Customer Sentiment Analysis and Insights Visualization for E-Commerce Using Machine Learning and Deep Learning Techniques

*Hemant Satwal[1], Ambily Balaram[1], and Manikandakumar M[1]

[1]Department of Artificial Intelligence and Data Science Engineering, CHRIST (Deemed to be University), Bengaluru, Karnataka, India
*hemantsingh1909@gmail.com, ambily.balaram@christuniversity.in, manikandakumar.m@christuniversity.in

Abstract. Customer sentiment analysis will make us aware of their thoughts on online shopping websites and make more informed decisions. Due to the growth of online marketplaces, we require automatic tools to discover valuable insights when working with very large volumes of customer comments. In this paper we have sentimentally classified Amazon Fine Food Reviews records of data using standard machine learning and deep learning architectures, including logistic regression, Naïve Bayes, XGBoost, random forest, decision tree, single-layer LSTM, LSTM with a dropout layer, and bidirectional LSTM.

Keywords: Sentiment Analysis · E-commerce · Machine Learning · Deep Learning · BiLSTM · Text Classification

1 Introduction

E-commerce has altered the tone with which businesses are communicating with their clients, and the reviews have become one of the primary sources of feedback and opinions. A lot of information is found in reviews about the degree of customer satisfaction, what they like, and what concerns them; therefore, it is worth analyzing them. Reviews contain a lot of unstructured data, which complicates analysis. This issue is addressed by sentiment analysis, one of the tasks in natural language processing (NLP), which automatically identifies which feelings the text represents.

Deep Learning (DL) and machine learning have become extremely popular in sentiment detection due to their ability to extract patterns in a large text. The text sentiment tasks have been successful with traditional machine learning approaches that include logistic regression, naive Bayes, XGBoost, random forest, and decision trees. Deep learning neural networks such as Long Short-Term Memory (LSTM) and Bidirectional LSTM (BiLSTM) have gained large influence in the past years, successfully capturing the context and word order in text.

R. Vasanth Kumar Mehta et al. (eds.), *Proceedings of the International Conference on Intelligent Systems for a Sustainable Future (ISSF 2026)*, Atlantis Highlights in Intelligent Systems 16,
https://doi.org/10.2991/978-94-6239-693-7_94

To enhance precision, numerous text processing procedures are followed, like the elimination of stop words, the division of text into tokens, the simplification of words to their stem words (lemmatization), and the transformation of words into numbers (TF-IDF). Models are compared by different measures of evaluation, such as accuracy, precision, recall, and F1-score, to identify the best sentiment analysis approach in e-commerce.

2 Related Work

Researchers have explored several approaches to enhance the performance of traditional machine learning models, including optimization techniques, selective feature extraction, and ensemble learning. Jiang [1] demonstrated that meta-heuristic Improved Particle Swarm Optimization (IPSO) can significantly improve the performance of Support Vector Machines (SVMs). Similarly, ensemble-based methods such as Random Forest and XGBoost outperform individual classifiers by capturing diverse sentiment patterns [2,3]. Feature selection techniques, including Chi-Square and Mutual Information, help reduce computational complexity while improving classification accuracy [4, 5].

With the rapid growth of e-commerce big data, deep learning techniques have gained substantial attention due to their ability to preserve contextual and sequential information in text. Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) models generally outperform classical machine learning approaches [6]. Hybrid models that combine CNNs and LSTMs achieve higher accuracy by learning both local word patterns and long-range dependencies [7]. Additionally, word embedding techniques such as Word2Vec and GloVe provide semantically rich vector representations, further improving sentiment analysis performance [8].

Recent studies have shifted focus toward document-level sentiment analysis and aspect-based sentiment analysis, which extracts opinions related to specific product attributes [9]. Multilingual and code-mixed sentiment analysis highlights the need for language-aware preprocessing techniques and robust representations in global e-commerce platforms [10]. Furthermore, several studies emphasize that effective text preprocessing plays a critical role in improving the performance of both machine learning and deep learning models [11].

3 Dataset Description

This study utilizes the Amazon Fine Food Reviews dataset for the experiments. This data contains customer feedback of over a decade. It contains 568454 reviews posted between 19th October 1999 and 19th October 2012. Such reviews are of 256,059 various customers and of 74,258 products. Within each review there are a few pieces of information, and these are Product ID, User ID, Profile Name, how useful was the review, a rating out of 1 to 5, the time it was considered.

The sentiment analysis indicates that the data is extremely skewed, with 5- and 4-star reviews comprising about 70% of the data. The features that were not useful were eliminated, and missing values were filled in before optioning the models to be trained. Our sentiment labels depended on review ratings: 4-star and 5-star reviews received a positive label (1), and 1-, 2-, and 3-star reviews received a negative label (0).

4 Methodology

4.1 Data Preprocessing

We applied several preprocessing techniques on the raw reviews of texts in order to enhance the data quality and minimize noise to then train the model. To begin with, We converted the rating into a binary sentiment label: Positive reviews were checked with 1, and negative reviews were checked with 0. We also eliminated the unwanted data fields and have corrected the missing values in order to maintain data consistency. Then we filtered the review text by removing punctuation, special characters, and numbers and converted all the words to lowercase. We cut the text into words, reduced the words to their simplest forms, and left only important words produced by the removal of common words like *"the"*, *"is"*, *"and"*, and *"so"*. Lastly, the cleaned text was transformed into numbers with TF-IDF so that the model could access it.

4.2 Feature Representation

The following are the techniques used for feature extraction:

TF–IDF: This technique represents text as TF-IDF vectors based on the frequency of word occurrences. It assigns weights to words based on their distribution in the documents.

4.3 Classification Models

The paper examines older forms of machine learning and more recent models of deep learning:

Machine Learning Models:

- Logistic Regression
- Naïve Bayes
- Decision Tree
- Random Forest
- Support Vector Machine
- XGBoost

Deep Learning Models:

- Bidirectional LSTM (Single layer)
- Bidirectional LSTM (Dropout layer)
- BERT (fine-tuned)

4.4 Evaluation Metrics

We carry out the accuracy, precision, recall and F1-score to measure the performance of the model to produce a balanced output.

4.5 Implementation Details

Dataset Processing: Tests performed on the Amazon Fine Food Reviews dataset were done in the case of a large amount of customer reviews gathered over an extended period of time and commonly utilized in sentiment analysis. During preprocessing we eliminated redundant features and addressed the missing data that made the data credible. We categorized sentiments according to the ratings of the review: reviews rated 4 or 5 were classified as positive (1), and reviews rated 1, 2, or 3 were classified as negative (0).

Feature Extraction:

- TF–IDF vectorization was used to convert text into numerical representations.
- The vocabulary size was limited to the 5,000 most frequent words for efficiency.

Model Training and Evaluation: The pre-processed data was divided into a training set and a testing set. 80 % of the information was policy-trained, and the rest of the 20 % was retained to assess its performance.

4.6 Performance Evaluation of Machine Learning Models

The machine learning models were evaluated on the test set, and their performance metrics are summarized in Table 1.

Table 1. Performance Comparison of Machine Learning Models

Model	Accuracy (%)	Precision	Recall	F1-Score
Logistic Regression	89.06	0.89	0.89	0.89
XGBoost	87.78	0.88	0.88	0.87
Naïve Bayes	83.48	0.84	0.83	0.80
Random Forest	91.14	0.91	0.91	0.91
Decision Tree	85.70	0.85	0.86	0.85

4.7 Performance Evaluation of Deep Learning Models

The deep learning methods were applied with LSTM models of varying configurations. Table 2 indicates the numbers of the performances and their comparison of performance numbers against the baseline accuracies.

Table 2. Performance Comparison of Deep Learning Models

Model	Base Paper Accuracy (%)	My Accuracy (%)	Improvement (%)
LSTM Single Layer	87	88.54	1.54
LSTM Dropout Layer	88	88.98	0.98
Bidirectional LSTM (Bi-LSTM)	87	87.52	0.52

4.8 Detailed Performance Metrics

Results of Logistic Regression: The model was found to have the general accuracy of 89.06 and equal performance in both classes. In the case of negative reviews (class 0), precision was 0.80, recall was 0.66, and the F1 score was 0.73. In positive reviews (class 1), the values of precision, recall, and F1-score were 0.91, 0.95, and 0.93, respectively.

XGBoost Results: The model had an accuracy of 87.78%. In the case of negative reviews, there was a precision of 0.85, a recall of 0.54, and an F1 score of 0.66. In the positive reviews, the precision was 0.88, the recall was 0.97, and the F1-score was 0.93.

Naïve Bayes Findings: The model got 83.48% accuracy. In the case of negative reviews, the precision was 0.89, the recall was 0.28, and the F1 was 0.43. In the case of positive reviews, the precision was 0.83, the recall was 0.99, and the F1 score was 0.90.

Random Forest Results: The best performance was that of Random Forest. It had a 91.14% accuracy and scored fairly on all measures. In the case of negative reviews, precision was 0.94, recall 0.64, and F1-score was 0.76. In the case of positive reviews, the precision was 0.91, the recall was 0.99, and the F1-score was 0.95.

Decision Tree Results: It got 85.70% accurate. In the case of negative reviews, precision, recall, and F1-score were 0.74, 0.53, and 0.62, respectively. Precision, recall, and F1-score ranged at 0.88, 0.95, and 0.91, respectively, in the case of positive reviews.

LSTM Single Layer Results: Single-layer LSTM achieved an accuracy of 88.54, a precision of 91.03, a recall of 94.59, and an F1-score of 92.77, as reported in Fig. 1.

LSTM Dropout Layer Results: Among all the deep learning models, the LSTM with dropout achieved the top accuracy of 88.98%. It also possessed precision, recall of 90.68, and an F1-score of 93.14, as reflected in Fig. 2.

Bi-LSTM Results: Bidirectional LSTM model achieved the accuracy of 87.52. It was also precise (92.46), recalled (91.56), and F1- perceived (92.01) as well. Fig. 3 indicates these results.

4.9 Key Observations

- Random Forest was most accurate with the value of 91.14, and then Logistic Regression at 89.06%.
- The worst result was that of decision tree with 85.70 accuracy, which was likely to overfit the training data.

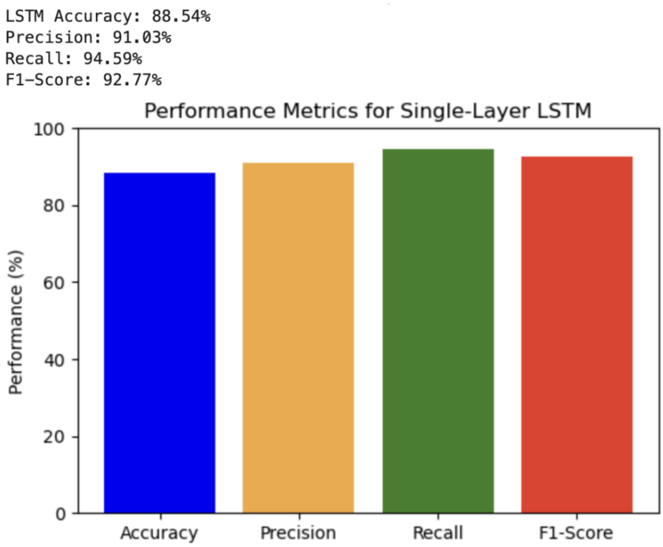

Fig. 1. LSTM Single Layer Report

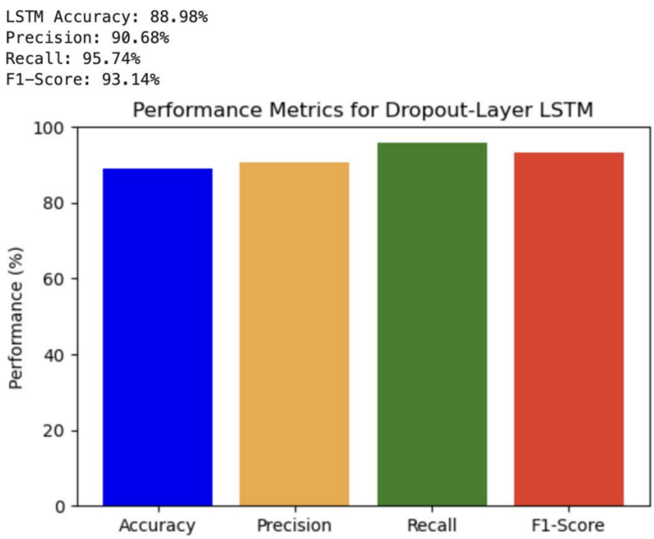

Fig. 2. LSTM Dropout Layer Report

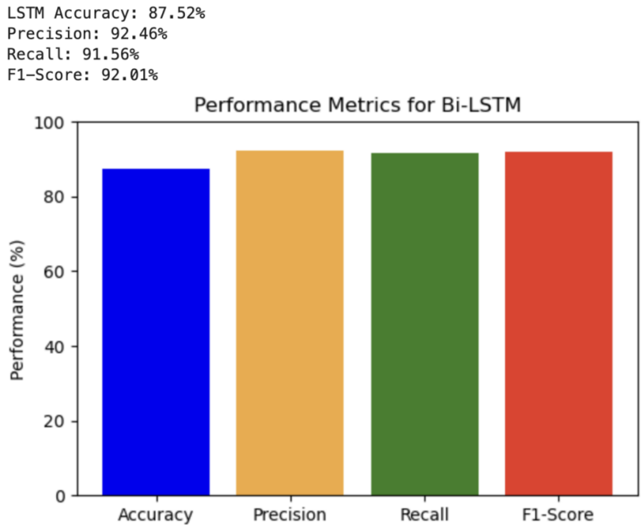

Fig. 3. Bi-directional LSTM Report

– The accuracy of deep learning models was better than that of the base paper.

4.10 Sentiment Trend Analysis

– The plot of sentiment distribution was made by matplotlib and seaborn.
– The positive reviews were also higher (score 4 or higher).
– The use of sentiment plots showed how the products were reviewed seasonally.

5 Challenges and Limitations

Positive gains were obtained, however this research has certain limitations. First, the data is not balanced: there are plenty of positive reviews as compared to the negative ones. This may make the outcomes favour the majority group. Second, the text data were difficult to classify because of the nature of the data. Due to the presence of mixed feelings in various reviews, some were positive regarding the piece, whereas others claimed issues with delivery. Last but not least, the level of computations used was a significant issue. Random Forest and XGBoost take more time to run since they are making complex decisions, and this could limit its capability to scale in real-time applications.

6 Conclusion and Future Scope

This study studied consumer attitudes towards e-commerce sites based on basic machine-learning algorithms on the complete data of the Fine Food Reviews on

Amazon. We would have called every review as positive or negative using fast text cleaning, TF-IDF, and standard learning algorithms.

These learning algorithms, according to the results, could provide good results when it comes to sentiment spotting. The best one was the random forest with a score of 91.14% correct. It was nearly that with logistic regression and XGBoost, and it goes to show that even with simple models, it is possible to perform well, provided you select good features. The decision tree performed worse, which is likely due to its poor handling of lots of words and its susceptibility to overfitting. In addition to the number of accuracy, the sentiment trends depicted helpful hints regarding the customer's moods as time passed by.

Concisely, this paper demonstrates that machine-learning sentiment analysis is effective and can assist e-commerce businesses to optimize the quality of their products, customer satisfaction, and business planning.

Acknowledgments The authors would like to thank Christ University, Bangalore, for providing the resources and academic guidance necessary to conduct this research.

Disclosure of Interests The authors have no competing interests to declare that are relevant to the content of this article.

References

1. X. Jiang, "Sentiment classification using Improved Particle Swarm Optimization and SVM," *Journal of Machine Learning Research*, 2022.
2. P. Kawale *et al.*, "Sentiment analysis on e-commerce reviews using ensemble learning," in *Proc. Int. Conf. AI & Data Science*, 2024.
3. K. Wassana *et al.*, "Machine learning-based sentiment analysis of Amazon product reviews," in *Proc. ACM Int. Conf. AI & E-Commerce*, 2021.
4. G. Sapthami *et al.*, "Feature selection in sentiment analysis for customer product reviews," *IEEE Trans. Artificial Intelligence*, 2023.
5. A. Jemai *et al.*, "Feature extraction methods and their impact on sentiment classification in e-commerce reviews," *IEEE Trans. Machine Learning*, 2021.
6. S. Gope *et al.*, "Comparative analysis of machine learning and deep learning models for sentiment analysis of Amazon product reviews," *Expert Systems*, Elsevier, 2022.
7. M. Juyal, "Hybrid and ensemble methods for improved sentiment classification accuracy," *Journal of Artificial Intelligence & Applications*, 2022.
8. A. Iqbal *et al.*, "Sentiment analysis of consumer reviews using deep learning," *Journal of Artificial Intelligence & Data Science*, 2023.
9. D. Prakash and V. Sharma, "Aspect-based sentiment analysis of Amazon reviews using Partitioning Around Medoids clustering," *IEEE Trans. Data Mining*, 2023.
10. A. Awatramani *et al.*, "Sentiment analysis of multilingual reviews using NLP and deep learning techniques," *Journal of Computational Linguistics*, 2021.
11. M. Hafeez and S. Kathirisetty, "The impact of data preprocessing on sentiment analysis performance using ML and deep learning models," *International Journal of Data Science*, 2022.