IJAREEIE

e-ISSN: 2278-8875
p-ISSN: 2320-3765

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

Volume 13, Issue 6, June 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.317

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.317|| A Monthly Peer Reviewed & Referred Journal |

||Volume 13, Issue 6, June 2024||

|DOI:10.15662/IJAREEIE.2024.1306015|

Sentiment Analysis on Amazon Customer Reviews

Abhishek Patel, Pankaj Yadav, Anurag Mishra, Nimish More, Dr. Sarika Bukkawar

B.E Student, Department of ECS, Shah and Anchor Kutchhi Engineering College, Chembur, Mumbai, India

B.E Student, Department of ECS, Shah and Anchor Kutchhi Engineering College, Chembur, Mumbai, India

B.E Student, Department of ECS, Shah and Anchor Kutchhi Engineering College, Chembur, Mumbai, India

B.E Student, Department of ECS, Shah and Anchor Kutchhi Engineering College, Chembur, Mumbai, India

Assistant Professor, Department of ECS, Shah and Anchor Kutchhi Engineering College, Chembur, Mumbai, India

ABSTRACT: In the realm of e-commerce, customer reviews play a pivotal role in shaping purchasing decisions. However, the sheer volume of feedback presents a formidable challenge for businesses seeking to extract actionable insights. Sentiment analysis emerges as a potential solution, automating the classification of reviews into positive or negative sentiments. This paper introduces a sentiment analysis tool developed using Python and Flask powered by the robust XGBoost model. The tool offers a user-friendly interface for analyzing individual reviews or bulk review data, enabling users to glean valuable insights into customer sentiments. Through automation, the tool streamlines the analysis process and empowers businesses to respond effectively to customer feedback, enhancing the overall user experience.

KEYWORDS: Sentiment analysis, e-commerce, customer reviews, XGBoost, Python, Flask

I.INTRODUCTION

In the digital age of online marketplaces like Amazon, customer reviews serve as a vital resource for informing purchasing decisions. However, the sheer volume of reviews poses a challenge for businesses to extract meaningful insights manually. To address this, sentiment analysis emerges as a powerful tool, automating the categorization of reviews into positive or negative sentiments [1]. This paper introduces a sentiment analysis tool developed for customers, leveraging Python and Flask to provide a user-friendly interface [2]. Powered by the XGBoost model, the tool achieves high accuracy in sentiment prediction, focusing solely on distinguishing between positive and negative sentiments [3]. With functionalities for both single review analysis and bulk review prediction, the tool enables users to gain actionable insights from large-scale review data, enhancing their understanding of customer perceptions and facilitating informed decision-making on the platform. Through automation, this tool not only streamlines the analysis process but also empowers businesses to respond effectively to customer feedback in real-time, ultimately enhancing the overall user experience.

Problem statement

In the realm of e-commerce, the exponential growth of customer reviews presents a formidable challenge for businesses like Amazon to manually extract actionable insights. To mitigate this issue, there's a pressing need for automated sentiment analysis tools capable of accurately categorizing reviews as positive or negative. This study aims to develop and implement a user-friendly sentiment analysis tool, leveraging Python and Flask, to streamline the process of extracting sentiments from large-scale review data.

By harnessing the power of the XGBoost model [4], the tool seeks to provide businesses with a scalable solution for gaining deeper insights into customer perceptions, facilitating informed decision-making, and enhancing overall user experience on the platform.

Problem solution

To address the challenge, our proposed approach harnesses the capabilities of Natural Language Processing (NLP) [5]. Specifically, we employ the XGBoost model, a robust rule- based sentiment analysis tool renowned for its efficacy in discerning text sentiment as positive or negative. By leveraging this model, we aim to swiftly and accurately assess the sentiments expressed within the reviews. This enables businesses to gain valuable insights into customer perceptions without manual intervention, facilitating quicker decision-making and response to feedback.

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.317|| A Monthly Peer Reviewed & Referred Journal |

||Volume 13, Issue 6, June 2024||

|DOI:10.15662/IJAREEIE.2024.1306015|

In addition to sentiment classification, our solution incorporates a visual representation of sentiments through a pie chart. This graphical depiction offers stakeholders a straightforward and intuitive overview of sentiment distribution across product reviews, encompassing positive and negative sentiments. This visualization empowers stakeholders to rapidly comprehend customer sentiments at a glance, facilitating informed decision-making and strategic response to feedback.

II.METHODOLOGY

The sentiment analysis methodology employed in this study consists of several sequential steps, as illustrated in the [Fig.1] flowchart. Initially, data collection was performed by retrieving raw customer review data from a reputable source, specifically Kaggle. Pandas, a powerful data manipulation library in Python, along with Porter Stemmer and part of speech tagging, was utilized to preprocess the acquired data, ensuring its readiness for subsequent analysis [6]. Following data collection and preprocessing, the next step involved text preprocessing, an essential task in sentiment analysis. Natural Language Toolkit (NLTK) was employed for tasks such as tokenization, stopword removal, and stemming, thereby enhancing the quality of the text data [7]. Additionally, the regular expression (re) library was utilized for pattern matching and replacement, facilitating further text cleaning and removal as needed [8].

Subsequently, sentiment classification was conducted using a machine learning model, specifically XGBoost. This model was trained and serialized using the pickle library, ensuring efficient loading and utilization during the classification process. Flask, a lightweight web framework, was utilized for handling HTTP requests and responses, enabling seamless integration of the sentiment analysis model into the application architecture [9].

For each customer review, the text features were transformed into a numerical format using techniques like CountVectorizer, enabling the model to process the data effectively. The transformed features were then scaled using a scaler, ensuring consistency in feature values across different reviews. Subsequently, sentiment probabilities were predicted using the trained XGBoost model, and sentiment classification was performed based on these probabilities, resulting in the categorization of reviews as either positive or negative. Furthermore, analysis and insights were derived from the predicted sentiments, and visualization techniques using the matplotlib.Pyplot library were employed to create informative pie charts representing sentiment results [10]. By following this comprehensive methodology, accurate sentiment analysis was conducted, enabling valuable insights into customer perceptions and facilitating data- driven decision-making

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.317|| A Monthly Peer Reviewed & Referred Journal |

||Volume 13, Issue 6, June 2024||

|DOI:10.15662/IJAREEIE.2024.1306015|

Flow chart

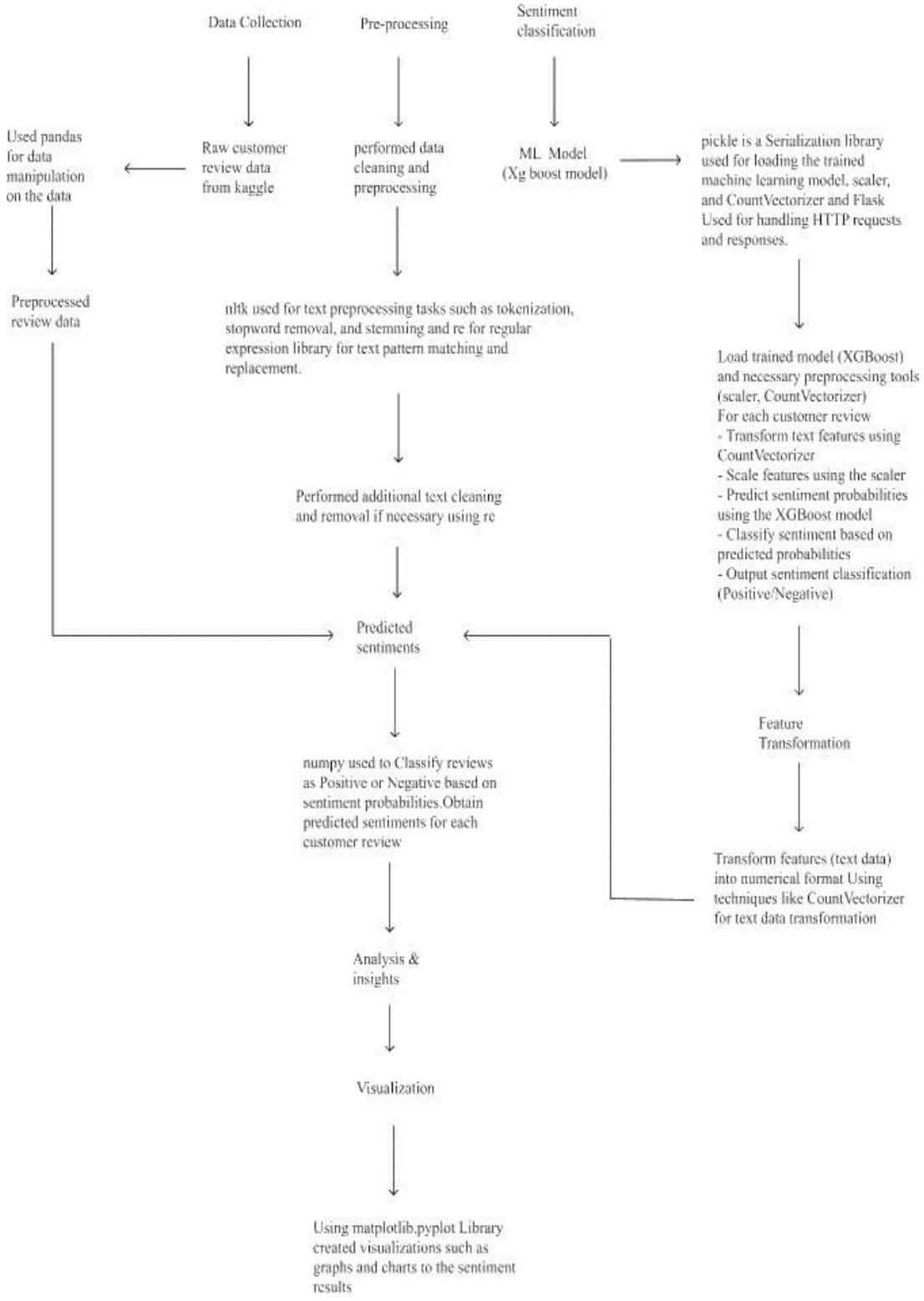


Fig. 1.Flowchart for sentiment analysis.

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.317|| A Monthly Peer Reviewed & Referred Journal |

||Volume 13, Issue 6, June 2024||

|DOI:10.15662/IJAREEIE.2024.1306015|

III. OUTPUTS

Fig. 2. Flask application for sentiment analysis

When initiating the Flask server for the sentiment analysis project, users are greeted with the user interface (UI) designed using Flask and Python, providing an intuitive platform for sentiment analysis as shown in Fig. 2.

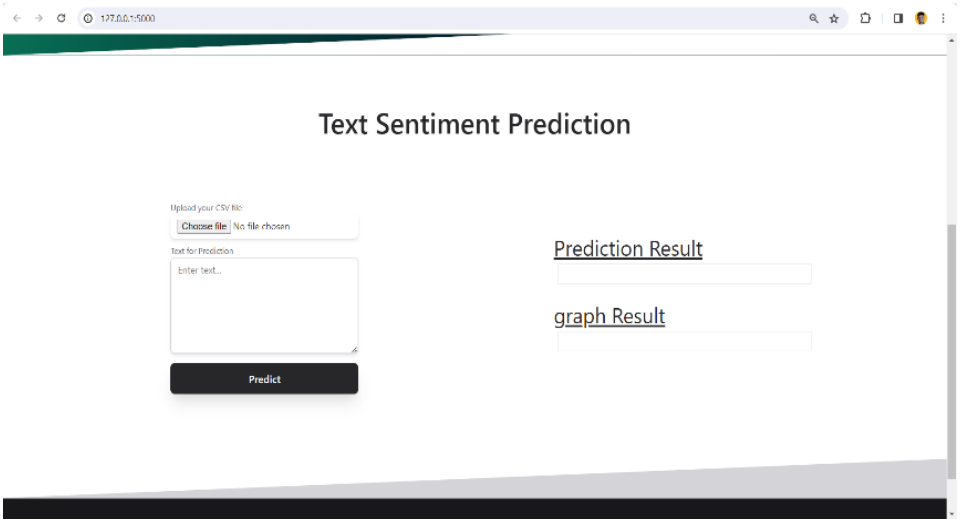


Fig. 3. Sentiment analysis for single and bulk text reviews

The sentiment prediction UI, powered by Flask and Python, facilitates both single review and bulk review analysis. Leveraging the XGBoost model and preprocessing methods, users can input individual reviews or upload bulk datasets for efficient sentiment analysis as shown in Fig. 3.

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.317|| A Monthly Peer Reviewed & Referred Journal |

||Volume 13, Issue 6, June 2024||

|DOI:10.15662/IJAREEIE.2024.1306015|

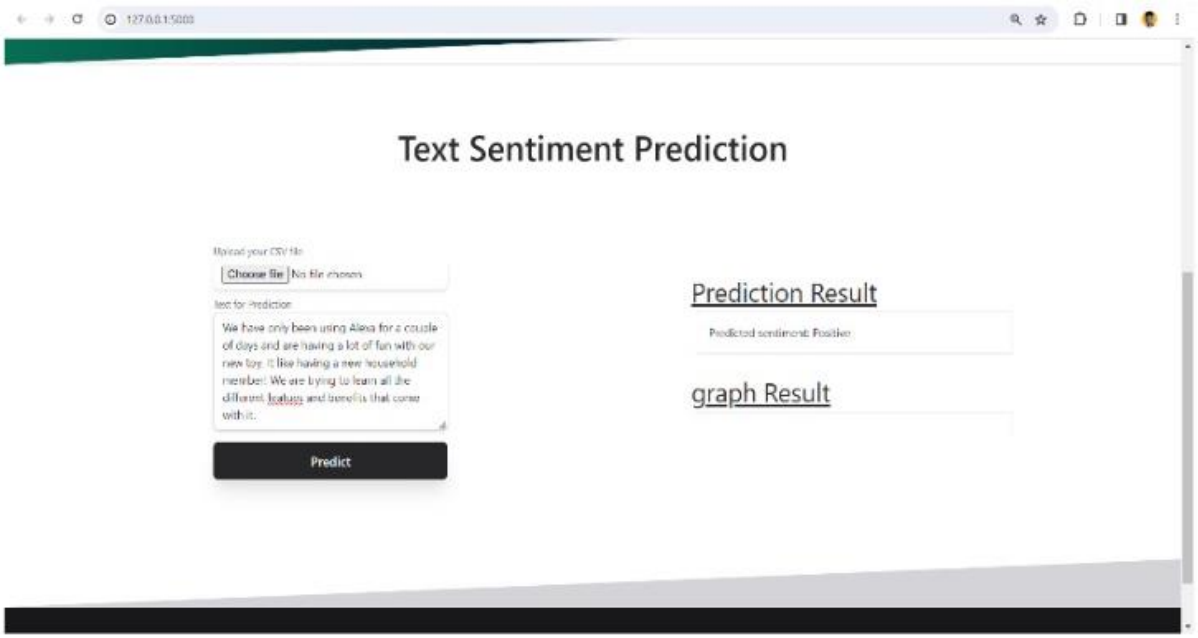


Fig. 4. Sentiment analysis for positive single prediction.

Utilizing the XGBoost model and text preprocessing techniques, the app accurately predicts the sentiment of single text reviews. Upon inputting positive text, the prediction result field reflects the sentiment as positive, demonstrating the efficacy of the sentiment analysis methodology as shown in Fig. 4.

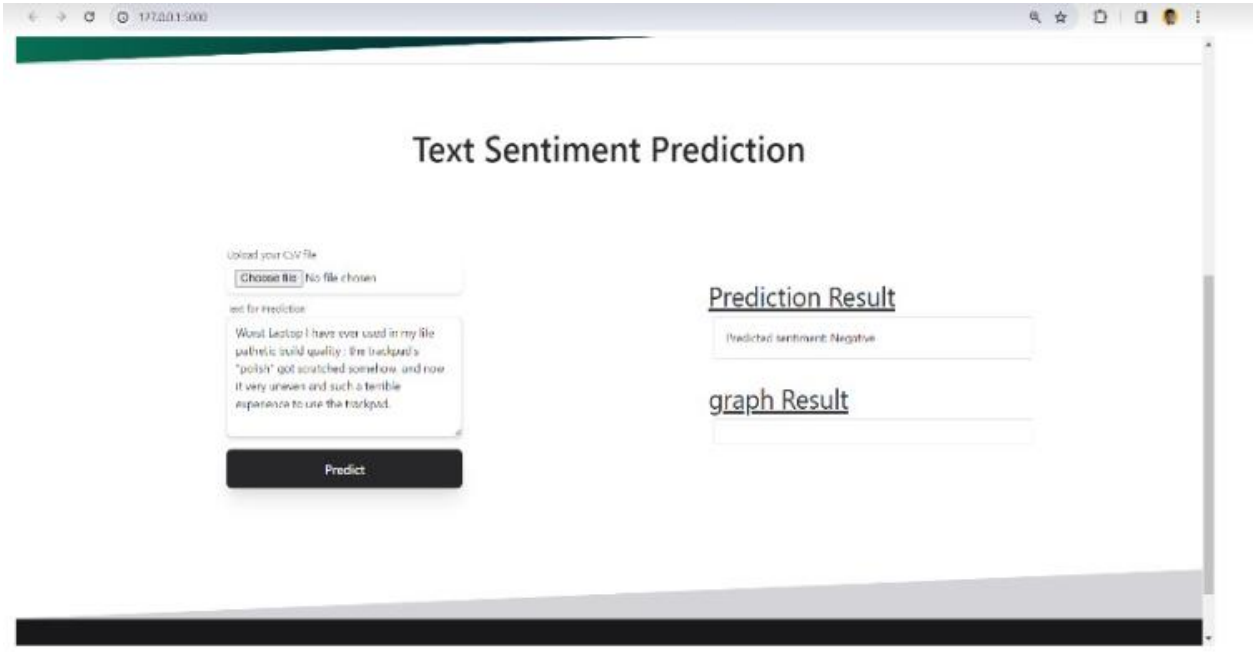


Fig. 5. Sentiment analysis for negative single prediction

Utilizing advanced text preprocessing techniques, including tokenization, stopword removal, and stemming with NLTK, the application accurately identifies negative sentiment in single text reviews. Upon inputting negative text, the prediction result field reflects the sentiment as negative as shown in Fig. 5.

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.317|| A Monthly Peer Reviewed & Referred Journal |

||Volume 13, Issue 6, June 2024||

|DOI:10.15662/IJAREEIE.2024.1306015|

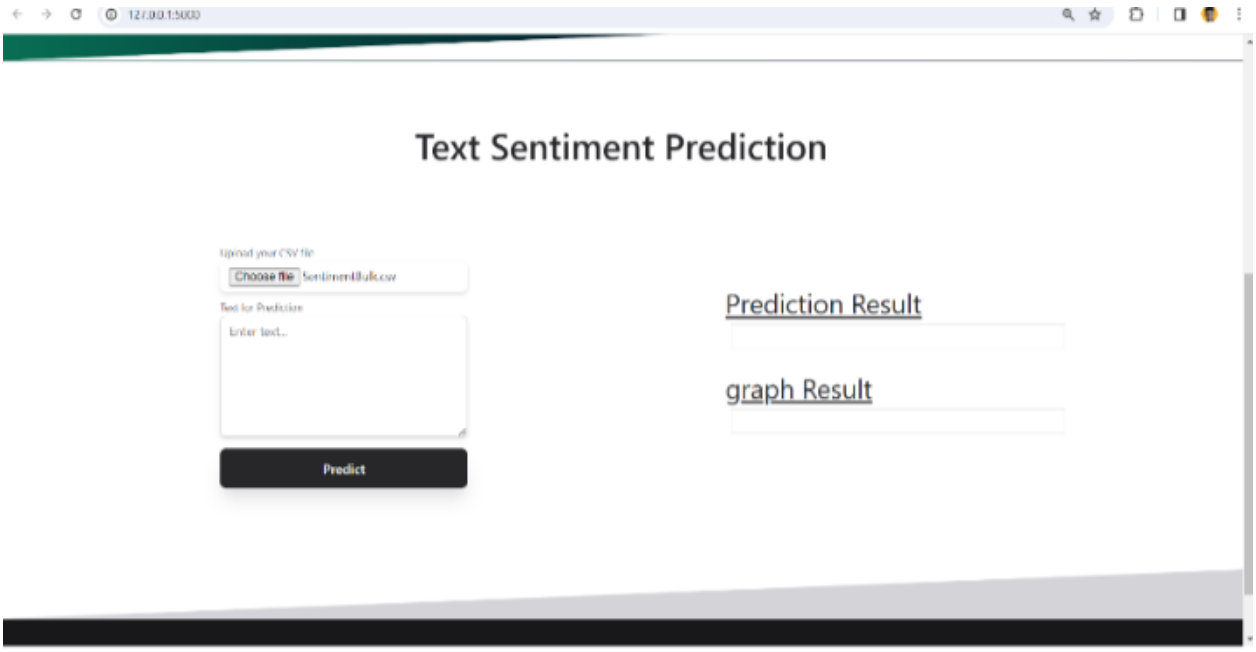


Fig. 6. Bulk analysis for bulk text reviews

By implementing CountVectorizer for text data transformation and XGBoost model for sentiment classification, the application enables bulk prediction functionality. Users can upload the sentimentBulk.csv file, facilitating efficient sentiment analysis of multiple reviews simultaneously as shown in Fig. 6.

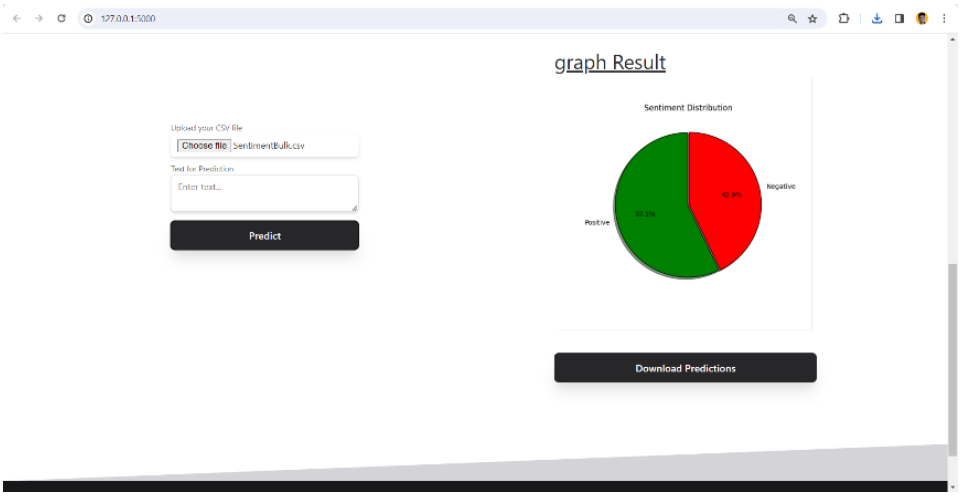


Fig. 7. Generated sentiment pie chart with a download prediction button

Utilizing the matplotlib.Pyplot library, the application generated a pie chart based on the sentiment predictions derived from the bulk reviews uploaded by the user. This visual representation offers stakeholders a clear overview of the distribution of positive and negative sentiments, enhancing the interpretability of the sentiment analysis results as shown in Fig. 7.

| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.317|| A Monthly Peer Reviewed & Referred Journal |

||Volume 13, Issue 6, June 2024||

|DOI:10.15662/IJAREEIE.2024.1306015|

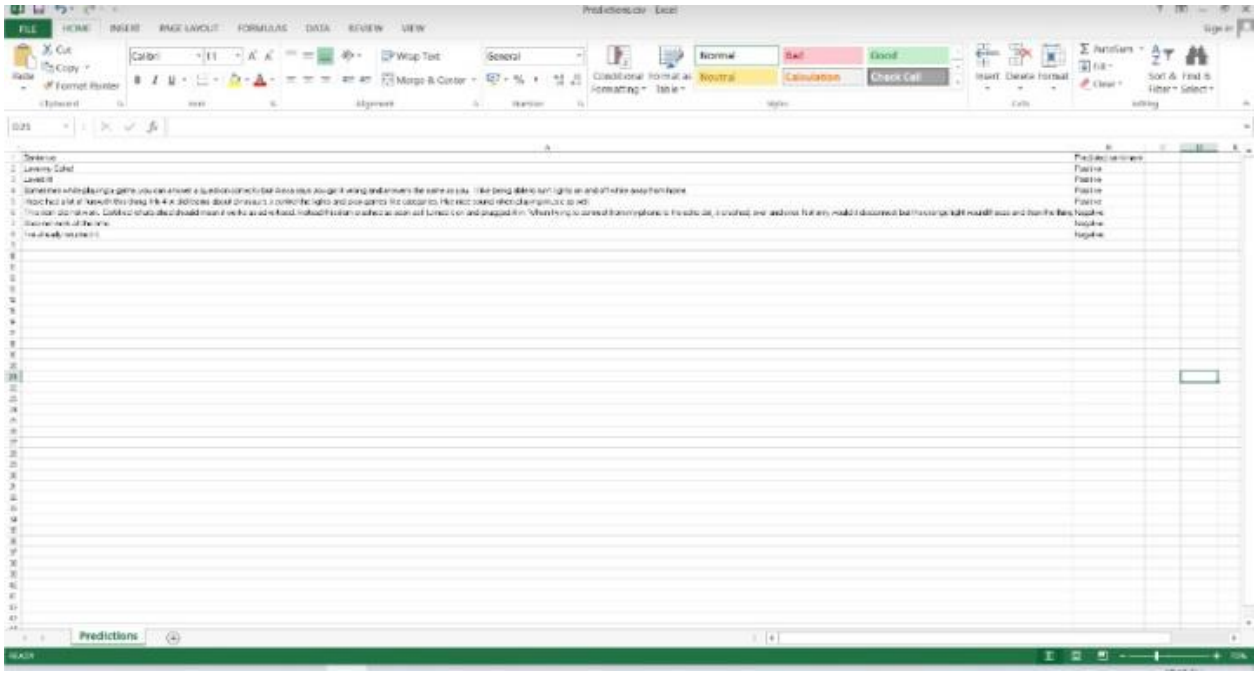


Fig. 8. Generated prediction.csv file which contain positive and negative reviews

Employing the serialization capability of the Python pickle library, the application generated a predictions.csv file upon user request, containing classifications of each review as either positive or negative. This functionality enhances user accessibility to sentiment analysis results, providing actionable insights into the sentiments expressed within the bulk reviews as shown in Fig. 8.

IV.RESULTS

In assessing the performance of our XGBoost model for sentiment analysis, we achieved an accuracy of approximately 94.18%. Leveraging this accuracy alongside the derived confusion matrix, which exhibited a distribution of [31, 854, 47, 13] for true positive (TP), true negative (TN), falsepositive (FP), and false negative (FN) predictions, respectively, we computed several performance indicators. Precision, measuring the accuracy of positive predictions, yielded a value of around 70.45%, while recall, indicative ofthe model's ability to identify positive instances, stood at approximately 39.74%. Furthermore, the model exhibited high specificity, accurately identifying negative instances at arate of roughly 98.50%. Balancing precision and recall, the F1score, calculated at approximately 50.89%, provides a holisticmeasure of the model's predictive capability. These metrics collectively underscore the robustness and effectiveness of our XGBoost model

V.CONCLUSION

In this study, we endeavored to develop a robust sentiment analysis system tailored for e-commerce platforms, with a specific focus on Amazon customer reviews. LeveragingPython's extensive ecosystem of libraries, we embarked on a comprehensive journey encompassing data preprocessing, machine learning model training, and web application deployment. Challenges emerged during various stages of the project, including data cleaning complexities, model serialization intricacies, and web application deployment hurdles. To overcome these challenges, a thorough strategy was adopted, utilizing pandas for data manipulation, NLTK for text preprocessing, and Flask for web application development. Furthermore, we addressed compatibility issues through diligent version management and resolved deployment complexities through iterative debugging and refinement. Ultimately, our project culminated in a sophisticated sentiment analysis tool capable of accurately predicting sentiments for customer reviews, thereby empowering businesses to glean valuable insights and enhance customer satisfaction in the dynamic landscape of e-commerce.

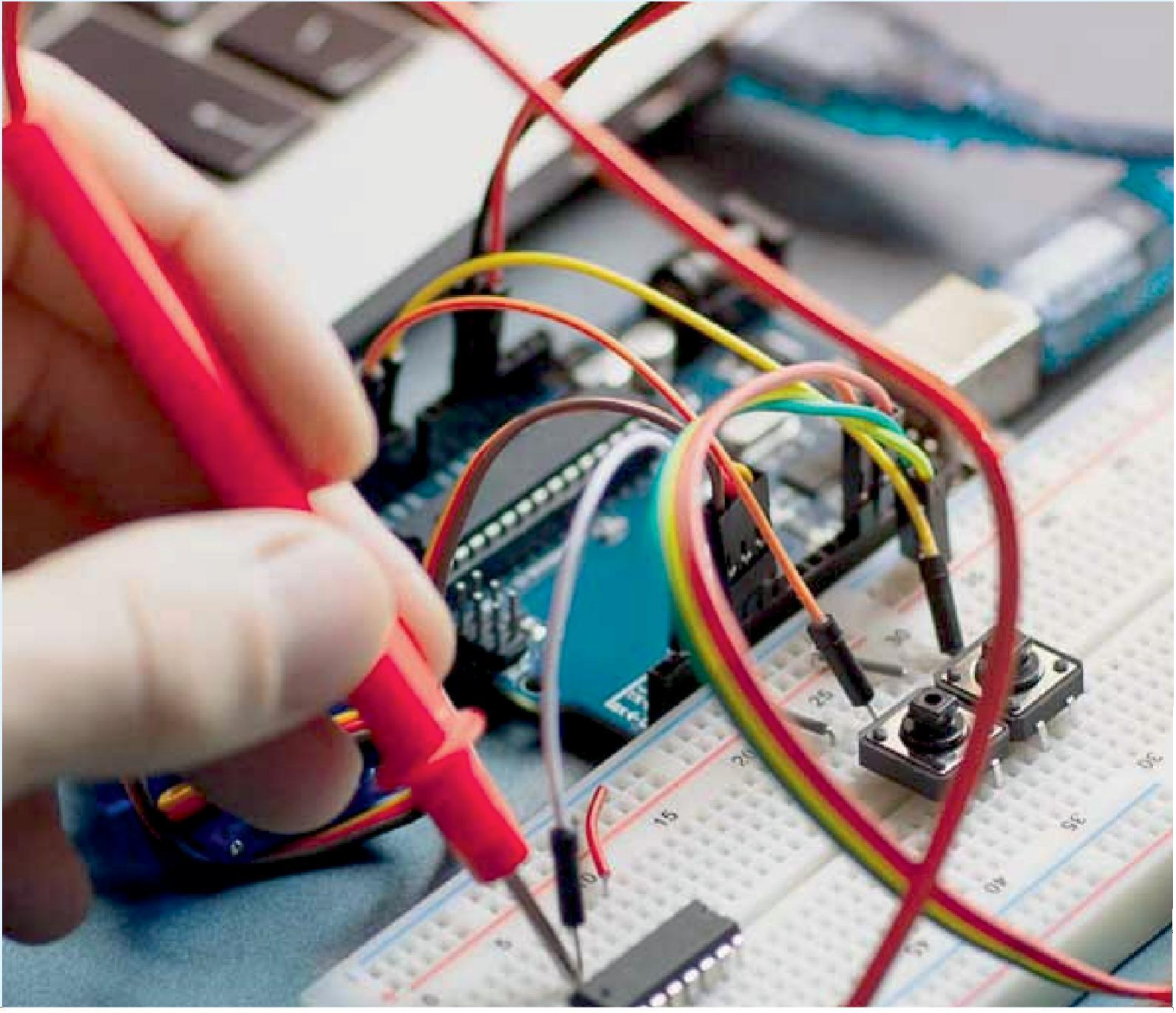
| e-ISSN: 2278 – 8875, p-ISSN: 2320 – 3765| www.ijareeie.com | Impact Factor: 8.317|| A Monthly Peer Reviewed & Referred Journal |

||Volume 13, Issue 6, June 2024||

|DOI:10.15662/IJAREEIE.2024.1306015|

REFERENCES

1. Binju saju, Siji jose, Amal anotny, "Comprehensive Study onSentiment Analysis", IEEE year 2020. https://ieeexplore.ieee.org/document/9213209
2. Prajval Sudhir, Varun Deshakulkarni Suresh. "Comparative study of various approach, application for sentiment analysis", IEEE year 2021
3. https://www.sciencedirect.com/science/article/pii/S2666285X210003 27?via%3Dihub
4. https://xgboost.readthedocs.io/en/stable/tutorials/model.html
5. Amina Samih, Abderrahim Ghadi, Abdelhadi Fennan. "Enhanced sentiment analysis based on improved word embeddings and XGboost", IJECE year 2022 https://https://www.researchgate.net/publication/367097372_Enhanced_sentiment_analysis_based_on_improved_word_embeddings_and_ XGboost
6. https://www.analyticsvidhya.com/blog/2021/06/nlp-sentiment- analysis/
7. Tirath Prasad Sahu, Sanjeev Ahuja, "Sentiment Analysis of MovieReviews", IEEE year 2021.
8. https://ieeexplore.ieee.org/document/7522583
9. https://www.nltk.org/
10. https://www.w3schools.com/python/python_regex.asp
11. https://flask.palletsprojects.com/en/3.0.x/
12. https://matplotlib.org/



INNO SPACE
SJIF Scientific Journal Impact Factor

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

International Journal of Advanced Research

in Electrical, Electronics and Instrumentation Engineering

9940 572 462 6381 907 438 ijareeie@gmail.com

www.ijareeie.com

Scan to save the contact details