Sentiment Analysis of MacBook Amazon Reviews Using NLP Techniques

Abstract

This paper presents a comprehensive analysis of MacBook Air Amazon reviews using various Natural Language Processing (NLP) techniques. The primary objective is to categorize the reviews into three sentiment groups positive, neutral, and negative. This will give an in-depth understanding of what customers anticipate. The project also aims to extract significant aspects from the reviews, providing information about particular characteristics that can have impact on customer satisfaction. Models such as BERT, Naive Bayes, Support Vector Machine (SVM), and Logistic Regression were used for this project. Evaluation metrics like accuracy, precision, recall, and f1-score were used to assess each model's performance. The selection of Logistic Regression and Naive Bayes was based on their ease of use and efficiency, while SVM was selected due to its resilience in high-dimensional environments and BERT's enhanced ability to comprehend contextual data. Aspects like camera, screen, battery, performance service were identified as critical factors that have a significant impact on customer satisfaction. The analysis also reveals overall sentiment trends in MacBook reviews, revealing both positive and negative sentiments expressed by customers. These results have important ramifications for retailers and manufacturers looking to improve customer service tactics and product features. This study offers a nuanced understanding of customer sentiments and practical insights for improved decision-making in product development and marketing by utilizing cutting-edge NLP techniques.

1.Introduction

Understanding customer opinions and enhancing goods and services requires an analysis of customer reviews, which offer vital insights into consumer satisfaction and product quality. The goal of this project is to categorize sentiments and determine key aspects affecting customer satisfaction by examining reviews for the MacBook Air. With applications in customer service and marketing, sentiment analysis, also known as opinion mining, involves categorizing reviews into positive, neutral, and negative sentiments. This project evaluates the success rates of various methods for sentiment classification using Natural Language Processing (NLP) techniques and machine learning models like Logistic Regression, Naive Bayes, Support Vector Machine (SVM), and BERT. The project uses these models to identify frequently mentioned aspects such as performance, MacBook, camera, price and quality to highlight overall sentiment trends. Manufacturers and retailers can improve customer satisfaction and loyalty by implementing the insightful insights obtained from the findings to improve product features and customer service tactics.

2. Background and Literature Review

The present era is a fast-paced world where the products are purchased only after going through reviews .Customer reviews always provide critical insights into consumer satisfaction and product quality, making their analysis essential for understanding customer opinions and improving products and services. Sentiment analysis, also known as opinion mining, involves categorizing reviews into positive, neutral, and negative sentiments. This technique has applications in various domains such as marketing, customer service, and product development, enabling businesses to make data-driven decisions. The growth of e-commerce has led to an abundance of online reviews, which present both opportunities and challenges for sentiment analysis. Traditional approaches like lexicon-based methods rely on predefined dictionaries of positive and negative words but often fail to capture context and nuances in language. Machine learning models such as Logistic Regression and Naive Bayes have been widely used for sentiment classification, providing a more data-driven approach by learning patterns from labeled datasets. However, these models may struggle with complex language structures and context. More advanced techniques, including Support Vector Machine (SVM), offer improved handling of high-dimensional data but can be computationally intensive.

The advent of deep learning and transformer-based models like BERT (Bidirectional Encoder Representations from Transformers) has revolutionized NLP by providing a deeper understanding of context and semantics in text, significantly enhancing the accuracy of sentiment analysis. Previous studies have demonstrated the effectiveness of these models in various applications. For instance, lexicon-based methods have been used for basic sentiment classification tasks, while machine learning models like Logistic Regression and Naive Bayes have been applied to larger, more complex datasets. SVM has shown robustness in handling high-dimensional feature spaces, making it suitable for text classification tasks. BERT, with its ability to capture bidirectional context, has set new benchmarks in NLP tasks, including sentiment analysis, by understanding the nuances and subtleties of human language. This project focuses on analyzing Amazon reviews for the MacBook Air, by implementing combination of these models to evaluate their effectiveness in sentiment classification and to identify key aspects mentioned by customers. By leveraging these NLP techniques, the project aims to provide valuable insights that can help manufacturers and retailers enhance product features and customer service strategies. This comprehensive approach not only highlights overall sentiment trends but also uncovers specific aspects such as product quality, performance, and customer service that significantly impact customer satisfaction.

3. Research Design and Methodology

3.1 Data Collection

The dataset used in this project was sourced from Kaggle that contains reviews of the MacBook Air. The dataset includes various fields such as product details, review date, rating, and review body focusing on reviews for MacBook products.

Each review includes information such as the reviewer's ID, product ID, rating, review text, and the date of the review. The dataset was preprocessed to remove any duplicates or irrelevant entries, resulting in a clean dataset suitable for analysis.

3.2 Data Preprocessing

Data preprocessing is crucial for preparing text data for analysis. The preprocessing steps are

- **Cleaning Text:** Removing special characters and converting text to lowercase.
- **String Conversion:** Converting the ‘body’ column to string type
- **Tokenization:** Splitting text into individual words.
- **Lemmatization:** Reducing words to their base forms to ensure uniformity.
- **Stopword Removal:** Eliminating common words that do not contribute to the sentiment analysis.

3.3 Defining Label

Labels for sentiment classification were derived from the ratings of the reviews. Reviews with 4-5 stars were labeled as positive, 3-star reviews as neutral, and 1–2-star reviews as negative. These labels provided the basis for training and evaluating the classification models.

3.4 Vectorization

After deriving the labels, data preprocessing and dataset splitting for machine learning tasks are implemented. It begins by importing necessary libraries such as pandas for data manipulation and scikit-learn for machine learning functionalities. Then, it initializes a TF-IDF vectorizer, configuring it to consider a maximum of 5000 features (words or phrases) including unigrams and bigrams. The reviews data are stored in the 'clean text' column of a data frame are transformed into a TF-IDF matrix, while the corresponding labels are extracted. Finally, the dataset is split into training and validation sets with 80% of the data used for training and 20% for validation. This division enables the training of a machine learning model on one portion of the data and the evaluation of its performance on another, facilitating effective model development.

4. Exploratory Data Analysis (EDA)

EDA involves visualizing data to understand its distribution and key characteristics. Key analyses performed include:

- **Rating Distribution**: Analyzing how ratings are distributed over the years.
- **Word Clouds**: Generating word clouds to visualize common terms in positive and negative reviews.
- **Sentiment Distribution**: Categorizing reviews into positive, neutral, and negative sentiments based on their ratings.

5. Models and Implementation

5.1 Logistic Regression

Logistic Regression is a linear model for binary classification that can also be extended to multi-class classification. It was implemented using the Scikit-learn library, and the model was trained on the feature vectors. The performance was evaluated using accuracy, precision, recall, and F1-score. The model achieved an accuracy of 86%.

5.2 Naive Bayes

Naive Bayes is a probabilistic classifier based on Bayes' theorem. The Multinomial Naive Bayes variant was used, which is suitable for text classification. The model was trained and evaluated similarly to Logistic Regression, and results were compared. The model achieved an accuracy of 86%.

5.3 Support Vector Machine (SVM)

SVM is a powerful classifier that finds the hyperplane that best separates the data into classes. A linear kernel was used for SVM, and the model was trained and evaluated using the same metrics as the previous models. It provided robust performance with an accuracy of 89%, highlighting its strength in handling high-dimensional data.

5.4 BERT (Bidirectional Encoder Representations from Transformers)

BERT is a state-of-the-art transformer model for NLP tasks. The pre-trained BERT model was fine-tuned on the sentiment analysis task using the Hugging Face Transformers library. The model's performance was evaluated on the validation set, and results were compared with other models. The model significant accuracy of 91%, demonstrating its superior capability in understanding contextual information.

6. Topic Modeling Latent Dirichlet Allocation (LDA)

It is used to uncover most common words used in the reviews. It is executed in the following steps

- **Cleaning Title Column**: Title column was cleaned and tokenized.
- **Vectorization**: Title column was vectorized using Count Vectorizer.
- **LDA Model**: Fitted to the vectorized text data to identify topics.
- **Top Words**: Identified top words for each topic to understand the main themes discussed by consumers.

7. Aspect-Based Analysis

Aspect-based analysis was conducted to analyze frequency, sentiments for specific aspects of the MacBook Air.

- **Aspect Extraction**: Nouns and noun phrases were extracted from the reviews.
- **Frequency Analysis**: The frequency of each aspect was calculated.
- **Sentiment Scoring**: Sentiment scores were calculated for each aspect using the Sentiment Intensity Analyzer (SIA).

8. Results and Discussion

8.1 Evaluation Methods

The performance of each classification model was evaluated using metrics such as accuracy, precision, recall, and F1-score. A 10-fold cross-validation was applied to ensure reliable estimates of model performance.

8.1.2 Logistic Regression

Logistic Regression showed reasonable performance with an accuracy of 86%. The precision, recall, and F1-score were higher for the positive class compared to the neutral and negative classes, indicating a bias towards predicting positive reviews.

8.1.3 Naive Bayes

Naive Bayes performed similarly to Logistic Regression, with an accuracy of 86%. The model struggled with neutral reviews, as indicated by the lower precision and recall for the neutral class.

8.1.4 Support Vector Machine (SVM)

SVM outperformed Logistic Regression and Naive Bayes, achieving an accuracy of 89%. The precision, recall, and F1-score for the negative and positive classes were higher, while the neutral class still posed a challenge.

8.1.5 BERT

The BERT model achieved the highest accuracy of 91%. It showed superior performance across all metrics, particularly excelled in identifying positive reviews. The model struggled with neutral reviews, like other models.

8.2 Confusion Matrix and Classification Report

Confusion matrices were plotted for each model to visualize the classification performance. The BERT model's confusion matrix showed fewer misclassifications compared to the other models. Classification reports provided detailed insights into precision, recall, and F1-score for each class.

9. Limitations

The study faced limitations in accurately classifying neutral reviews, as these often contain mixed sentiments and subtle nuances that are challenging for models to interpret correctly. While high performance was achieved for positive and negative reviews, the neutral category proved more difficult due to its inherent ambiguity.

10.Future Scope

Future work will address these limitations by incorporating additional features and enhancing the detection of implicit sentiments. This will involve advanced feature engineering techniques, such as using sentiment lexicons, context-specific embeddings, and syntactic dependencies, to better capture the complexities of neutral sentiments. Furthermore, improving the models' ability to infer sentiments from context and expanding the dataset to include a more diverse range of reviews will also be critical areas of focus.

10. Conclusion

This paper presents a comprehensive approach to sentiment analysis of MacBook Amazon reviews using various NLP techniques. The study highlights the effectiveness of different classification models and provides insights into customer satisfaction. The study also successfully extracted aspects from the reviews with the help of topic modelling and further analyzed sentiments of those aspects. Despite certain limitations, the results demonstrate the potential of sentiment analysis in understanding consumer opinions and improving product development.

11. References

(1)Diekson, Z. A., Prakoso, M. R. B., Putra, M. S. Q., Syaputra, M. S. A. F., Achmad, S., & Sutoyo, R. (2023). Sentiment analysis for customer review: Case study of Traveloka. Procedia Computer Science, 216, 682–690. https://doi.org/10.1016/j.procs.2022.12.184

(2)Selvakumar, B., & Lakshmanan, B. (2022). Sentimental analysis on user's reviews using BERT. Materials Today: Proceedings, 62, 4931–4935. https://doi.org/10.1016/j.matpr.2022.03.678

(3)Singh, A., & Kumar, P. (2022). Aspect based sentiment analysis for customer reviews. IEEE Conference Publication. https://ieeexplore.ieee.org/document/9761302