SENTIMENT ANALYSIS FOR MARKETTING

Ramya Sree Manubramha
University of New Haven
New Haven, CT 06511
rmanu4@unh.newhaven.edu

Chethan Chakradhar Mangala
University of New Haven
New Haven, CT 06511
cmang3@unh.newhaven.edu

Sri Sai Durga Myneni
University of New Haven
New Haven, CT 06511
smyne3@unh.newhaven.edu

Abstract

This paper presents a sentiment analysis approach leveraging Long Short-Term Memory (LSTM) networks for classifying text sentiment as positive, negative, or neutral. The model is trained on textual data preprocessed using tokenization, lemmatization, and padding to address variations in text length. Word embeddings are utilized to capture semantic relationships between words, enhancing the model's ability to discern context. The LSTM architecture is selected due to its strength in handling sequential data and its capacity for capturing long-term dependencies. The study evaluates model performance using metrics such as accuracy and F1-score, highlighting the effectiveness of LSTMs in sentiment classification tasks. Results demonstrate the model's robustness, achieving competitive accuracy and demonstrating the potential for deployment in real-world applications like customer feedback analysis and social media monitoring. Future work includes exploring pre-trained transformer-based models for further performance improvement.

1. Introduction

The capacity to decipher and evaluate textual data has grown in importance for companies, scholars, and politicians in a data-driven world. The natural language processing (NLP) task of sentiment analysis, sometimes referred to as opinion mining, seeks to ascertain the sentiment or emotional tone conveyed in a particular text. It is essential for bettering decision-making procedures, raising client satisfaction, and comprehending public opinion.

In order to categorize feelings in textual data as positive, negative, or neutral, this project focuses on using machine learning techniques, particularly Long Short-Term Memory (LSTM) networks. Because it can capture long-term dependencies, LSTM, a form of recurrent neural network (RNN), is well-suited for sequential data analysis and is therefore perfect for text-based sentiment classification applications.

To manage variations in text structure and length, the procedure starts with data preprocessing, which includes padding, lemmatization, and tokenization. GloVe is a word embedding technique that preserves contextual meaning and semantic links by representing words in a dense vector space. Following data processing, the LSTM model is trained and assessed using performance metrics such as accuracy and F1-score.

This project's main goals are to develop a strong sentiment analysis model, show that it can be applied to actual datasets, and assess how well it performs. This project seeks to offer insights that can be used in a variety of fields, such as market research, social media monitoring, and customer feedback analysis, by examining sentiment in textual data.

2. Related Work

Sentiment analysis has been a prominent area of research within natural language processing (NLP), with various methods evolving over time to improve performance and generalizability. Early approaches relied on rule-based systems and lexicon-based methods, such as SentiWordNet, where predefined dictionaries of sentiment-bearing words were used to classify text. While these methods provided interpretability, their reliance on handcrafted rules limited scalability and accuracy in diverse contexts.

Machine learning-based techniques have since revolutionized sentiment analysis, employing algorithms such as Naïve Bayes, Support Vector Machines (SVM), and Logistic Regression. These models, often paired with bag-of-words or term frequency-inverse document frequency (TF-IDF) representations, improved sentiment classification accuracy by leveraging statistical patterns in data. However, their inability to capture contextual information and semantic relationships between words posed challenges for complex linguistic structures.

In recent years, deep learning methods have emerged as the state-of-the-art for sentiment analysis, particularly with the use of recurrent neural networks (RNNs) and their variants. Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs) have shown remarkable success in capturing long-term dependencies in text sequences. Models utilizing pre-trained word embeddings such as GloVe, Word2Vec, and FastText have further enhanced performance by embedding semantic similarity into vector representations.

Additionally, transformer-based architectures like BERT, RoBERTa, and GPT have set new benchmarks for sentiment analysis. By leveraging self-attention mechanisms, these models can process entire text sequences in parallel and understand nuanced language patterns. Although transformer models outperform RNN-based methods in many scenarios, their computational requirements can make them less accessible for resource-constrained applications.

This project builds on the advancements in deep learning for sentiment analysis by employing LSTM networks, which strike a balance between computational efficiency and the ability to handle sequential data. While transformer-based models remain a potential area for exploration, this study demonstrates that LSTMs, combined with effective preprocessing and word embedding techniques, provide a robust and accessible approach for sentiment classification tasks.

3. Dataset

The dataset used in this project was sourced from Kaggle, titled *Amazon Earphones Reviews* (source link). The dataset contains reviews of earphone products, with four columns:

1. **ReviewTitle**: Title of the review.

2. **ReviewBody**: Detailed text of the review.
3. **ReviewStar**: Rating provided by the user (on a scale of 1 to 5).
4. **Product**: Name of the product being reviewed.

For this study, only the **ReviewBody** and **ReviewStar** columns were utilized. The **ReviewBody** was used as the input text data for sentiment analysis, while the **ReviewStar** served as the corresponding label, with the following sentiment mapping:

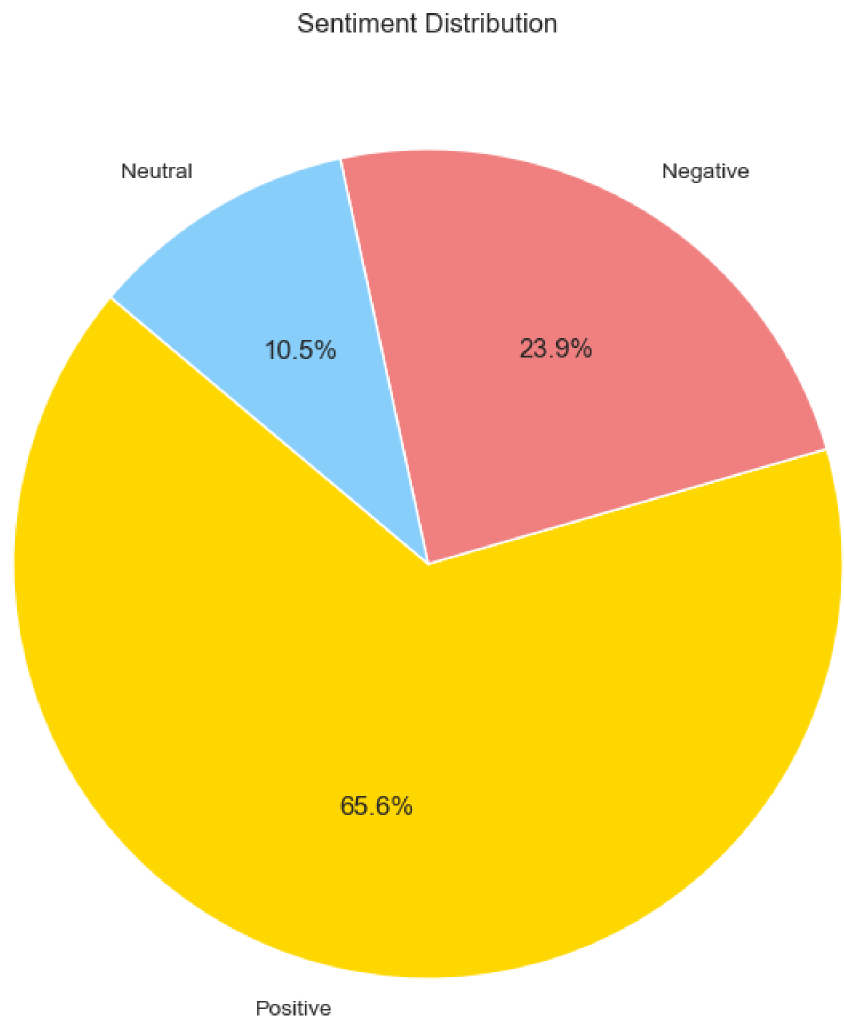
- **1-2 stars**: Negative sentiment.
- **3 stars**: Neutral sentiment.
- **4-5 stars**: Positive sentiment.

Sentiment Distribution

The dataset contained the following distribution of sentiments:

- **65.6% Positive Reviews**
- **23.9% Negative Reviews**
- **10.5% Neutral Reviews**

This distribution is illustrated in the pie chart below:



Data Cleaning Process

To prepare the data for model training, the following cleaning steps were performed:

1. **Handling Missing Data**: Reviews with missing text or star ratings were removed.
2. **Text Preprocessing**:
   - Lowercasing all text to maintain uniformity.
   - Removing special characters, numbers, and unnecessary punctuation.
   - Eliminating stopwords to reduce noise and focus on meaningful content.
3. **Balancing the Dataset**: Ensuring that all sentiment classes (negative, neutral, positive) were adequately represented to prevent model bias.
4. **Tokenization and Padding**: The cleaned text was tokenized into sequences, and padding was applied to maintain a consistent length for all input data.

These preprocessing steps ensured that the dataset was in a structured format suitable for training the LSTM model effectively.

4. Methodology

The sentiment analysis project was implemented using a Long Short-Term Memory (LSTM) network, a type of Recurrent Neural Network (RNN) that is well-suited for handling sequential data such as text. The methodology consisted of the following steps:

4.1. Model Architecture

The sentiment analysis task was implemented using a Long Short-Term Memory (LSTM) network. LSTM is a specialized type of Recurrent Neural Network (RNN) designed to capture long-term dependencies in sequential data, making it highly suitable for text analysis.

The architecture of the LSTM model is as follows:

1. Embedding Layer: Converts input tokens into dense vector representations of 100 dimensions using pre-trained embeddings.
2. LSTM Layers:
   - The first LSTM layer outputs a sequence of 64-dimensional vectors for each timestep.
   - A second LSTM layer extracts further temporal dependencies and outputs a 64-dimensional vector.
3. Dense Layers:
   - A fully connected layer with 10 units applies feature transformation.

- The final output layer, with 3 units, uses a softmax activation function to classify the input into three sentiment categories: positive, negative, or neutral.

Model Summary

The table below provides the details of the model architecture:

Model: "sequential"

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 100, 100)	892,700
lstm (LSTM)	(None, 100, 64)	42,240
lstm_1 (LSTM)	(None, 64)	33,024
dense (Dense)	(None, 10)	650
dense_1 (Dense)	(None, 3)	33

Total params: 968,647 (3.70 MB)
Trainable params: 75,947 (296.67 KB)
Non-trainable params: 892,700 (3.41 MB)

- Total Parameters: 968,647
  - Trainable Parameters: 75,947
  - Non-trainable Parameters: 892,700

The embedding layer uses pre-trained weights, contributing to the non-trainable parameters. The model design ensures efficient learning of contextual and sequential patterns in text data.

4.2. Model Training

The model was trained using the processed dataset with the following configurations:

- **Loss Function**: Categorical cross-entropy, suitable for multi-class classification tasks.
- **Optimizer**: Adam optimizer was used to update model weights efficiently.
- **Evaluation Metrics**: Accuracy and F1-score were used to evaluate model performance on both training and validation datasets.

4.3. Model Evaluation and Validation

The trained LSTM model was evaluated using a validation dataset. The following steps were performed:

- **Confusion Matrix**: To visualize the model's predictions across different sentiment classes.
- **Accuracy and F1-Score**: To measure the overall effectiveness and balance of the model's predictions.

4.4. Deployment Potential

The results demonstrated that the LSTM model effectively captured sentiment from textual reviews, making it suitable for real-world applications such as product feedback analysis and social media sentiment monitoring.

5. Results

The performance of the LSTM model was evaluated on the test dataset using precision, recall, F1-score, and accuracy as metrics. The results are summarized below:

Sentiment Class	Precision	Recall	F1-Score	Support
Negative (0)	0.60	0.56	0.58	686
Neutral (1)	0.00	0.00	0.00	301
Positive (2)	0.77	0.91	0.83	1,881

5.1 Overall Metrics

- **Accuracy**: 73%
- **Macro Average**:
  - Precision: 0.46
  - Recall: 0.49
  - F1-Score: 0.47
- **Weighted Average**:
  - Precision: 0.65
  - Recall: 0.73
  - F1-Score: 0.68

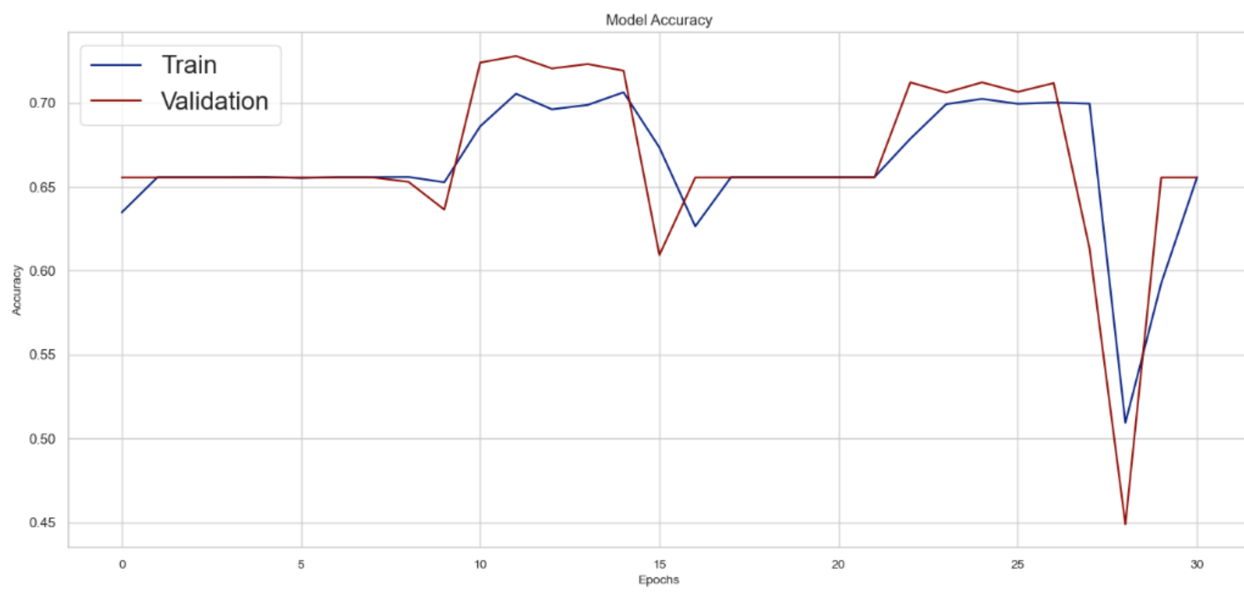
5.2 Observations

1. **Positive Sentiment**:
   - The model performed well on positive sentiment reviews, achieving an F1-score of 0.83.
   - This class had the largest support (1,881 examples), which likely contributed to better performance due to more training examples.
2. **Negative Sentiment**:
   - The model showed moderate performance for negative sentiment, with an F1-score of 0.58.
   - Precision and recall were balanced, but there is room for improvement.

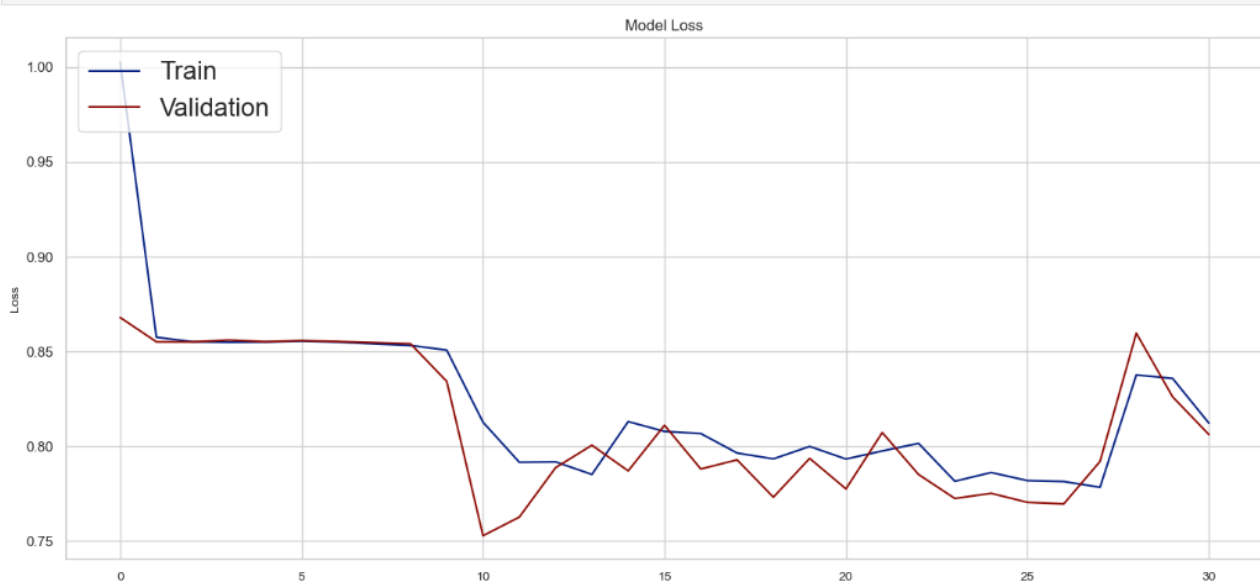
3. **Neutral Sentiment**:
    - The model struggled with the neutral class, achieving a precision, recall, and F1-score of 0.00.
    - The low support (301 examples) and possible class imbalance may have affected the performance for this category.

5.3 Training and Validation Graphs:

1. **Model Accuracy**:
    - The accuracy of the model improved steadily for both training and validation, peaking around epochs 10–20.
    - Notable dips and spikes were observed, particularly around epochs 25–30, indicating some instability during training.



2. **Model Loss**:
    - The loss decreased significantly in the initial epochs, stabilizing at a lower range for both training and validation after epoch 10.
    - A sudden increase in validation loss is observed near epoch 25, hinting at possible overfitting or instability



6. Conclusion

While the model demonstrates acceptable overall accuracy (73%) and strong performance on the positive sentiment class, its performance on the neutral class is poor. This highlights a potential issue with class imbalance or insufficient representation of the neutral category in the training data. Future improvements could involve:

- Balancing the dataset.
- Experimenting with different architectures or hyperparameters.
- Incorporating additional data for the underperforming class.

7. Limitations and Future Work

Limitations

1. **Class Imbalance**:
   The dataset was highly imbalanced, with the positive sentiment class having significantly more examples compared to the negative and neutral classes. This imbalance negatively impacted the model's ability to classify the neutral sentiment, as reflected in the poor F1-score for this class.
2. **Neutral Sentiment Performance**:
   The model failed to effectively identify neutral reviews, achieving an F1-score of 0.00. This indicates that the model struggled to learn meaningful features for the neutral class, likely due to insufficient training examples or overlapping features with other classes.

Future Work

1. **Addressing Class Imbalance**:
   Techniques such as oversampling the minority classes, undersampling the majority class, or using class-weighted loss functions can be implemented to improve the model's performance on underrepresented classes, especially neutral sentiments.
2. **Data Augmentation**:
   Synthetic data generation or augmentation techniques can be used to increase the diversity of examples for the minority classes, helping the model better understand their characteristics.
3. **Enhanced Preprocessing**:
   Advanced text preprocessing techniques, such as handling synonyms, semantic normalization, and domain-specific stopword removal, can be incorporated to better prepare the data.
4. **Incorporating Transformer Models**:
   Models like BERT, RoBERTa, or DistilBERT could be explored as

alternatives to LSTM, as they are state-of-the-art for text classification and sentiment analysis tasks.

5. **Cross-Domain Evaluation**:
   Future studies can explore the model's generalizability by applying it to reviews from different domains, such as electronics, books, or fashion. Fine-tuning the model on such datasets would help assess its robustness.
6. **Explainability and Interpretability**:
   Implementing explainable AI techniques, such as attention mechanisms or SHAP (SHapley Additive exPlanations), can help interpret the model's decisions and provide deeper insights into its predictions.

By addressing these limitations and implementing the proposed improvements, the model's performance can be enhanced, making it more robust and generalizable for real-world sentiment analysis tasks.

8.Acknowledgement

I would like to express my heartfelt gratitude to everyone who supported me during this project.

First and foremost, I would like to thank my professor and academic advisor for their invaluable guidance, constructive feedback, and encouragement throughout the project. Their expertise and insights have been instrumental in shaping the scope and execution of this work.

Special thanks to Kaggle and the contributors of the *Amazon Earphones Reviews* dataset, which formed the foundation of this study. Their work in curating and sharing the dataset was essential to this research.

Lastly, I want to extend my appreciation to my peers and friends for their constant support, brainstorming sessions, and motivation during challenging times. This project would not have been possible without their encouragement and collaboration.

References

- Baktha, K., & Tripathy, B. K. (2017, April 1). *Investigation of recurrent neural networks in the field of sentiment analysis*. IEEE Xplore. https://doi.org/10.1109/ICCSP.2017.8286763
- Katić, T., & Milićević, N. (2018, September 1). *Comparing Sentiment Analysis and Document Representation Methods of Amazon Reviews*. IEEE Xplore. https://doi.org/10.1109/SISY.2018.8524814
- Gope, J. C., Tabassum, T., Mabrur, M. Md., Yu, K., & Arifuzzaman, M. (2022, February 1). *Sentiment Analysis of Amazon Product Reviews Using*

Machine Learning and Deep Learning Models. IEEE Xplore. https://doi.org/10.1109/ICAEEE54957.2022.9836420

- Sentiment Analysis of *Amazon Smartphone Reviews Using Machine Learning & Deep Learning* | IEEE Conference Publication | IEEE Xplore. (n.d.). Ieeexplore.ieee.org. https://ieeexplore.ieee.org/document/9915917
- Hossain, A., Umme Hani Konok, Islam, R., Mohd, R., Ruhadina Musfikin, Uddin, M. M., Hossain, S., & Rashedul Amin Tuhin. (2023). *Utilizing GloVe Embeddings for Deep Learning-Based Analysis of Research Paper Abstracts*. 1–6. https://doi.org/10.1109/hora58378.2023.10156746
- Zheng Hengyu. (2020). *Improved SMOTE algorithm for imbalanced dataset*. https://doi.org/10.1109/cac51589.2020.9326603\