PROJECT REPORT

On

Sentiment Analysis using Machine Learning & Deep Learning (Amazon Music Instruments Reviews)

Submitted in partial fulfillment of the requirements for the degree of

BACHELOR OF TECHNOLOGY

Computer Science & Engineering (DS)

Submitted by
Adarsh Singh (FET/BCE/2021-25/031)
Shrestha Kumar Agarwal (FET/BCE/2021-25/040)

Guided by
Dr. Deepak Sahoo

SRI SRI UNIVERSITY
LEARN • LEAD • SERVE

FACULTY OF ENGINEERING & TECHNOLOGY

SRI SRI UNIVERSITY

Ward No. 3, Sandhapur, Godisahi, Odisha 754006

APRIL 2025

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to our project guide, Dr. Deepak Sahoo, for his invaluable guidance, continuous support, and insightful feedback throughout the development of this project. His expertise and encouragement were instrumental in helping us overcome numerous challenges.

We are particularly grateful for the time he dedicated to reviewing my work, providing detailed suggestions, and ensuring the project maintained high academic standards. His approach to problem-solving and attention to detail has not only contributed to the success of this project but has also helped me grow as a researcher.

We would also like to extend my appreciation to my colleagues and friends for their support, constructive discussions, and collaboration throughout this journey. Their enthusiasm and willingness to share knowledge have made this experience both enriching and enjoyable.

The completion of this project would not have been possible without the support and encouragement of all these individuals. Thank you for being an integral part of this academic endeavor.

Adarsh Singh
Shrestha Kumar Agarwal

ABSTRACT

This technical paper presents a comprehensive study on sentiment analysis applied to customer reviews of musical instruments available on Amazon. The primary objective is to evaluate and compare the effectiveness of traditional machine learning algorithms and deep learning models in accurately classifying textual sentiment. The project is driven by the increasing reliance on customer feedback for product evaluation and the growing significance of natural language processing (NLP) in extracting meaningful insights from unstructured data.

To achieve this, the dataset underwent a series of preprocessing steps, including text normalization, tokenization, stopword removal, and vectorization. Feature extraction techniques such as TF-IDF were applied to convert textual data into numerical representations suitable for model training. Several supervised learning algorithms were explored, including Logistic Regression, Support Vector Machine (SVM), and Random Forest, along with a deep learning model based on a Bidirectional Long Short-Term Memory (Bi-LSTM) architecture.

The models were evaluated using standard performance metrics such as accuracy, precision, recall, and F1-score. Among the traditional classifiers, Logistic Regression emerged as the most effective, achieving an accuracy of 95.27% along with strong precision and recall across sentiment classes. The Bi-LSTM model, while slightly less accurate at 82%, demonstrated competitive performance and a better understanding of the context in some cases, particularly for complex sentiment patterns.

This study highlights the advantages and limitations of both machine learning and deep learning approaches in sentiment classification tasks. It also offers valuable insights into consumer perceptions and preferences within the musical instruments category on Amazon. The findings underscore the practical implications of sentiment analysis in e-commerce and emphasize the importance of model selection based on specific use-case requirements and data characteristics.

TABLE OF CONTENTS

1. Introduction

1.1 Background and Motivation

The exponential growth of e-commerce platforms has generated vast amounts of user-generated content in the form of product reviews. These reviews contain valuable insights into customer satisfaction, product quality, and market trends. Sentiment analysis allows businesses to automatically process and understand customer opinions at scale, providing actionable intelligence for product improvement and marketing strategies.

The music instrument industry represents a particularly interesting domain for sentiment analysis due to the subjective nature of musical equipment quality and the technical expertise often reflected in reviews. Understanding the nuanced feedback provided by musicians and audio professionals can help manufacturers improve product design and retailers enhance their offerings.

1.2 Problem Statement

Customer reviews on e-commerce platforms like Amazon contain valuable sentiment information that, when extracted and analyzed properly, can provide insights into product performance and customer satisfaction. However, manual analysis of thousands of reviews is impractical. This research aims to develop and evaluate automated sentiment analysis techniques to accurately classify Amazon music instrument reviews as positive, negative, or neutral.

1.3 Aim and Objectives

The primary objectives of this study are:

1. To implement and compare various machine learning algorithms for sentiment classification on Amazon music instrument reviews.
2. To analyze the effectiveness of different feature extraction techniques for sentiment analysis
3. To evaluate the performance of traditional machine learning versus deep learning approaches (including Bi-LSTM) for sentiment classification.
4. To identify linguistic patterns and key terms associated with different sentiment categories in music instrument reviews.

2. Literature Review

2.1 Previous Work in Sentiment Analysis

Sentiment analysis has evolved significantly over the past decade. Early approaches relied heavily on lexicon-based methods that assigned sentiment scores to words based on pre-compiled dictionaries. Pang and Lee (2008) pioneered machine learning approaches to sentiment analysis by treating it as a classification problem. Later, Taboada et al. (2011) demonstrated the effectiveness of semantic orientation calculators for determining text polarity.

2.2 Machine Learning Approaches

Traditional machine learning methods have shown considerable success in sentiment analysis tasks. Pang et al. (2002) compared Naive Bayes, Maximum Entropy, and Support Vector Machines (SVM) for sentiment classification of movie reviews. Dave et al. (2003) applied similar techniques to product reviews. Recent work by Uysal and Gunal (2014) demonstrated that ensemble methods like Random Forests can outperform individual classifiers in certain sentiment analysis contexts.

2.3 Deep Learning Approaches

Deep learning has revolutionized sentiment analysis in recent years. Socher et al. (2013) introduced Recursive Neural Tensor Networks for sentiment analysis. Later, Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, became popular due to their ability to capture sequential dependencies in text (Wang et al., 2016). Bidirectional LSTMs (Bi-LSTMs) further improved performance by processing sequences in both forward and backward directions (Zhou et al., 2016). Transformer-based models like BERT (Devlin et al., 2019) have recently achieved state-of-the-art results on various sentiment analysis benchmarks.

3. Dataset Description

3.1 Amazon Musical Instruments Reviews Overview

This study utilized the Amazon Music Reviews dataset available on Kaggle, which contains customer reviews for musical instruments sold on Amazon. The dataset includes the following key fields:

- **reviewerID**: ID of the reviewer
- **asin**: ID of the product
- **reviewerName**: Name of the reviewer
- **helpful**: Helpfulness rating of the review
- **reviewText**: Text of the review
- **overall**: Product rating (1-5 stars)
- **summary**: Summary of the review
- **unixReviewTime**: Time of the review (unix time)
- **reviewTime**: Time of the review (raw format)

The dataset contains a diverse range of reviews covering various musical instruments including guitars, keyboards, recording equipment, and accessories.

3.2 Data Preprocessing Steps

Several preprocessing steps were implemented to prepare the text data for analysis:

1. Handling missing values: Missing review texts were replaced with empty strings
2. Text combination: The review text and summary fields were combined to form a comprehensive review
3. Text cleaning:
    - Conversion to lowercase
    - Removal of punctuation
    - Removal of numerical values
    - Removal of URLs and hyperlinks
    - Removal of newline characters
4. Text processing:
    - Tokenization of words using NLTK
    - Removal of stopwords (except for "not" to preserve negations)
    - Lemmatization using WordNetLemmatizer to reduce words to their base forms

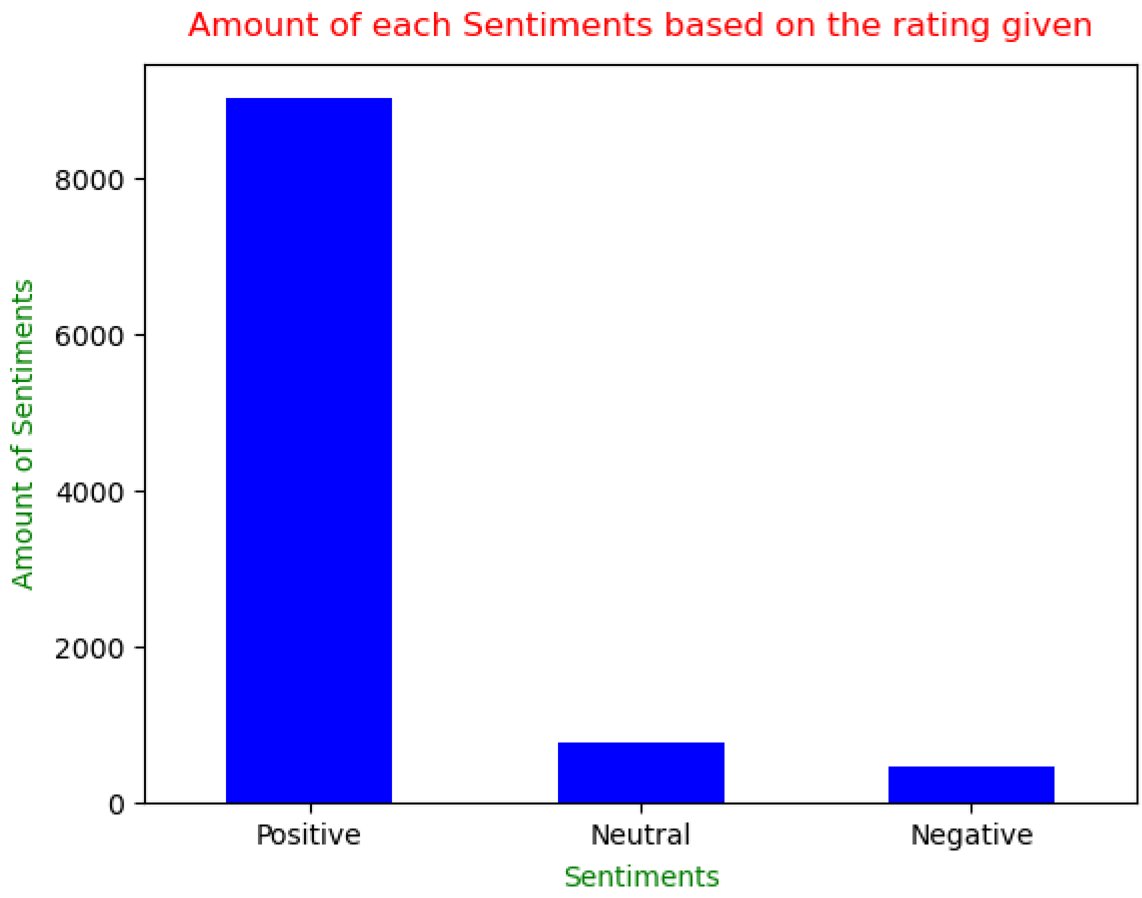
3.3 Exploratory Data Analysis

The exploratory data analysis revealed several important characteristics of the dataset:

Rating Distribution: The majority of products received high ratings, with an average score of 4.48 out of 5. This indicated a potential class imbalance issue, with positive reviews significantly outnumbering negative ones.

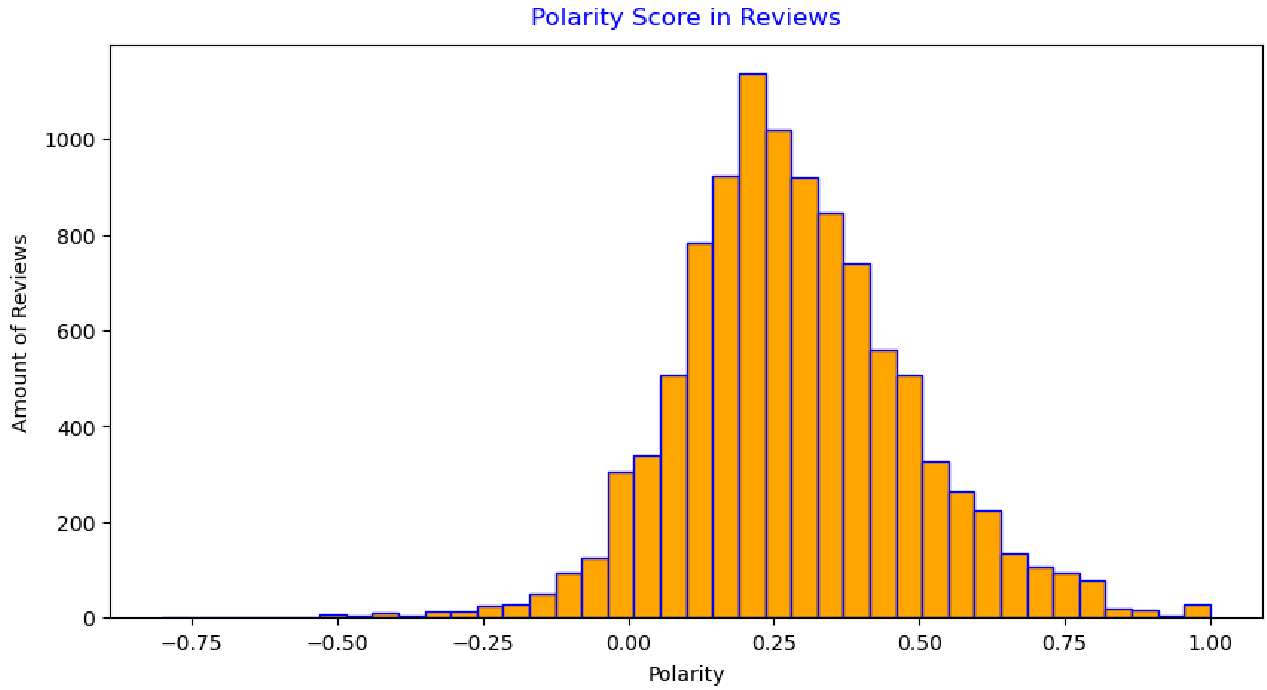
Sentiment Distribution: Reviews were categorized as:

- Positive: Ratings > 3
- Neutral: Ratings = 3
- Negative: Ratings < 3

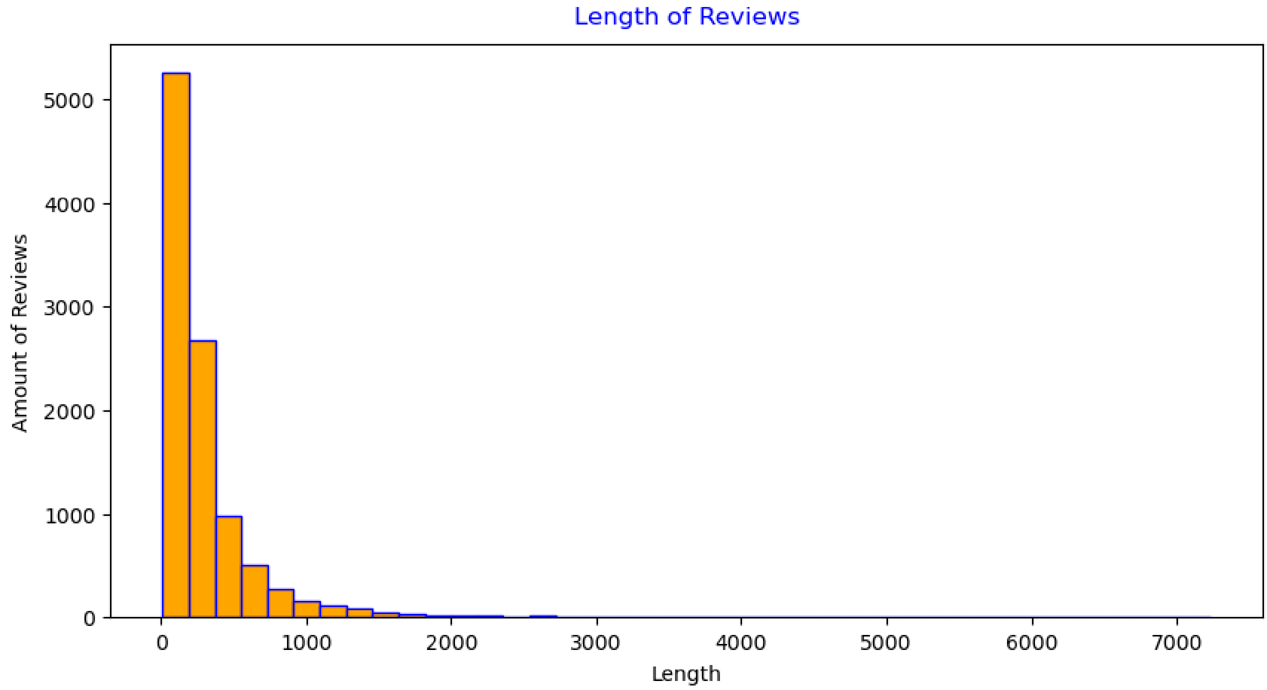


The analysis confirmed a substantial imbalance with positive reviews dominating the dataset.

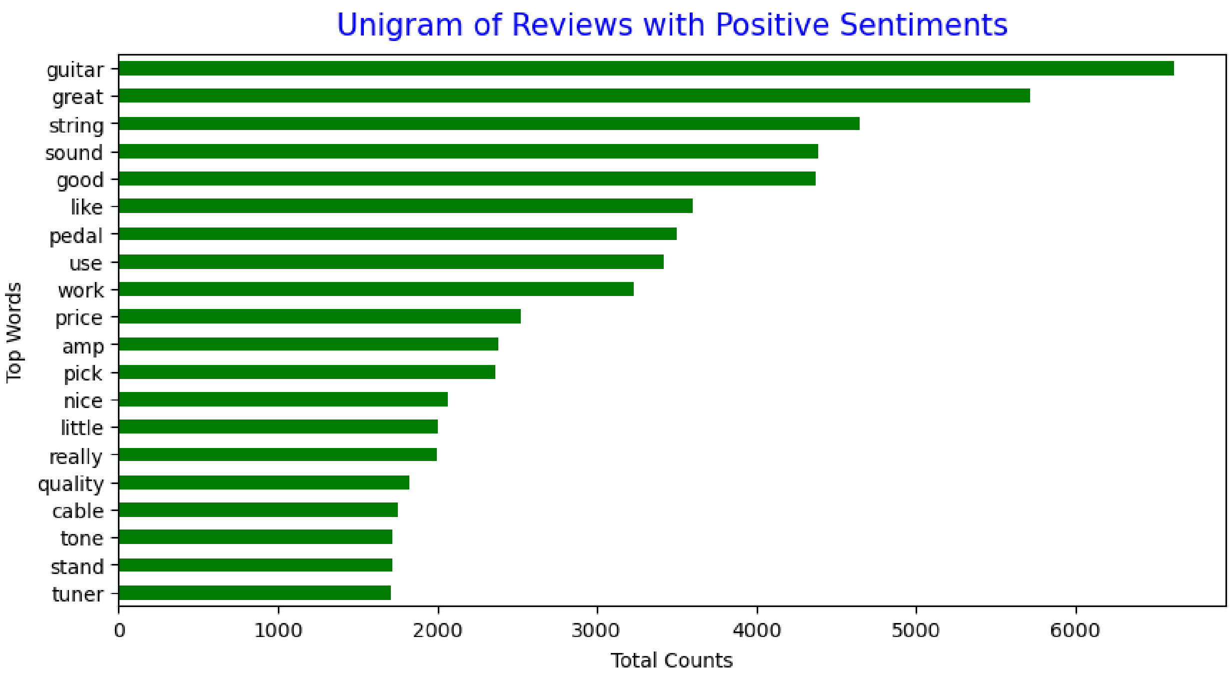
Polarity Analysis: Using TextBlob, the sentiment polarity of each review was calculated. The histogram of polarity scores showed a normal-like distribution with a positive skew, confirming the prevalence of positive sentiments.



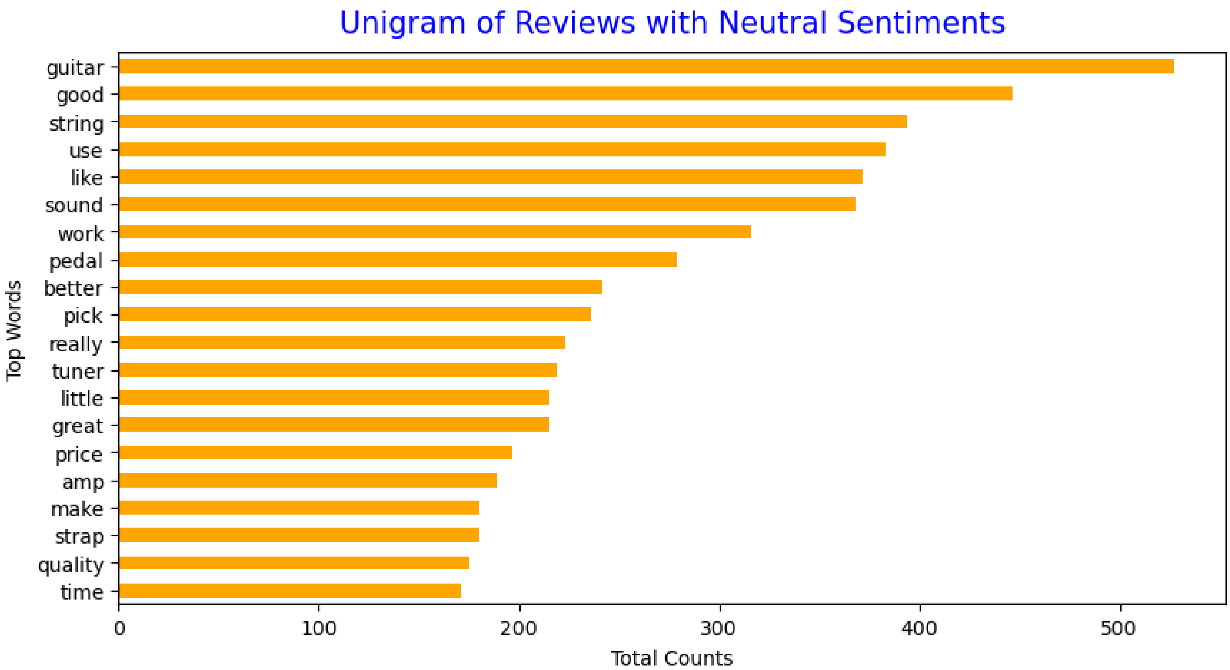
Text Length Analysis: Review length analysis revealed that most reviews contained between 0-200 words, with a right-skewed distribution.



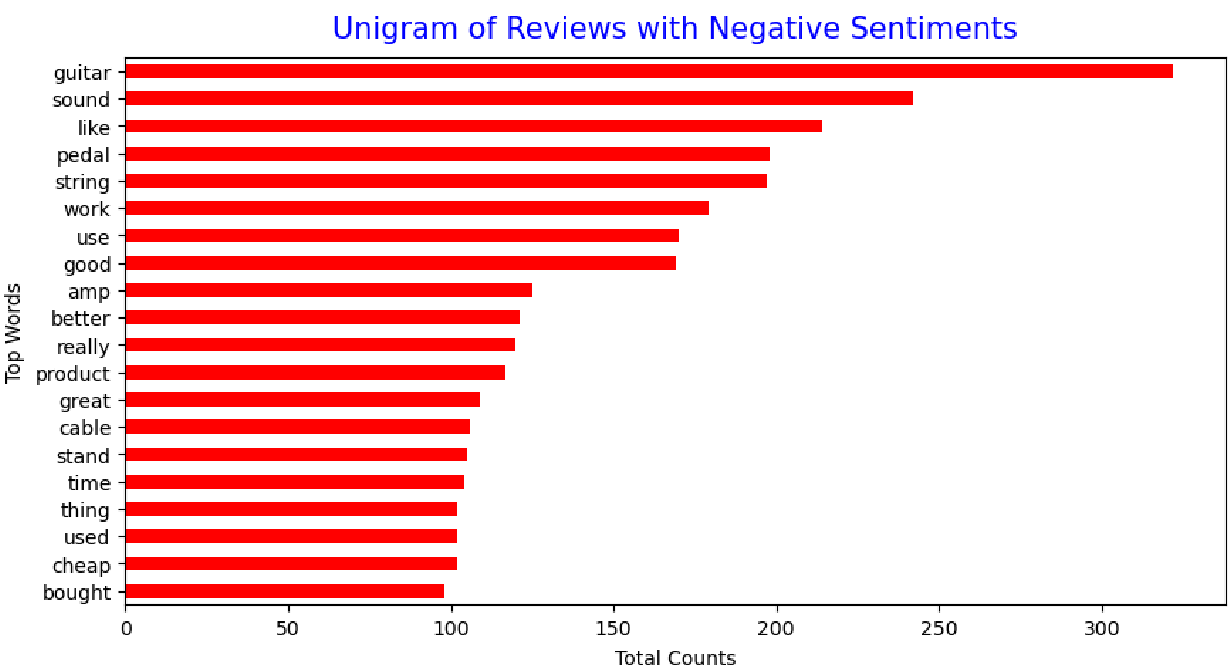
N-gram Analysis: Unigram, bigram, and trigram analyses were performed for positive, neutral, and negative reviews to identify frequently occurring terms in each sentiment category. Word clouds were also generated to visualize the most prominent terms.



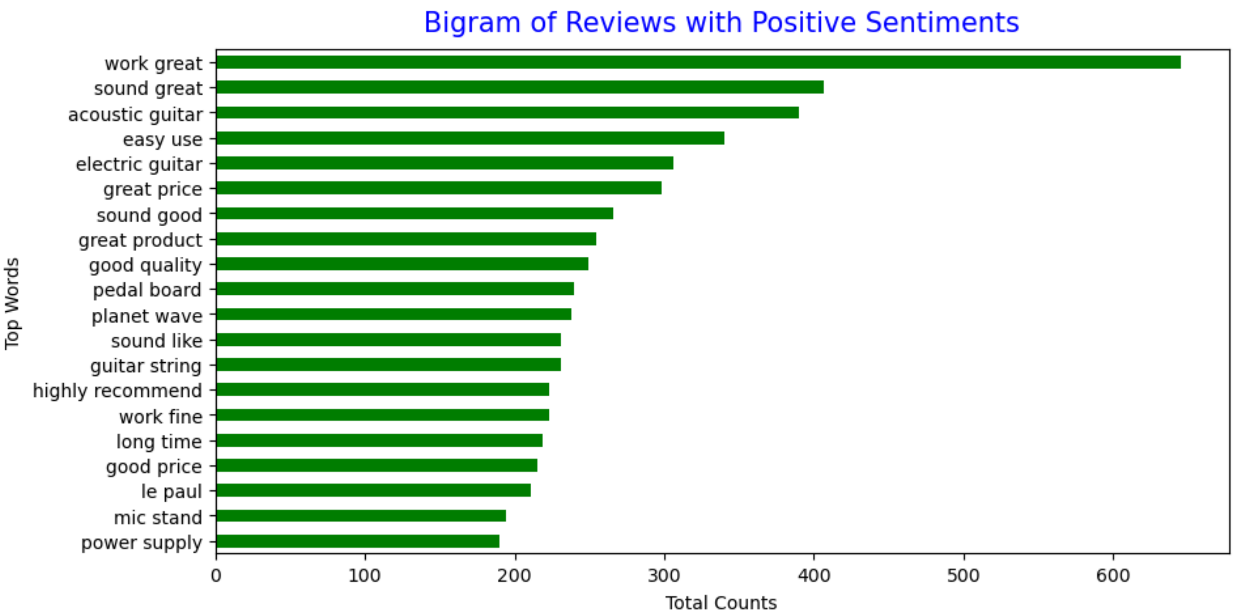
Unigram of Reviews with Positive Sentiments
Top Words
guitar
great
string
sound
good
like
pedal
use
work
price
amp
pick
nice
little
really
quality
cable
tone
stand
tuner
0
1000
2000
3000
4000
5000
6000
Total Counts

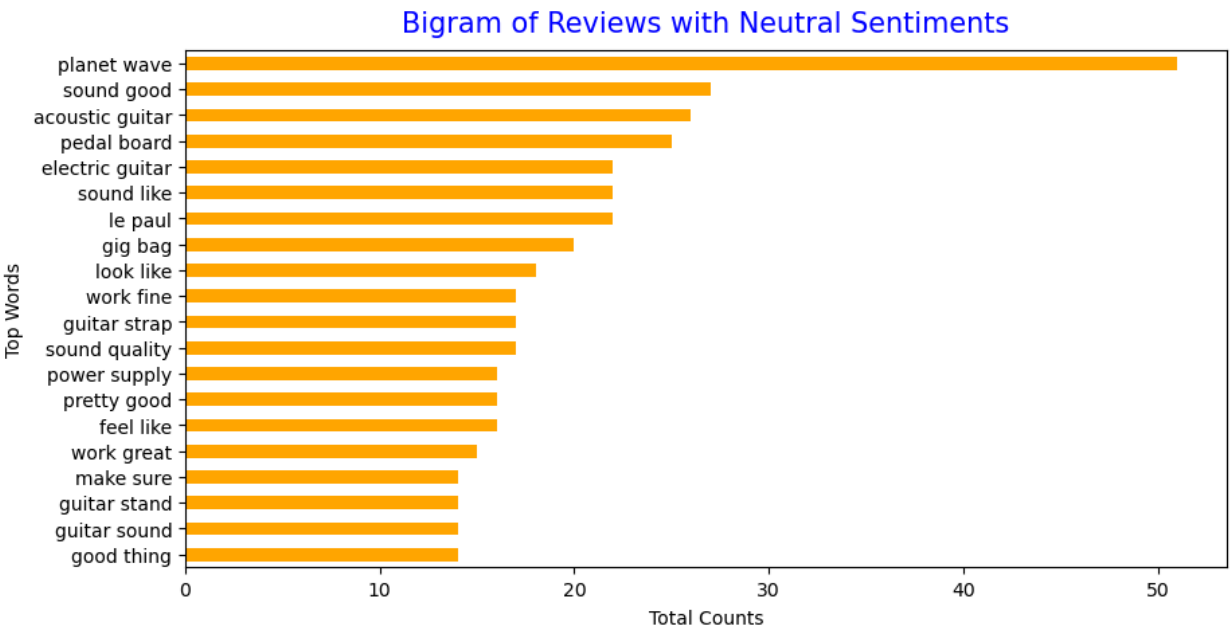


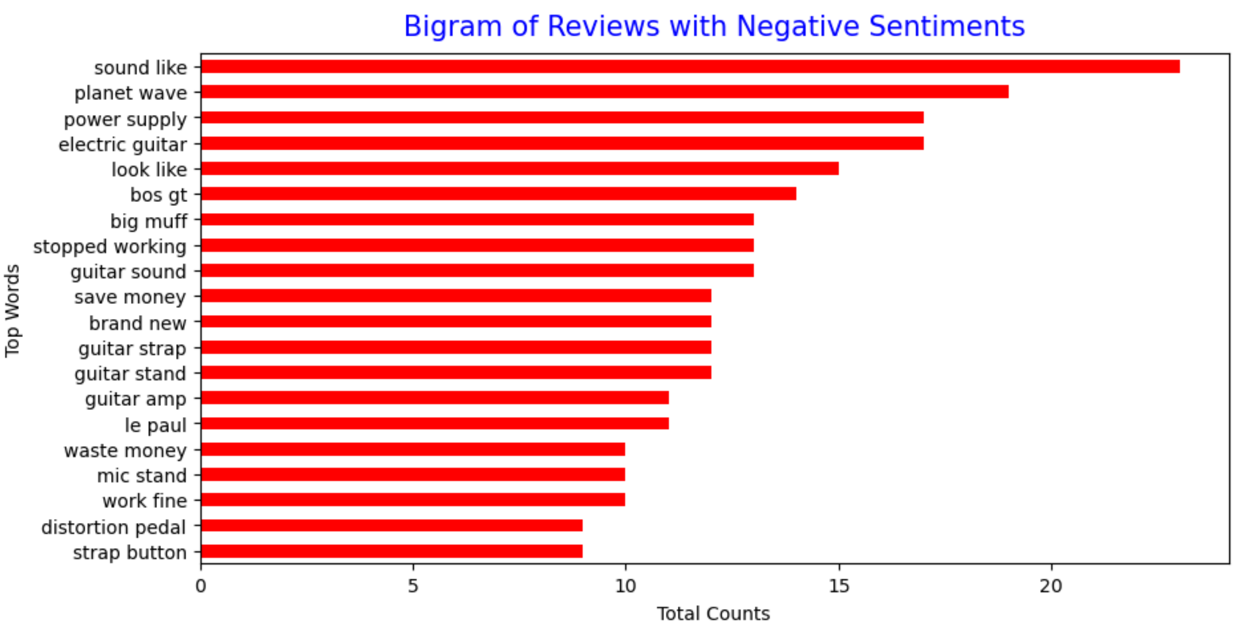
Unigram of Reviews with Neutral Sentiments
Top Words
guitar
good
string
use
like
sound
work
pedal
better
pick
really
tuner
little
great
price
amp
make
strap
quality
time
0
100
200
300
400
500
Total Counts

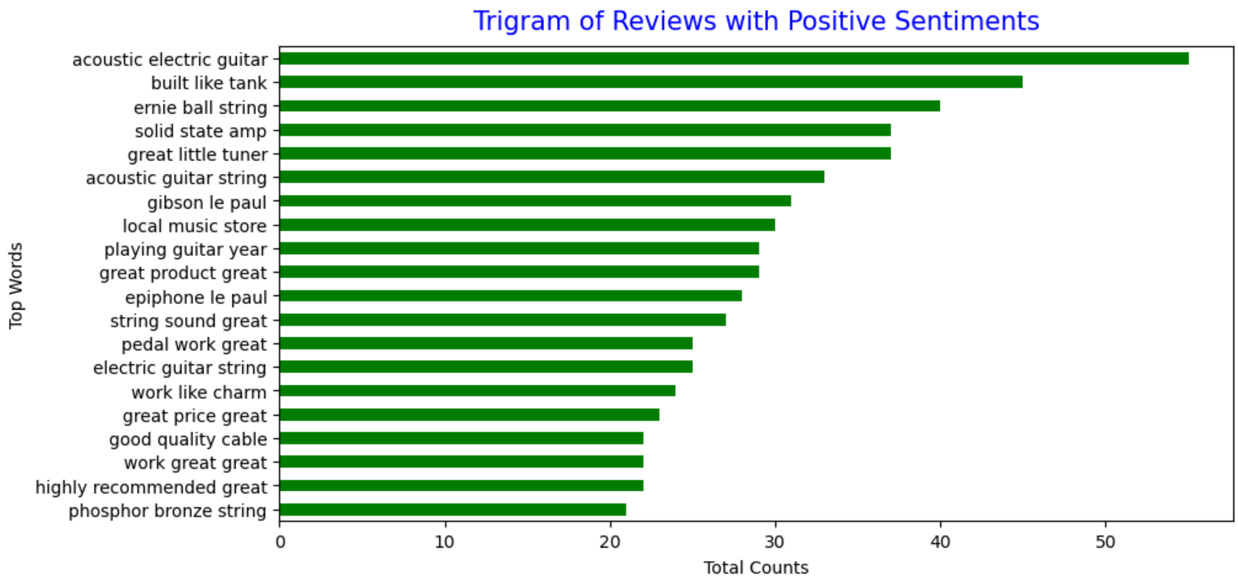


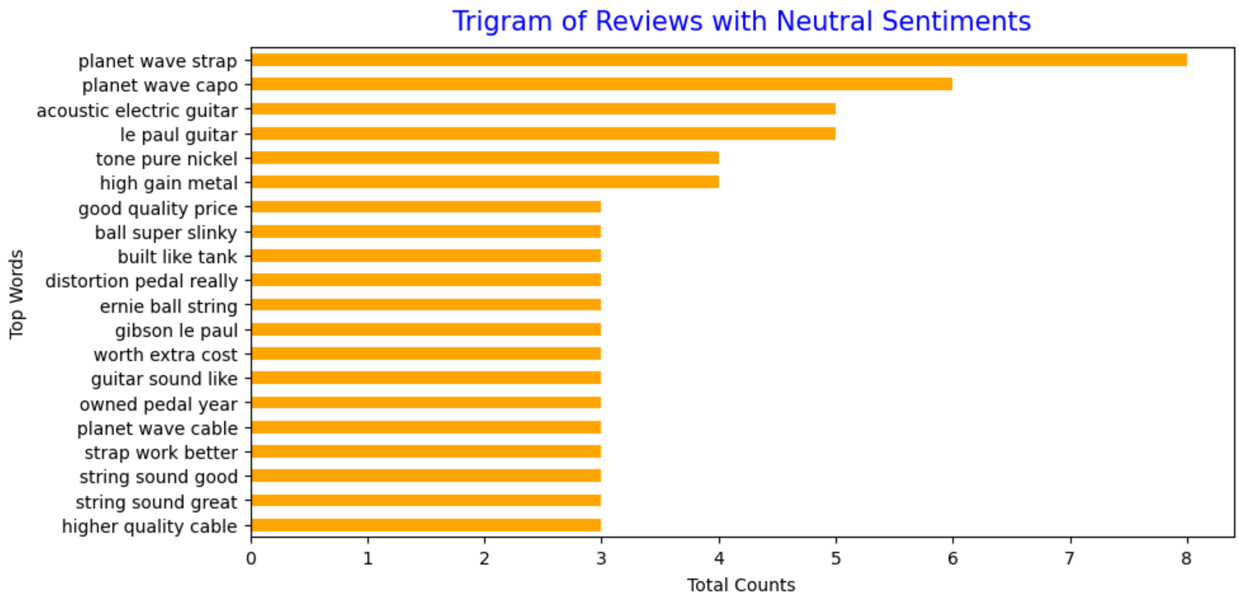
Unigram of Reviews with Negative Sentiments
Top Words
guitar
sound
like
pedal
string
work
use
good
amp
better
really
product
great
cable
stand
time
thing
used
cheap
bought
0
50
100
150
200
250
300
Total Counts

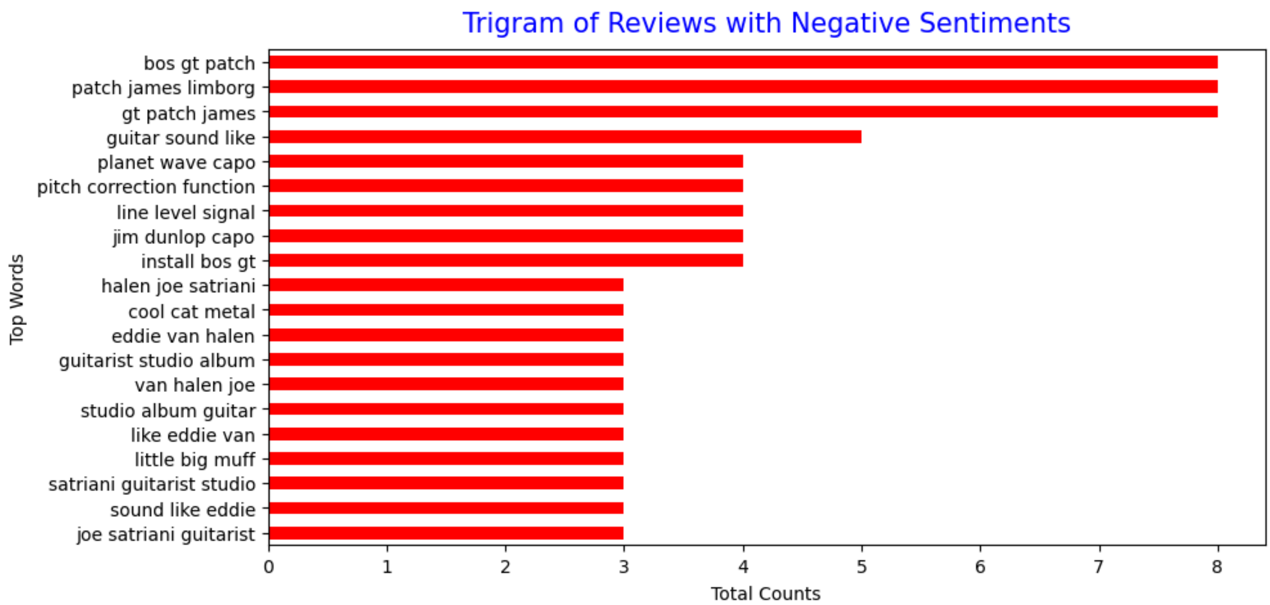

Bigram of Reviews with Positive Sentiments
work great
sound great
acoustic guitar
easy use
electric guitar
great price
sound good
great product
good quality
pedal board
planet wave
sound like
guitar string
highly recommend
work fine
long time
good price
le paul
mic stand
power supply
Top Words
0
100
200
300
400
500
600
Total Counts


Bigram of Reviews with Neutral Sentiments
planet wave
sound good
acoustic guitar
pedal board
electric guitar
sound like
le paul
gig bag
look like
work fine
guitar strap
sound quality
power supply
pretty good
feel like
work great
make sure
guitar stand
guitar sound
good thing
Top Words
0
10
20
30
40
50
Total Counts

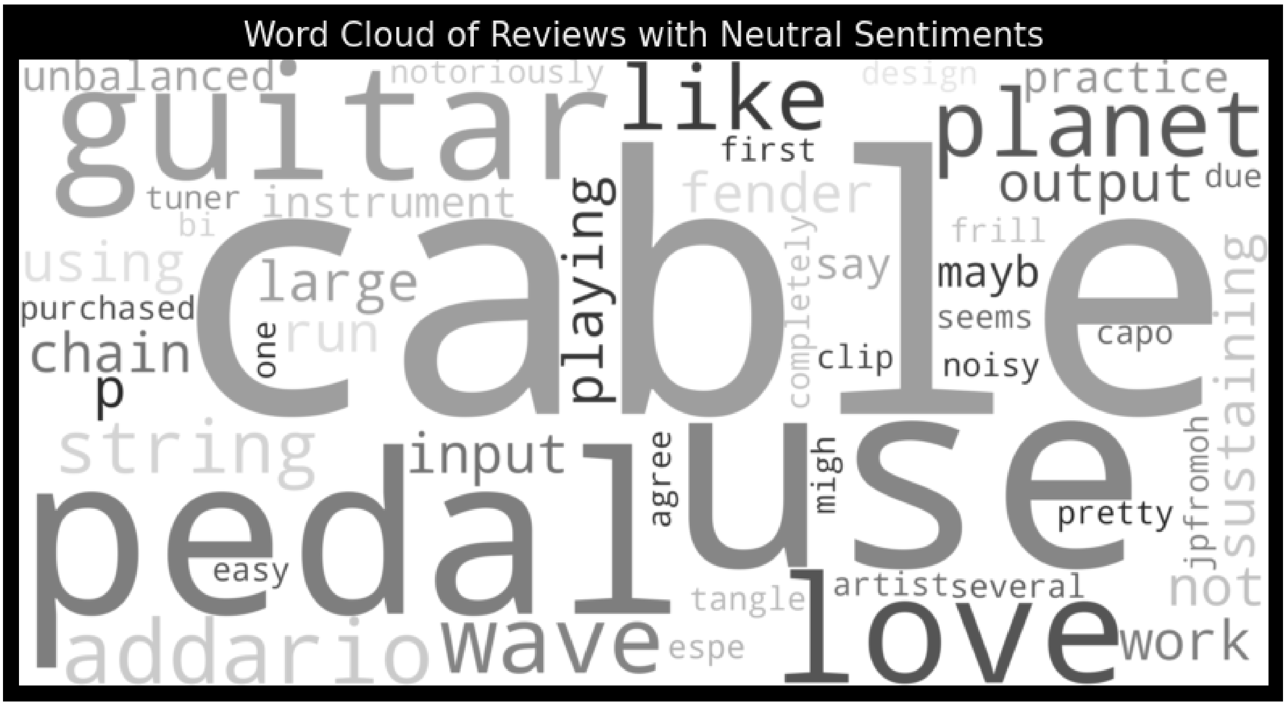

Bigram of Reviews with Negative Sentiments
sound like
planet wave
power supply
electric guitar
look like
bos gt
big muff
stopped working
guitar sound
save money
brand new
guitar strap
guitar stand
guitar amp
le paul
waste money
mic stand
work fine
distortion pedal
strap button
Top Words
0
5
10
15
20
Total Counts

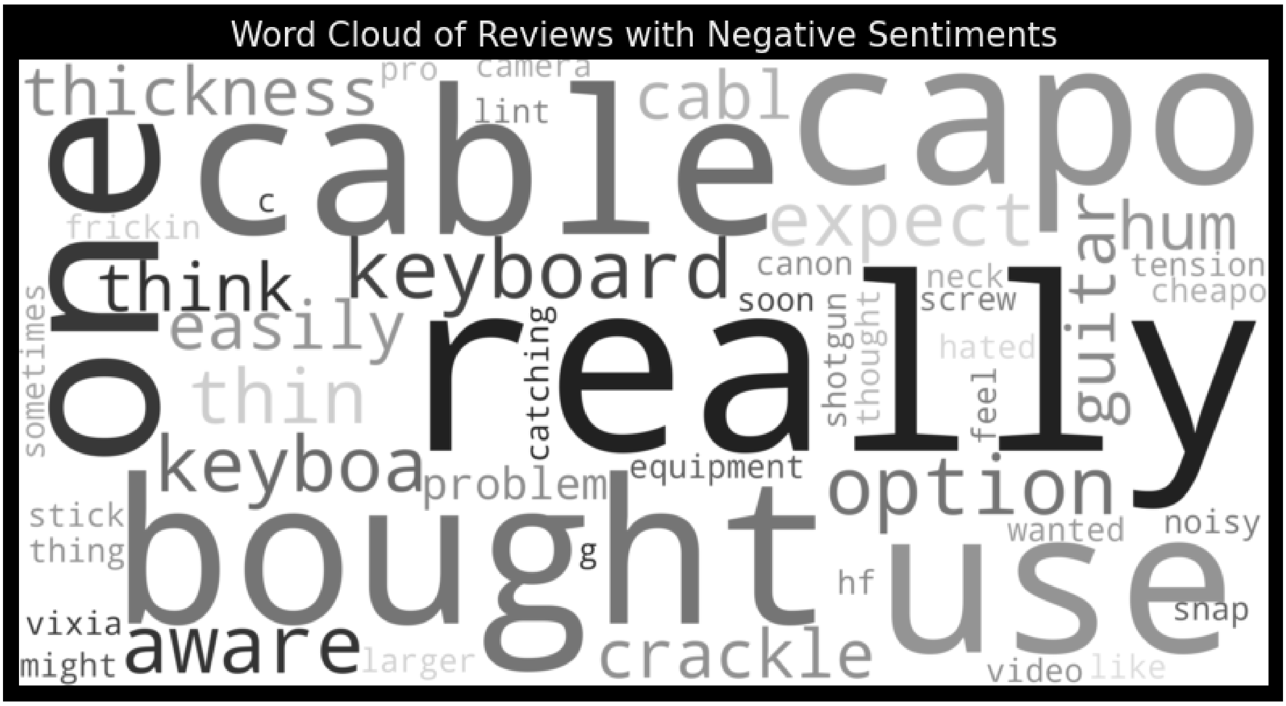

Trigram of Reviews with Positive Sentiments
acoustic electric guitar
built like tank
ernie ball string
solid state amp
great little tuner
acoustic guitar string
gibson le paul
local music store
playing guitar year
great product great
epiphone le paul
string sound great
pedal work great
electric guitar string
work like charm
great price great
good quality cable
work great great
highly recommended great
phosphor bronze string
Top Words
0
10
20
30
40
50
Total Counts


Trigram of Reviews with Neutral Sentiments
planet wave strap
planet wave capo
acoustic electric guitar
le paul guitar
tone pure nickel
high gain metal
good quality price
ball super slinky
built like tank
distortion pedal really
ernie ball string
gibson le paul
worth extra cost
guitar sound like
owned pedal year
planet wave cable
strap work better
string sound good
string sound great
higher quality cable
Top Words
0
1
2
3
4
5
6
7
8
Total Counts


Trigram of Reviews with Negative Sentiments
bos gt patch
patch james limborg
gt patch james
guitar sound like
planet wave capo
pitch correction function
line level signal
jim dunlop capo
install bos gt
halen joe satriani
cool cat metal
eddie van halen
guitarist studio album
van halen joe
studio album guitar
like eddie van
little big muff
satriani guitarist studio
sound like eddie
joe satriani guitarist
Top Words
0
1
2
3
4
5
6
7
8
Total Counts

Word Cloud Visualization: *Positive reviews* emphasize words like *"string," "great,"* and *"affordable,"* reflecting satisfaction. *Neutral reviews* focus on product-related terms like *"cable" and "pedal,"* with minimal emotional tone. *Negative reviews* feature words such as *"really," "problem," and "expect,"* highlighting dissatisfaction. These visuals complement the N-gram analysis by illustrating frequently discussed aspects across sentiment categories.





4. Methodology

4.1 Feature Extraction Techniques

Two main feature extraction approaches were used in this study:

1. **TF-IDF Vectorization**: For traditional machine learning models, text data was transformed into numerical features using TF-IDF (Term Frequency-Inverse Document Frequency) vectorization. The implementation focused on bigrams rather than unigrams to capture contextual relationships between words. The vectorizer was configured to extract the 5,000 most important bigrams from the review corpus.

2. **Word Embeddings**: For the Bi-LSTM model, reviews were tokenized and converted to sequences of word indices, which were then mapped to dense embedding vectors to capture semantic relationships between words.

4.2 Addressing Class Imbalance

To address the class imbalance problem identified during exploratory analysis, SMOTE (Synthetic Minority Oversampling Technique) was applied. This technique generated synthetic samples for the underrepresented classes (neutral and negative) to create a balanced dataset for model training.

4.3 Traditional Machine Learning Models

Several machine learning algorithms were implemented and compared:

1. Decision Tree Classifier
2. Logistic Regression
3. Support Vector Machine (SVM)
4. Random Forest Classifier
5. Bernoulli Naive Bayes
6. K-Nearest Neighbors Classifier

4.4 Deep Learning Model: Bi-LSTM

A Bidirectional Long Short-Term Memory (Bi-LSTM) network was implemented to capture sequential patterns in the review text. The Bi-LSTM architecture processes text in both forward and backward directions, allowing it to capture context from both past and future words.

4.5 Evaluation Metrics

The models were evaluated using the following metrics:

- Accuracy: Overall correctness of classification
- Precision: Ratio of true positives to all positive predictions
- Recall: Ratio of true positives to all actual positives
- F1-score: Harmonic mean of precision and recall
- Confusion Matrix: Visual representation of classification performance
- ROC Curve: Visualization of true positive rate vs. false positive rate

5. Implementation Details

5.1 Technical Requirements and Libraries

The implementation utilized the following Python libraries:

- pandas and numpy for data manipulation
- matplotlib for visualization
- NLTK for natural language processing
- scikit-learn for machine learning models and evaluation
- textblob for sentiment polarity extraction
- wordcloud for word cloud visualization
- imbalanced-learn for implementing SMOTE
- tensorflow and keras for Bi-LSTM implementation

5.2 Model Architecture and Parameters

The traditional machine learning models were implemented using scikit-learn with various hyperparameters:

Logistic Regression: After hyperparameter tuning via GridSearchCV, the optimal parameters were:

- C: 10000.0 (regularization parameter)
- penalty: l2 (L2 regularization)

Bi-LSTM Architecture:

- Embedding layer (max_features=10000, embedding_dim=100)
- First Bidirectional LSTM layer (64 units, return_sequences=True)
- Second Bidirectional LSTM layer (32 units)
- Dropout layer (0.2) to prevent overfitting
- Dense output layer with softmax activation for 3-class classification
- Adam optimizer with categorical crossentropy loss function

5.3 Training Process

The dataset was split into training (75%) and testing (25%) sets with stratification to maintain class distribution. For traditional models, 10-fold cross-validation was used to ensure robust performance evaluation. The Logistic Regression model was further optimized using GridSearchCV to find the optimal hyperparameters.

For the Bi-LSTM model, the training process involved:

- Tokenization and sequence padding to ensure uniform input dimensions
- Batch training with a size of 32
- Training for 10 epochs with 20% validation split
- Adam optimizer with learning rate decay

6. Results and Analysis

6.1 Cross-Validation Performance

The 10-fold cross-validation results for the traditional machine learning models were:

- Decision Tree: 82.18% accuracy
- Logistic Regression: 88.20% accuracy
- SVM: 88.05% accuracy
- Random Forest: 87.58% accuracy
- Naive Bayes: 80.99% accuracy
- K-Neighbors: 84.77% accuracy

Logistic Regression emerged as the top-performing traditional model with 88.20% accuracy.

6.2 Hyperparameter Tuning

Grid search for the Logistic Regression model yielded the following optimal parameters:

- C: 10000.0
- penalty: l2

With these optimized parameters, the training accuracy improved to 94.73%.

6.3 Final Model Performance

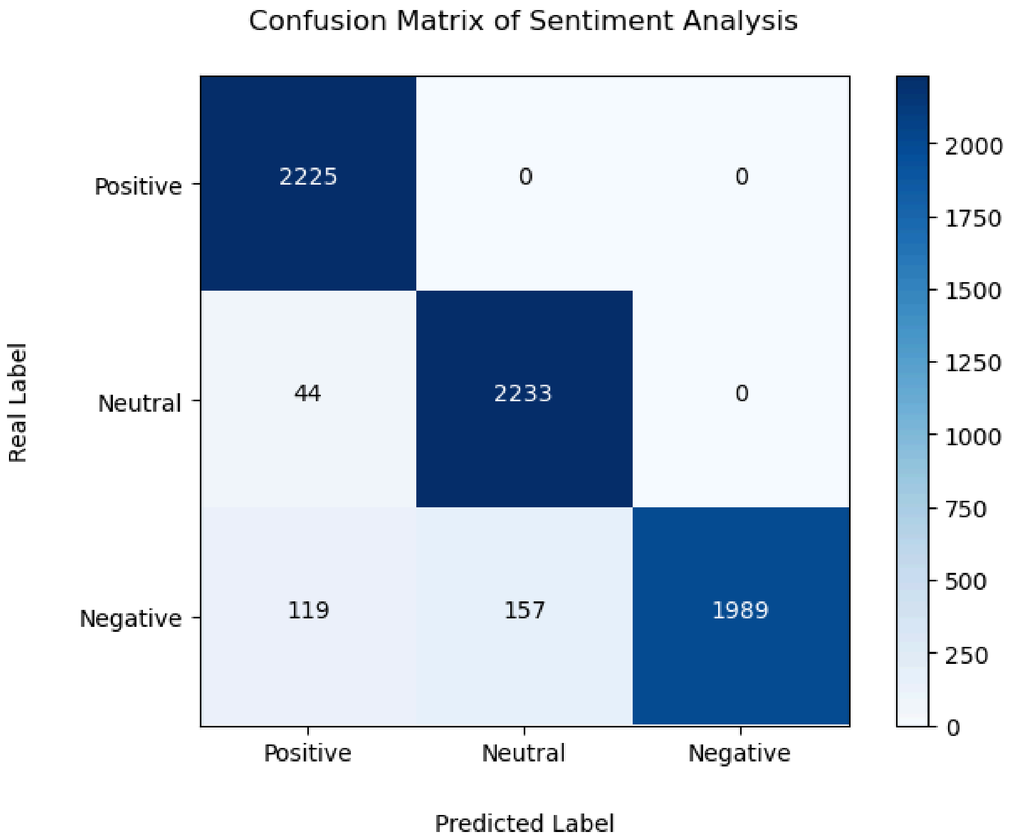
The final Logistic Regression model achieved an impressive 95.27% accuracy on the test set. The detailed performance metrics were:

Precision, Recall, and F1-Score:

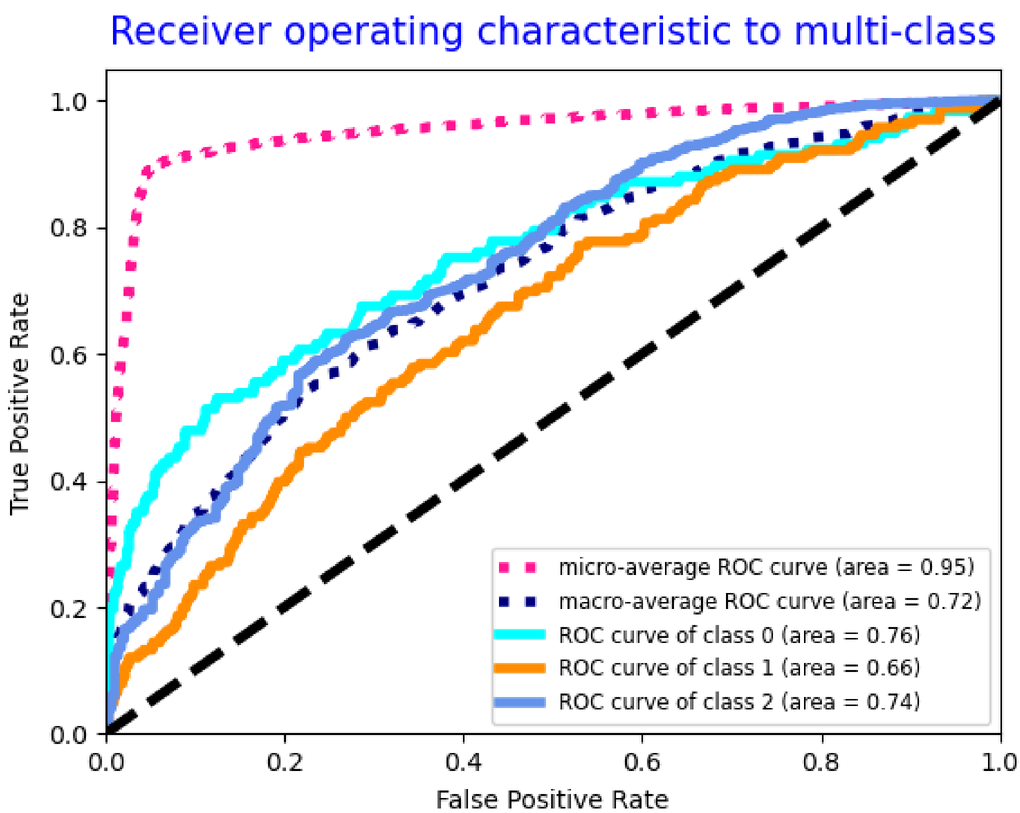
- Negative class (0): 0.93 precision, 1.00 recall, 0.96 F1-score
- Neutral class (1): 0.93 precision, 0.98 recall, 0.96 F1-score
- Positive class (2): 1.00 precision, 0.88 recall, 0.94 F1-score

Overall macro-average F1-score: 0.95

Confusion Matrix Analysis: The confusion matrix revealed that the model performed exceptionally well in classifying positive and neutral sentiments. It showed some challenges in correctly identifying negative sentiments, likely due to the initial class imbalance in the original dataset. However, the SMOTE technique helped mitigate this issue significantly.



ROC Curve Analysis: The ROC curves showed excellent classification performance for positive and negative classes with high area under the curve values. The analysis suggested that a threshold between 0.6-0.8 would provide optimal true positive rate (TPR) and false positive rate (FPR) balance.



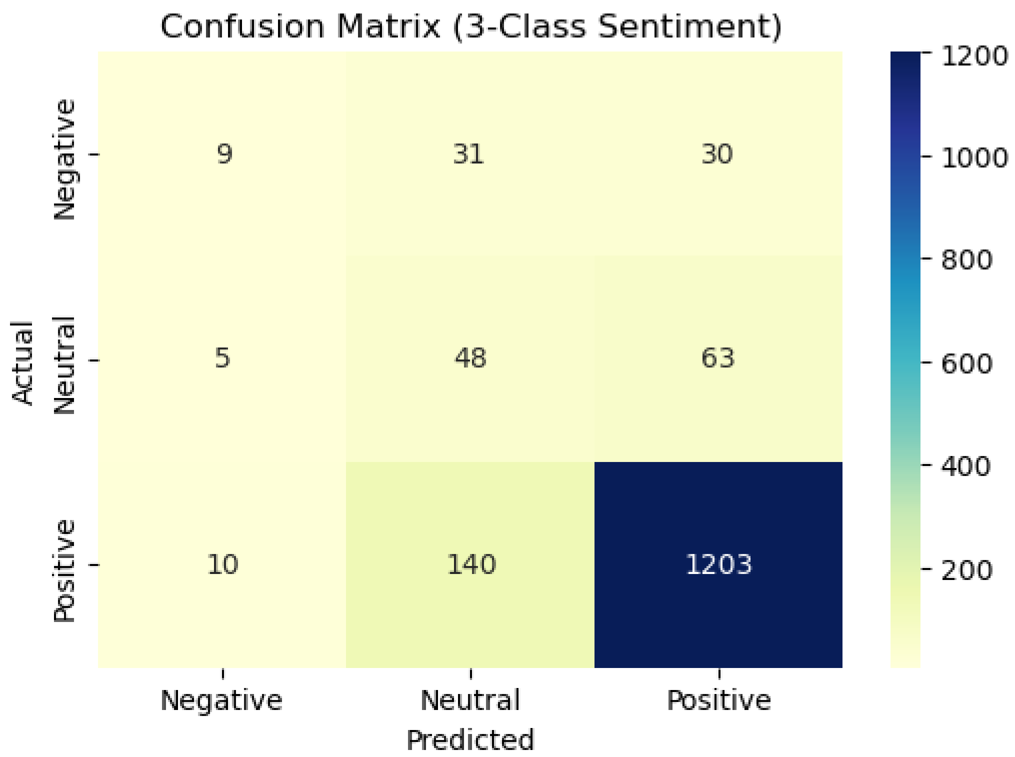
6.4 Bi-LSTM Performance

The Bi-LSTM model achieved 82% overall accuracy on the test set, with the following detailed metrics:

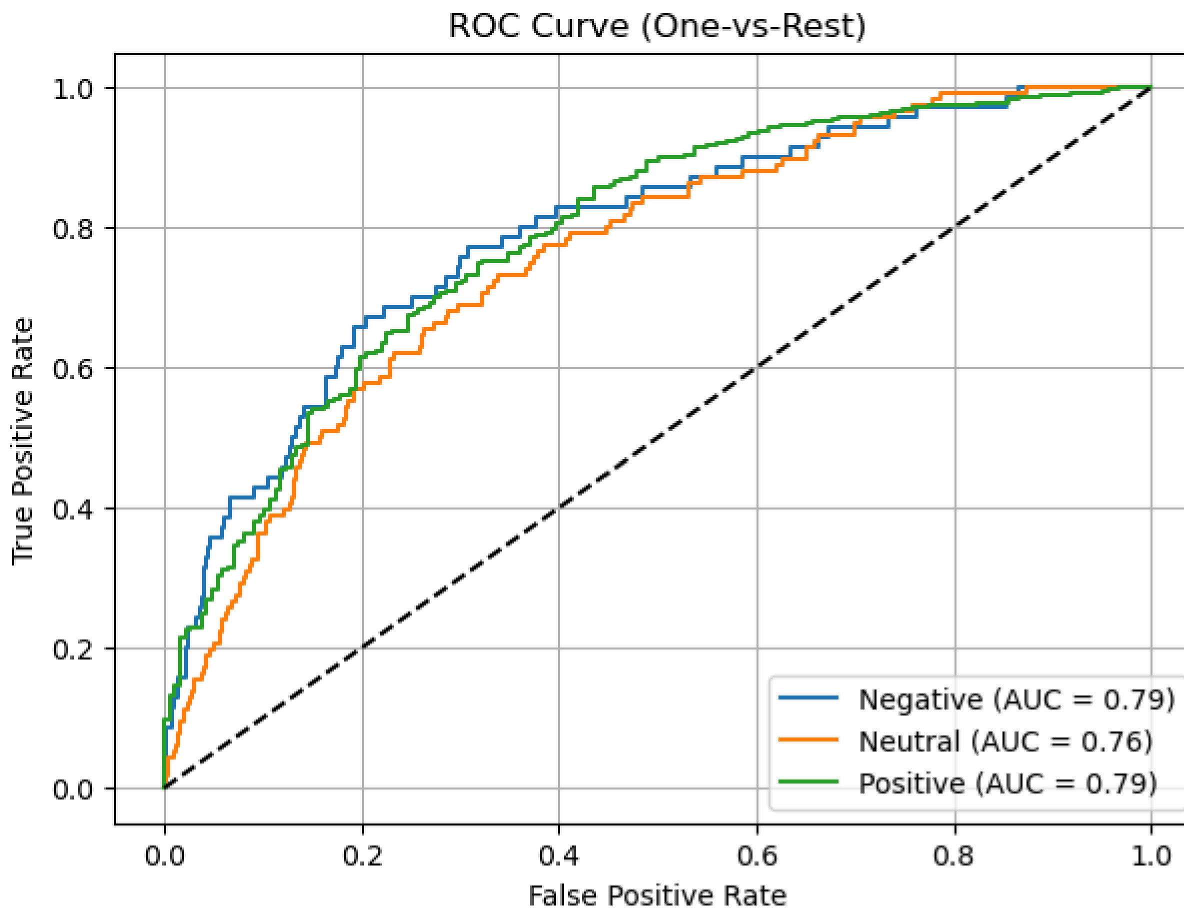
Classification Report for Bi-LSTM:

Sentiment	**Precision**	**Recall**	**F1-Score**	**Support**
Negative	0.38	0.13	0.19	70
Neutral	0.22	0.41	0.29	116
Positive	0.93	0.89	0.91	1353
Accuracy			**0.82**	1539
Macro Avg	0.51	0.48	0.46	1539
Weighted Avg	0.85	0.82	0.83	1539

Confusion Matrix Analysis: The bi-LSTM model excels at positive sentiment classification but struggles with negative sentiment. The model shows a systematic bias toward positive predictions, misclassifying many negative (30) and neutral (63) instances as positive. This imbalance suggests potential training data distribution issues or difficulties in capturing negative linguistic features.



ROC Curve Analysis: ROC curves demonstrate solid discriminative capability with AUC values of 0.79 for negative and positive classes, and 0.76 for neutral. The steep initial curve portions (0.0-0.2 FPR range) indicate the model achieves high true positive rates with minimal false positives at conservative thresholds. Despite classification challenges shown in the confusion matrix, the comparable AUC values suggest the model ranks predictions well, though threshold optimization may be needed, particularly for negative sentiment detection.



The Bi-LSTM model showed strong performance on positive sentiment classification (91% F1-score) but struggled significantly with negative (19% F1-score) and neutral (29% F1-score) sentiments. This performance discrepancy across classes reflects the challenges posed by the inherent class imbalance in the dataset, which appears to have impacted the deep learning model more severely than the traditional machine learning approach with SMOTE.

7. Discussion

7.1 Interpretation of Results

The high performance of the Logistic Regression model (95.27% accuracy) demonstrates that for this particular dataset, a well-tuned traditional machine learning approach with appropriate handling of class imbalance can achieve excellent results. This suggests that the lexical features captured by TF-IDF vectorization contain sufficient information for accurate sentiment classification of music instrument reviews.

The effectiveness of bigram features indicates the importance of capturing word pairs rather than individual words in sentiment analysis. This is particularly relevant for music instrument reviews where technical terminology and context-dependent descriptions are common.

7.2 Model Comparison

A significant performance gap was observed between the Logistic Regression model (95.27% accuracy) and the Bi-LSTM model (82% accuracy). This disparity can be attributed to several factors:

1. **Class Imbalance Handling**: The traditional machine learning pipeline incorporated SMOTE for balancing class distribution, while the deep learning approach may have been more sensitive to the original imbalance.

2. **Feature Representation**: TF-IDF bigrams may have captured sentiment-indicative patterns more effectively than sequential word embeddings for this specific dataset.

3. **Training Data Size**: Deep learning models typically require larger training datasets to achieve optimal performance. The limited number of negative and neutral examples may have hindered the Bi-LSTM's ability to learn distinguishing features for these classes.

4. **Model Complexity**: The Bi-LSTM model's complexity may have led to some overfitting on the dominant positive class, as evidenced by its high performance on positive reviews (91% F1-score) but poor performance on negative (19% F1-score) and neutral (29% F1-score) classes.

7.3 Feature Importance

Analysis of the Logistic Regression coefficients revealed the most influential bigrams for each sentiment class. Positive reviews frequently contained terms like "great sound," "easy use," and "highly recommend." Negative reviews were characterized by phrases like "waste money," "poor quality," and "customer service." These insights provide valuable information about which aspects of products and services most impact customer satisfaction.

7.4 Challenges and Limitations

Several challenges and limitations were encountered in this study:

1. **Class Imbalance**: The original dataset had a significant imbalance with positive reviews far outnumbering negative ones. While SMOTE helped address this issue for traditional models, its effect was less pronounced for the Bi-LSTM model.

2. **Ambiguous Expressions**: Some reviews contained mixed sentiments or used sarcasm, making classification challenging.

3. **Domain-Specific Terminology**: Musical instrument reviews often contain technical jargon that general-purpose NLP tools may not handle optimally.

4. **Review Length Variability**: The wide range of review lengths posed challenges for some models, particularly those sensitive to input dimensionality.

5. **Deep Learning Limitations**: Despite the theoretical advantages of Bi-LSTM in capturing sequential dependencies, its performance on minority classes was suboptimal, suggesting that more advanced techniques such as class weighting or specialized architectures might be necessary for highly imbalanced text datasets.

8. Conclusion and Future Work

8.1 Conclusion

This study successfully demonstrated the application of various machine learning and deep learning techniques for sentiment analysis on Amazon music instrument reviews. The optimized Logistic Regression model achieved excellent performance (95.27% accuracy) with high precision and recall across all sentiment classes. The study highlighted the importance of proper text preprocessing, feature extraction, and addressing class imbalance in sentiment analysis tasks.

The analysis of frequently occurring terms in different sentiment categories provided insights into the aspects of products and services that drive customer satisfaction or dissatisfaction in the musical instrument domain.

While the Bi-LSTM model showed promising results for the majority class (positive reviews), its performance on minority classes was limited, underscoring the challenges of applying deep learning to imbalanced datasets without specialized techniques.

8.2 Future Work

Several directions for future research can be explored:

1. **Improved Deep Learning Approaches**:
    - Implementing class weighting or specialized loss functions for the Bi-LSTM model to better handle class imbalance
    - Exploring attention mechanisms to focus on sentiment-indicative parts of reviews
    - Utilizing transfer learning with pre-trained language models like BERT or RoBERTa

2. **Ensemble Methods**: Combining predictions from traditional machine learning and deep learning models to leverage the strengths of both approaches.

3. **Aspect-Based Sentiment Analysis**: Extending the current approach to identify specific aspects of products (sound quality, durability, price, etc.) mentioned in reviews and the associated sentiments.

4. **Data Augmentation Techniques**: Exploring more sophisticated data augmentation methods beyond SMOTE, such as back-translation or contextual word replacement.

5. **Temporal Analysis**: Studying how sentiment patterns evolve over time to identify trends in customer satisfaction.

6. **Cross-Domain Transfer Learning**: Investigating how models trained on music instrument reviews perform when applied to reviews from other product categories.

9. References

1. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. NAACL-HLT.
   Available at: https://arxiv.org/abs/1810.04805 [Accessed 17 August 2021].
2. Gao, Z., Feng, A., Song, X., & Wu, X. (2019). Target-Dependent Sentiment Classification with BERT. IEEE.
   Available at: https://ieeexplore.ieee.org/document/8864964 [Accessed 17 August 2021].
3. Sun, C., Huang, L., & Qiu, X. (2019). Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence. arXiv.
   Available at: https://arxiv.org/pdf/1903.09588.pdf [Accessed 17 August 2021].
4. Ni, J., Li, J., & McAuley, J. (2019). Justifying Recommendations using Distantly-Labeled Reviews and Fine-Grained Aspects. EMNLP.
   Available at: https://cseweb.ucsd.edu/~jmcauley/pdfs/emnlp19a.pdf [Accessed 17 August 2021].
5. Ni, J. (2019). Amazon review data.
   Available at: https://nijianmo.github.io/amazon/index.html [Accessed 17 August 2021].
6. Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A., & Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. EMNLP.
   Available at: https://www.aclweb.org/anthology/D13-1170.pdf [Accessed 17 August 2021].
7. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1-2), pp.1-135.
   Available at: https://www.cs.cornell.edu/home/llee/omsa/omsa-published.pdf
8. Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp.79-86.
   Available at: https://www.aclweb.org/anthology/W02-1011.pdf [Accessed 17 August 2021].
9. Dave, K., Lawrence, S., & Pennock, D. M. (2003). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. Proceedings of the 12th international conference on World Wide Web (WWW), pp.519-528.
   Available at: https://dl.acm.org/doi/10.1145/775152.775226 [Accessed 17 August 2021].
10. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. Computational Linguistics, 37(2), pp.267-307.
   Available at: https://www.aclweb.org/anthology/J11-2001.pdf [Accessed 17 August 2021].
11. Uysal, A. K., & Gunal, S. (2014). The impact of preprocessing on text classification. Information Processing & Management, 50(1), pp.104-112.
   Available at: https://doi.org/10.1016/j.ipm.2013.08.006 [Accessed 17 August 2021].
12. Wang, X., Liu, Y., Sun, C., Wang, B., & Wang, X. (2016). Predicting polarities of tweets by composing word embeddings with long short-term memory. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, pp.1343-1353.
   Available at: https://aclanthology.org/P16-1127.pdf [Accessed 17 August 2021].
13. Zhou, P., Shi, W., Tian, J., Qi, Z., Li, B., Hao, H., & Xu, B. (2016). Attention-based bidirectional long short-term memory networks for relation classification. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, pp.207-212.
   Available at: https://aclanthology.org/P16-2034.pdf [Accessed 17 August 2021].
14. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, pp.321-357.
   Available at: https://www.jair.org/index.php/jair/article/view/10302 [Accessed 17 August 2021].
15. Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5(4), pp.1093-1113.
   Available at: https://www.sciencedirect.com/science/article/pii/S2090447914000550

16. Gupte, A., Joshi, S., Gadgul, P., Kadam, A., & Gupte, A. (2014). Comparative study of classification algorithms used in sentiment analysis. International Journal of Computer Science and Information Technologies, 5(5), pp.6261-6264.
Available at: https://www.ijcsit.com/docs/Volume%205/vol5issue05/ijcsit20140505110.pdf
17. Samal, B., Behera, A., & Panda, M. (2017). Performance analysis of supervised machine learning techniques for sentiment analysis. IEEE.
Available at: https://ieeexplore.ieee.org/document/8071579 [Accessed 16 August 2021].
18. Rathee, N., Joshi, N., & Kaur, J. (2018). Sentiment Analysis Using Machine Learning Techniques on Python. IEEE. Available at: https://ieeexplore.ieee.org/document/8663224
19. M S, N., & R, R. (2013). Sentiment analysis in twitter using machine learning techniques. IEEE.
Available at: https://ieeexplore.ieee.org/document/6726818 [Accessed 26 August 2021].
20. Gautam, G., & Yadav, D. (2014). Sentiment analysis of twitter data using machine learning approaches and semantic analysis. IEEE.
Available at: https://ieeexplore.ieee.org/document/6897213 [Accessed 26 August 2021].
21. Palanisamy, P., Yadav, V., & Elchuri, H. (2013). Serendio: Simple and Practical lexicon-based approach to Sentiment Analysis. Proceedings of the Second Joint Conference on Lexical and Computational Semantics (*SEM), pp.543-548.
Available at: https://aclanthology.org/S13-2091.pdf [Accessed 16 August 2021].
22. Ghosh, S., & Gunning, D. (2019). Natural Language Processing Fundamentals. Packt Publishing.
23. Liu, B. (2012). Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies,5(1),pp.1-167. Available at: https://www.morganclaypool.com/doi/abs/10.2200/S00416ED1V01Y201204HLT016
24. Shwartz, S. (2021). 12 Twitter Sentiment Analysis Algorithms Compared. AI Perspectives.
Available at: https://www.aiperspectives.com/twitter-sentiment-analysis [Accessed 1 July 2021].
25. Wolff, R. (2021). Quick Introduction to Sentiment Analysis. Medium. Available at: https://towardsdatascience.com/quick-introduction-to-sentiment-analysis-74bd3dfb536c
26. Devlin, J., Chang, M. W., Lee, K. and Toutanova, K., 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of NAACL-HLT.
Available at: https://aclanthology.org/N19-1423.pdf [Accessed 22 April 2025].
27. Pang, B. and Lee, L., 2008. Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval, 2(1-2), pp.1–135.
Available at: https://www.cs.cornell.edu/home/llee/omsa/omsa.pdf [Accessed 22 April 2025].
28. Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A. and Potts, C., 2013. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. In Proceedings of EMNLP.
Available at: https://aclanthology.org/D13-1170.pdf [Accessed 22 April 2025].
29. Taboada, M., Brooke, J., Tofiloski, M., Voll, K. and Stede, M., 2011. Lexicon-Based Methods for Sentiment Analysis. Computational Linguistics, 37(2), pp.267–307.
Available at: https://aclanthology.org/J11-2001.pdf [Accessed 22 April 2025].
30. Chawla, N. V., Bowyer, K. W., Hall, L. O. and Kegelmeyer, W. P., 2002. SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, 16, pp.321–357.
Available at: https://www.jair.org/index.php/jair/article/view/10302 [Accessed 22 April 2025].