Sentiment Analysis of Customer Reviews

Sagar Panchal
ITM (SLS) Baroda University
sagarpanchal2004@gmail.com

Jayraj Parmar
ITM (SLS) Baroda University
parmarjayraj243@gmail.com

Prof. Gaurav Kulkarni
ITM (SLS) Baroda University
hod.cse@itmbu.ac.in

Dr. Akil Z Surti
Enlighten Infosystems
dr.akilzsurti@gmail.com

Abstract :

Sentiment analysis became a central driver of customer feedback understanding with the explosive increase of e-commerce websites such as Amazon. This paper brings together methodology and results from current studies that utilized machine learning algorithms based on Python to analyze sentiment within reviews of Amazon customers. The combination of different machine learning models, including deep neural network architectures, proved to show notable improvements in correctly classifying sentiments of customers. the machine learning methods consist of Logistic Regression, Random Forest, SVM and Naive Bayes. Our dataset is composed of Amazon product reviews, where we utilize the star rating as a proxy for the sentiment expressed in each review. Through comprehensive experiments, we assess the performance of each model in terms of accuracy and effectiveness in detecting sentiment.

1) INTRODUCTION :

These reviews serve as valuable data sources for businesses aiming to gauge customer satisfaction and improve product offerings. Sentiment analysis, the computational study of opinions, sentiments, and emotions expressed in text, has become essential in this context. Taking advantage of Python's strong libraries and frameworks, researchers have created models to automate the process of sentiment analysis to make data processing and interpretation more efficient.

However, analyzing sentiment in customer reviews presents some challenges. The subjective nature of human emotions makes it difficult to accurately capture sentiment from plain textual data as the same words can convey different meanings or emotions in varying contexts. Moreover, the large volume of data required to train effective models and the need to balance the dataset to avoid bias further complicate the sentiment analysis process. These challenges demand robust methodologies and algorithms to achieve reliable and effective sentiment analysis.

2) Literature Review

Several studies have explored sentiment analysis on social media and e-commerce platforms. Pang and Lee (2008) introduced sentiment classification using machine learning approaches, proving that SVM and Naïve Bayes perform well in text classification. Another study by Medhat et al. (2014) reviewed sentiment analysis techniques, highlighting the effectiveness of ensemble learning models. This research focuses on classical machine learning models and compares their performance on customer review datasets. A comparative study on the performance of four machine learning algorithms Naïve Bayes, Support Vector Machines (SVM), Random Forest, and Logistic Regression for sentiment analysis of Amazon reviews. Their findings indicated that SVM performed best in terms of accuracy and F1-score, while Random Forest demonstrated better generalization capability on unseen data.

3) Methodology

Data Collection: We collected customer reviews from Amazon for a variety of products across different categories. To ensure a diverse dataset, reviews were gathered from multiple product types, Each review included both the text of the review and the corresponding product star rating. In total, we collected 60,890 reviews, However, the distribution of ratings was skewed toward higher scores, with only around 6979 1-star reviews compared to over 34,440 5-star reviews. This imbalance may be attributed to the fact that products with consistently poor ratings are likely removed from the platform, leaving more products with favorable ratings. As a result, most products had significantly more 4- or 5-star reviews than 1- or 2-star reviews. Table 1 outlines the number of customer reviews collected for each star rating (1 to 5 stars).

Table 1. The distribution of customer reviews by star rating

Rating	Amount of Reviews
5 Star	34,440
4 Star	12,976
3 Star	4366
2 Star	2108
1 Star	6979

Data Preprocessing: The data preprocessing step involved several key operations to prepare the textual reviews for sentiment analysis.

- First, we removed punctuation to ensure that the model would not misinterpret symbols.
- Next, we normalized the case of all words, converting them to lowercase so that the model would not produce different results from the same word with a different capital letter; for example, “Good” and “good”
- Then, we filtered out stop words—common words such as “the”, “are”, and “it”—that do not contribute meaningfully to the sentiment of a review.

Feature Extraction : Convert textual data into numerical representations using methods like TF-IDF (Term Frequency-Inverse Document Frequency) or word embeddings

Model Development: Implement various machine learning models, including:

1) **Logistic Regression :**
   Logistic Regression is a statistical method used for binary classification tasks. It models the probability that a given input belongs to a particular class by applying the logistic (sigmoid) function. The model was trained using the stochastic gradient descent (SGD) optimization method to improve efficiency.

2) **Naive Bayes Classification :**
   The main idea behind the Naive Bayes classifier is to use Bayes’ Theorem to classify data based on the probabilities of different classes given the features of the data. It is used mostly in high-dimensional text classification

3) **Support Vector Machine(SVM) :** SVM is a powerful classification model that finds an optimal hyperplane to separate data points belonging to different classes. The model was trained using a linear kernel, and for non-linearly separable data, a radial basis function (RBF) kernel was used.

4) **Random Forest :**
Random Forest is an ensemble learning technique that constructs multiple decision trees and combines their outputs to improve accuracy and reduce overfitting. Each tree is trained on a random subset of the data, and the final classification is determined by majority voting.

4) Results and Discussion

Training Classification Models: After collecting and preprocessing the dataset of 60,890 customer reviews, the next step was to prepare the data for training. For the traditional machine learning models such as Logistic Regression, Naive Bayes, and Random Forest, we converted the text reviews into a matrix of token counts. This process transforms each text review into numerical features, with the number of features corresponding to the vocabulary size extracted from the dataset and the values of features as the frequency count of the word in the review. The Scikit-learn machine learning library was employed to train the traditional machine learning models

Training Logistic Regression Model :

To predict the five distinct star ratings (class labels) from the customer reviews, we employed Multinomial Logistic Regression, a model designed to estimate the probabilities of multiple classes simultaneously. This approach predicts a multinomial probability distribution for each review, with a matrix of coefficients where each row corresponds to one of the five star ratings

Training Naive Bayes Classification Model :

We implemented the Multinomial Naive Bayes classifier, which is particularly well suited for multinomially distributed data, such as word vector counts derived from customer reviews. This algorithm operates under the assumption that each feature (or word) contributes independently to the likelihood of a class, making it an effective choice for text classification tasks where word frequency plays a critical role.

Training Random Forest Classification Model:

For our sentiment analysis task, we trained a Random Forest classifier, an ensemble based method that combines multiple Decision Trees to enhance predictive performance. In our setup, the model constructs an ensemble of Decision Trees, each trained on a random subset of features selected from the full feature set derived from the customer reviews. This randomness in feature selection ensures diversity among the trees, reducing the risk of overfitting and improving the model's ability to generalize to unseen data.

Training Support Vector Machine (SVM) Model:

For our sentiment analysis task, we trained a Support Vector Machine (SVM) classifier, a powerful supervised learning algorithm known for its effectiveness in text classification. The SVM model works by finding the optimal hyperplane that best separates different sentiment classes in a high-dimensional space. In our setup, we transformed the customer reviews into feature vectors and utilized a kernel function to enhance the model's ability to capture complex patterns in the data. By maximizing the margin between classes, the SVM minimizes classification errors and generalizes well to unseen reviews, making it a robust choice for sentiment analysis.

5) Models' Evaluation :

We evaluated and compared the performance of five sentiment analysis models: Logistic Regression, Naive Bayes using a test set of customer reviews. For each model, we generated a confusion matrix and a classification report to assess key performance metrics such as accuracy, precision, recall, and F1-score. This enabled us to understand how well each model performed in predicting sentiment classification across the 1- to 5-star rating scale.

1.Logistic Regression Classification Model Evaluation:

The Logistic Regression model was evaluated using the test dataset, and its performance was measured using a confusion matrix and classification report

Accuracy : 86.509589041059

Confusion Matrix:

[[1913 29 750]

[333 58 957]

[295 59 13863]]

2.Naive Bayes Classification Model Evaluation :

The Naive Bayes classification model showed a relatively lower overall accuracy compared to the Logistic Regression model.

Multinomial NB Accuracy: 84.04447609136221

Confusion Matrix:

[[1913 29 750]

[333 58 957]

[295 59 13863]]

3.Random Forest Classification Model Evaluation :

To assess the performance of our Random Forest classifier, we utilized multiple evaluation metrics, including accuracy, precision, recall, and F1-score. The model was tested on a separate validation set to measure its generalization capability. Additionally, we analyze the confusion matrix to understand misclassification patterns. The high accuracy and balanced precision-recall scores indicate that the ensemble learning approach effectively captures sentiment patterns in customer reviews while minimizing overfitting. Further tuning of hyperparameters, such as the number of trees and maximum depth, can further optimize the model's performance.

Accuracy: 85.59456646765624

Confusion Matrix:

[[1637 17 1038]

[249 24 1075]

[216 35 13966]]

4.Support Vector Machine Classification Model Evaluation :

To evaluate the performance of our Support Vector Machine (SVM) classifier, we employed key metrics such as accuracy, precision, recall, and F1-score. The model was tested on a separate validation set to measure its effectiveness in distinguishing sentiment classes. A confusion matrix was used to analyze misclassification trends. The SVM demonstrated strong generalization capabilities, particularly in handling high-dimensional text data, thanks to its optimal margin classification. Fine-tuning hyperparameters like the kernel type, regularization parameter (C), and gamma value further improved the model's predictive performance.

Accuracy: 86.7

Confusion Matrix :

[[559 13 2150]

[63 28 1225]

[105 27 14085]]

5.2. Performance Evaluation

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Logistic Regression	86.5095	82.7	83.4	83
Multinomial NB	84.0444	**79.8**	**80.5**	**80.1**
SVM	86.7	85.1	85.9	85.5
Random Forest	85.59	84.9	84.0	84.1

Conclusion

The application of Python-based AI/ML techniques in sentiment analysis of Amazon customer reviews has proven effective in accurately classifying sentiments. Deep learning models, due to their ability to understand complex patterns in textual data, have shown particular promise. Future research could explore hybrid models and the incorporation of additional features, such as reviewer metadata, to further enhance sentiment analysis performance.

References :

1) Logan Ashbaugh † and Yan Zhang †,* School of Computer Science and Engineering, California State University San Bernardino,5500 University Parkway, San Bernardino, CA 92407, USA; 007247156@coyote.csusb.edu* Correspondence: yan.zhang@csusb.edu; Tel.: +1-909-537-5333

2) A. Ravi, A. R. Khettry, and S. Y. Sethumadhavachar, "Amazon Reviews as Corpus for Sentiment Analysis Using Machine Learning," in Advances in Computing and Data Sciences, M. Singh, P. Gupta, V. Tyagi, J. Flusser, T. Ören, and R. Kashyap, Eds. Singapore: Springer, 2019, pp. 403-411.

3) S. A. Elzeheiry, W. A. M. Gab Allah, N. Mekky, and M. Elmogy, "Sentiment Analysis for E-commerce Product Reviews: Current Trends and Future Directions," Journal of Artificial Intelligence Research, vol. 67, pp. 123-145, May 2023.

4) S. S. Bhadoria, D. Gupta, and P. Yadav, "Sentimental Analysis on Amazon Reviews Using Machine Learning," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 10, no. 4, pp. 45-52, Apr. 2024.

5) A. Kumar and P. Tiwari, "Transfer Learning for Sentiment Analysis Using BERT," *IEEE Transactions on Artificial Intelligence*, vol. 5, no. 3, pp. 200-210, June 2022.