© June 2025 | IJIRT | Volume 12 Issue 1 | ISSN: 2349-6002

Sentiment Analysis of Amazon Review

Kiran A. Salunkhe[1], Kalyani R. Jadhav[2], Ketan Jadhav[3], Divya P. Jadhav[4], Sayali S. Kadam[5], Aishwarya M. Chavan[6]

[1,2,3,4,5,6] *Department of Computer Science Engineering, Yashoda Technical Campus, Satara, Maharashtra, India*

***Abstract*—This sentiment analysis tool processes Amazon product reviews to classify them as positive, negative, or neutral using advanced NLP models. It supports input via text, CSV uploads, or direct Amazon URLs. The app cleans review data through preprocessing steps to improve accuracy. Visualizations like bar charts and word clouds help interpret sentiment trends. This enables businesses to efficiently understand customer feedback and improve their products**.

***Index Terms*—NLP, CSV, AMAZON URL.**

I. INTRODUCTION

With the rapid expansion of e-commerce, customer reviews have become a critical source of information influencing both consumer decisions and business strategies. Amazon, as one of the largest online retail platforms, generates a massive volume of product reviews daily, reflecting a wide range of customer experiences and sentiments. However, manually analyzing this vast amount of textual data is time-consuming, inefficient, and prone to errors. Sentiment analysis, a branch of Natural Language Processing (NLP), provides an automated approach to classify reviews into positive, negative, or neutral categories, thereby extracting valuable insights from unstructured text. This research presents a Streamlit-based application that employs advanced NLP models such as RoBERTa, DistilBERT, and BiLSTM with attention mechanisms to perform accurate sentiment classification on Amazon product reviews.

The application supports multiple input methods, including direct text entry, CSV file uploads, and extraction from Amazon product URLs. It also incorporates essential preprocessing steps like tokenization, stopword removal, and lemmatization to enhance prediction performance. Furthermore, the tool offers visualizations like bar charts and word clouds to facilitate a clear understanding of sentiment trends and frequently mentioned keywords. By automating the sentiment analysis process, this application enables businesses to efficiently analyze large volumes of customer feedback, supporting informed decision-making and product improvement strategies.

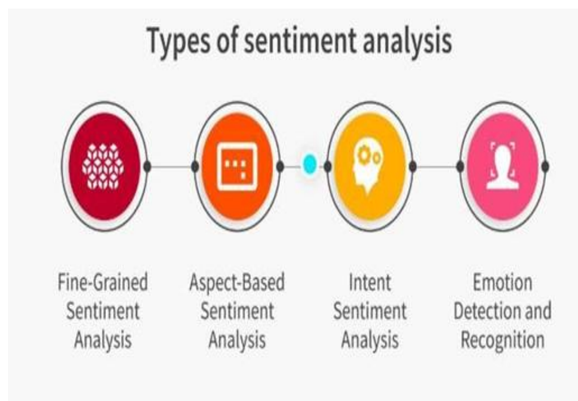


fig 1. Types of sentiment analysis

Sentiment analysis can be categorized based on the level of detail and the granularity of the analysis performed on text data. The main types include:

1. Fine-Grained Sentiment Analysis

This type assigns detailed sentiment scores or labels, often on a scale (e.g., very positive, positive, neutral, negative, very negative). It provides a nuanced understanding of opinions rather than simple positive or negative classifications.

2. Polarity-Based Sentiment Analysis

The most common form, this type classifies text into broad polarity classes such as positive, negative, or neutral.

3. Aspect-Based Sentiment Analysis

Instead of analyzing the overall sentiment of a review, aspect-based analysis focuses on specific components or features of a product (e.g., battery life, camera quality in a smartphone). It identifies sentiment related to each aspect, helping businesses pinpoint strengths and weaknesses.

4. Emotion Detection
Beyond polarity, this type classifies text into specific emotions like happiness, anger, sadness, surprise, or fear. It provides deeper insight into the emotional tone behind the text.

II. LITRATURE SURVAY

Archana. P et al [6] Proposed the methodology of tourism to analyze that is spent on the ticket, places, foods,
and other reasons. In every day millions of peoples traveling in every place, they visit countries, business meetings, and etc. tourism is an information flow based on two types first flow is the information to provide consumer or tourists, the information is consumed at hotel rooms, foods, tickets, and facilities. Another one is information to service benefactors which follows information is reverse order consists of tourists' aggregate information.
Norjihan Abdul Ghani et al [8] Described social media analytics in big data. The data produced by the customers using social media boards is the effect of the addition of their contextual details and daily activities. The existing works are to obtain a broad perception of the big social media analytics for research subjects. In this work is classified by the dictionary based on significant features.
Meena Rambocas et al [9] To proposed online marketing for sentiment analysis. They found to highlight the individuality of action-oriented marketing by online sentiment inquiry. They discussed the request of sentiment inquiry in marketing. The research and suggestions reference to statement the tasks for researchers opposed in using this technique. The study delivers instructors and practitioners with a broad review of the request on online sentiment inquiry within the advertising correction.
Nitesh Sharma et al [2] To described web-based application, allows for the conception of present sentiments linked with a keyword (phrase or word, hashtag) of Twitter comments by conspiracy them on a map. It allows customers to not only quantify the sentiment but also map its strength in terms of layout. The main incentive behind building this application is to provide a single automated platform, which helps as a whole end-to-end system for sentiment study of Twitter comments beside theirs.
Oscar Araque et al [10] present in semantic similarity-based perspective of effect in lexicon methods. Many semantics techniques through word embedding. The semantic relationship metric is divided into text arguments and lexical vocabulary. This method provides a sentiment cataloging model that usages of the semantic relationship measure in the combination with surrounding representations. The semantic relationship method developed the performance of sentiment analysis to a strong baseline and improving the statistically significant.
Devika et al [11] described a study of different sentiment analysis approaches. Different approaches and levels are analyzed in sentiment analysis. A first method is a machine learning through SVM, NB, and maximum entropy techniques were described. Some techniques can be improved in one process to another process to be described. Another one technique were explained like lexical methods.
Rudy Prabowo et al [12] described rule base classification in sentiment analysis. They described rule-based classification, supervised knowledge and machine knowledge into new combined techniques. This method is analyzed the movie and product reviews, and user comments. This method improved the better efficiency in terms of the micro and macro-averaged classifier.
They generate the induction algorithm to analyze deeper
learning like real-world scenarios. It generates two sets of rules that are original set and another one is the induced rule set.
Fabio Clarizia et al [13] described the study of e-learning and sentiment analysis. E-learning is the most actual training approach. E-learning platforms are easily sympathetic to students and their learning skills. They described the use of mixed graph and acquired by the Latent Dirichlet allocation (LDA) method tools for sentiment organization.

III. METHADOLOGY

The sentiment analysis of Amazon reviews follows a systematic approach involving data collection, preprocessing, model selection, and result visualization:

1. Data Collection:

Users can input Amazon product reviews through three modes: manual text input, CSV file uploads containing multiple reviews, or by providing Amazon product URLs from which reviews are automatically extracted. This flexible data intake ensures comprehensive coverage of review sources.

2. Data Preprocessing:

The collected review texts undergo essential preprocessing steps to enhance model performance. This includes tokenization to split text into meaningful units, removal of stopwords to eliminate common non-informative words, and lemmatization to reduce words to their base forms. These steps help reduce noise and improve the quality of input data.

3. Model Selection and Training:

The system employs state-of-the-art Natural Language Processing models such as RoBERTa, DistilBERT, and BiLSTM with attention mechanisms. These models are chosen for their proven effectiveness in capturing contextual nuances and handling the complexity of natural language. The models are fine-tuned on labeled datasets to classify review sentiments into positive, negative, or neutral categories.

4. Sentiment Classification:

After preprocessing, the cleaned review data is fed into the selected model, which predicts the sentiment polarity for each review. The model's output reflects the overall sentiment conveyed by the customer's feedback.

5. Visualization and Interpretation:

To facilitate better understanding, the application generates visual summaries including bar charts depicting sentiment distribution and word clouds highlighting frequently mentioned keywords. These visualizations aid users in quickly grasping sentiment trends and identifying key themes within the reviews.

6. Output Utilization:

The analyzed sentiment data can be used by businesses to assess customer satisfaction, identify product strengths and weaknesses, and make data-driven improvements to their offerings.

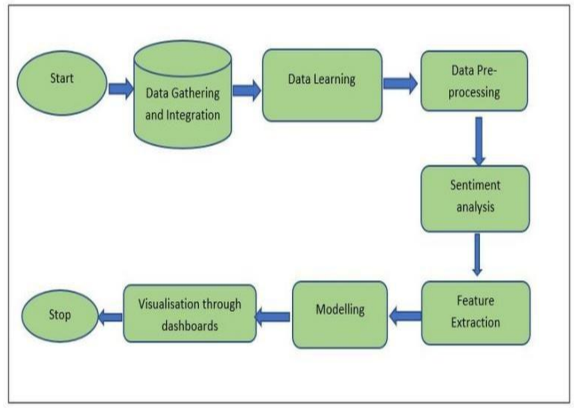


fig2. work flow

Above fig2 shows the workflow:

1. Start
- This is the initiation point of the pipeline.

2. Data Gathering and Integration
- Purpose: Collect raw data from multiple sources (e.g., social media, databases, APIs, CSV files).
- Integration: Combine data into a unified format or system for consistency.
- Outcome: A structured dataset ready for analysis.

3. Data Learning
- Often refers to exploratory data analysis (EDA) or understanding the structure, types, and patterns in data.
- This can involve statistical summaries, visualizations, and identifying data quality issues.

4. Data Pre-processing
- This step prepares the data for analysis and modeling.
- Tasks include:
  - Removing missing or duplicate values
  - Tokenization (in NLP)
  - Normalization or scaling
  - Converting text to lowercase, removing stop words, etc.

5. Sentiment Analysis
- **Goal**: Determine the sentiment expressed in text data— positive, negative, or neutral.

- Utilizes natural language processing (NLP) and possibly pre-trained models.
- Example: Analyzing tweets or customer reviews.

6. Feature Extraction
- Convert raw data (especially text) into numerical features that models can understand.
- Techniques:
  - TF-IDF
  - Word embeddings (Word2Vec, BERT)
  - Frequency counts, part-of-speech tagging, etc.

7. Modelling
- Train machine learning models using the extracted features.
- Common algorithms: Logistic Regression, SVM, Random Forest, or deep learning for text.

8. Visualisation through dashboards
- Use tools like Power BI, Tableau, or Matplotlib/Seaborn (in Python) to:
  - Present results
  - Show sentiment trends
  - Monitor model performance
- Helps stakeholders interpret and make decisions based on the analysis.

9. Stop
- Marks the completion of the process.

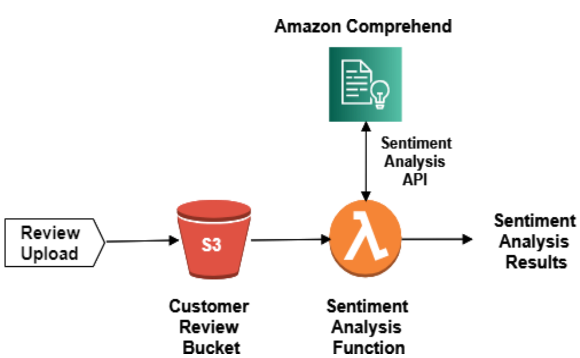


Fig3.Block Daigram

Above fig3 shows The Block Daigram of sentiment analysis on Amazon reviews begins with data input, where users can provide review text manually, upload multiple reviews via CSV files, or directly fetch reviews using Amazon product URLs. Once the reviews are collected, they undergo preprocessing to prepare the data for accurate analysis. This preprocessing includes tokenization, which breaks down text into individual words or tokens; removal of stopwords to eliminate common, non-informative words; and lemmatization to convert words to their base forms, reducing linguistic variation. The cleaned text is then passed through advanced Natural Language Processing models such as RoBERTa, DistilBERT, or BiLSTM with attention mechanisms, which analyze the context and semantics of the reviews to classify them into positive, negative, or neutral sentiments. Finally, the results are presented through intuitive visualizations like bar charts and word clouds, allowing users to easily interpret overall sentiment trends and identify frequently discussed product features. This structured workflow enables efficient and scalable analysis of large volumes of Amazon reviews, facilitating better understanding of customer opinions and supporting informed business decisions.

IV. RESULT AND OUTPUTS

The output is displayed on the streamlit app generator website:

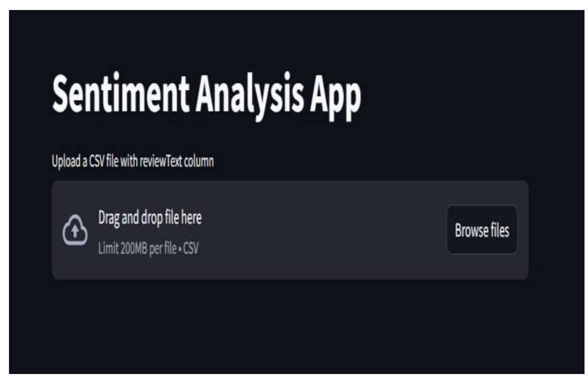


Fig.4 first interaction page

The Above page shows the first landing page of our application which is on the the sentiment analysis on Amazon Review.

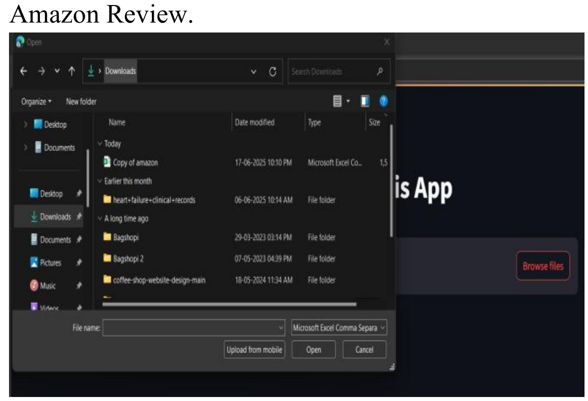
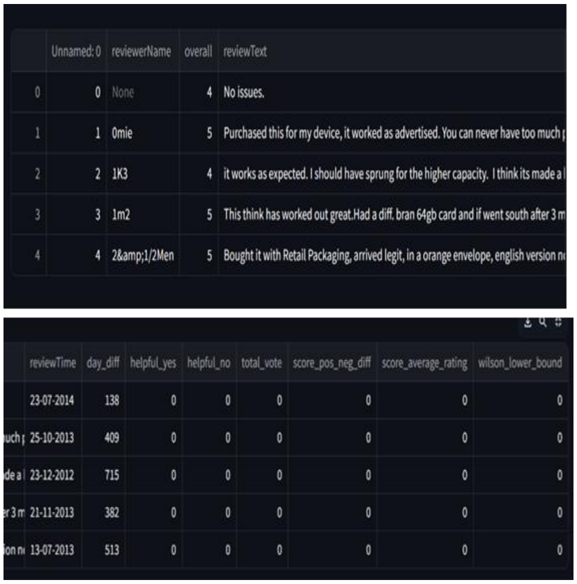


Fig.5 Selecting dataset

The next step after the app generation is to slecting the dataset to clicking on the browse button given.



	Unnamed: 0	reviewerName	overall	reviewText
0	0	None	4	No issues.
1	1	0mie	5	Purchased this for my device, it worked as advertised. You can never have too much
2	2	1K3	4	it works as expected. I should have sprung for the higher capacity. I think its made a l
3	3	1m2	5	This think has worked out great.Had a diff. bran 64gb card and if went south after 3 m
4	4	2&1/2Men	5	Bought it with Retail Packaging, arrived legit, in a orange envelope, english version n

reviewTime	day_diff	helpful_yes	helpful_no	total_vote	score_pos_neg_diff	score_average_rating	wilson_lower_bound
23-07-2014	138	0	0	0	0	0	0
25-10-2013	409	0	0	0	0	0	0
23-12-2012	715	0	0	0	0	0	0
21-11-2013	382	0	0	0	0	0	0
13-07-2013	513	0	0	0	0	0	0

Fig.6 Raw Data

After this the dataset is generated means the raw dataset is generated.

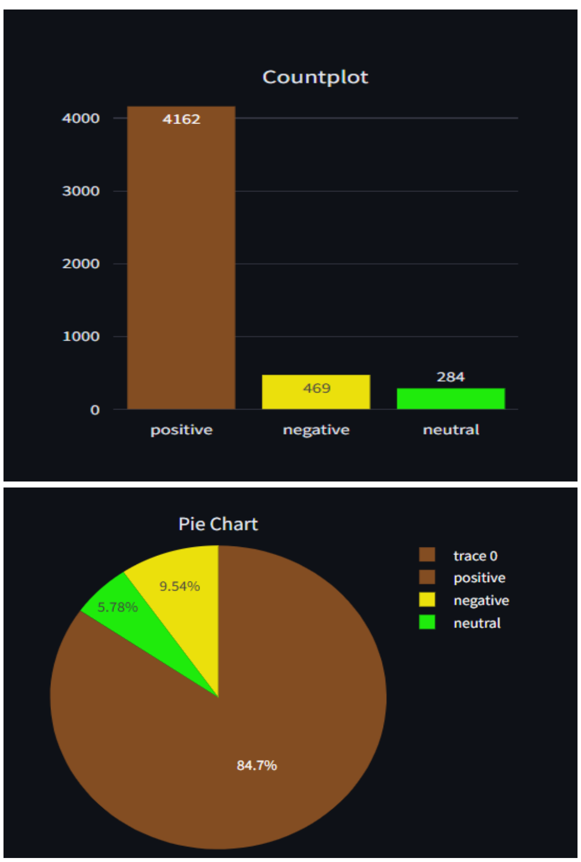
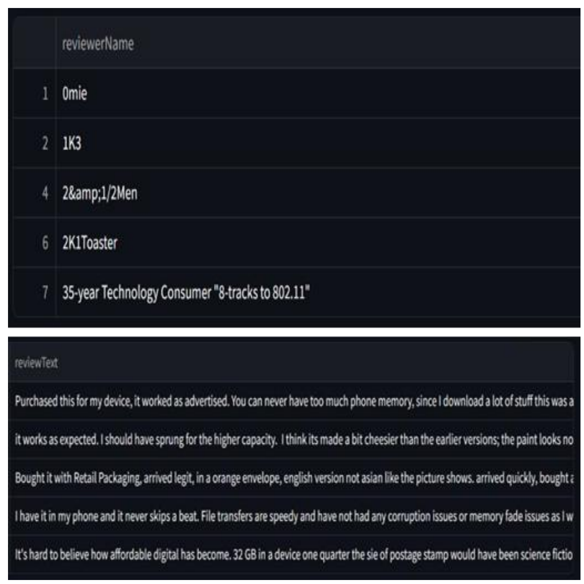


Fig7.Sentiment distribution

Given data shows how the sentiment is distributed with the given revive data.

Fig.8 Word Cloud for Positive Reviews

The dataset gives us the above given data Word Cloud for Positive Reviews.



	reviewerName
1	0mie
2	1K3
4	2&1/2Men
6	2K1Toaster
7	35-year Technology Consumer "8-tracks to 802.11"

reviewText
Purchased this for my device, it worked as advertised. You can never have too much phone memory, since I download a lot of stuff this was a
it works as expected. I should have sprung for the higher capacity. I think its made a bit cheesier than the earlier versions; the paint looks no
Bought it with Retail Packaging, arrived legit, in a orange envelope, english version not asian like the picture shows. arrived quickly, bought a
I have it in my phone and it never skips a beat. File transfers are speedy and have not had any corruption issues or memory fade issues as I w
It's hard to believe how affordable digital has become. 32 GB in a device one quarter the sie of postage stamp would have been science fictio

Fig.9 top positive review

That was the final answer or the final output of the given data

V.CONCLUSION

The provided Amazon review dataset contains a total of 4,915 entries, each representing individual customer feedback on a product. Most reviews have high overall ratings, indicating general customer satisfaction. However, a significant number of reviews have zero helpfulness votes, suggesting limited engagement from other users in rating the usefulness of the reviews. The dataset also includes time-related data, which helps analyze trends in customer feedback over time. Overall, the dataset is well-structured and suitable for deeper analysis, such as sentiment classification, feature extraction, and modeling for recommendation systems. With appropriate visualization and modeling techniques, this data can offer meaningful insights into customer behavior and product performance.

VI. FUTURE SCOP

The Amazon review dataset offers significant future scope for advanced analysis and application. It can be used to perform sentiment analysis, build personalized recommendation systems, and detect fake or biased reviews. Time-based trends and topic modeling can help businesses understand customer preferences and improve products. Additionally, the data can support interactive dashboards and customer segmentation, making it valuable for marketing and strategic decision-making.

REFERENCES

[1] Shahid Shayaa, Noor is awaiting Jaafar, Shamshul Bahri, Ainin Sulaiman, Phoong Seuk Wal, Yeong Wai Chung, Arsalan Zahid Piprani, and Mohammed Ali Al-garadi "Sentiment analysis of big data: Methods, applications and open challenges" IEEE Access, PP:2169- 3536, Year: 2018.

[2] Nitesh Sharma, Ussama Yaqub, Vijayalakshmi Atluri, Soon Ae Chun, Rachit Pabreja, and Jaideep Vaidya "Web-based Application for Sentiment Analysis of live Tweets" ACM, DOI: htttps://doi.org/10.1145/3209281.3209402, Year: 2018.

[3] Alexander Pak and Patrick Paroubek "Twitter as a corpus for sentiment analysis and opinion mining" LREC, PP: 1320-1326, Year: 2010.

[4] Oumayma El Ansari, Hajar Mousannif and, Jihad Zahir "Context- Based Sentiment Analysis: A Survey," DOI: https://doi.org/10.1007/978-3-030-02852-2-7_8, pp: 91-97, Year: 2018

[5] B.Vamshi Krishna, A. P. Siva Kumar and, Ajeet Kumar Pandey "Feature-Based Opinion Mining and sSentiment Analysis Using Fuzzy Logic," Cognitive science and Artificial Intelligence, Springer Briefs in Forensic and Medical Bioinformatics, PP: 79-89, Year: 2018.

[6] Archana. P, Periyasamy. P, and PIyyappan. R "Sentiment Analysis in Tourism," International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 9, PP: 443- 450, Year: 2014.

[7] Alaei, Ali Reza, Stantic, Becken, Susanne, and Bela "Sentiment Analysis in Tourism: Capitalizing on Big Data," Journal of Travel Research, PP: 1-17, Year: 2017.

[8] Norjihan Abdul Ghani, Suraya Hamid, Ibrahim Abaker Targio Hashem, Ejaz Ahmed "Big Social Media Data Analytics: A Survey," DOI: 10.1016/j.chb.2018.08.039, Year: 2018.

[9] Meena Rambocas and Barney G. Pacheco "Online sentiment analysis in marketing research: a review" Journal of Research in Interactive Marketing, DOI: https://doi.org/10.1108/JRIM-05-2017-0030, Year: 2018.

[10] Oscar Araque, Carlos A. Iglesias, and Ganggao Zhu "A semantic similarity-based perspective of affect lexicons for sentiment analysis" Elsevier, Knowledge-Based Systems, PP: 346–359, Year: 2018.

[11] Devika M D, Amal Ganesh and, Sunitha C "Sentiment Analysis: A Comparative Study on Different Approaches" Fourth International Conference on Recent Trends in Computer Science & Engineering. PP: 44-49, Year: 2016

[12] Rudy Prabowo and Mike Thelwall "Sentiment analysis: A combined approach" Journal of Informetrics, PP-143-157, Year: 2009.