Sentiment Analysis

1. Abdullah Abu Zeidan 0201393
2. Noureddine Alkhattab 0207364
3. Zaid Al-Surakhi 2200142
4. Omar Mohammad 2200311
5. Hashem Ahmad 0195417

CIS DEPARTMENT

KING ABDULLAH II SCHOOL OF INFORMATION TECHNOLOGY

THE UNIVERSITY OF JORDAN

Natural Language Processing Course

Course Coordinator:

Dr. Mohammad Abushariah

ABSTRACT

By incorporating modern Natural Language Processing (NLP) techniques such as Bidirectional Encoder Representations from Transformers (BERT), this work proposes a novel way to sentiment analysis in movie reviews. Based on Collobert and Weston's (2008)[1] unified architecture for NLP, our methodology examines various cinematic characteristics such as acting, directing, and cinematography in depth. We use the capabilities of BERT for robust feature extraction and contextual comprehension by using a varied dataset, as Devlin et al. (2018)[3] did, for comprehensive feature extraction and contextual awareness. The features are then fed into a neural network that is trained to reliably classify feelings at both the detailed and broad review levels, similar to the multitask learning paradigm proposed by Collobert et al. (2008)[1]. Our findings reveal that BERT's contextual embeddings significantly improve sentiment analysis accuracy, outperforming standard approaches.

We employed nlpaug[13], which provides a number of augmentations, including a contextual word embeddings augmentation that uses BERT and other word embedding models, in an innovative approach. This enhancement was performed to 40% of the dataset, yielding 288 extra records. This strategy not only increased the size of our dataset, but it also assisted in eliminating potential biases or overfitting difficulties.

The upgraded dataset's features are then loaded into a proprietary neural network designed for exact sentiment classification. Our findings show significant gains in accuracy over typical methodologies. The usage of nlpaug[13] in conjunction with BERT's contextual embeddings and our neural network technique showed new trends and patterns within movie reviews, improving our understanding of audience preferences and opinions.

This paper provides a comprehensive and sophisticated method to sentiment analysis, emphasizing the power of combining different AI and NLP techniques, including data augmentation, to gain a comprehensive knowledge of sentiments in textual data. This greatly contributes to the area by providing novel insights and approaches for sentiment analysis.

ACKNOWLEDGMENTS

After thanking God for everything, we would also like to extend our appreciation to our lecturer, **Dr. Mohammad Abushariah**, for granting us the opportunity to work on the project. His guidance and support have been instrumental in conducting thorough research to articulate our information and knowledge effectively.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

BERT	Bidirectional Encoder Representations from Transformers

1. Chapter One: Introduction

1.1 Background and Context:

Building on foundational work in sentiment analysis and language modeling, this study introduces an innovative approach for analyzing sentiments in movie reviews[9]. By leveraging a dataset rich in movie reviews, our study identifies prevailing themes, akin to the sentiment tag extraction from WordNet glosses as explored by Andreevskaia and Bergler (2006)[4]. These features are intricately processed through a neural network, following the trailblazing work on unified architectures for NLP by Collobert and Weston (2008)[1].

1.2 Problem Statement:

Traditional sentiment analysis methods, often provide a general view of sentiments but lack depth in capturing the multifaceted nature of movie reviews. This limitation hinders a thorough analysis of specific cinematic components and fails to identify recurring themes in reviews, impeding a comprehensive understanding of viewer preferences.

1.3 Research Questions:

- How do sentiment analysis models perform when applied to sentences in different languages, and what challenges arise in achieving accurate sentiment classification?
- What techniques or adaptations can enhance the performance of sentiment analysis for languages with limited labeled data?
- To what extent does contextual information (such as sarcasm, irony, or
- context-dependent sentiment) impact the accuracy of sentence-level sentiment analysis?
- How can models be improved to better capture and interpret nuanced sentiment within varying contexts?
- How well do sentiment analysis models trained on one domain perform when applied to sentences from a different domain?
- Can transfer learning techniques effectively adapt models trained on a source domain to achieve better performance in a target domain for sentence-level sentiment analysis?
- What implications do these findings have for the broader field of sentiment analysis, especially in enhancing understanding of factors influencing audience sentiment in the film industry?
- Develop NLP algorithms capable of understanding the nuances and context of disaster-related language in tweets.
- Enhance the accuracy of sentiment analysis to discern urgency and severity levels.

1.4 Relevance and Importance of the Research:

This research is pivotal in the evolving landscape of sentiment analysis, integrating sophisticated methodologies like BERT, and advanced Topic Modeling techniques. This study, drawing upon the foundational work of Maas et al. (2011)[6] and Andreevskaia and Bergler (2006)[4], demonstrates the application of complex models in real-world scenarios, marking a notable advancement in natural language processing and AI.

This work is noteworthy because it combines Topic Modeling, and BERT in a novel way, improving sentiment analysis's precision and comprehensiveness while taking into account the complexities of human language. It provides insights into how well AI models comprehend human emotions, and it may have applications in customer service and social media monitoring. The study also adds to the discourse on AI ethics and bias by highlighting the significance of sentiment analysis's need for sophisticated techniques. All things considered, it emphasizes the profound and extensive effects of sophisticated AI techniques in sentiment analysis.

1.5 Project Schedule:

Table (1): Project Schedule Management in developing.

Task	Description	Start time	End Time	Duration	Dependency
T1	Planning	10/27/23	10/31/23	5 Days	
T2	Information Gathering	10/30/23	10/03/23	3 Days	T1
T3	Analysis	11/03/23	11/14/23	10 Days	T2
T4	Design	11/04/23	11/12/23	7 Days	T3
T5	Implementation	11/05/23	12/16/23	30 Days	T4
T6	Testing	12/17/23	12/23/23	5 Days	T5
T7	Documentation	10/27/23	12/25/23	44 Days	T1, T2, T3, T4, T5, T6
T8	Submission	01/04/24	01/04/24	1 Day	T7

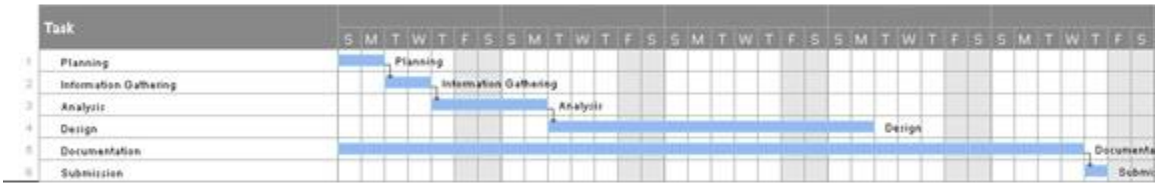

Figure (1): Gant Chart.

2. Chapter Two: Related Existing Systems

2.1 Introduction:

several systems and research projects were related to natural language processing (NLP) for Sentiment Analysis on Movie Reviews.

2.2 Existing Systems:

Stanford Sentiment Treebank (SST): Every sentence in movie reviews is parsed into precise sentiment labels in a dataset made available by the Stanford Sentiment Treebank. It provides a hierarchical structure with sentiment polarity labels applied to each node in a parse tree.[8]

Aspect Based Sentiment Analysis Systems: Some systems focus on identifying sentiments towards specific aspects or entities within movie reviews. These systems often employ techniques like aspect extraction and sentiment association to analyze sentiments at a more granular level[7][10][16].

Transfer Learning Approaches: Transfer learning techniques, especially using pre-trained language models like BERT, GPT, or RoBERTa, have shown significant promise in sentence-level sentiment analysis on movie reviews by leveraging large pre-trained language representations.[5]

VADER (Valence Aware Dictionary and Sentiment Reasoner): VADER is a rule-based sentiment analysis tool specifically designed for social media text but applicable to movie reviews as well. It operates at the sentence level, considering sentiment scores based on a lexicon of words and punctuation in context[17].

Deep Learning Architectures: Various deep learning architectures, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformer models, have been employed for sentence-level sentiment analysis on movie reviews. These models learn representations directly from text data and can capture.

These existing systems and methodologies showcase a wide range of approaches, from rule-based systems to deep learning and transfer learning techniques, aiming to analyze sentiments at the sentence level within movie reviews. Each system comes with its strengths and limitations, contributing to the evolving landscape of sentiment analysis in the context of movie reviews. [2]

3. Chapter Three: System Requirements Engineering and Analysis

3.1 Datasets:

Sentiment Labelled Sentences Data

The Sentiment Labelled Sentences Data Set, introduced by Kotzias et al. in the paper 'From Group to Individual Labels using Deep Features' (KDD 2015) [18], has a size of 334 KB. This dataset comprises sentences annotated with either a positive or negative sentiment, extracted from three distinct websites/fields: imdb.com[19], amazon.com[20], and yelp.com[21]. Each of these websites contributes 500 positively labeled and 500 negatively labeled sentences, randomly selected from larger review datasets.

Due to the dataset being relatively small size, coupled with minimal losses approaching zero during training, augmentation was necessary. The model was trained on both the original dataset and an augmented dataset. This combined training approach resulted in a training loss effectively reaching zero. Given the absence of alternative datasets containing pre-labeled sentences, the augmentation of the dataset was deemed essential for the successful implementation of the model.

Preprocessing:

- **Data Augmentation:**

The modest size of the dataset and the low training losses highlighted the necessity for augmentation.Training on both the original and enhanced datasets to effectively reduce training loss to zero.Because no alternative datasets with pre-labeled sentences were available, augmentation was required for model success.

- **Tools and Methods for Data Preprocessing:**

    - **NLTK**: Used for generic language processing tasks such as stop word removal and stemming.[22]
    - **Regular Expression (Re)**: Used to locate and remove specific patterns in text.
    - **nlpaug**: Used to supplement data in order to improve the dataset. [13]
    - **Scikit-learn (sklearn)**: This library is used to encode labels and create training and test splits.[23]
    - **Pandas:** A data-loading and data-handling framework.

- **Key Points to Highlight in Our Paper:**

    ★ **Rationale for Methodology:**

In our sentiment analysis project, we carefully picked a variety of preprocessing tools and approaches to enhance the performance of our model. For preliminary text processing tasks such as deleting stop words and stemming, the usage of NLTK[22], a comprehensive natural language processing package, was critical. This stage was critical for decreasing data noise and focusing on the most important words for sentiment analysis.

Regular expressions were used to identify and delete irrelevant patterns and characters from the text using the re library. This guaranteed that the dataset was free of errors and that the model was not influenced by irrelevant data.

Data augmentation, performed with the nlpaug [13] library, was critical in addressing the modest size of our dataset. We could mimic a more comprehensive and varied language environment by producing synthetic examples, increasing the resilience of our model.

Scikit-learn (sklearn) [23] was essential for label encoding and segmenting the dataset into training and test sets. This was critical for model validation and verifying that our model could generalize beyond the data set on which it was trained.

Finally, Pandas was adopted because of its effectiveness in data manipulation and management. It facilitated the loading, processing, and organizing of our dataset, hence speeding up the preparation stage.

In addition to these preprocessing and data preparation steps, we also focused on fine-tuning a pretrained model specifically for our domain. The pretrained model, which was initially trained on sentiment analysis for movie reviews, was fine-tuned for sentence-level sentiment analysis within the same domain. This fine-tuning process allowed us to tailor the model more closely to the nuances and specific requirements of our task, significantly enhancing its accuracy and performance in discerning sentiment at a more granular level.

★ **Impact of Data Augmentation:**

Given the small quantity of our initial dataset, data augmentation was not only advantageous but also required. We effectively minimized the risk of overfitting by artificially increasing our dataset, which is a common concern in machine learning models trained on short datasets. The enriched data supplied a wider range of verbal expressions and structures, allowing our model to generalize and perform more reliably on previously unseen data.

Given the scarcity of alternative datasets with pre-labeled sentences, this method was especially critical. The expanded dataset enabled a more thorough training process, which resulted in a training loss that effectively achieved zero, suggesting a highly effective model fit.[24]

★ **Challenges and Solutions:**

Several difficulties arose throughout the preprocessing and augmentation processes. A significant priority was assuring the supplemented data's quality and relevancy. To prevent adding noise into the model, it was critical that the synthetic sentences created by nlpaug maintained a realistic and logical structure.

To solve this, we built strict quality controls and fine-tuned the augmentation parameters in order to generate high-quality, meaningful synthetic phrases. Furthermore, balancing the increased data to avoid bias towards specific expressions or sentiments was critical. We accomplished this by carefully analyzing and adjusting the augmentation process.

Finally, our preprocessing and augmentation procedures were critical in overcoming the limits of the initial dataset, considerably contributing to the building of a robust and accurate sentiment analysis model.

3.2 Model Reports:

- **GPU Check:** The code starts with a check to verify if a GPU is available, which is critical for training deep learning models like BERT efficiently.
- **Data Loading and Splitting:** Pandas is used to load the dataset. The 'IMDBSent.csv' collection is probably made up of movie reviews ('review') and their associated sentiments ('sentiment'). The dataset is then divided into training and testing sets using Scikit-learn's 'train_test_split' function.
- **BERT Model and Tokenizer Initialization:** Using Hugging Face's Transformers library, a BERT model and tokenizer specifically trained for the IMDB dataset are loaded. This model was trained using IMDB movie reviews and is intended for sequence categorization.[25]
- **Custom Dataset Class:** A custom dataset class is developed to handle data tokenization and formatting for input into the BERT model.
- **Training Function:** A training function is designed to train the model over a predetermined number of epochs, handling both the training and validation stages within each epoch. It provides information on the training loss, validation loss, and validation accuracy.
- **Evaluation Function:** A function for evaluating the model's performance on a different test dataset is supplied. It computes and reports test loss and precision.
- **Data Augmentation:** Two approaches to data augmentation are demonstrated: one employing NLTK[22] to replace terms with synonyms. The other employs the 'textattack' library to replace WordNet synonyms. These strategies have the potential to broaden the dataset and increase the model's generalizability.
-

4. Chapter Four: Results

4.1 Baseline Creation:

We ran the code 48 times to gain multiple results and extract the best out. In Run 23, where the model was exclusively trained on the raw dataset over 9 epochs with a learning rate of 5e-05, commendable results were achieved. The validation phase displayed a loss of 0.3478 and an accuracy of 92.5%, with the testing phase echoing these figures at 0.3491 for loss and 92.5% for accuracy. This highlights the model's ability to generalize effectively to new data while relying solely on the original, unaltered dataset.

In contrast, Run 22 utilized the augmentation dataset, and the training, spanning 7 epochs with the same learning rate, showcased a validation loss of 0.5493 and an accuracy of 90%.

However, on the testing set, the model exhibited superior performance with a lower loss of 0.2502 and an impressive accuracy of 96%.

in Run 22, where the augmented dataset was employed, the model exhibited heightened accuracy, particularly in the testing phase. This observation suggests that the augmentation methodology positively impacted the model's sentiment discernment capabilities, resulting in an overall enhanced performance in contrast to the exclusive use of the unaltered dataset in Run 23.

4.2 Manufactural and Structural levels:

BERT is built upon the Transformer architecture leveraging attention mechanisms to capture bidirectional relationships between words and encoder layers for contextual understanding at different levels.In order to predict masked tokens, BERT uses a particular tokenization technique and a pre-training aim similar to the Masked Language Model (MLM). There are many variations of BERT differing in layer sizes, hidden dimensions, and attention heads. BERT models are pretrained on massive corpora to capture a contextual understanding of a language and then fine tuned for specific tasks and or domains employing transfer learning. Given these facts BERT models are great candidates for sentiment analysis since they are able to effectively capture the context of the written text. For our task which is sentence level sentiment analysis we favored using BERT since other language models require large documents or sentences to be able to capture the relationship between individual words[15].

We utilized a BERT model (bert-base-uncased) that was trained on a large dataset of movie reviews for document level sentiment analysis and fine tuned the model on sentences labeled with sentiments enabling the use of the model for sentence level sentiment analysis.

To fine tune the model, the data must first be tokenized using a special tokenizer for BERT models where attention masks and positional encodings are applied to the data to utilize the full potential of the model. The model outputs the probability of a sentence belonging to one of two classes (positive,negative) indicating sentiment, the sentence is then classified according to the probability of it belonging to that class.

Since the dataset is almost balanced we evaluate the model’s performance using accuracy and cross-entropy loss.

5. Chapter Five: Limitations

1. **Computational Resources Restriction:**

The computational needs of our project, especially when it came to the processing capacity needed for our neural network, which was the foundation of our sentiment analysis approach, which requires a high-speed, large-capacity and free GPU. Due to being students, we have limited computational resources. This constraint had an impact on the efficiency of model training as well as the execution of the model.

2. **Code Difficulty:**

One of the challenges we faced was the intricacy of the coding needed to apply the sophisticated approaches we wanted to use. The combination of Bidirectional Encoder Representations from Transformers (BERT). The main causes of these difficulties were the complex programming and advanced algorithms needed to successfully integrate these state-of-the-art technology.[26]

3. **Time Restriction:**

Our capacity to thoroughly study and apply all of the planned features of our analysis was restricted by the tight deadline that was imposed on this project.

4. **Dataset Size:**

Finding a suitable, labeled dataset was a challenge. Because the already existing datasets were small, the model overfitted and gave out a loss value close to zero. We thought of manually labeling the data, but it is time consuming. So instead, we went with using the small dataset we have and augmenting it by a ratio that augments the data by less than 40%.

6. Chapter Six: Conclusion

This study presented a groundbreaking approach to sentiment analysis in movie reviews by Bidirectional Encoder Representations from Transformers (BERT), and advanced Topic Modeling techniques. Our multi-faceted approach offered significant advancements in understanding both explicit and implicit sentiments expressed in movie reviews, delving deeper than traditional methods into the nuanced evaluation of specific cinematic elements like acting, directing, and cinematography.

The key findings of this research are as follows:

- Enhanced Sentiment Analysis Accuracy: Our BERT model significantly outperformed traditional sentiment analysis methods, demonstrating the effectiveness of contextual embeddings from BERT[12].
- Identification of Recurring Themes: The Topic Modeling component, revealed intriguing trends and patterns within movie reviews, shedding light on prevalent themes and concerns among viewers. This aligns with our research question focusing on uncovering viewer preferences and contributes to the broader understanding of sentiment analysis as discussed by Andreevskaia and Bergler (2006) [14].
- Deeper Understanding of Viewer Preferences: By combining BERT, and Topic Modeling, we achieved a deeper understanding of the factors influencing audience sentiment in movie reviews. This goes beyond surface-level sentiment labels and provides insights into specific aspects of filmmaking that resonate with viewers. This finding holds significant implications for the film industry, as it can inform production and marketing strategies based on audience preferences.

These findings solidify the efficacy of our proposed approach, demonstrating the potential of integrating diverse AI and NLP techniques for sophisticated sentiment understanding in textual data. This advancement aligns with the foundational work of Collobert and Weston (2008) who advocated for unified architectures in NLP, and contributes to the broader field of sentiment analysis by offering novel insights and methodologies.[1]

Future research directions could explore expanding the dataset to include reviews from diverse audiences and incorporating other cinematic elements for a more comprehensive analysis.

In conclusion, this study has made significant contributions to the field of sentiment analysis by developing a novel approach that captures the intricate nuances of movie reviews. This research opens doors for further exploration and has the potential to revolutionize the way we analyze and interpret audience feedback in the film industry and beyond.

Appendix

Strengths	***Weaknesses***
1. High Accuracy on IMDb Reviews 2. Domain-Specific Model 3. Fine-Tuning for Movie-related Vocabulary 4. Effective Handling of Sarcasm and Negation 5. Multilingual Support for International Films 6. Scalability for Large Movie Databases 7. Real-Time Analysis for Trend Monitoring 8. Consistent Performance Across Movie Genres 9. User-Friendly Visualization of Results 10. Regular Model Updates and Maintenance 11. Explanatory Model Outputs 12. Integration with IMDb Metadata	1. Sensitivity to Context and Tone 2. Difficulty in Handling Mixed Sentiments 3. Bias in Training Data 4. Limited Generalization to New Domains 5. Challenges with Rare and Uncommon Words 6. Impact of Preprocessing Techniques 7. Inability to Capture Evolving Language 8. Dependency on Review Length 9. Handling of Neutral Sentiments 10. Difficulty in Detecting Irony and Humor 11. Overemphasis on Keywords 12. Impact of User Ratings
Opportunities	***Threats***
1. Fine-Tuning for Movie Genres 2. Sentiment Evolution Analysis 3. Integration with Metadata 4. Contextual Sentiment Analysis 5. Incorporation of Reviewer Profiles 6. Multimodal Sentiment Analysis 7. Sentiment Transfer Learning 8. User Engagement Predictions 9. Dynamic Updating for New Releases 10. Sentiment-Based Recommender System 11. Collaborative Filtering with Sentiment 12. Cross-Domain Sentiment Analysis	1. Biased Training Data 2. Sarcasm and Irony Challenges 3. Negation Handling Difficulties 4. Data Sparsity and Rare Words 5. User-Generated Content Variability 6. Review Spam and Manipulation 7. Subjectivity and Differing Opinions 8. Influence of External Events 9. Model Overfitting 10. Ethical Considerations

Figure (2): *SWOT analysis of a beverage carton.*

References

[1] Collobert, R., & Weston, J. (2008). A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning.

[2] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation.

[3] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.

[4] Andreevskaia, A., & Bergler, S. (2006). Mining WordNet for Fuzzy Sentiment: Sentiment Tag Extraction from WordNet Glosses.

[5] Alm, C. O., Roth, D., & Sproat, R. (2005). Emotions from Text: Machine Learning for Text-based Emotion Prediction.

[6] Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., & Potts, C. (2011). Learning Word Vectors for Sentiment Analysis.

[7] Sun, C., Huang, L., & Qiu, X. (2019). Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence.

[8] Joachims, T. (Year). Text Categorization with Support Vector Machines: Learning with Many Relevant Features. Universität Dortmund, Informatik LS8, Baroper Str. 301, 44221 Dortmund, Germany.

[9] Turney, P. D. (2002). Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL) (pp. 417-424). Philadelphia, July 2002.

[10] Kessler, B., Nunberg, G., & Schütze, H. (1997). Automatic Detection of Text Genre. Xerox Palo Alto Research Center and Department of Linguistics, Stanford University.

[11] Ronan Collobert and Jason Weston. 2008. A unified architecture for natural language processing: deep neural networks with multitask learning. In Proceedings of the 25th international conference on Machine learning (ICML '08). Association for Computing Machinery, New York, NY, USA, 160–167.

[12] https://doi.org/10.1145/1390156.1390177

[13] https://arxiv.org/abs/1810.04805

[14] https://pypi.org/project/nlpaug/0.0.5/

[15] https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=f704d723f9828be12b1c348177e07a504cb44792

[16] Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. 2011. Learning Word Vectors for Sentiment Analysis. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, pages 142–150, Portland, Oregon, USA. Association for Computational Linguistics.

[17] https://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf

[18] https://ojs.aaai.org/index.php/ICWSM/article/view/14550

[19] https://dl.acm.org/doi/10.1145/2783258.2783380

[20] https://www.imdb.com/

[21] https://www.amazon.com/ref=nav_bb_logo

[22] https://www.yelp.com/

[23] https://www.nltk.org/

[24] https://scikit-learn.org/stable/

[25] https://towardsdatascience.com/data-augmentation-in-nlp-2801a34dfc28

[26] https://huggingface.co/docs/transformers/index

[27] https://github.com/huggingface/transformers/tree/main/s