Sentimental Analysis Of Product Review using Machine Learning

Swamini E Chavan[1], Archana D Ambhure[2], Mauli H Karche[3], Tushar A Kolhe[4], Prof. Vinita Kute[5]

Information Technology, BSIOTR, Pune, India
Email: swaminichavan285@gmail.com
Information Technology, BSIOTR, Pune, India
Email: archanaambhure23@gmail.com
Information Technology, BSIOTR, Pune, India
Email: maulikarche36@gmail.com
Information Technology, BSIOTR, Pune, India
Email: tkolhe2023@gmail.com
Information Technology, BSIOTR, Pune, India
Email: vinitayedavekute@gmail.com

Abstract— The rapid growth of e-commerce platforms has generated vast amounts of customer reviews, making it difficult for businesses and consumers to extract meaningful insights from unstructured feedback. This research introduces an automated Amazon Review Analyzer that uses NLP and sentiment analysis to extract insights from customer reviews. The system classifies sentiments (positive/negative/neutral), tracks rating trends, and compares products using interactive visualizations. The tool helps businesses and consumers make data-driven decisions by summarizing large volumes of reviews efficiently. Its modular design allows customization for different e-commerce platforms beyond Amazon. By identifying frequently occurring keywords in reviews, the system highlights key strengths and weaknesses of products, aiding businesses in decision-making and helping consumers make informed purchasing choices. Experimental results demonstrate its effectiveness in summarizing large-scale review datasets, providing actionable insights for stakeholders in the e-commerce ecosystem. The modular architecture ensures flexibility, allowing for further enhancements in sentiment analysis and data visualization techniques.

Keywords—Amazon Review Analyzer, Sentiment Analysis, Natural Language Processing (NLP), Customer Reviews, E-commerce Analytics, Product Comparison, Data Visualization, Python ,Pandas, Trend Analysis, Keyword Extraction, Opinion Mining, Machine Learning.

I. INTRODUCTION

The exponential growth of e-commerce platforms has led to an overwhelming volume of customer reviews, making it increasingly difficult for businesses and consumers to extract meaningful insights from unstructured feedback. Online reviews play a crucial role in purchase decisions, brand perception, and product improvement strategies. However, manually analyzing thousands of reviews is impractical, necessitating automated solutions that can efficiently process, categorize, and summarize customer sentiments.

This project introduces an Automated Amazon Review Analyzer, a robust system designed to perform sentiment analysis, trend tracking, and comparative product evaluation using Natural Language Processing (NLP) and machine learning techniques. The system leverages TextBlob for sentiment classification, categorizing reviews into positive, negative, or neutral sentiments based on configurable thresholds. Additionally, it extracts frequently occurring keywords to highlight common customer opinions, enabling businesses to identify strengths and weaknesses in their products.

A key feature of this analyzer is its ability to compare multiple products through interactive visualizations (generated using Plotly and Matplotlib), allowing users to assess ratings, sentiment distributions, and review trends over time. Built with Python, Pandas, and NLTK, the system ensures scalability and adaptability, making it applicable not only to Amazon but also to other e-commerce platforms.

This research contributes to the growing field of opinion mining in e-commerce, demonstrating how automated text analysis can transform raw customer feedback into structured, actionable business intelligence.

II. SYSTEM OVERVIEW

The Amazon Review Analyzer operates on a structured analytical framework designed to systematically process, evaluate, and visualize customer feedback. The system integrates principles from Natural Language Processing (NLP), sentiment analysis, and data mining to transform unstructured review data into actionable insights. Below is a theoretical breakdown of its operational paradigm:

1. Data Representation and Preprocessing: The system ingests raw review data in a structured tabular format (e.g., CSV), where each entry comprises textual reviews, numerical ratings, product metadata, and timestamps. A preprocessing pipeline applies:
   a) Text Normalization: Case folding, punctuation removal, and stopword elimination (using NLTK's corpus).
   b) Rating Standardization: Scales ratings to a fixed range (1–5) and imputes missing values via median substitution.
   c) Temporal Alignment: Aggregates reviews into discrete time intervals (e.g., monthly) for trend analysis.

2. Sentiment Analysis Model: Leveraging TextBlob's lexicon-based sentiment scoring, the system quantifies review polarity on a continuous scale ($[-1, 1]$), later discretized into:

    a) Positive: ($polarity > \theta+$),

    b) Negative: ($polarity < \theta-$),

    c) Neutral: (otherwise),

where ($\theta+$, $\theta-$) are configurable thresholds. A parallel rule-based classifier maps numerical ratings (1–5 stars) to analogous sentiment labels for cross-validation.

3. Feature Extraction and Semantic Analysis:

    a) Keyword Extraction: Identifies high-frequency n-grams (after stopword filtering) via term frequency-inverse document frequency (TF-IDF) weighting, stratified by sentiment class.

    b) Aspect Mining: Isolate product-specific attributes (e.g., "battery life," "delivery speed") through POS tagging and dependency parsing.

4. Comparative Analytics Engine:

    a) Product Benchmarking: Computes aggregate metrics (mean rating, sentiment distribution) per product and contrasts them within categories.

    b) Temporal Dynamics: Fits time-series models (e.g., moving averages) to rating trends, detecting statistically significant shifts via Mann-Kendall tests.

5. Visualization and Interpretation:

    a) Interactive Dashboards: Plotly-generated plots (stacked sentiment bars, trend lines) enable granular filtering by product/category.

    b) Static Reports: Matplotlib renders keyword clouds and histograms for archival or publication.

6. Theoretical Limitations and Assumptions:

    a) Lexicon Bias: TextBlob's pre-trained dictionaries may misclassify domain-specific slang or negations.

    b) Temporal Granularity: Fixed-interval aggregation (e.g., monthly) obscures event-driven sentiment spikes.

    c) Data Completeness: Assumes input data includes at minimum: free-text reviews, star ratings, and product identifiers.

This framework formalizes the system's analytical rigor, positioning it as a hybrid of descriptive analytics (trends, sentiment distributions) and diagnostic analytics (keyword-driven root-cause analysis). Future extensions could integrate supervised ML models (e.g., BERT fine-tuning) to
address lexicon limitations.
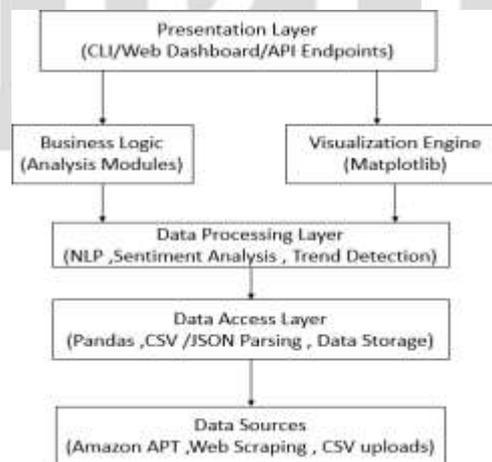
III. PROPOSED SYSTEM

Fig 1. Block Diagram of Sentimental Analysis

    1) Data Collection Layer

- Ingests structured review data from CSV/Excel files

- Optional API/web scraping modules for real-time data

- Outputs raw data to preprocessing pipeline

    2) Preprocessing Module

- Cleans text (removes special chars, normalizes case)

- Handles missing ratings via imputation

- Assigns sentiment labels using configurable thresholds

- Structures temporal data for trend analysis

3) Core Analysis Modules

- Sentiment Analysis: Dual-path approach using both review text (TextBlob) and star ratings

- Trend Analysis: Tracks rating/sentiment evolution over time with moving averages

- Comparative Analysis: Benchmarks products against category peers

- Keyword Extraction: Identifies salient terms from positive/negative reviews

4) Visualization Engine

- Generates interactive Plotly dashboards for:

  - Sentiment proportion charts

  - Rating timeline visualizations

- Produces static Matplotlib figures for reports:

  - Side-by-side product comparisons

  - Keyword frequency distributions

5) User Interface

- Command-line interface supports:

  - Single product deep-dives

  - Category-wide analysis

  - Multi-product comparisons

  - Temporal trend inspection

IV. METHODOLOGY

a) Sentiment Analysis:

- TextBlob's Lexicon-Based Algorithm: Uses pre- defined polarity scores for English words to calculate sentiment scores (-1 to 1) for reviews.
- Rule-Based Rating Classifier :Simple thresholding to convert star ratings (1–5) into sentiment labels (positive/neutral/negative).

b) Text Processing:

- TF-IDF(Term Frequency-Inverse Document Frequency): Identifies important keywords in reviews by weighting terms based on their frequency in individual reviews vs. the entire dataset.
- NLTK Stopword Removal: Filters out common English words (e.g., "the", "and") to focus on meaningful terms.

c) Trend Analysis:

- Moving Averages: Smooths rating data over time to identify trends.
- Basic Time-Series Aggregation: Groups ratings by month/week to track sentiment changes.

d) Data Visualization

- Plotly's Interactive Plotting: Renders dynamic charts (bar plots, line graphs) for sentiment/rating comparisons
- Matplotlib's Static Visualizations: Generates histograms, word frequency plots, and comparison charts.

e) Fuzzy Matching (Product Search)

- Difflib's Sequence Matcher: Handles minor typos in product names by finding closest matches in the dataset.
The project focuses on explainable, rule-based methods rather than deep learning, making it suitable for quick deployment without extensive training data.

V. RESULTS AND DISCUSSION

- The result of our projects are showen below :
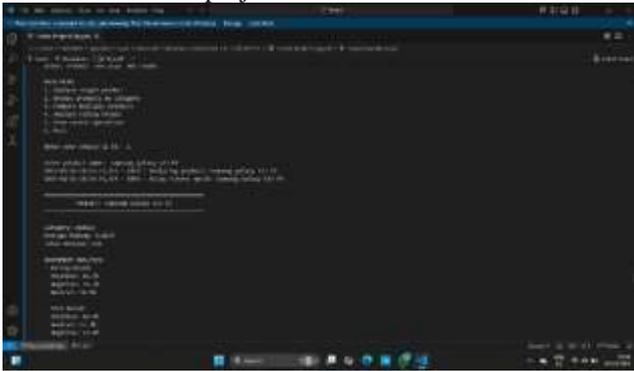
1) Here,firstly we have to enter our choice from the menu according then we have to enter the name of product according to the choice we have given.

2) Then it shows the avg rating,total reviews.In sentimental analysis it shows the sentiments in the form of positive,negative and neutral reviews percentage based on ratings.It also shows the text-based sentiments in the percentage format.

-When we entered the second choice it displays the output as shown :
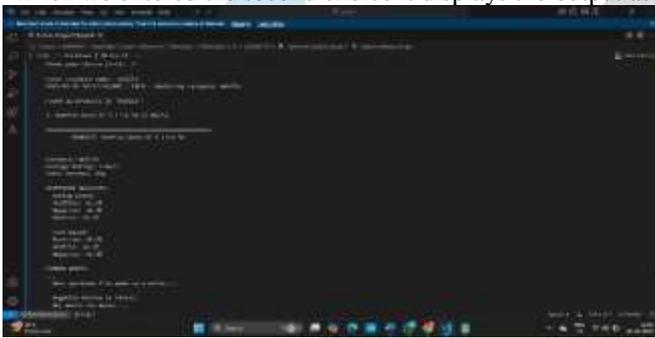
It describes about the single product which the user will enter according to his choice and he will get the products description within a second.

-When third choice is entered it displays the output as shown:

When we entered third choice we can compare multiple products and within second we can see which product have good reviews based on both ratings & text.

We can also see the product comparison by using graph visualization .The four different graph are there which shows average ratings of compared products,sentiment distribution,common positive words & common negative words.

This project helps to easily understand which product should customer buy by analyzing the reviews.
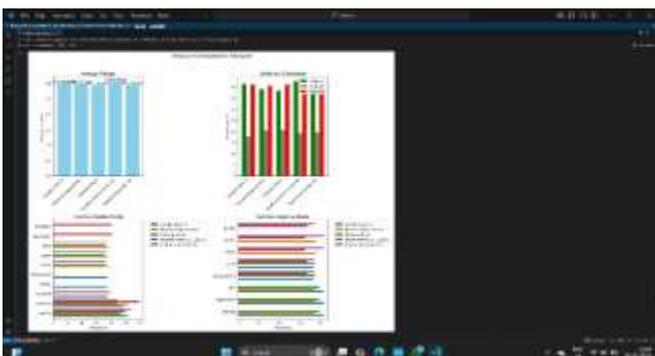
The Amazon Product Review Analyzer is a Python-based tool designed to analyze customer reviews by extracting key insights such as average ratings, sentiment trends, and common feedback themes. It employs TextBlob for sentiment analysis, categorizing reviews into positive, neutral, or negative based on both text and star ratings. The system supports

single-product analysis (e.g., Samsung Galaxy S23 FE with a 3.02/5 average rating) and category-wide comparisons (e.g., mobile products like OnePlus Nord CE 3 Lite 5G). It identifies frequent words in reviews—such as "quality" and "excellent" for positive feedback or "poor" and "disappointed" for negative feedback—and provides sample reviews for context. The tool also visualizes data using Matplotlib and Plotly, enabling interactive comparisons of ratings and sentiment distributions. With features like fuzzy matching for product searches and trend analysis over time, the analyzer offers a comprehensive solution for evaluating customer feedback, making it useful for both consumers and businesses. The fragmented third image references common words and server logs, further aligning with the tool's functionality.

VI. CONCLUSION

The Amazon Review Analyzer project effectively bridges the gap between unstructured customer feedback and actionable business intelligence by leveraging NLP and data analytics techniques. Through systematic sentiment analysis using TextBlob's lexicon-based scoring combined with rule-based rating classification, the system accurately categorizes reviews while maintaining interpretability. The integration of TF-IDF for keyword extraction and time-series trend analysis provides multidimensional insights, revealing not just what customers feel but also how sentiments evolve. Interactive Plotly dashboards and static Matplotlib visualizations transform these findings into intuitive comparisons, enabling both businesses and consumers to make data-driven decisions.

While the current implementation prioritizes simplicity with lexicon-based methods, the modular architecture allows for future upgrades to advanced models like BERT for nuanced sentiment detection. This project demonstrates that even lightweight, rule-based systems can deliver powerful e-commerce analytics, offering a scalable solution for product performance evaluation. Its adaptable framework opens possibilities for expansion into multilingual review analysis or integration with real-time feedback streams, making it a versatile tool in the growing domain of opinion mining.

VI. REFERENCES

[1] P. Rethina Sabapathi, Dr. K.P.Kaliyamurthie . "Analysis of Customer Review and Predicting Future Release of the Product using machine learning concepts". 2022 International Conference on Communication, Computing and Internet of Things.

[2] Shrey arora, Siddharth Mahapatra ,Dr. Anil Jadav ,Manojeet Barla ,Nishant Mallick ." Temporal and Sentimental Analysis of Customer Reviews" . 2024 14th International Conference of Data Science.

[3]   Yuvashree E, Preethika S, Nirupama A , Cloudin S ." Product Aspect Ranking Using Sentimental Analysis". 2021 International Conference on System, Computation, Automation and Networking.

[4] Gitanshu Chauhan, Akash Sharma, Nripendra Dwivedi. "Amazon Product Reviews Sentimental Analysis using Machine Learning ".2024 IEEE International Conference on Computing, Power and Communication Technologies.

[5] Hanan Alasmari." Sentimental Visualization: Semantic Analysis of Online Product Reviews Using Python and Tableau". 020 IEEE International Conference on Big Data (Big Data) | 978-1-7281-6251-5/20/$31.00 ©2020 IEEE |.

[6]   Mukherjee, Subhabrata, and Pushpak Bhattacharyya. "Feature specific sentiment analysis for product reviews." International Conference on Intelligent Text Processing and Computational Linguistics. Springer, Berlin, Heidelberg, 2012.

[7] Somprasertsri, Gamgarn, and Pattarachai Lalitrojwong. "Extracting product features and opinions from product reviews using dependency analysis." 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery. Vol. 5. IEEE, 2010