Sentiment Analysis on Amazon Product Reviews

Namia

***Abstract*—In the dynamic e-commerce landscape, Amazon strives for seamless shopping experiences. However, extracting insights from vast textual data, especially concise reviews, poses challenges. This paper employs sentiment analysis, a natural language processing subfield, to categorize text into positive, negative, and neutral sentiments. Analyzing Amazon customer feedback aims to provide actionable insights and product recommendations. [1] The study presents a comparative analysis of machine learning models, utilizing techniques like Bag-of-Words, Tf-Idf, Word2Vec for vector representation, and training with classification algorithms such as Decision Trees, Naive Bayes, Random Forest, AdaBoost, Gradient Boosting, XGB Boosting, Multinomial Naive Bayes, and deep learning algorithms like RNN and LSTM.**

***Index Terms*—Sentiment analysis, Tf-Idf, Word2Vec**

I. Introduction

Nowadays, Social media significantly influences e-commerce product opinions, with customer reviews driving purchasing decisions. However, managing the vast and varied feedback poses challenges, particularly for platforms like Amazon. Effective sentiment analysis is crucial for extracting actionable insights from this feedback and improving user experiences. [2]

A. Problem Statement

The challenge lies in processing vast customer feedback to derive meaningful insights, hindering efforts to enhance service quality and product offerings. Robust sentiment analysis methodologies are needed to categorize sentiments and drive data-driven decisions.

B. Relevance of problem statement

Accurate sentiment analysis is essential for tailoring product recommendations and addressing user concerns, highlighting the need for effective methodologies to drive improvements in service quality and customer satisfaction.

C. Aim of the problem statement

This research aims to conduct sentiment analysis on Amazon product reviews to predict polarity. I used **Deep learning** along with traditional **Machine-learning** models like Logistic Regression, Na¨ıve Bayes, Random Forest, and LSTM with **diverse Feature extraction** methods to perform the sentiment analysis tasks to understand customer satisfaction on products.

II. Previous Works and Challenges Faced

Prior studies in sentiment analysis have primarily utilized conventional techniques such as Bag-of-Words and Tf-Idf for vector representation and classification. While effective in identifying basic emotions within text, these methods often overlook semantic nuances and overall tone in feedback. This study aims to address these limitations by exploring a wider array of factors, including semantic relationships within the text. By integrating deep learning methodologies like RNNs and LSTMs alongside traditional approaches, we seek to enhance the depth and accuracy of sentiment analysis. [3] One of the primary **challenges** encountered in previous research lies in the reliance on traditional methods, which may overlook subtle nuances in user sentiment. Additionally, the complexity of **analyzing semantic relationships and overall tone** poses a significant obstacle. To overcome these challenges, this study aims to leverage advanced deep learning techniques alongside traditional approaches, thereby striving for a more comprehensive understanding of user sentiment in text data.

III. Proposed Framework

A. Dataset

The dataset consists of 400,000 amazon reviews with two columns. **Label** which provides an high level understanding on whether the review is negative (0) or positive (1) and **Review** for Customer provided feedback.

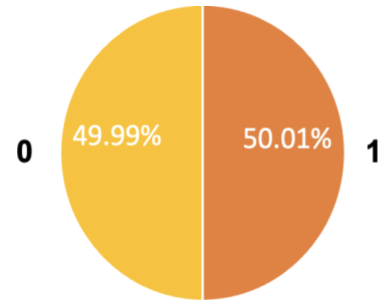

Fig. 1. Balanced dataset on positive and negative reviews

B. Approach

In this proposed paper, our objective is to enhance model accuracy by refining hyperparameters using **Grid Search CV** and implementing **cross-validation** testing across all models.

Additionally, we intend to augment the dataset size and incorporate deep learning methodologies such as **RNN** and **LSTM**, alongside **traditional classification** techniques. We will also leverage **word2vec** for clustering similar features, allowing us to capture semantic nuances within the chosen domain more effectively. This comprehensive approach will enable us to achieve a deeper understanding of the semantic intricacies within the domain under study. [3], [4]

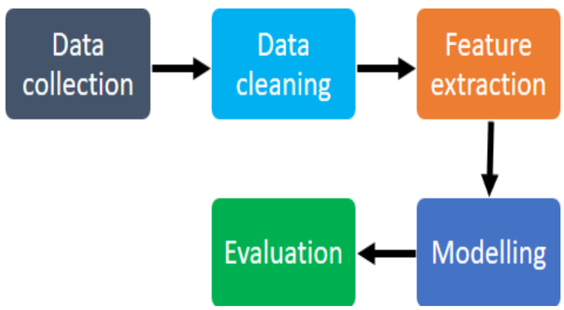

Fig. 2. NLP Flow chart

a) DATA CLEANING: To clean and prepare our data for analysis, the following steps were undertaken:

- **Tokenization**: I broke down the text into smaller parts, like individual words or smaller word fragments, to make it easier to work with.
- **Lowercasing**: I made sure that all the text was in the same format by converting everything to lowercase letters. This helps in treating words with different cases (like "good" and "Good") as the same word.
- **Removing punctuation**: I simplified the text and focused on the important words by getting rid of punctuation marks like commas, periods, and exclamation points.
- **Removing special characters and numbers**: I got rid of any special characters and numbers in the text because they usually don't add any meaning to the analysis.
- **Removing stop words**: I filtered out common words, known as stop words, such as "the," "is," and "and," because they appear frequently in the text but don't carry much significance for sentiment analysis. This helps reduce noise in the data and makes the analysis better.
- **Lemmatization over stemming**: Even though, lemmatization and stemming simplify words. Stemming choped off word endings to find a root word, but it created non-real words. Lemmatization considered the word's context and meaning, and gave a meaningful root word. Although lemmatization took longer, I prefer lemmatization for precise word normalization in this project as it gives more accurate results.
- **Vectorization**
  1) **TF-IDF**: I used TF-IDF to understand word importance in a document compared to a whole dataset. It highlighted words that are common in a particular review but rare across all reviews. TF-IDF helped me to identify unique terms specific to certain products or topics which made my sentiment analysis easier to understand the polarity of the review.
  2) **CountVectorizer:** It gave me the frequency of words in a review. I got the popular product features that attracts customers and common customer concerns based on word frequencies but it did not help me in classification of my reviews.
  3) **Word2Vec:** Using this technique I could find the connection between words which was useful for understanding the meaning of reviews. It represents words in a continuous vector space, allowing for detailed analysis of similarities and differences between product descriptions. Word2Vec is helpful for tasks like grouping similar reviews or finding related products based on their descriptions but not an added advantage for sentiment analysis.

After trying TF-IDF, CountVectorizer, and Word2Vec, I decided to use TF-IDF because it's better at highlighting important words in reviews and helps in deciding polarity . The TF-IDF model performed better than the others.

b) FEATURE EXTRACTION: Feature selection techniques are used to choose the most relevant features for model training while discarding irrelevant or redundant ones. This helps in reducing dimensionality and improving model performance. I utilize a word cloud to analyze the predominant features present in both positive and negative comments. [2]

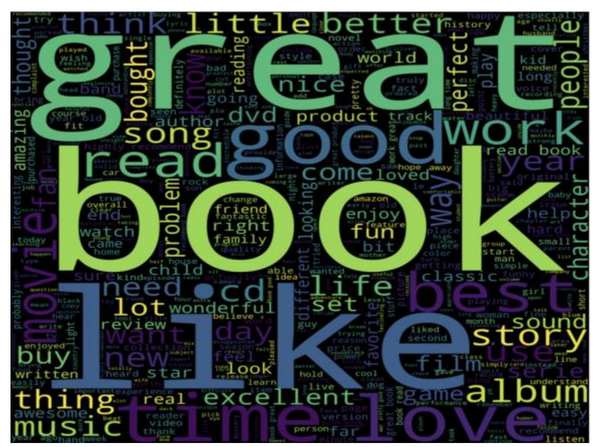

Fig. 3. Word cloud - Positive reviews

IV. Experiments Conducted

Initially, I experimented with six classification algorithms: DecisionTreeClassifier, RandomForestClassifier, MultinomialNB, AdaBoostClassifier, GradientBoostingClassifier, and XGBClassifier. I applied these algorithms to lemmatized data using two different vectorizers: Count Vectorizer and TF-IDF. Among these, the top three performing models were MultinomialNB_Lemmatized_CV, MultinomialNB_Lemmatized_TF-IDF, and AdaBoostClassifier_Lemmatized_TF-IDF. [4]

Later, I explored another vectorization technique called Word2Vec, along with incorporating deep learning models like RNN and LSTM. Deep learning models (RNN and LSTM) did not yield an optimal fit due to limitations in the dataset size and complexity. In the final implementation, I ran all classification and deep learning algorithms using all three vectorization techniques. Surprisingly, the top three models remained the same: MultinomialNB_Lemmatized_CV, MultinomialNB_Lemmatized_TF-IDF, and AdaBoostClassifier_Lemmatized_TF-IDF. [5]

To optimize these top models further, I employed GridSearchCV. The best model obtained was GSV_ADA_LEM_TFIDF, which used **AdaBoostClassifier with lemmatized data using TF-IDF Vectorizer**. It achieved a train accuracy of 0.86 and a test accuracy of 0.84.

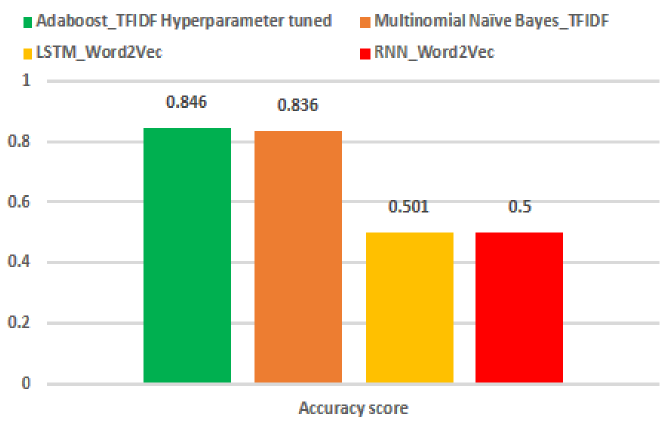

Fig. 4. Comparison of models (BEST AND WORST)

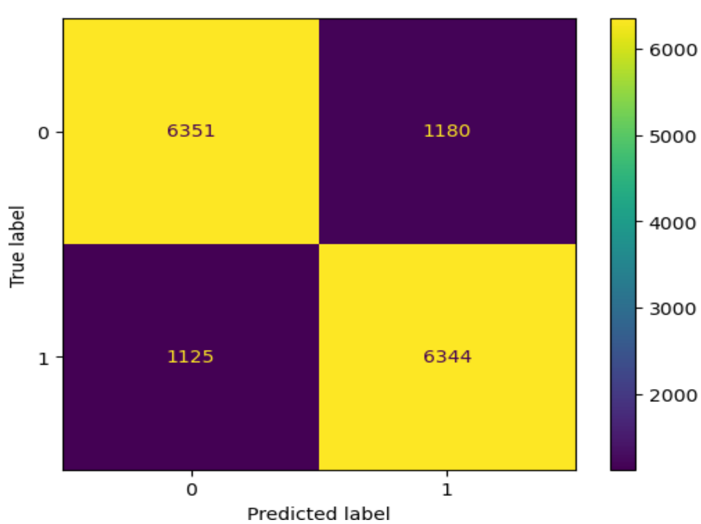

Fig. 5. Confusion Matrix of the Best fit model

The above confusion matrix provides an overview on how accurately my best model **(HYPERPARAMETER TUNED ADABOOST CLASSIFIER WITH TF-IDF)** works in predicting the positive and negative sentiments on the customer review.

V. Conclusion

This sentiment analysis project has successfully implemented a robust recommendation system based on user feedback and reviews. Through categorizing sentiments into positive, negative, or neutral categories, the system effectively determines overall user sentiment. For positive sentiments, the system suggests products or services with positive reviews, aiming to reinforce positive experiences and enhance user satisfaction. As for negative sentiments, the system recommends solutions to address user concerns and improve the user experience. Also, the recommendation system incorporates personalization and contextual relevance factors to tailor recommendations to individual users' preferences and specific scenarios. Continuous improvement mechanisms ensure that the system adapts to evolving user sentiments, providing increasingly relevant and valuable suggestions over time, thereby enhancing overall user satisfaction and engagement.

VI. Future Scope

In the future, enhancements in sentiment analysis could involve refining deep learning models like RNNs and LSTMs through architecture tuning and attention mechanisms. Integration of pre-trained language models like BERT via transfer learning for better accuracy and generalization is also on the horizon. Exploring semantic analysis techniques to understand subtleties such as sarcasm and cultural context holds promise for improving sentiment interpretation. Additionally, developing methods for analyzing sentiment across multiple modalities (text, images, audio) could provide a more comprehensive understanding of user feedback. Finally, implementing personalized recommendations and marketing strategies based on sentiment analysis insights could significantly enhance user experience and engagement. [6], [7]

References

[1] R. M. R. G. Malhar Anjaria, "A novel sentiment analysis of social networks using supervised learning," *Journal of Theoretical and Applied Information Technology*, vol. 4, 2013.

[2] S. Maurya and V. Pratap, "Sentiment analysis on amazon product reviews," in *2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON)*, vol. 1, 2022, pp. 236–240.

[3] Pankaj, P. Pandey, Muskan, and N. Soni, "Sentiment analysis on customer feedback data: Amazon product reviews," in *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, 2019, pp. 320–322.

[4] Pankaj, P. K. Pandey, Muskan, and N. Soni, "Sentiment analysis on customer feedback data: Amazon product reviews," *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, pp. 320–322, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:204232170

[5] A. Alsayat, "Improving sentiment analysis for social media applications using an ensemble deep learning language model," *Arabian Journal for Science and Engineering*, vol. 47, no. 2, pp. 2499–2511, 2022.

[6] A. Rashid and C.-Y. Huang, "Sentiment analysis on consumer reviews of amazon products," *International Journal of Computer Theory and Engineering*, vol. 13, pp. 35–41, 01 2021.

[7] K. Richard, H. Rim, and H. H. Choi, "Fine-tuning bert on twitter and reddit data in luganda and english," 12 2023.