International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
IJISAE
ISSN:2147-6799 www.ijisae.org Original Research Paper

Optimizing Machine Learning & Deep Learning Models for Sentiment Classification of Online Product Reviews

[1]Rajendra Kumar, [2]Manish Kumar Goyal

Submitted: 05/01/2025 **Revised:** 15/02/2025 **Accepted:** 20/02/2025

Abstract: The rapid expansion of e-commerce has led to an overwhelming amount of user-generated content in the form of online product reviews. Sentiment classification of these reviews is critical for businesses and consumers alike, enabling better decision-making and enhancing customer satisfaction. This paper explores various machine learning (ML) and deep learning (DL) techniques for optimizing sentiment classification models specifically tailored for online product reviews. The methodology integrates various NLP techniques like Bag of Words (BoW), Term Frequency-Inverse Document Frequency (TF-IDF), and Word2Vec embeddings. These were combined with machine learning models such as Multinomial Naive Bayes, Logistic Regression, Random Forest, and Long Short-Term Memory (LSTM) neural networks. We conduct a comparative analysis of traditional ML algorithms and deep learning architectures in terms of performance, efficiency, and interpretability. Various techniques such as feature engineering, hyperparameter tuning, and model optimization are evaluated. The results demonstrate that deep learning models, particularly transformer-based architectures, outperform conventional ML methods in terms of accuracy, robustness, and scalability, though ML models remain relevant due to their efficiency and interpretability.

Keywords*: Machne Learning, Deep Learning, LSTM, TF-IDF, BoW, Word2Vec*

Introduction

With the proliferation of e-commerce platforms such as Amazon, eBay, and Alibaba, online product reviews have become a valuable source of information for both consumers and businesses [1]. Sentiment analysis, a subset of natural language processing (NLP), aims to classify the sentiment expressed in textual reviews as positive, negative, or neutral. The challenge lies in optimizing ML and DL models to handle the complexity and variability of human language effectively [2].

The task of sentiment classification, however, is not trivial. Online reviews often exhibit complex linguistic phenomena, including:

Sarcasm and Irony: Reviews may contain sarcastic or ironic statements that contradict the literal meaning of the words.

Ambiguity and Negation: The presence of ambiguous phrases and negation can significantly alter the sentiment expressed.

Domain-Specific Jargon: Reviews may contain technical terms and jargon specific to the product domain.

Spelling Errors and Slang: Online reviews often contain informal language, spelling errors, and slang, making it challenging for traditional NLP techniques.

Imbalanced Datasets: Datasets might have more positive reviews than negative, leading to biased models.

This paper aims to provide a comprehensive overview of techniques for optimizing ML and DL models for sentiment classification of online product reviews. We will explore various feature engineering strategies, model architectures, and optimization techniques, focusing on addressing the challenges mentioned above.

Literature Review

Previous research has explored various approaches to sentiment classification. Early work relied on lexicon-based methods, which assign sentiment scores to words and phrases based on predefined dictionaries. However, these methods often struggle with context and ambiguity.

Machine learning approaches, such as Support Vector Machines (SVMs), Naive Bayes, and Logistic Regression, have shown promising results [3, 4, 5]. These methods typically rely on feature engineering techniques, such as:

Bag-of-Words (BoW): Represents text as a collection of words, ignoring grammar and word order.

Term Frequency-Inverse Document Frequency (TF-IDF): Weights words based on their frequency in the document and their inverse frequency in the corpus.

[1]*Mtech Scholar, Department of CSE,Apex institute of engineering and Jaipur, Rajasthan*
[2]*Assistant Professior,Department of CSE,Apex institute of engineering and Jaipur, Rajasthant*
rajapex86@gmail.com, coe@apexcollege.in

N-grams: Captures word sequences of length n, providing some context.

Deep learning models, particularly Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs), have achieved state-of-the-art performance in sentiment classification. RNNs, such as Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRUs), are well-suited for capturing sequential dependencies in text [6]. CNNs, on the other hand, can effectively extract local features and patterns. Transformer models, such as BERT, RoBERTa, and DistilBERT, have further advanced the field by leveraging attention mechanisms to capture long-range dependencies and contextual information [7].

2.1 Traditional Machine Learning Approaches

Traditional ML approaches rely on statistical and rule-based techniques. The most commonly used ML models include [8]:

Naïve Bayes: A probabilistic classifier based on Bayes' theorem, effective for text classification tasks but limited in handling complex language structures.

Support Vector Machine (SVM): A linear classifier that maximizes margin separation, providing robust performance but requiring feature engineering.

Random Forest: An ensemble learning method that improves classification accuracy but may struggle with interpretability.

XGBoost: A gradient boosting algorithm known for its efficiency and high performance in structured data classification.

2.2 Deep Learning Approaches

Deep learning models have demonstrated superior performance by leveraging large-scale datasets and learning complex patterns in text. Key architectures include [9]:

Convolutional Neural Networks (CNNs): Extract local features from text, effective for short sentences but lacking sequential understanding.

Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM): Capture sequential dependencies, improving performance in long-text classification.

Bidirectional Encoder Representations from Transformers (BERT): A state-of-the-art transformer-based model that understands contextual relationships in text, achieving high accuracy but requiring extensive computational resources.

2.3 Optimization Techniques for Sentiment Classification

Feature Engineering: Utilizing TF-IDF, Word2Vec, GloVe, FastText, and BERT embeddings to improve text representation [10].

Hyperparameter Tuning: Applying grid search and Bayesian optimization to refine model performance [11].

Regularization Techniques: Implementing dropout and L2 regularization to prevent overfitting [12].

Data Augmentation: Generating synthetic data through back-translation to improve model generalization [13].

Transfer Learning: Fine-tuning pre-trained models like BERT and GPT-3 for domain-specific sentiment classification [14].

Methodology

This section outlines the methodology employed in this study, including dataset selection, feature engineering techniques, model architectures, and evaluation metrics [15].

3.1. Dataset Selection

We will utilize publicly available dataset of online product reviews, including:

Amazon Product Reviews Dataset: A large dataset containing reviews from various product categories.

3.2. Feature Engineering

We will explore various feature engineering techniques, including:

Text Preprocessing: Cleaning the text by removing punctuation, stop words, and special characters. We will also perform stemming and lemmatization to reduce words to their root form.

Word Embeddings: Representing words as dense vectors, capturing semantic relationships. We will use pre-trained word embeddings, such as Word2Vec, GloVe, and FastText, as well as fine-tuned embeddings from transformer models.

TF-IDF and N-grams: Extracting features based on term frequency and word sequences.

Sentiment Lexicons: Utilizing sentiment dictionaries, such as VADER and SentiWordNet, to extract sentiment scores for words and phrases.

Part-of-Speech (POS) Tagging: Identifying the grammatical role of words, which can provide valuable contextual information.

3.3. Model Architectures

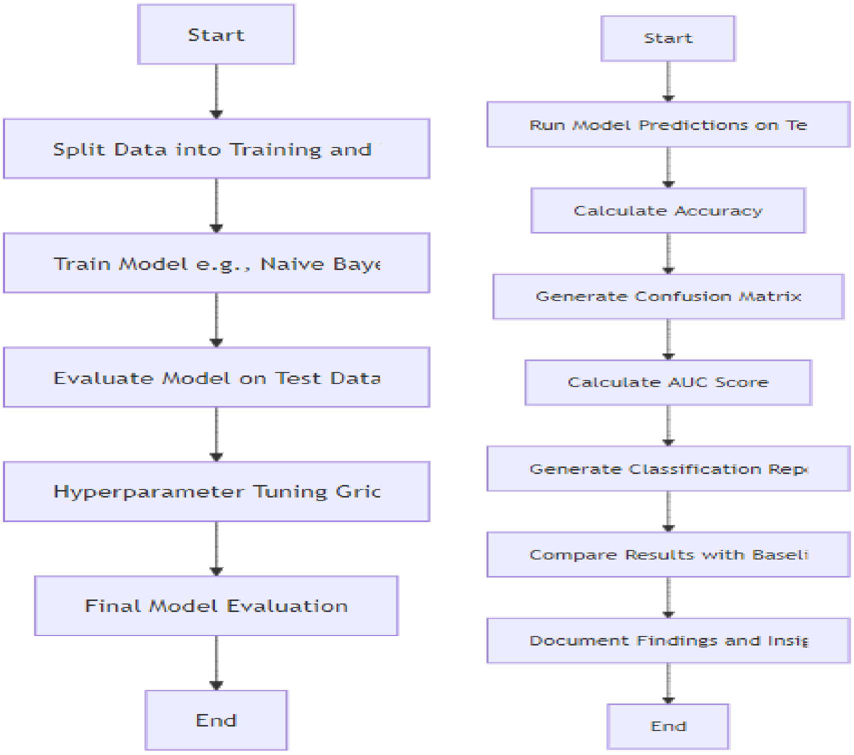
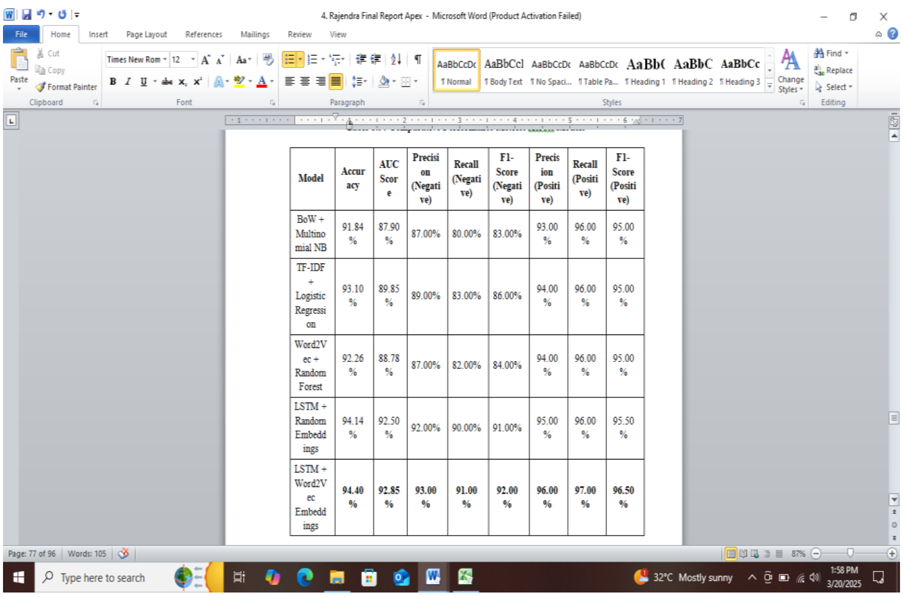


Figure 1: Model Development and Performance Analysis

Result Analysis

A comprehensive analysis of the results obtained from applying various Natural Language Processing (NLP) techniques and machine learning models to sentiment analysis of Amazon product reviews. The dataset comprises 400,000 reviews of unlocked mobile phones sold on Amazon.com. The key objective is to classify these reviews into positive or negative sentiments based on the textual content. The methodology involved the use of Bag of Words (BoW), Term Frequency-Inverse Document Frequency (TF-IDF), Word2Vec embeddings, and Long Short-Term Memory (LSTM) neural networks. This chapter discusses the results obtained from each approach, along with a detailed examination of the model performance metrics and insights derived from the analysis.

Table 1: Comparative Performance Metrics across Models



Model	Accuracy	AUC Score	Precision (Negative)	Recall (Negative)	F1-Score (Negative)	Precision (Positive)	Recall (Positive)	F1-Score (Positive)
BoW + Multinomial NB	91.84 %	87.90 %	87.00%	80.00%	83.00%	93.00 %	96.00 %	95.00 %
TF-IDF + Logistic Regression	93.10 %	89.85 %	89.00%	83.00%	86.00%	94.00 %	96.00 %	95.00 %
Word2Vec + Random Forest	92.26 %	88.78 %	87.00%	82.00%	84.00%	94.00 %	96.00 %	95.00 %
LSTM + Random Embeddings	94.14 %	92.50 %	92.00%	90.00%	91.00%	95.00 %	96.00 %	95.50 %
LSTM + Word2Vec Embeddings	**94.40 %**	**92.85 %**	**93.00 %**	**91.00 %**	**92.00 %**	**96.00 %**	**97.00 %**	**96.50 %**

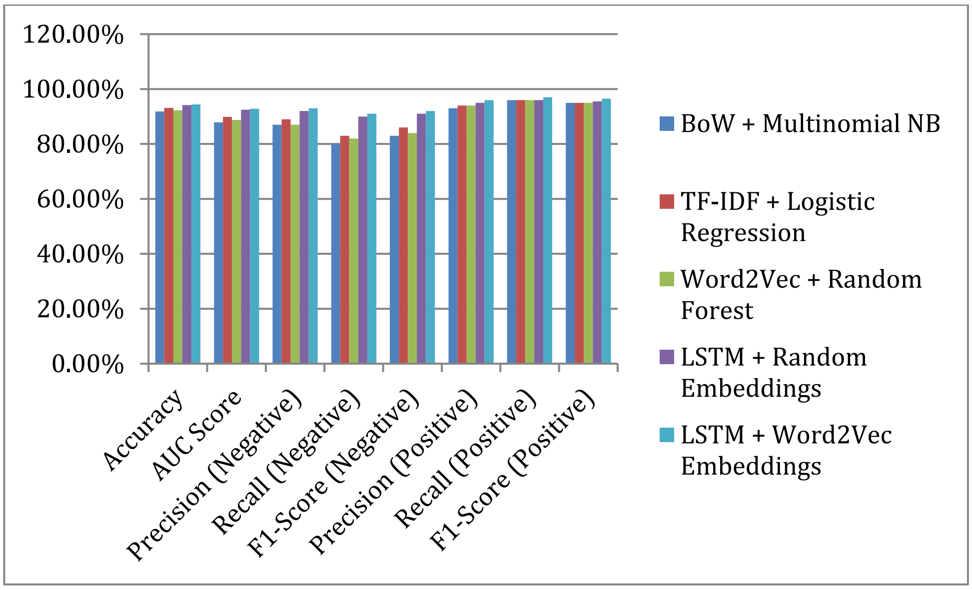


Figure 2: Comparative Analysis of Different Models

The comprehensive analysis of sentiment in Amazon product reviews using various NLP techniques and machine learning models has provided valuable insights into the strengths and limitations of each approach. The LSTM model with Word2Vec embeddings emerged as the most effective, offering the highest accuracy and AUC scores while maintaining balanced performance across positive and negative sentiments.

The study underscores the importance of advanced word embeddings and deep learning architectures in capturing the nuances of sentiment in text data. However, it also highlights the need for careful consideration of computational resources, model interpretability, and domain-specific challenges.

Future research could explore more sophisticated models, such as transformers, and address the challenges of handling sarcasm, class imbalance, and generalization to other domains. The potential applications of sentiment analysis are vast, and with continued advancements in NLP, the ability to understand and act on customer feedback will only improve, driving better business outcomes and enhanced consumer experiences.

Conclusion

This study undertook the task of analyzing sentiment in Amazon product reviews, specifically focusing on unlocked mobile phones, using various Natural Language Processing (NLP) techniques and machine learning models. The overarching goal was to classify reviews into positive, neutral, or negative sentiments, leveraging both traditional models like Bag of Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF), as well as more advanced methods such as Word2Vec embeddings and Long Short-Term Memory (LSTM) neural networks.

References

[1] Abdullah, R., Suhariyanto., & Sarno, R. (2021). Aspect Based Sentiment Analysis for Explicit and Implicit Aspects in Restaurant Review using Grammatical Rules, Hybrid Approach, and SentiCircle. *International Journal of Intelligent Engineering and Systems*, *14*(5), 294-305.

[2] Abu Alfeilat, H. A., Hassanat, A., Lasassmeh, O., Tarawneh, A. S., Alhasanat, M. B., Eyal Salman, H. S., & Prasath, V. (2019). Effects of Distance Measure Choice on K-Nearest Neighbor Classifier Performance: A Review. *Big data*, *7*(4), 221–248.

[3] Akhtar, N., Zubair, N., Kumar, A., & Ahmad, T. (2017). Aspect-based Sentiment Oriented Summarization of Hotel Reviews. *In Proceedings of the 7th International Conference on Advances in Computing & Communications, ICACC-2017, Procedia Computer Science*, *115*, 563–571.

[4] Akhtar, S. Md., Garg, T., & Ekbal, A. (2020). Multi-task Learning for Aspect Term Extraction and Aspect Sentiment Classification. *Neurocomputing*, 1-24.

[5] Akhtar, S. Md., Gupta, D., Ekbal, A., & Bhattacharyya, P. (2017). Feature Selection and Ensemble Construction: A Two-step Method for Aspect Based Sentiment Analysis. *Knowledge-Based Systems*, 1-40.

[6] AlGhamdi, N., & Khatoon, S. (2021). Improving Sentiment Prediction using Heterogeneous and Homogeneous Ensemble Methods: A Comparative Study. *Procedia Computer Science,194*, 60-68.

[7] Ali, M., & Kumar, D. (2021). A Combination between Deep learning for feature extraction and Machine Learning for Recognition. *In the Proceedings of the 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 01-05.

[8] Alqaryouti, O., Siyam, N., Monem, A.A., & Shaalan, K. (2019). Aspect-Based Sentiment Analysis Using Smart Government Review Data. *Applied Computing and Informatics*, 1-12.

[9] Al-Smadi, M., Al-Ayyoub, M., Jararweh, Y., & Qawasmeh, O. (2018). Enhancing Aspect-Based Sentiment Analysis of Arabic Hotels' reviews using morphological, syntactic and semantic features. *Information Processing and Management*, 1-12.

[10] Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Al-Shamma, O., Santamaria, J., Fadhel, A. M., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data, 8*(53), 1-74.

[11] Amplayo, K. R., Lee, S., & Song, M. (2018). Incorporating Product Description to Sentiment Topic Models for Improved Aspect-based Sentiment Analysis. *Information Sciences*, 1 - 39.

[12] Anand, D., & Naorem, D. (2016). Semi-supervised Aspect Based Sentiment Analysis for Movies using Review Filtering. *In Proceedings of the 7th International Conference on Intelligent Human-Computer Interaction, IHCI 2015, Procedia Computer Science, 84*, 86 – 93.

[13] Antony, P.J. (2011). Parts Of Speech Tagging for Indian Languages: A Literature Survey. *International Journal of Computer Applications, 34*(8), 22-29.

[14] Asghar, MZ., Khan, A., Ahmad, S., Qasim, M., & Khan, IA. (2017). Lexicon-enhanced sentiment analysis framework using rule-based classification scheme. *PLoS ONE*, *12*(2), 1-22.

[15] Awad, M., Khanna R. (2015). Support Vector Machines for Classification. *In: Efficient Learning Machines. Apress, Berkeley, CA*, 39-66.

[16] Bansal, A., & Kumar, N. (2021). Aspect-Based Sentiment Analysis Using Attribute Extraction of Hospital Reviews. *New Gener. Comput*, 1-20. Batra, M., & Jyoti, V. (2019). Feature-Based Opinion Mining. *IGI Global*, 20-39.

[17] Behdenna, S., Barigou, F., & Belalem, G. (2018). Document Level Sentiment Analysis: A survey. *EAI Endorsed Transactions on Context- aware Systems and Applications, 4*(13), 1-8.

[18] Belyadi, H., & Haghighat, A. (2021). Chapter 5 - Supervised learning. *Machine Learning Guide for Oil and Gas Using Python, Gulf Professional Publishing*, 169-295.

[19] Berrar, D. (2019). Bayes' Theorem and Naive Bayes Classifier. *Encyclopedia of Bioinformatics and Computational Biology, Academic Press*, *1*, 403-412.

[20] Bhamare, R. B., Jeyanthi, P., & Subhashini, R. (2019). Aspect Category Extraction for Sentiment Analysis using Multivariate Filter Method of Feature Selection. *International Journal of Recent Technology and Engineering (IJRTE)*, *8*(3), 2138-2143.