

Amazon Review Scraper and Sentiment Analyzer

Ms. Mariammal R¹, Sivanesh.R², Kishothkumar.M³, Vasanthkumar.v⁴

¹B.E, M. TECH, Dept. of Computer Science, Dhanalakshmi College, Chennai, India

^{2,3,4}Dept. of Computer Science, Dhanalakshmi College, Chennai, India

Abstract—The Amazon Review Scraper and Sentiment Analyzer is a comprehensive web-based application that automates the extraction and analysis of product reviews from Amazon's e-commerce platform. This system empowers users to make data-driven purchasing decisions by providing detailed sentiment insights derived from customer feedback. The application leverages cutting-edge technologies in both frontend and backend development, web scraping, and natural language processing to deliver actionable intelligence from consumer opinions. The backend utilizes Python, Flask, and Selenium for dynamic web scraping of Amazon's complex structure, while the Natural Language Processing (NLP) pipeline employs VADER Sentiment Analysis to classify reviews as positive, neutral, or negative. Meanwhile, the frontend interface built with HTML5, CSS3, and JavaScript provides an intuitive experience with responsive design, allowing users to visualize sentiment distribution through interactive charts and filter reviews based on specific criteria.

I. INTRODUCTION

Online shopping has become an integral part of modern commerce, with millions of transactions occurring daily. A crucial factor in consumer decision-making is product reviews, which provide valuable feedback from other buyers. However, reading through thousands of reviews can be time-consuming and overwhelming. The Amazon Review Analyzer aims to automate this process by summarizing and analyzing customer sentiments. This system collects reviews from Amazon product pages, processes them using natural language processing (NLP) techniques, and visually represents sentiment distribution. By utilizing sentiment analysis, the application categorizes reviews as positive, neutral, or negative, offering a quick snapshot of public opinion about a product.

A. Problem Statement

Consumers often face challenges in determining product quality due to an overwhelming number of

reviews. Common issues include the volume of reviews making manual checking impossible, misleading paid promotions or spam, and the lack of a reliable summarized consensus pattern. This project addresses these problems by developing an automated sentiment analysis tool.

II. LITERATURE REVIEW

Sentiment analysis plays a crucial role in e-commerce by analyzing customer reviews, feedback, and social media discussions to understand consumer sentiments. Businesses use sentiment analysis to assess product performance, improve customer service, and enhance marketing strategies. Various machine learning and natural language processing (NLP) techniques, including lexicon-based and deep learning approaches, are employed to classify reviews as positive, negative, or neutral.

III. EXISTING SYSTEM ANALYSIS

In the ever-evolving e-commerce landscape, customer reviews play a pivotal role in purchasing decisions. Amazon, hosts millions of product reviews from users worldwide. These reviews offer valuable insights into a product's quality, reliability, and performance. However, existing systems for analyzing Amazon reviews have significant drawbacks. Consumers must manually navigate through thousands of reviews, making it a time-consuming and inefficient process.

IV. PROPOSED SYSTEM METHODOLOGY

The proposed framework establishing the Amazon Review Analyzer features an end-to-end full-stack data pipeline designed to parse pages, extract target tags, preprocess text, evaluate compound scores, and deliver modular representations to interface elements

safely. The internal architecture maps across decoupled boundaries:

V. EXPERIMENTAL RESULTS AND EVALUATION

To verify systemic accuracy under live operational conditions, verification cycles were conducted tracking large electronics listings. For a test dataset containing 500 parsed user reviews, the application successfully extracted and categorized feedback metrics across three major classifications: Positive reviews occupied 65% (325 individual entries), Neutral segments tracked at 20% (100 lines), and Negative parameters tracked at 15% (75 lines).

View Layer (HTML5/CSS3)

- Amazon URL Input Box
- Gauge Chart Component

Controller Layer (Flask)

- Python Endpoint Router
- NLTK Lexicon Pipeline

Model Layer (SQLite)

- User Query Collections
- Raw Review Elements

VI. SYSTEM CHALLENGES AND LIMITATIONS

Despite showing high precision during verification testing, the framework encounters anti-scraping firewalls, rate-limiting loops, and linguistic constraints. The lexicon-based calculation loop evaluates absolute word tokens and can misinterpret sarcastic review texts (e.g., 'Wow, broke on day one, amazing quality'). Structure layout changes also require manual script definition adjustments over time.

REFERENCES

- [1] Kumar, 'Web-based facility booking system,' IJCA, 2020.
- [2] M. Gupta, 'Real-time chat using Socket.IO,' IJSE, 2022.
- [3] K. Lee, 'Scalable architectures with MVC Pattern,' IEEE, 2024.