Journal of Computer Science

Research Article

Twitter Sentiment Analysis Using Machine Learning and Deep Learning Techniques

[1]Safa Khalid Al Sarayrah, [2]Noor Mahmoud Alkudah and [1]Hesham Yousef Al-kharabsheh

[1]*Department of Software Engineering, Faculty of Information Technology, The World Islamic Sciences and Education University, Amman, Jordan*
[2]*Department of Computer Science, Faculty of Information Technology, The World Islamic Sciences and Education University, Amman, Jordan*

Article history
Received: 24-10-2024
Revised: 05-03-2025
Accepted: 18-03-2025

Corresponding Author:
Safa Khalid Al Sarayrah
Department of Software Engineering, Faculty of Information Technology, The World Islamic Sciences and Education University, Amman, Jordan
Email: safaa.sariereh@wise.edu.jo

Abstract: This research investigates the use of Machine Learning (ML) and Deep Learning, including BiLSTM approaches, for Sentiment Analysis (SA) of consumer reviews on social media sites. Businesses are increasingly depending on online reviews to determine customer satisfaction due to social media's explosive growth. We used three classification models, the assessment of these attitudes uses Naive Bayes (NB), Support Vector Machine (SVM), and a BiLSTM model. Customer reviews categorized as neutral, negative, or positive feelings made up the dataset used for this study, implying that positive reviews are related to satisfied customers. Text cleaning, tokenization, Bert, and TF-IDF feature extraction were among the preprocessing procedures. Our results show that MNB and CNB had accuracy rates of 80.21 and 81.44%, respectively, whereas linear and RBF SVM had slightly higher accuracy rates of 84.28 and 88.64%, respectively. Also, BiLSTM achieves 87.72%. Besides, this research sheds light on unbalanced dataset issues. Therefore, we apply random oversampling to the research dataset. The evaluated outcomes show that sentiment classification and knowledge extraction through machine learning models and deep learning methods yield beneficial insights that enable businesses to understand their customers better.

Keywords: Sentiment Analysis, Machine Learning, Deep Learning, Techniques, Twitter

Introduction

Information and communication technology has grown in significance recently, both in terms of delivering service offerings and affecting consumer experiences (Luo *et al.*, 2021). Sustaining the level of customer satisfaction is crucial to the continued success of the industry (Ara *et al.*, 2020). Therefore, industries always try to understand and analyze customer reviews. Over half of the population on the planet uses social media (Alharbi & Qamar, 2022). They utilize social media for information, marketing, and other online activities in addition to enjoyment.

Textual polarity is identified by sentiment analysis. Whether a Text is favorable, negative, or neutral is determined by it. Because it extracts the viewpoint or stance of the speaker, it is also known as Opinion Mining (Devika *et al.*, 2016). Identifying the features or markers that enable us to correctly classify the various classes is one of the challenges in any text-based classification process. Classes for positive or negative sentiment analysis (Hung & Alias, 2023).

As information and communication technology has developed, the field of consumer sentiment analysis has recently been engaged in customer feedback research case studies that utilize data obtained from online reviews (Ara *et al.*, 2020; Zhou *et al.*, 2019). Text mining, as well as natural language processing models and techniques, can receive raw data and transform it into a structured form ready for data processing, also explain how this transformation occurs (Jain *et al.*, 2021). On the topic of review mining, a substantial number of materials, most especially those using machine learning techniques, have been published (Jagdale *et al.*, 2019; Singh *et al.*, 2020). In the past, consumer opinion analysis has employed data mining and text classification (Alhojely, 2016). amounts to the computerized interpretation of the consumers' opinions, beliefs, and attitudes regarding particular products or services (Pang & Lee, 2016). In essence, consumer sentiment analysis is a computer application that allows spotting and understanding the perspectives, experiences, and strategies of customers applied in online reviews available on various websites (Jain *et al.*, 2021).

SCIENCE Publications

© 2025 Safa Khalid Al Sarayrah, Noor Mahmoud Alkudah and Hesham Yousef Al-kharabsheh. This open-access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.

DOI: 10.3844/jcssp.2025.1785.1794

Assume that any customer wants to purchase a good or service in this day and age. In those circumstances, the abundance of user reviews and discussion forums about goods and services available online eliminates the need to consult friends and family for firsthand accounts of past customers' experiences. Because customer experience is readily available in Internet evaluations, there is also no need for any service provider to hold customer surveys, opinion polls, or focus groups to gather this data. As of right now, we've discovered that openly posting online reviews on social media has aided and changed the course of business growth. Labor costs and time can be decreased by automating the labor-intensive task of analyzing every customer review (Verma *et al.*, 2023). Machine learning is capable of extracting consumer attitudes toward a product from website reviews. Reviews contain certain themes around which they are praising or criticizing, whereas reviews reflect either positive feedback or negative feedback. As regards answering the users' problems, sentiment analysis of the consumers is understood in a broad sense as a collection of these results (Park *et al.*, 2021).

Related Works

Sentiment analysis (SA) refers to the use of technology that is capable of automatically detecting and analyzing the sentiment conveyed in textual materials (Gunasekaran, 2023). Opinion mining has garnered significant attention in recent studies. Through their conversations on social media platforms, customers' sentiments about reviews are discovered through sentiment analysis (Hamad *et al.*, 2021). One of the areas of natural language processing research that has drawn the interest of scholars is sentiment analysis, and more and more research papers are being produced in this area (Cui *et al.*, 2023). Rajasekaran *et al.* (2019) analyzed restaurant reviews sentimentally using SVM and Naïve Bayes, and the accuracy rate was 72.2%. Additionally, SVM was used to assess restaurant reviews with a 94.56% accuracy rate (Krishna *et al.*, 2019). It is now essential that the data on social media platforms—like Facebook, Instagram, WhatsApp, and Twitter—convey interesting information about people's opinions, moods, and sentiments regarding any ideas, policies, or products, given the widespread adoption of these platforms in the communication space (Yi & Liu, 2020). A deep convolutional neural network serves as the foundation for analyzing Twitter streams according to scientific research. Twitter data analysis enabled a deep CNN system to integrate its features before using them for sentiment prediction during training sessions (Jianqiang *et al.*, 2018). For the social network Twitter, the importance of text pre-processing is unquestionable. Comments, reviews, blog posts, and posts in microblogs directed by users towards the quality of food, speed of meal delivery, and others can be analyzed for sentiments. Artificial intelligence in natural language processing shows great promise in identifying both good and bad feelings and no sentiment in existing datasets (Geler *et al.*, 2021). Simply, analytical models are created in an automated manner using pertinence, which is a subdivision of machine learning (Adak *et al.*, 2022). Using different combinations and techniques to review the opinions regarding the restaurants has also been done with the help of K-Nearest Neighbors (KNN), Principal Component Analysis (PCA), and Support Vector Machine (SVM). The support vector machine model was hence the best-performing model with nearly 96 percent accuracy, according to the authors, which is higher than the other models for classification algorithms (Verma *et al.*, 2023).

The multinomial Naive Bayes method reached the best performance level with 77% accuracy, and the Support Vector Machine (SVM) achieved 76% accuracy. The research utilized reviews obtained from Amazon, IMDB, and Yelp, yet it excluded all studies without user ratings when conducting its analysis (Gupta *et al.*, 2023). Multiple machine learning techniques, including Naïve Bayes (NB), Random Forest, Bi-LSTM, and BERT, were used to analyze the sentiment of Amazon mobile phone reviews, which were labeled as positive, negative, and neutral. The Bi-LSTM, together with BERT, offered the most efficient performance in achieving the classification objective. The accuracy rate achieved by Random Forest surpassed Naive Bayes by reaching 90% for multiclass classification and 94% for binary (AlQahtani, 2021).

The majority of research on Twitter sentiment analysis can be divided into supervised and unsupervised methodologies (da Silva *et al.*, 2014; Hagen *et al.*, 2015; Saif *et al.*, 2012). Also, lexicon-oriented methods (Montejo-Ráez *et al.*, 2014; Paltoglou & Thelwall, 2012; Thelwall *et al.*, 2012). While the training of classifiers forms an integral part of supervised methods, which utilize features including Part of Speech (POS), word N-grams, and tweet context information features including hashtags, retweets, emoticons, and capitalized words, among others. (Jianqiang *et al.*, 2018). SentiWordNet and other lexicons with established sentiments are employed by lexicon-based algorithms to determine the net sentiment of a particular text (Baccianella *et al.*, 2010).

The Deep Learning method that researchers use is Bidirectional Long Short-Term Memory (BiLSTM). The bidirectional long short-term memory method (BiLSTM) functions as a framework called “Bidirectional” due to its processing capability that moves through sequences forwards and backward (Rana *et al.*, 2024). A study classified 56,351 user reviews from ten well-known over-the-top apps as either positive, neutral, or negative. Sentiment nuances and textual features were extracted using ML and DL models. BiLSTM outperformed others, achieving 92% accuracy (Ryan *et al.*, 2024). Moreover, the goal of the study is to use the BiLSTM approach to classify Twitter review data into three sentiment classes: positive, neutral, and negative. the BiLSTM model achieved a 58.56% accuracy rate, using 25 epochs, word2vec CBOW (Pratiwi *et al.*, 2024). In addition to introducing BiLSTM and Attention mechanisms, the model processes text vectors using a pre-trained BERT model and audio vectors using a collaborative speech analysis approach. By combining the attention information of the text and audio modes with the global and local information of the text and audio, the feature vector's weight is dynamically changed. After the fusion features are self-focused, the Softmax layer provides the emotion analysis results. According to the experimental findings, the model's accuracy on the CMU-MOSI dataset is 85.3% (Cheng & Li, 2024).

The machine learning approaches, such as SVM and NB, produce the greatest results when crucial criteria like performance, efficiency, and accuracy are taken into account, and sentiment analysis has received the majority of research attention (Devika *et al.*, 2016). Moreover, the linear support vector machine (SVM) operates as a strong, efficient machine learning approach for data classification. The main goal consists of locating the most suitable hyperplane that establishes category distinctions within the feature space (Nagelli *et al.*, 2025). SVM functions for both regression and classification responsibilities, as well as distinguishing between multiple data clusters (Alsemaree *et al.*, 2024).

In this area, supervised machine-learning techniques are widely used because of their superior accuracy and simplicity. In sentiment analysis, classification utilizing the NB and SVM algorithms is frequently employed (Wankhade *et al.*, 2022).

Materials

The balanced and unbalanced datasets, which contain text classified as positive, neutral, or negative, are used in this work for sentiment analysis. This dataset is publicly available under the name "Sentiment Analysis Dataset"; and is available at Kaggle.com. The text is cleaned by applying stemming using the Porter Stemmer from NLTK using Python code on Jupyter, converting it to lowercase, and eliminating stop words. After that, TF-IDF is used to extract features for ML, and Bert for deep learning.

To categorize the feelings, four machine learning models are employed: multinomial Naive Bayes, linear SVM, SVM with RBF kernel, and complement Naive Bayes, and BiLSTM as deep learning. Following training and testing, the models' performance is assessed using F1-score, accuracy, precision, and recall.

Methodology

The steps of our work are displayed in Figure 1. Two stages make up the analysis that was done for this study. Sentiment analysis is done in the first phase, and reviews are classified in the second phase.

Dataset

The training dataset utilized in this study is publicly available under the name "Sentiment Analysis Dataset". The dataset is available at Kaggle.com. It comprises 27,480 tweets in English. Each tweet is labeled as Positive, Negative, or Neutral. The dataset includes various features, such as textID, text, selected text, and sentiment, but this study focuses specifically on the selected text and sentiment features. While training the classifier, we discovered a serious problem with class imbalance. More than 40% of the training cases were neutral classes, which could have biased the classifiers. To address this problem, we apply a random oversampling technique using the resample function from the sklearn.utils to increase the minor classes to achieve a dataset with balanced classes. Hence, Classical performance benefits directly from random oversampling implementations to handle training instance imbalances among different information categories. (Muhibullaev & Kim, 2016). Besides, the main goal of random oversampling is to help mitigate the issue of class imbalance (Moreo *et al.*, 2016). Moreover, in order to ensure equal representation and weight for both classes during model training, random oversampling duplicates minority class samples at random (Thomas & Jeba, 2024). In contrast, SMOTE creates comparable instances from numeric vectors that reflect the phrases, whereas random oversampling merely duplicates already existing instances. We believed that because the examples were brief sentences, the synthesized vectors might not accurately represent additional sentences in the same class, ultimately leading to a drop in performance (Muhibullaev & Kim, 2016).

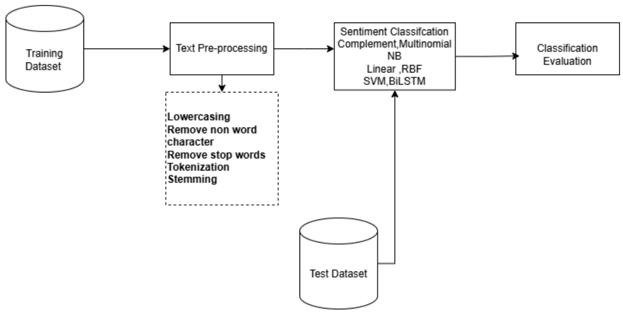

Fig. 1: Research Methodology

Table 1: Dataset Distribution of Sentiment Classes Before and After Oversampling

Sentiment	Before Oversampling	After Oversampling
Neutral	11,118	11,118
Positive	8,582	11,118
Negative	7,780	11,118

Table 1 shows the distribution of the classes before and after applying the oversampling technique. In random oversampling, instances from the minority class are duplicated at random (Sneha & Annappa, 2024) This technique for dataset rebalancing possesses simple implementation rules that are easy to grasp, according to (Mohammed *et al.*, 2020). The correct classification of imbalanced datasets represents a vital step when working in text mining and machine learning applications The combination of SVM with a linear kernel produces optimal results for accuracy, together with recall performance on datasets processed by ADASYN as well as SMOTE techniques (Mujahid *et al.*, 2024). Also, Researchers created oversampling techniques that might not result in fewer instances from the majority class and addressed the problem of class imbalance by reproducing examples from the minority class (Fernández *et al.*, 2018).

Preprocessing

The first and foremost step of the analysis is data pretreatment, which is required because customers can have their feedback in natural language grammar (Zahoor *et al.*, 2020). At this point, specifically in this stage, we converted the entire text into lowercase and eliminated numbers and special characters in whitespace. In addition, we did tokenization. After that, we eliminated non-English words and stop words (such as ‘and’, ‘but’ ‘or’) and then applied stemming, using the Porter stemmer, in order to go into the core of the word (such as interesting becomes interest). Besides, for training the BiLSTM classifier we used BERT as a tokenizer to convert text after the preprocessing process, then fed it to the BiLSTM classifier. Likewise, BERT is a context-aware model that uses bidirectional pre-trained embeddings to encode text into contextualized vectors, excelling at understanding complex language patterns such as context, syntax, and sentiment polarity (Teotia *et al.*, 2023). However, employing the appropriate technique is essential for achieving reliable results in sentiment analysis, especially when working with huge and diverse text data, like Twitter. The Bidirectional Encoder Representations from Transformers (BERT) model stands as one of the newest natural language models developed by Google. BERT provides the main advantage of understanding text information from both forward and backward directions, which takes into account both the words that come before and after a sentence (Angdresey *et al.*, 2025). Combined with only an extra output layer, the pre-trained BERT model cannot only fulfill a range of state-of-the-art tasks, including question answering, but does so without significant modifications to its architecture (Angdresey *et al.*, 2025). By using pre-training on sizable datasets, producing contextual embeddings, and offering bidirectional context, Bert excels at text classification (Angdresey *et al.*, 2025). This enables Bert to effectively express the text's complex relationships and meanings. Furthermore, by creating word clouds for both positive and negative sentiments and surpassing conventional techniques in extracting review semantics, the Bert-BiLSTM model enhances sentiment analysis efficacy. It does this by combining BERT's contextual word embeddings to prevent tokenization ambiguity and BiLSTM to capture bidirectional semantic dependencies (Du *et al.*, 2024). Notably, a model achieves 93.98% accuracy in sentiment analysis of Chinese stock evaluations by using BERT to encode stock review text for improved semantic representation, BiLSTM to boost contextual understanding of review sequences, and an Attention mechanism to focus on important textual information (Li *et al.*, 2024).

All data is divided into training and testing components, while the training portion receives 80 percent of the data and the testing gets the remaining 20 percent. The process of text data conversion to numerical values utilizes vectorization through TF-IDF (Term Frequency-Inverse Document Frequency). The TF-IDF (Term Frequency-Inverse Document Frequency) method displays word weight measurements according to (Zahoor *et al.*, 2020); weight increases when a word appears, but IDF calculates a weight that denotes term rarity in the word index. The main function of TF-IDF stands as an indicator to determine which words hold significance within documents (Wisky *et al.*, 2024).

Sentiments Classification

After the preprocessing phase, the comments will be classified based on negative, positive, and neutral. We tested two different kinds of classifiers for sentiment analysis. In the first stage, a probabilistic model that is effective in the classification of text was employed: Multinomial Naive Bayes. Complement Naïve Bayes was also used to avoid the potential negative effects of such problems as class imbalance. In addition, Support Vector Machines (SVM) were also worth investigating, with RBF and linear basis function (RBF) accordingly used as kernel functions to differentiate different patterns within the data. Also, we used Deep learning techniques such as Bilstm classifier. These models are trained on the TF-IDF features of the training data and then evaluated on the test data. SVM proves superior to NB and LR when it comes to classifying both texts and documents in English. The evaluation process and measurement involved three selected metrics known as precision, recall, and F1 value (Luo, 2021). The combination of Count Vectorizer with Rotten Tomatoes and IMDB datasets produced the highest accuracy for Bernoulli Naive Bayes. Multinomial Naive Bayes produced superior performance in sentiment analysis of IMDB movie reviews through the use of TF-IDF features. (Danyal *et al.*, 2024). The effectiveness of SVM techniques at categorizing customer evaluations stands at 90%, whereas Naive Bayes reaches 77% and KNN operates at 66% accuracy (İnan, 2024). The SVM classifier proved superior to all other models through its use of 2-gram TF-IDF features in sentiment analysis to analyze consumer opinions. The model featured 86% accuracy as its optimal performance compared to all other techniques, including NB, LSTM, and GRU (Azrir *et al.*, 2024). Two frequently used and successful classification methods for sentiment analysis are SVM and Naïve Bayes (Abbasi *et al.*, 2008; Saranya & Jayanthy, 2002). Furthermore, based on several metrics, including accuracy, precision, recall, and F1, it was discovered that SVM performs better than the other machine learning algorithms in sentiment analysis for Amazon reviews for electronics. It produces an accuracy of 87.07%, while a Decision Tree yields the lowest accuracy of 82.05% (Nagelli *et al.*, 2025).

Methods for sentiment analysis include evaluation techniques. Sentiment analysis is evaluated using a wide variety of evaluation models. The following are the most widely used evaluation models (Padmaja & Fatima, 2013; Sokolova *et al.*, 2006):

Accuracy: It calculates the percentage of accurate forecasts to all instances examined.

$$Accuracy = \frac{Correct\ Prediction}{Total\ Number\ of\ Instance}$$

Precision: It calculates the proportion of accurately predicted positive patterns in a positive class out of all the expected patterns.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

Recall: It calculates the proportion of correctly categorized positive patterns.

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

F1-score gives a single statistic that combines recall and precision into a single figure.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Performance can be assessed through evaluation metrics such as Precision, Recall, Accuracy score, F1 score, and confusion matrix. Such an analysis provides a comprehensive framework for evaluating sentiment classification performance as a whole.

Results and Discussion

The Naive Bayes classifier together with Support vector machine (SVM) and Bilstm serves for sentiment analysis. Precision, recall, F1-score together with total accuracy served as the measurement criteria for evaluating experimental outcomes. An evaluation using Complement Naive Bayes achieved 81.44% accuracy in sentiment evaluation. Complement Naive Bayes achieved a higher score than Multinomial NB in classifying Tweets between positive, negative, and neutral categories. The Complement Naïve Bayes Model showcases success within positive and negative sentiment classification according to Ray *et al.* (2024) findings. Table 2 displays the results obtained.

Table 2: Evaluation Metrics for Both Multinomial and Complement NB

Model Type	Accuracy	Precision	Recall	F1-Score
Complement Naive Bayes	81.44%	82.00%	81.00%	81.00%
Multinomial Naive Bayes	80.21%	81.00%	80.00%	80.00%

On the flip side, the RBF SVM and Linear SVM classifiers performed better than NB with accuracies of 88.64% and 8428%, respectively. Similarly, with a testing accuracy of 8974%, the SVM (RBF kernel) outperforms all other classifiers, including Random Forest, Multinomial Naive Bayes, and LSTM, as the best model on a Twitter data set for evaluating user sentiment (Patil *et al.*, 2024). While the Support Vector Machine (Linear Kernel) achieves the same accuracy measurement of 77% as the Support Vector Machine, because both algorithms perform identically. (RBF Kernel) for identifying and analyzing public opinions by categorizing them into positive and negative reviews (Idris & Mussalimun, 2024). Evaluation metrics for RBF SVM and Linear SVM are presented in Table 3.

Table 3: Evaluation Metrics for Both Linear and RBF SVM

Model Type	Accuracy	Precision	Recall	F1-Score
RBF SVM	88.64%	89.00%	89.00%	89.00%
Linear SVM	84.28%	84.00%	84.00%	84.00%

Overall, the SVM model demonstrated superior performance to Naive Bayes regarding accuracy levels and balanced sentiment category classification. This can be explained by the fact that the SVM effectively deals with non-linear patterns, and is particularly efficient in high-dimensional spaces, which is especially important in tasks involving the classification of text.

The BiLSTM classifier achieves a strong performance during the training and test processes. We noticed a high improvement in the 5 epoch, the accuracy increased from 69.97% to 93.66%, which indicates that the model has a good generalization to unseen data. Besides, the loss decreased from 0.6830 to 0.1895, recommending efficient learning that avoids overfitting. Table 4 shows the results of the BiLSTM classifier.

Table 4: BiLSTM Results

Metric	Training	Validation	Test
Accuracy	93.66%	88.16%	87.72%
Loss	0.1895	0.3862	-
Precision	0.88		
Recall	0.88		
F1-score	0.88		

Insights into the model's performance can be gained from looking at the Complement Naive Bayes confusion matrix presented in Figure 2. 1824 negative, 1775 neutral, and 1834 positive cases were accurately classified by the model. However, it misclassifies 225 positive cases as neutral and 343 negative examples as neutral. Overall, the model does well, although it has some difficulty differentiating between neutral and negative attitudes. This is probably because the linguistic patterns and contexts of these classes are similar.

While Multinomial NB shows 1654 negative, 1870 neutral, and 1827 positive cases were accurately classified by the model. Yet, 156 neutral occurrences are

DOI: 10.3844/jcssp.2025.1785.1794

incorrectly classified as positive, and 514 negative instances as neutral. The model does well overall, although it struggles to differentiate between comparable groups, such as positive and neutral.

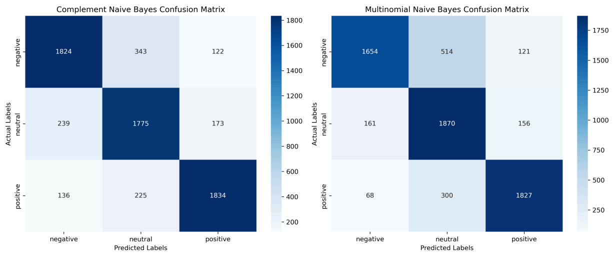

Fig. 2: Confusion Matrix for Multinomial and Complement NB

Figure 3 presents the confusion matrix of SVM, which displays essential performance metrics about the model. The RBF SVM model accurately identified 1935 positive, 1957 neutral, and 2021 negative cases. It did, however, make a few mistakes, misclassifying 129 positive cases as neutral and 181 negative ones as neutral. The model does well overall, although it has a little trouble telling the difference between neutral and negative attitudes. This is probably because of comparable contexts or overlapping linguistic patterns. While the Linear SVM accurately categorized 1878 positive, 1807 neutral, and 1937 negative cases. Nevertheless, made a few mistakes, misclassifying 176 positive instances as neutral and 237 negative ones as neutral. Although the model does well overall, it struggles to differentiate between neutral and negative attitudes, most likely there is a context similar or overlapping linguistic patterns.

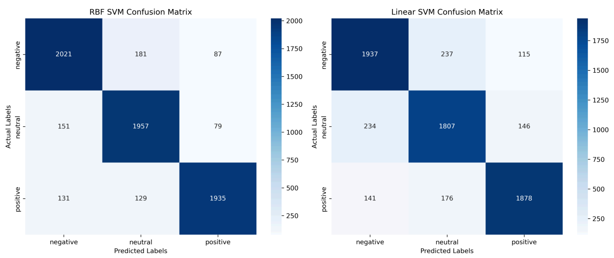

Fig. 3: Confusion Matrix for Linear and RBF SVM

The BiLSTM model shows its ability to categorize attitudes into positive, neutral, and negative categories, is demonstrated by the BiLSTM Confusion Matrix. Instances of 1981 being negative, 1954 being neutral, and 1921 being positive were all accurately predicted by the model. It did, however, make a few mistakes, misclassifying 204 negative cases as neutral and 153 neutral ones as negative. The model does well overall, although it has a little trouble telling the difference between neutral and negative attitudes. This is probably because of comparable contexts or overlapping linguistic patterns. Analysis of BiLSTM model results can be found in Figure 4, which displays its confusion matrix.

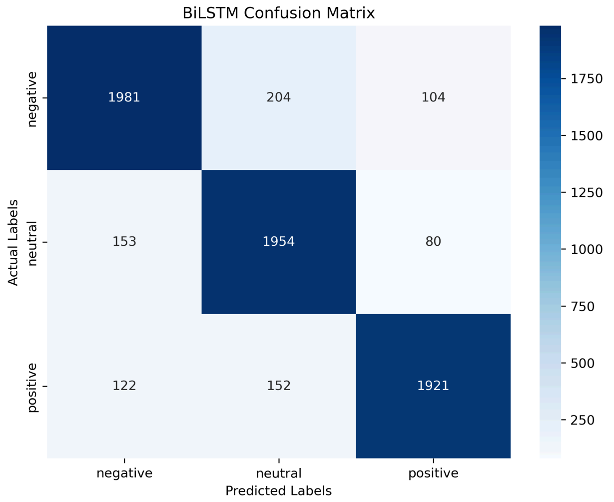

Fig. 4: BiLSTM Confusion Matrix

Conclusion

The research explored sentiment analysis effectiveness in customer review classification through Naive Bayes and Support Vector Machine (SVM) algorithm implementation. The analysis revealed that SVM outperformed Naive Bayes in terms of overall accuracy. Especially, RBF SVM outperformed other classifiers. The SVM classifier was found to be a great choice for classifying consumer reviews. Moreover, BiLSTM achieves impressive performance in dealing with classifying customer reviews.

Overall, Research shows BiLSTM represents a strong possibility as a method for sentiment analysis of customer review text. The research needs further improvement by implementing advanced deep learning algorithms along with the analysis of multiple datasets and text categorization methods for customer satisfaction evaluation. These results generally enhance the retail business that is interested in preserving its customers. Since customer reviews play an essential role in online shopping, as they offer both buyers and businesses insightful information (Panduro-Ramirez, 2024). Thus, Sentiment analysis can help retail businesses enhance customer retention by understanding customer feedback. Also, these reviews provide insightful information for prospective customers looking for recommendations and aid in raising the caliber of goods and services (Rana *et al.*, 2024).

Acknowledgment

I can acknowledge that I have not received any funding. I also acknowledge that I have not received administrative and technical support.

Funding Information

The research project did not acquire any outside financial support.

Author's Contributions

The section defines the work each author contributed to preparing and publishing this manuscript, including its development process.

Ethics

Every ethical problem that emerges following manuscript publication must be discussed by the authors.

References

Abbasi, A., Hsinchun, C., Thoms, S., & Tianjun, F. (2008). Affect Analysis of Web Forums and Blogs Using Correlation Ensembles. In *IEEE Transactions on Knowledge and Data Engineering* (Vol. 20, Issue 9, pp. 1168-1180). https://doi.org/10.1109/tkde.2008.51

Adak, A., Pradhan, B., & Shukla, N. (2022). Sentiment Analysis of Customer Reviews of Food Delivery Services Using Deep Learning and Explainable Artificial Intelligence: Systematic Review. *Foods*, *11*(10), 1500. https://doi.org/10.3390/foods11101500

Alharbi, L. M., & Qamar, A. M. (2022). Arabic Sentiment Analysis of Eateries' Reviews Using Deep Learning. In *Ingénierie des systèmes d information* (Vol. 27, Issue 3, pp. 503-508). https://doi.org/10.18280/isi.270318

Alhojely, S. (2016). Sentiment Analysis and Opinion Mining: A Survey. *International Journal of Computer Applications*, *150*(6), 22-25. https://doi.org/10.5120/ijca2016911545

AlQahtani, S. M. A. (2021). Product Sentiment Analysis for Amazon Reviews. *International Journal of Computer Science and Information Technology*, *13*(3), 15-30. https://doi.org/10.5121/ijcsit.2021.13302

Alsemaree, O., Alam, A. S., Gill, S. S., & Uhlig, S. (2024). Sentiment analysis of Arabic social media texts: A machine learning approach to deciphering customer perceptions. *Heliyon*, *10*(9), e27863. https://doi.org/10.1016/j.heliyon.2024.e27863

Angdresey, A., Sitanayah, L., & Tangka, I. L. H. (2025). Sentiment Analysis for Political Debates on YouTube Comments using BERT Labeling, Random Oversampling, and Multinomial Naïve Bayes. *Journal of Computing Theories and Applications*, *2*(3), 342-354. https://doi.org/10.62411/jcta.11668

Ara, J., Hasan, M. T., Al Omar, A., & Bhuiyan, H. (2020). Understanding Customer Sentiment: Lexical Analysis of Restaurant Reviews. *2020 IEEE Region 10 Symposium (TENSYMP)*. 2020 IEEE Region 10 Symposium (TENSYMP), Dhaka, Bangladesh. https://doi.org/10.1109/tensymp50017.2020.9230712

Azrir, A. H. I. bin A., Naveen, P., & Haw, S.-C. (2024). Sentiment Analysis using Machine Learning Models on Shopee Reviews. *Journal of System and Management Sciences*, *14*(2). https://doi.org/10.33168/jsms.2024.0213

Baccianella, S., Esuli, A., & Sebastiani, F. (2010). Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. *Lrec*, 2200-2204.

Cheng, Z.-W., & Li, S.-X. (2024). Attention Information Fusion Emotion Analysis Model Based on BiLSTM-BERT. *2024 4th International Conference on Consumer Electronics and Computer Engineering (ICCECE)*. 2024 4th International Conference on Consumer Electronics and Computer Engineering (ICCECE), Guangzhou, China. https://doi.org/10.1109/iccece61317.2024.10504175

Cui, J., Wang, Z., Ho, S.-B., & Cambria, E. (2023). Survey on sentiment analysis: evolution of research methods and topics. *Artificial Intelligence Review*, *56*(8), 8469-8510. https://doi.org/10.1007/s10462-022-10386-z

da Silva, N. F. F., Hruschka, E. R., & Hruschka, E. R. (2014). Tweet sentiment analysis with classifier ensembles. *Decision Support Systems*, *66*, 170-179. https://doi.org/10.1016/j.dss.2014.07.003

Danyal, M. M., Khan, S. S., Khan, M., Ullah, S., Ghaffar, M. B., & Khan, W. (2024). Sentiment analysis of movie reviews based on NB approaches using TF-IDF and count vectorizer. *Social Network Analysis and Mining*, *14*(1), 87. https://doi.org/10.1007/s13278-024-01250-9

Devika, M. D., Sunitha, C., & Ganesh, A. (2016). Sentiment Analysis: A Comparative Study on Different Approaches. *Procedia Computer Science*, *87*, 44-49. https://doi.org/10.1016/j.procs.2016.05.124

Du, J., Zhao, H., Ye, Z., & Li, M. (2024). Bert-BiLSTM Model for Sentiment Analysis Using Contextual Embeddings and Bidirectional Dependencies. *International Symposium on Internet of Things and Smart Cities*. International Symposium on Internet of Things and Smart Cities, Nanjing, China. https://doi.org/10.1109/isitsc64373.2024.00022

Fernandez, A., Garcia, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary. *Journal of Artificial Intelligence Research*, *61*, 863-905. https://doi.org/10.1613/jair.1.11192

Geler, Z., Savić, M., Bratić, B., Kurbalija, V., Ivanović, M., & Dai, W. (2021). Sentiment prediction based on analysis of customers assessments in food serving businesses. *Connection Science*, *33*(3), 674-692. https://doi.org/10.1080/09540091.2020.1870436

Gunasekaran, K. P. (2023). Exploring Sentiment Analysis Techniques in Natural Language Processing: A Comprehensive Review. *ArXiv:2305.14842*.

Gupta, K., Jiwani, N., & Afreen, N. (2023). A Combined Approach of Sentimental Analysis Using Machine Learning Techniques. *Revue d'Intelligence Artificielle*, *37*(1), 1-6. https://doi.org/10.18280/ria.370101

Hagen, M., Potthast, M., Büchner, M., & Stein, B. (2015). *Twitter Sentiment Detection via Ensemble Classification Using Averaged Confidence Scores*. 741-754. https://doi.org/10.1007/978-3-319-16354-3_81

Hamad, M. M., Salih, M. A., & Jaleel, R. A. (2021). Sentiment Analysis of Restaurant Reviews in Social Media using Naïve Bayes. *Applications of Modelling and Simulation*, *5*, 166-172.

Hung, L. P., & Alias, S. (2023). Beyond Sentiment Analysis: A Review of Recent Trends in Text Based Sentiment Analysis and Emotion Detection. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, *27*(1), 84-95. https://doi.org/10.20965/jaciii.2023.p0084

Idris, M. & Mussalimun (2024). Sentiment Analysis of Google Play Store Reviews using Support Vector Machines. *International Journal of Applied Information Systems (IJAIS)*, *12*(42), 48-53.

İnan, H. E. (2024). Comparison of Machine Learning Algorithms for Classification of Hotel Reviews: Sentiment Analysis of TripAdvisor Reviews. *GSI Journals Serie A: Advancements in Tourism Recreation and Sports Sciences*, *7*(1), 111-122. https://doi.org/10.53353/atrss.1327615

Jagdale, R. S., Shirsat, V. S., & Deshmukh, S. N. (2019). *Sentiment Analysis on Product Reviews Using Machine Learning Techniques*. 639-647. https://doi.org/10.1007/978-981-13-0617-4_61

Jain, P. K., Pamula, R., & Srivastava, G. (2021). A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. *Computer Science Review*, *41*, 100413. https://doi.org/10.1016/j.cosrev.2021.100413

Jianqiang, Z., Xiaolin, G., & Xuejun, Z. (2018). Deep Convolution Neural Networks for Twitter Sentiment Analysis. *IEEE Access*, *6*, 23253-23260. https://doi.org/10.1109/access.2017.2776930

Krishna, A., Akhilesh, V., Aich, A., & Hegde, C. (2019). *Sentiment Analysis of Restaurant Reviews Using Machine Learning Techniques*. 687-696. https://doi.org/10.1007/978-981-13-5802-9_60

Li, X., Chen, L., & Chen, B. (2024). BERT-BiLSTM-Attention model for sentiment analysis on Chinese stock reviews. *Research Square*. https://doi.org/10.21203/rs.3.rs-4023113/v1

Luo, J. M., Vu, H. Q., Li, G., & Law, R. (2021). Understanding service attributes of robot hotels: A sentiment analysis of customer online reviews. *International Journal of Hospitality Management*, *98*, 103032. https://doi.org/10.1016/j.ijhm.2021.103032

Luo, X. (2021). Efficient English text classification using selected Machine Learning Techniques. *Alexandria Engineering Journal*, *60*(3), 3401-3409. https://doi.org/10.1016/j.aej.2021.02.009

Mohammed, R., Rawashdeh, J., & Abdullah, M. (2020). Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results. *2020 11th International Conference on Information and Communication Systems (ICICS)*. 2020 11th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan. https://doi.org/10.1109/icics49469.2020.239556

Montejo-Ráez, A., Martínez-Cámara, E., Martín-Valdivia, M. T., & Ureña-López, L. A. (2014). A knowledge-based approach for polarity classification in Twitter. *Journal of the Association for Information Science and Technology*, *65*(2), 414-425. https://doi.org/10.1002/asi.22984

Moreo, A., Esuli, A., & Sebastiani, F. (2016). Distributional Random Oversampling for Imbalanced Text Classification. *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '16: The 39th International ACM SIGIR conference on research and development in Information Retrieval, Pisa Italy. https://doi.org/10.1145/2911451.2914722

Muhibullaev, B., & Kim, J. (2024). Accurate Information Type Classification for Software Issue Discussions With Random Oversampling. *IEEE Access*, *12*, 65373-65385. https://doi.org/10.1109/access.2024.3398732

Mujahid, M., Kına, E., Rustam, F., Villar, M. G., Alvarado, E. S., De La Torre Diez, I., & Ashraf, I. (2024). Data oversampling and imbalanced datasets: an investigation of performance for machine learning and feature engineering. *Journal of Big Data*, *11*(1). https://doi.org/10.1186/s40537-024-00943-4

Nagelli, A., Saleena, B., & Prakash, B. (2025). Comparative evaluation of deep learning and machine learning techniques for sentiment analysis of electronic product review data. *ITM Web of Conferences*, *74*, 03008. https://doi.org/10.1051/itmconf/20257403008

Padmaja, S., & Fatima, S. S. (2013). Opinion Mining and Sentiment Analysis - An Assessment of Peoples' Belief: A Survey. In *International Journal of Ad hoc, Sensor & Ubiquitous Computing* (Vol. 4, Issue 1, pp. 21-33). https://doi.org/10.5121/ijasuc.2013.4102

Paltoglou, G., & Thelwall, M. (2012). Twitter, MySpace, Digg. *ACM Transactions on Intelligent Systems and Technology*, *3*(4), 1-19. https://doi.org/10.1145/2337542.2337551

Panduro-Ramirez, J. (2024). Sentiment Analysis in Customer Reviews for Product Recommendation in E-commerce Using Machine Learning. *2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*. https://doi.org/10.1109/accai61061.2024.10602027

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, *2*, 1-135. https://doi.org/10.1561/1500000011

Park, S., Cho, J., Park, K., & Shin, H. (2021). Customer sentiment analysis with more sensibility. *Engineering Applications of Artificial Intelligence*, *104*, 104356. https://doi.org/10.1016/j.engappai.2021.104356

Patil, S., Subil, D., Nasar, N., Kokatnoor, S. A., Krishnan, B., & Kumar, S. (2024). Text Mining - A Comparative Review of Twitter Sentiments Analysis. *Recent Advances in Computer Science and Communications*, *17*(1). https://doi.org/10.2174/2666255816666230726140726

Pratiwi, R. W., Handayani, S. F., Dairoh, D., & Af'idah, D. I. (2024). Sentiment analysis on Twitter reviews data using the bidirectional long short-term memory (BiLTM). *AIP Conference Proceedings*. Proceedings of the Tegal International Conference on Applied Sciences 2022: Applied Science Research Post-covid-19 Pandemic, Tegal, Indonesia. https://doi.org/10.1063/5.0199238

Rajasekaran, R., Kanumuri, U., Siddhardha Kumar, M., Ramasubbareddy, S., & Ashok, S. (2019). Sentiment Analysis of Restaurant Reviews. *Smart Intelligent Computing and Applications*, 383-390. https://doi.org/10.1007/978-981-13-1927-3_41

Rana, M. R. R., Nawaz, A., Ali, T., Alattas, A. S., & AbdElminaam, D. S. (2024). Sentiment Analysis of Product Reviews Using Transformer Enhanced 1D-CNN and BiLSTM. *Cybernetics and Information Technologies*, *24*(3), 112-131. https://doi.org/10.2478/cait-2024-0028

Ray, A., Dutta, S. K., Dey, S., Roy, T., Mukherjee, P., Bandyopadhyay, S., & Thakur, S. S. (2024). Sentiment Analysis of Amazon Product Reviews: A Comprehensive Evaluation Using Naïve Bayes Classifiers. *Computing and Machine Learning*, 347-358. https://doi.org/10.1007/978-981-97-7571-2_27

Ryan, A. A., Arpita, H. D., Tabassum, A., Ahammad, Md. S., & Akram, M. A. (2024). A Comparative Analysis Between Deep Learning and Machine Learning Algorithms Based on User Review Sentiment Analysis from Various OTT Applications. 2024 International Conference on Computer, Electrical & Communication Engineering (ICCECE), Kolkata, India. https://doi.org/10.1109/iccece58645.2024.10497229

Saif, H., He, Y., & Alani, H. (2012). *Semantic Sentiment Analysis of Twitter*. 508-524. https://doi.org/10.1007/978-3-642-35176-1_32

Saranya, K., & Jayanthy, S. (2017). Onto-based sentiment classification using machine learning techniques. 2017 4th International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore. https://doi.org/10.1109/iciiecs.2017.8276047

Singh, N. K., Tomar, D. S., & Sangaiah, A. K. (2020). Sentiment analysis: a review and comparative analysis over social media. *Journal of Ambient Intelligence and Humanized Computing*, *11*(1), 97-117. https://doi.org/10.1007/s12652-018-0862-8

Sneha, H. R., & Annappa, B. (2024). Exploratory Analysis of Methods, Techniques, and Metrics to Handle Class Imbalance Problem. In *Procedia Computer Science* (Vol. 235, pp. 863-877). https://doi.org/10.1016/j.procs.2024.04.082

Sokolova, M., Japkowicz, N., & Szpakowicz, S. (2006). Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation. *AI 2006: Advances in Artificial Intelligence*, 1015-1021. https://doi.org/10.1007/11941439_114

Teotia, T., Sharma, M., & Elmiligi, H. (2023). A comprehensive study of machine learning approaches for keystroke dynamics authentication. In *International Journal of Biometrics* (Vol. 15, Issue 5, pp. 560-586). https://doi.org/10.1504/ijbm.2023.133148

Thelwall, M., Buckley, K., & Paltoglou, G. (2012). Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, *63*(1), 163-173. https://doi.org/10.1002/asi.21662

Thomas, R., & Jeba, J. R. (2024). A novel framework for an intelligent deep learning based product recommendation system using sentiment analysis (SA). In *Automatika* (Vol. 65, Issue 2, pp. 410-424). https://doi.org/10.1080/00051144.2023.2295148

Verma, P., Bhardwaj, T., Bhatia, A., & Mursleen, M. (2023). *Sentiment Analysis "Using SVM, KNN and SVM with PCA"* (pp. 35-53). Springer International Publishing. https://doi.org/10.1007/978-3-031-28581-3_5

Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A Survey on Sentiment Analysis Methods, Applications, and Challenges. *Artificial Intelligence Review*, *55*(7), 5731-5780. https://doi.org/10.1007/s10462-022-10144-1

Wisky, I. A., Defit, S., & Nurcahyo, G. W. (2024). Development of extraction features for Detecting Adolescent Personality with Machine Learning Algorithms. In *JOIV: International Journal on Informatics Visualization* (Vol. 8, Issues 3-2, p. 1606). https://doi.org/10.62527/joiv.8.3-2.3091

Yi, S., & Liu, X. (2020). Machine learning based customer sentiment analysis for recommending shoppers, shops based on customers' review. *Complex & Intelligent Systems*, *6*(3), 621-634. https://doi.org/10.1007/s40747-020-00155-2

Zahoor, K., Bawany, N. Z., & Hamid, S. (2020). Sentiment Analysis and Classification of Restaurant Reviews using Machine Learning. *2020 21st International Arab Conference on Information Technology (ACIT)*. 2020 21st International Arab Conference on Information Technology (ACIT), Giza, Egypt. https://doi.org/10.1109/acit50332.2020.9300098

Zhou, Q., Xu, Z., & Yen, N. Y. (2019). User sentiment analysis based on social network information and its application in consumer reconstruction intention. *Computers in Human Behavior*, *100*, 177-183. https://doi.org/10.1016/j.chb.2018.07.006