

Flipkart Product Recommendation Engine with Sentimental Analysis

Project Documention

By

Sripathi V R

Title: Web Scraping for Product Recommendation with Sentimental Analysis Based on Reviews from Flipkart (Live Data) Using LangChain and AWS Deployment.

1. Introduction

In today's competitive e-commerce landscape, understanding customer sentiment is crucial for product success. This project aims to leverage web scraping and sentiment analysis techniques to gather and analyze product reviews from Flipkart, leading to data-driven product recommendations. By employing a systematic approach to data collection and analysis, we can derive valuable insights that enhance customer experience and inform business strategies.

2. Skills Gained

- Web Scraping: Techniques for extracting data from web pages using libraries like Selenium and BeautifulSoup.
- Data Parsing and Structuring: Skills in organizing unstructured data into a structured format suitable for analysis.
- Sentiment Analysis: Ability to analyze textual data to gauge customer sentiments using libraries like TextBlob.
- Data Visualization: Techniques for creating visual representations of data to aid in understanding and decision-making.
- Python Programming: Proficiency in Python for data analysis, web scraping, and application development.
- LangChain Implementation: Experience in using LangChain for generating meaningful product recommendations.
- AWS Deployment: Skills in deploying applications to AWS for scalability and accessibility.

3. Project Domain

- E-commerce: Understanding consumer behavior and preferences through online reviews.
- Data Science: Applying analytical techniques to extract insights from large datasets.
- Machine Learning: Utilizing data-driven approaches to improve product recommendation systems.
- Deep Learning and NLP: Employing natural language processing techniques for sentiment analysis.
- LLM: Working with large language models to enhance recommendations and insights.

4. Problem Statement

The primary goal of this project is to develop an automated system for collecting and analyzing product reviews from Flipkart. The project encompasses several key tasks:

- **Web Scraping:** Developing a scraper to extract product reviews across various categories such as electronics, clothing, and home appliances.
- **Data Cleaning and Structuring:** Processing the scraped data to ensure its suitability for analysis.
- **Sentiment Analysis:** Implementing sentiment analysis on the collected reviews to classify customer sentiment as positive, negative, or neutral.
- **Product Recommendation:** Using sentiment analysis results to recommend top products within each category.
- **Visualization and Reporting:** Creating visual representations of the data and compiling a comprehensive report on the analysis, findings, and recommendations.

5. Approach

Program Descriptions

1. Data Scraping

This program initiates the project by scraping links for five selected mobile phones from Flipkart. Using the Selenium library, the program navigates through the Flipkart website to extract the URLs of product pages. The output is a comprehensive list of links to be utilized for subsequent data collection.

2. Preprocess Links

In this stage, the program processes the scraped links to ensure they are correctly formatted for further scraping. It validates the URLs and prepares them for review extraction, ensuring a smooth transition to the next phase.

3. Scraping Reviews and Ratings

Leveraging BeautifulSoup, this program extracts product reviews and ratings from the processed links. Key information such as product name, price, reviews, rating, and product URL are collected, yielding a structured dataset crucial for analysis.

4. Data Preprocessing

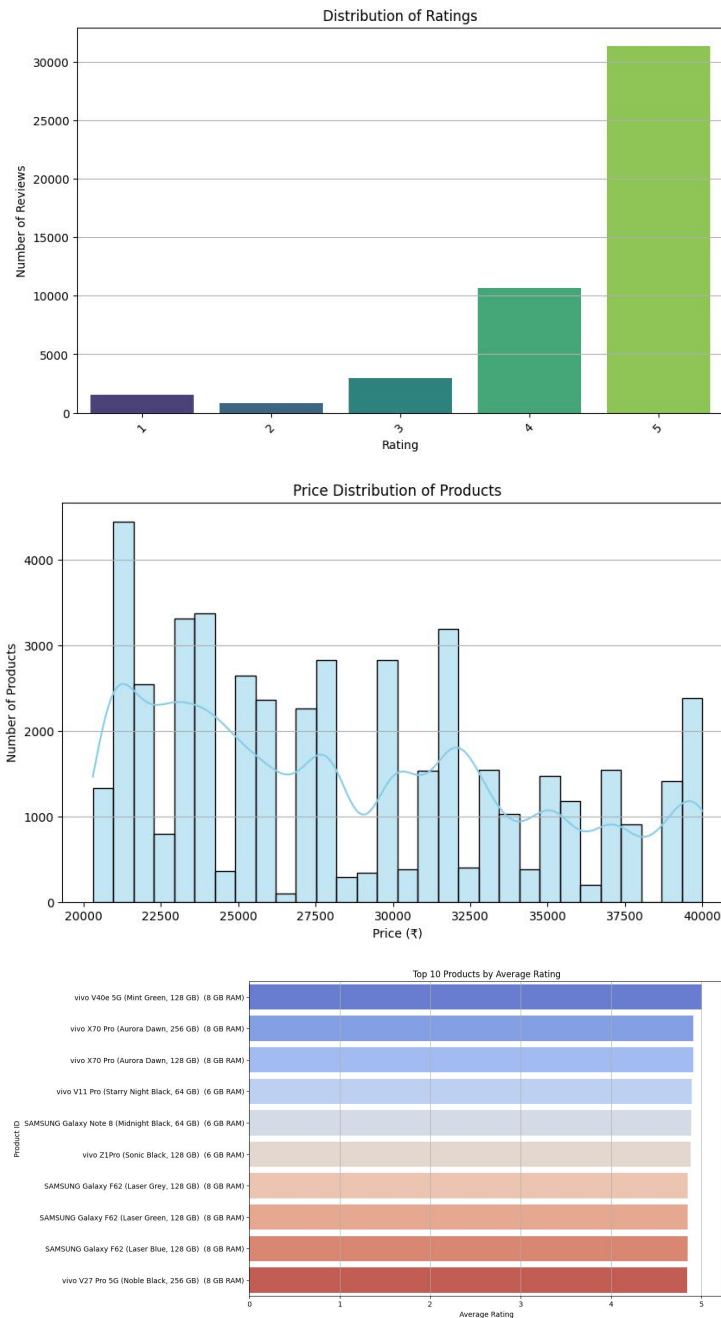
This program cleans the dataset by:

Removing duplicates and irrelevant data.

Normalizing the review text (eliminating HTML tags and unnecessary whitespace).

Tokenizing the review text for effective analysis.

Saving the cleaned data into a structured format (CSV), confirming that 55,056 reviews were processed.



5. Sentiment Analysis

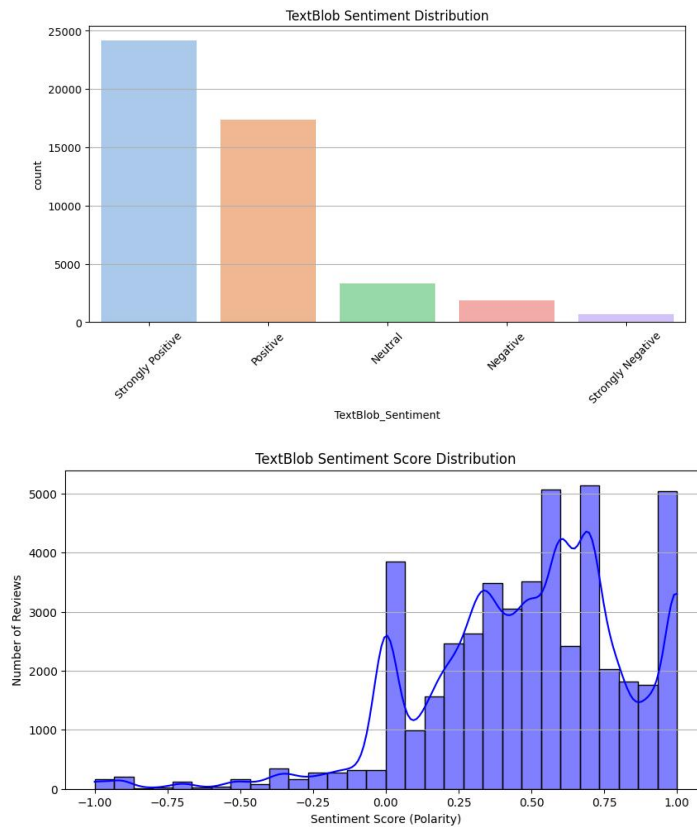
Utilizing the TextBlob library, this program performs sentiment analysis on the cleaned reviews. It classifies each review's sentiment as Strongly positive, positive, strongly negative, negative or neutral, and assigns a sentiment score. The results are saved in a new dataset, which includes:

Original product name

Cleaned review text

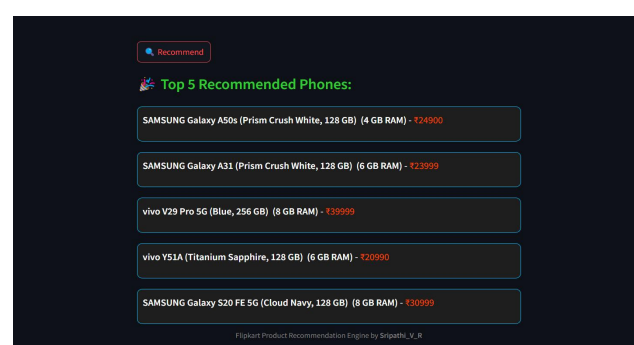
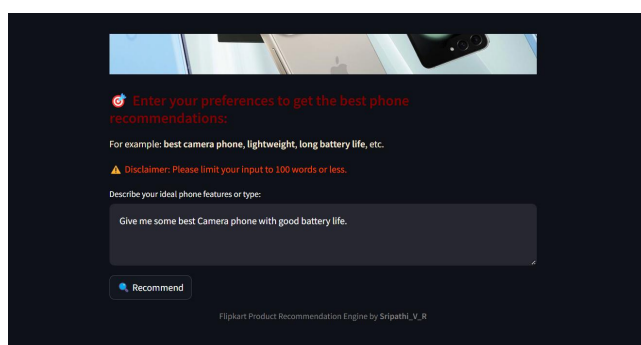
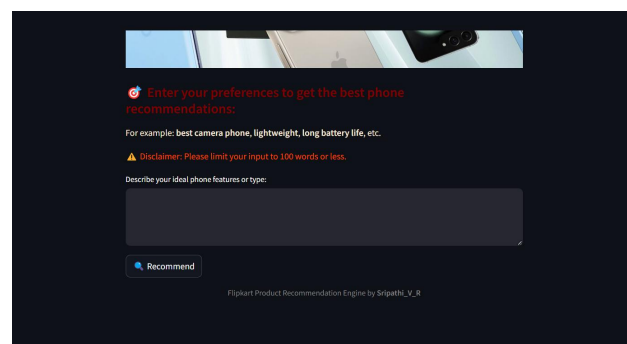
TextBlob sentiment classification

Sentiment score: The program generates a summary report of average sentiment scores for each product, identifying top products based on customer sentiment.



6. Streamlit Application for Recommendations

The final program implements a Streamlit application that serves as the user interface for interacting with the product recommendation system. By utilizing LangChain, the application generates product recommendations based on the sentiment analysis results. Users can access the deployed application on AWS, allowing them to explore product recommendations derived from real-time customer feedback.



6. Results

Scraped Dataset: A comprehensive dataset of reviews from Flipkart, including product IDs, review texts, and ratings.

Sentiment Analysis Output: Results indicating the sentiment (positive, negative, neutral) of each review, facilitating a deeper understanding of customer opinions.

Product Recommendations: A structured list of recommended products based on sentiment analysis, helping consumers make informed purchasing decisions.

Visualizations: Graphical representations (e.g., charts, graphs) depicting sentiment distribution and top product recommendations, enhancing the clarity of insights derived from the data.

Detailed Report: A comprehensive summary detailing the approach, analysis, findings, and recommendations generated throughout the project.

7. Project Evaluation Metrics

- **Accuracy of Sentiment Analysis:** Measures the percentage of correctly identified sentiments.
- **Completeness of Data Collection:** Evaluates the number of reviews successfully scraped and analyzed.
- **Quality of Recommendations:** Assesses the relevance and usefulness of the recommended products.
- **Clarity of Documentation:** Ensures the quality and comprehensiveness of the project report and code documentation.
- **Code Quality:** Reviews adherence to coding standards, use of version control, and overall readability and maintainability of the code.

8. Technical Tags

- i. Web Scraping
- ii. Data Analysis
- iii. Sentiment Analysis
- iv. E-commerce
- v. Selenium
- vi. Python
- vii. LangChain
- viii. BeautifulSoup
- ix. TextBlob
- x. Pandas
- xi. Cohere

9. Dataset

- Source: Flipkart product pages (strictly live pages).
- Products: Comparison of any 5 mobile phones in the price range between Rs 20,000 to 40,000.
- Format: Structured formats such as CSV or JSON.
- Variables: Product ID, Review Text, Rating, Sentiment Score (after analysis).

Dataset Explanation :

- The dataset comprises reviews scraped from Flipkart, including product IDs, review texts, and ratings. The preprocessing steps involved:
- HTML Tag Removal: Eliminating extraneous HTML content from review texts.
- Standardizing Ratings: Converting ratings to a uniform scale for consistency.
- Text Normalization: Tokenizing and normalizing review texts to prepare for sentiment analysis.

10. Conclusion

This project successfully demonstrates the effective application of web scraping and sentiment analysis techniques within the e-commerce domain. By automating the data collection process and analyzing customer sentiment, we can derive valuable insights that inform product recommendations and enhance customer experience. The use of LangChain for generating recommendations and AWS for deployment provides a robust framework for scalable and efficient applications in the e-commerce sector.

Through this project, we gain a deeper understanding of consumer behavior, allowing businesses to tailor their offerings to meet customer preferences effectively. The combination of technical skills acquired and the practical application of theories reinforces the importance of data-driven decision-making in today's market.