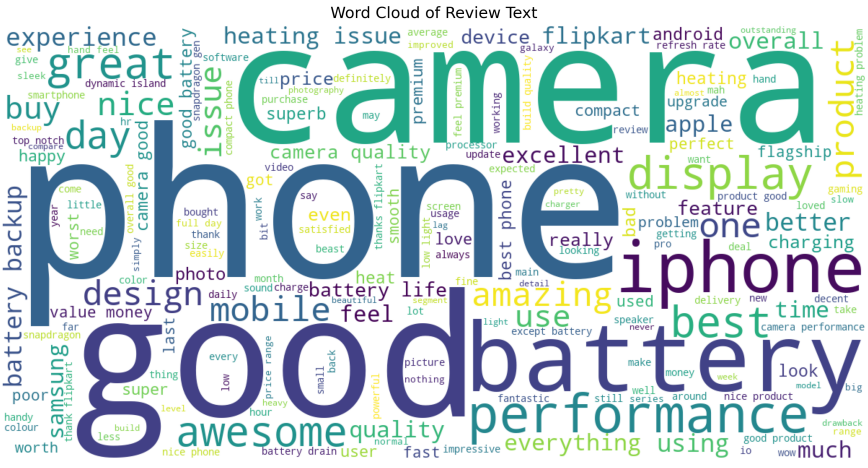
4.5 Review Length vs Sentiment



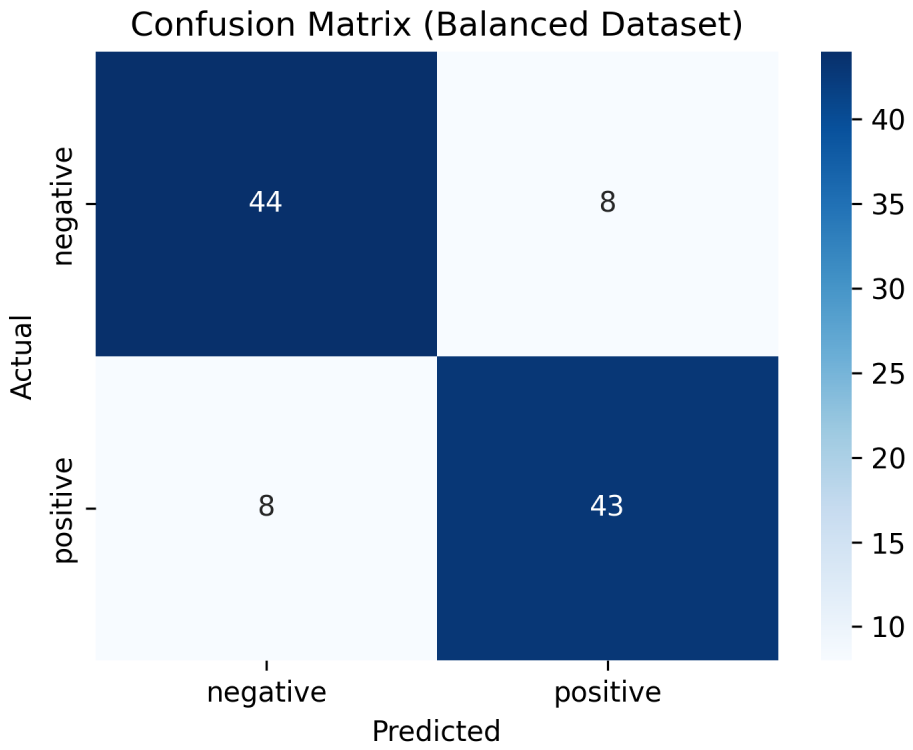
4.6 Helpful Upvotes vs Sentiment



4.7 Word Cloud



4.8 Confusion Matrix (Balanced Dataset)



5 Model Summary

The model used:

- TF-IDF Vectorizer
- Logistic Regression Classifier

Performance:

- Accuracy: 84%
- Balanced precision and recall
- F1-score: 84%

6 Business Insights

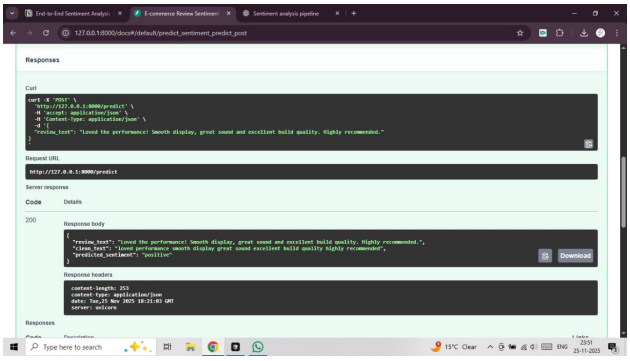
- Heating and battery issues are the most common complaints.

- iPhone reviews frequently praise camera quality.
- Verified buyers write more reliable and detailed reviews.
- Dataset shows a heavy positive bias.
- Balanced training improves detection of negative feedback.

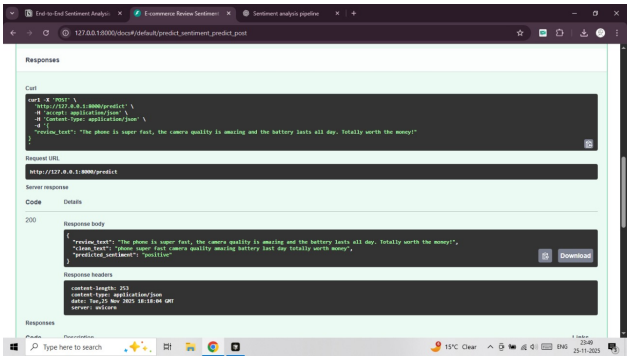
7 Model Testing on FastAPI Server

The deployed FastAPI model was tested using Swagger UI. Below are the actual request–response screenshots from the live API hosted locally.

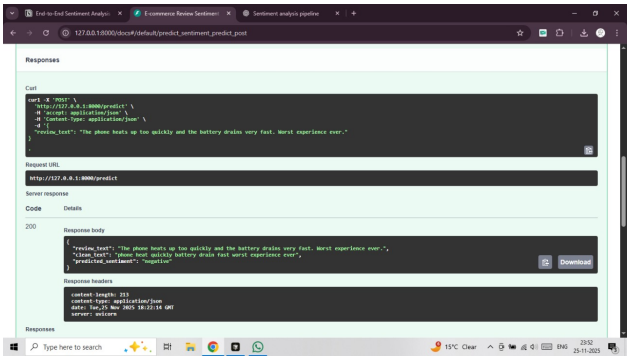
Test Case 1: Strongly Positive Review



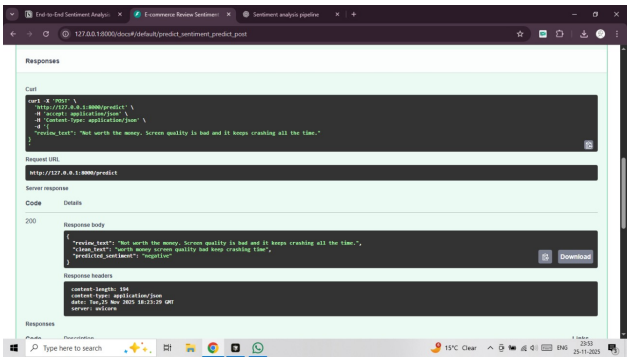
Test Case 2: Positive Experience Review



Test Case 3: Strongly Negative Review



Test Case 4: Negative Experience Review



8 Limitations

- **Manual data extraction**: Limited by Flipkart's anti-scraping policies.
- **Small number of negative reviews**: Required oversampling rather than natural distribution.
- **Rating-based sentiment labelling**: Some users give high ratings but write negative comments.
- **Model trained only on text**: No metadata (likes, reviewer profile, purchase history) included.
- **FastAPI not containerized on this system**: Docker virtualization unsupported on laptop hardware.
- **No neural models used**: Logistic Regression works well but misses complex contextual meaning.

9 Future Work

- Expand dataset with more naturally occurring negative reviews.
- Deploy the API on cloud platforms like AWS, Railway, or Render.
- Use BERT or other transformer models for deeper text understanding.
- Build aspect-based sentiment (battery, camera, delivery, pricing).
- Integrate dashboard using Streamlit / PowerBI for real-time monitoring.
- Add multilingual sentiment support (Hindi, Hinglish).

10 Conclusion

This project successfully demonstrates an end-to-end sentiment analysis pipeline with actionable insights for e-commerce platforms. The balanced model performs well on unseen reviews and the FastAPI endpoint enables easy integration into real-world applications.