Uncovering Thematic Sentiments in Literary Quotes from Goodreads

A. H. M. Rezaul Karim
Dept. of Computer Science
George Mason University
Fairfax, VA, USA
akarim9@gmu.edu

***Abstract*—In the evolving field of Natural Language Processing (NLP), the task of sentiment analysis is crucial for interpreting the intricate layers of human emotions and ideas in text. This study explores the nuanced field of thematic sentiment analysis, a key component in NLP for understanding complex human expressions. We utilize the "Goodreads Quotes" dataset, an extensive collection of 82,460 quotes each categorized under one of 27 thematic labels, to examine its effectiveness in NLP-focused literary analysis. The approach involved fine-tuning a pre-trained BERT model on this dataset, resulting in an accuracy of approximately 45%, thereby illustrating its capability to grasp detailed thematic sentiments. Additionally, we implemented a Support Vector Machine (SVM) model, which achieved around 37% accuracy. These findings emphasize the potential of applying advanced NLP methods that utilize high computational resources with respect to low computational complexity Machine Learning (ML) models for literary sentiment analysis. The study emphasizes the challenges of interpreting literary sentiment and the value of thematic analysis, offering new avenues for advancing NLP technologies.**

***Index Terms*—NLP, BERT model, Fine-tuning, SVM, Statistical Analysis.**

I. Introduction

The field of Natural Language Processing (NLP) has witnessed remarkable development in recent years, revolutionizing how machines understand and interact with human language. NLP, a critical intersection of computer science, artificial intelligence, and linguistics, facilitates the processing and analysis of vast quantities of natural language data. It has many applications, including healthcare, where it aids in patient data analysis [1], finance for sentiment analysis of market trends [2], and customer service for enhancing interaction through chatbots [3], demonstrating NLP's versatility in several fields.
Sentiment analysis, a subset of NLP, is focused on extracting opinions and emotions from text. This technique has evolved, ranging from basic lexicon-based approaches to complex machine learning algorithms [4]. Thematic sentiment analysis, a more advanced form, goes beyond general sentiment extraction to identify underlying themes and sentiments in a text leveraging NLP models to ascertain the emotions and nuances behind each quote or piece of text, offering a deeper understanding of the context.
While many Machine Learning (ML) models are computationally less demanding, the advent of advanced NLP models using Transformers [6] and other sophisticated techniques presents a new paradigm in language processing. These advanced models, though resource-intensive, offer significant improvements in accuracy and context understanding. The trade-off between computational expense and performance gain is critical. It's essential to evaluate whether the benefits of using these advanced NLP models outweigh the simplicity and lower resource requirements of traditional ML approaches, particularly in tasks such as thematic sentiment analysis where context and nuance are important.
In this study, the "Goodreads Quotes" dataset [15] was utilized to compare the ability of two different models to capture the thematic sentiments of different quotes. The BERT model [7], at the forefront of NLP technology, was considered as the computationally expensive advanced NLP model for this experiment, whereas, the Support Vector Machine (SVM) [5] was used for the comparison as one of the traditional machine learning model. After fine-tuning the BERT model on the dataset, an average accuracy of 45% was achieved, while the SVM model achieved a 37% accuracy after training on the data. A comprehensive statistical performance analysis was conducted to evaluate the outcomes. Using the ANOVA test, it was confirmed that there is a significant statistical difference between the performance of the two models. Further analysis with Approximate Visual Tests and Paired Observation tests led to the conclusion, both with a 99% confidence level, that the BERT model substantially outperforms the SVM in thematic sentiment analysis. This result not only highlights the advanced capabilities of modern NLP models but also offers conclusive evidence of their effectiveness over traditional ML models in complex linguistic tasks.

II. Related Work

The BERT(Bidirectional Encoder Representation from Transformers) model represents a significant leap in the field of NLP. It works by processing words in relation to all the other words in a sentence, rather than the one-by-one approach, which allows for a deeper understanding of the context. This model has been adopted for numerous NLP tasks [18], [19] due to its ability to handle a wide range of language processing tasks and its deep bidirectional nature. The working mechanism of the BERT model is demonstrated in figure 1. It can be seen that the model has two stages, namely, the Pre-training and the Fine-tuning. The pre-training stage uses

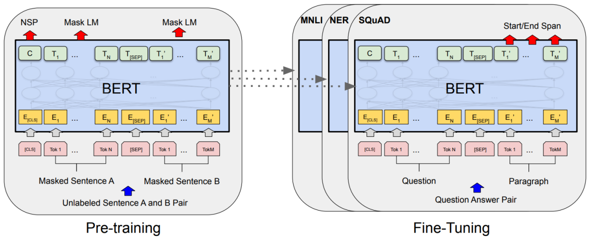

Fig. 1. Working mechanism of BERT model. This figure demonstrates the two stages of working with the model, Pre-Training and Fine-Tuning.

a large corpus of data to train the model and the fine-tuning stage makes the model ready for particular tasks.

The Support Vector Machine (SVM) is a supervised learning algorithm that is known for its effectiveness in classification tasks. It functions by finding the hyperplane that best divides a set of data points into classes. This model has been a popular choice for text classification [16], [17] due to its efficiency in high-dimensional spaces and its flexibility in handling various types of data.

Recent advancements in thematic sentiment analysis have seen various approaches and methodologies. Several studies have explored the use of machine learning techniques for sentiment analysis, setting a foundation for thematic analysis in text [8]–[10]. Many research works employed deep learning techniques for sentiment analysis, revealing the potential for a nuanced understanding of themes and sentiments in large datasets [11], [12].

Moreover, thematic sentiment analysis has been increasingly utilized in social media analysis and other sectors [13], [14] which highlights the evolving nature of sentiment analysis, moving from basic sentiment extraction to more complex thematic understanding, a trend that has been facilitated by the advent of advanced NLP models.

III. Problem Description

A. Description of Problem addressed

The primary challenge addressed in this study lies in the domain of thematic sentiment analysis within literary texts, a complex task that involves identifying the thematic sentiment conveyed in a quote. Traditional sentiment analysis focuses on classifying text into basic emotional categories like positive, negative, or neutral. However, this approach overlooks the thematic elements that are crucial for a comprehensive understanding of literary works. Themes in literature, such as love, betrayal, or courage, provide a deeper context to the sentiments expressed, making the analysis more intricate and nuanced.

Moreover, the problem extends to the technological realm, concerning the efficacy of NLP and ML models in accurately capturing and interpreting these nuanced expressions. While advanced models like BERT have shown promise in various NLP tasks, their effectiveness in thematic sentiment analysis, especially in the context of diverse and complex literary data, remains a subject of exploration. On the other hand, traditional ML models like SVM, known for their efficiency in classification tasks, face challenges in dealing with the high-dimensional and context-rich data typical of literary texts.

The comparison between the BERT model and the SVM model in order to identify the theme of a quote is the focus of this study. This experiment provides insights into the strengths and limitations of each approach in the context of thematic sentiment analysis. The outcome of this comparison is critical in guiding future research and development in NLP, particularly in enhancing the understanding and processing of literary content.

B. Research Techniques

A number of research methodologies and statistical tools have been utilized in this research which helped with getting the desired outcome and then analyzing the outcome to conclude which of the methods used is better. The techniques used can be listed as such:

- Finding the mean, median, and quartiles and using box plots and bar charts.
- Sampling the dataset.
- Using a 10-fold cross-validation technique to record the performance of the SVM model after training on the dataset.
- Fine-tuning a pre-trained BERT model on the dataset and recording the accuracy.
- Hypothesis Testing.
- Confidence Interval calculation.
- Analysis of Performance (ANOVA) test to check whether the models are significantly different in performance.
- Approximate Visual Test to find the better model.
- Paired Observation test to solidify the performance analysis.

The process of application of the techniques has been discussed in detail in the following sections.

IV. Methodology

A. Dataset

A dataset called the “Goodreads Quotes” [15] was collected from Kaggle [20]. The dataset contains ‘Quotes’ and the related ‘Tag’ for each of the quotes. These ‘Tags’ represent the themes of each quote, for example, if a quote is inspirational, romantic, or has a theme of war, etc. A sample of the dataset is provided in figure 2. A total of 82,460 quotes are present in the dataset which are each assigned a tag from a list of 27 themes.

quotes	tag
you won't even take your bow? are you planning to throttle a	humor
you realize that is you allow me to court you, all your opposition to	romance
dustfinger still clearly remembered the feeling of being in love for tl	happiness
their conversation ceased abruptly with the entry of an oddly-	books
the universe is a philosophical abyss.	philosophy
we need not take refuge in supernatural gods to explain our saints a	wisdom
to solitude	poetry
a flower knows, when its butterfly will return,	poetry
any intelligent person knows that life is a beautiful thing and that th	happiness
until we have begun to go without them, we fail to realize how unne	philosophy

Fig. 2. A sample of the dataset with two columns, Quote and Tag.

B. Dataset Preprocessing

A total of 27 different CSV files were in the dataset which contained the quotes and the name of the CSV files represented the themes of the respective quotes. All these data were compiled under a single file and named "Quotes Dataset" by adding an extra column named 'Tag' that indicated the theme of the quote in that row. Furthermore, any quotes that contained symbols other than those in the English Language were discarded.

C. Data Sampling

A random sampling of 10% of the data was used to create and test the experimental setup. The mean, median, and quartiles for the quote length were calculated. A box plot was used as seen in figure 4 to determine the outliers in the data that had a higher number of words and could result in creating complications for the models. All the data with more than 1.5 times the interquartile range was removed and thus a dataset was constructed with a relatively similar number of words for each quote across all the themes. To verify the average number of words in each quote, a bar chart was generated, which is presented in figure 5, which confirms that the average length of the quotes by the number of words is almost similar for each of the 27 themes.

D. Train-Test Set

An 80-20 split of the dataset was made using random sampling to create the Training set and the Validation set to be used in the experiment. The split is shown in figure 3, where the 63,838 data were split using the *'TrainTestSplit'* library of *'scikit'*.

E. Bidirectional Encoder Representations from Transformers (BERT)

The BERT model is based on the transformer architecture, which relies heavily on attention mechanisms, particularly self-attention, to draw global dependencies between input and output. Unlike the traditional models that read the text input sequentially, BERT reads the entire sequence of words at once, allowing the model to learn the context of a word based on all its surroundings. This model has two stages:

- Pre-training: Model is trained on a large corpus of text on unsupervised tasks.
- Fine-tuning: The pre-trained model is fine-tuned with labeled data for specific tasks.

In this experiment, we focus on fine-tuning an already pre-trained BERT model. The core of the BERT is the transformer, which uses an attention mechanism. The formula is the scaler dot-product attention, where Q is the query matrix, K is the key matrix, V is the value matrix and d_k is the dimension of the key.

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{1}$$

After that, BERT uses the Token Embedding to convert words to vectors. During the fine-tuning, the final hidden states of the transformer are used to predict output labels for the sentiment analysis task. Since this is a classification task of sort, the hidden state corresponding to the first token, [CLS], is used to predict the class label.

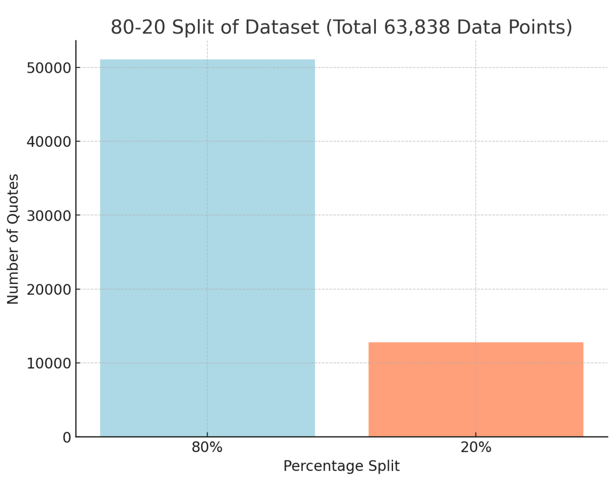

Fig. 3. 80-20 Split of the dataset to create the training set and the test set.

F. Support Vector Machine (SVM) Classifier

The basic principle behind SVM is to find a hyperplane in an N-dimensional space, where N is the number of features, that distinctly classifies the data points. The hyperplane is a decision boundary that separates data points of different classes. The support vectors are data points that are closest to the hyperplane and influence the position and orientation of the hyperplane. Using these support vectors, we plan to maximize the margin of the classifier. The margin is a gap between the two lines on the closest class points. This is calculated as the perpendicular distance from the line to the support vectors. In this experiment, we are using non-linearly separable data and so the kernel trick to transform the input space into a higher-dimensional space where it is easier to separate the data linearly is used. The radial basis function (RBF) kernel is as follows:

$$K(x_i, x_j) = exp(-\gamma||x_i - x_j||^2), \gamma > 0 \tag{2}$$

In classification, SVM outputs a class label for each input vector. The decision function is:

$$f(x) = sign(w.x + b) \tag{3}$$

Here, w is the weight vector, x is the feature vector and b is the bias.

G. ***Experimental Setup***

Using the small subset of the dataset, the experiment was set up. The research was performed in a machine with a core i7 12th-generation processor and NVIDIA RTX 4070 GPU.

1) ***Environment Setup:*** A virtual environment was created using Anaconda [21] where Python 3.9 [22] was used and various libraries were installed. Pytorch [23] was used as the framework for implementing the different models.

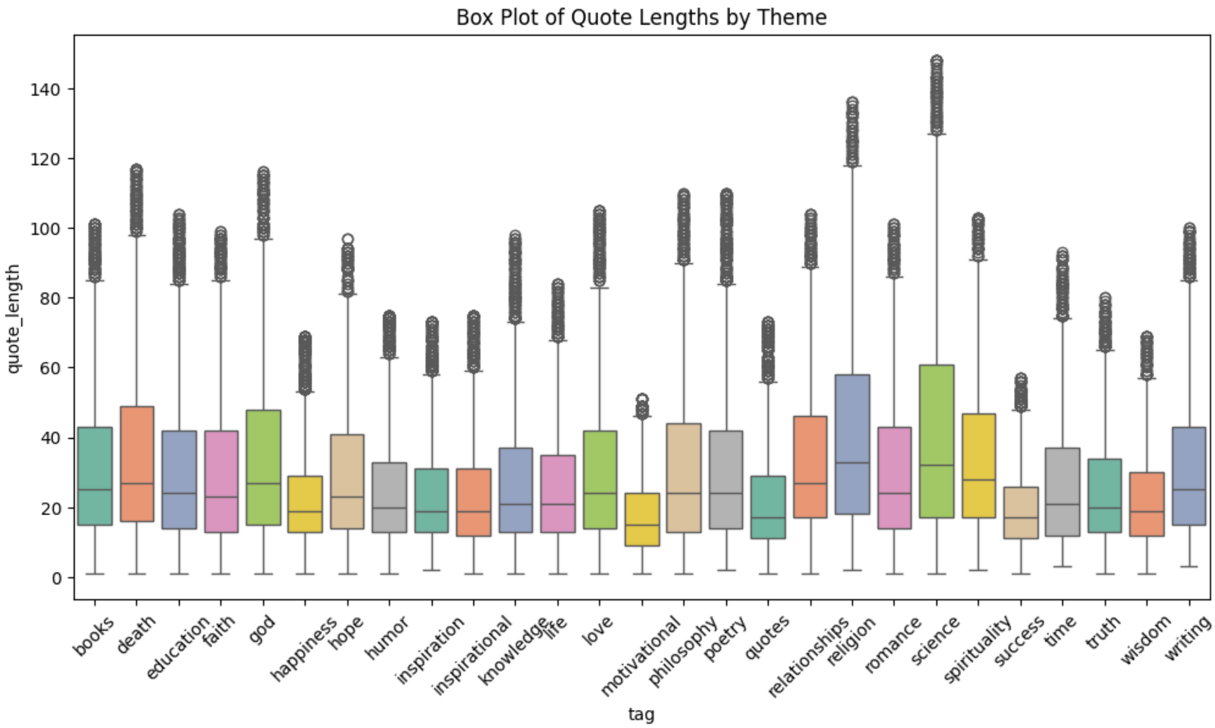

Fig. 4. Box-plot demonstrating the quote length by Tags. Using the Interquartile range, the outliers were discarded.

*2) **BERT model setup**:* The default tokenizer was used for the BERT model in order to tokenize the quotes. This tokenizer is designed to work well with the BERT model, handling aspects like word piece tokenization and adding special tokens like [CLS] and [SEP] necessary for BERT's input. The max length of the data or the token size was decided to be 200 after looking at the lengths of each quote by their words. The Huggingface Transformers library [24] was used to implement this classifier. The model was fine-tuned on 80% of the data and the rest 20% was used as the validation set. The model was tuned for 10 epochs and the accuracy was noted. The process was repeated 10 times with different training and validation sets and the accuracies were noted for each run.

*3) **TF-IDF tecnique setup**:* The SVM model cannot capture the intricate details of the words in the quotes. So, the text data needs to be converted into a format suitable for the SVM model. The Term Frequency-Inverse Document Frequency (TF-IDF) [25] technique was used to convert the quotes into a feature vector, highlighting the importance of words within the dataset, thus capturing the thematic essence of the quotes. The Term Frequency(TF) measures how frequently a term occurs in a document or quote in this case. The formula used for this is:

$$TF(t,q) = \frac{\text{Number of times term } t \text{ appears in quote } q}{\text{Total Number of terms in quote } q} \quad (4)$$

The Inverse Document Frequency(IDF) measures how important a term is with the entire corpus. The IDF value increases when the term is rare across the documents and decreases when it is more common. The IDF is calculated as follows:

$$IDF(t,Q) = \log\left(\frac{\text{Total number of quotes } |Q|}{1 + \text{Number of quotes containing term } t}\right) \quad (5)$$

Finally, we calculated the TF-IDF score as the product of the two.

$$\text{TF-IDF}(t,q,Q) = TF(t,q) \text{ x } IDF(t,Q) \quad (6)$$

The intuition behind this is that the terms that appear frequently in one document but not in many documents across the corpus are likely to be significant.

*4) **SVM model setup**:* A number of different kernels are available for the SVM model. However, after careful analysis, a Radial-basis function kernel was chosen for its effectiveness in high-dimensional spaces and its suitability for text classification tasks. The model's hyper-parameters, such as the regularization parameter and the margin of the hyperplane, were optimized using a grid search with cross-validation to find the best combination that maximizes the accuracy of the model. An 80-20 split of the dataset was done as seen in figure 3 to create the train-test sets to train and evaluate the model's performance.

V. Experimental Results and Analysis

After the experiment was setup as discussed in the previous section, the models were evaluated on the validation sets to get the accuracy of their performance. Both the models were tested 10 times each to get the accuracy of their prediction of the

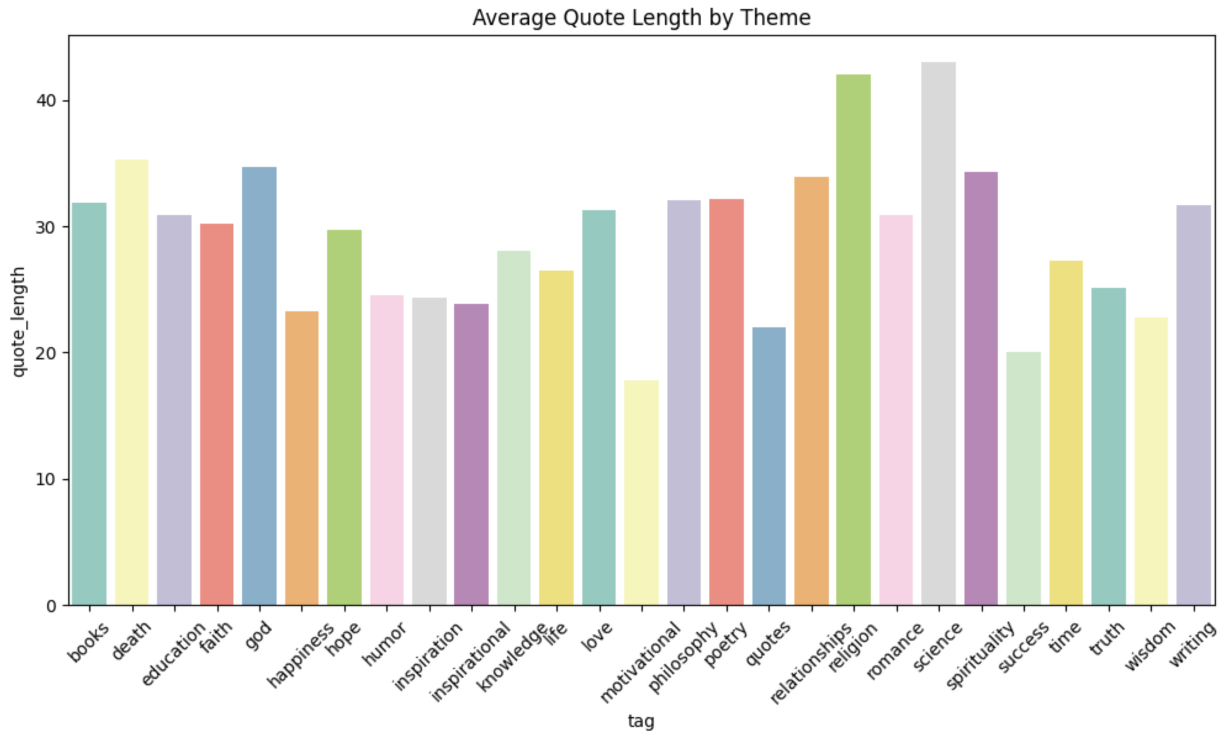

Fig. 5. Bar chart demonstrating the average length of the quotes in each theme after the removal of outliers.

themes of each quote with different validation sets. The results were noted as seen in table I and used for further analysis. The Accuracy is determined based on the following formula:

$$Accuracy = \frac{|\text{Correctly Classified Instances}|}{|\text{Total Number of Quotes}|} \tag{7}$$

TABLE I
LIST OF ACCURACIES OF BERT AND SVM MODELS WITH 10 DIFFERENT TRAIN-TEST SPLIT

Trial Number	Result of BERT	Result of SVM
1	0.460604	0.356961
2	0.473762	0.369884
3	0.471491	0.376542
4	0.466635	0.367926
5	0.451594	0.377521
6	0.443531	0.368709
7	0.441729	0.368905
8	0.440162	0.388291
9	0.434837	0.378304
10	0.438596	0.378108

A. Hypothesis Testing

In this case, let us assume a Null hypothesis that the performance of both the models is the same and that they provide the same accuracy metric as defined earlier. With respect to that, the alternative hypothesis would be that the models have a significant difference in their performance and one of the models performs better than the other.

- ***Null Hypothesis:*** Both the models have the same performance on the dataset provided for the defined task.
- ***Alternative Hypothesis:*** The models are different in terms of their performance on the provided dataset for the defined task and one is better than the other.

To reject the Null Hypothesis and accept the Alternative Hypothesis, a few tests were done on the results received after 10 different execution of the models and collecting their accuracies on each run. These tests are discussed in detail.

B. Analysis of Variance (ANOVA)

After performing the analysis of variance on the two models, it was clear that the two models have a significant statistical difference in terms of their performance on the presented task. The ANOVA test was performed on Microsoft Excel (single instance ANOVA test) and the results are shown in figure 6. We calculated the F-value of 218.58138 and received the P-Value as 1.64177e-11. Since the P-value is much less than the alpha (0.01), we can say with 99% confidence that the performance of the two models is significantly different.

C. Approximate Visual Test

In order to determine which one of the two models performs better, we lean towards the approximate visual test to check if we can get an answer. This is a test where we calculate the confidence interval of both the models' performance and

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Column 1	10	4.522941	0.452294	0.000215		
Column 2	10	3.731151	0.373115	7.21E-05		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.031347	1	0.031347	218.5814	1.64E-11	4.413873
Within Groups	0.002581	18	0.000143			
Total	0.033928	19				

Fig. 6. Performing ANOVA test to determine if the models are statistically different in terms of performance.

compare them to see which one has a higher confidence interval. We use the following formulas for this calculation:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{8}$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{9}$$

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{10}$$

$$CI = \bar{x} \pm t_{\frac{\alpha}{2},n-1}\left(\frac{s}{\sqrt{n}}\right) \tag{11}$$

We can only say that a model is better than the other with a certain confidence interval if the upper bound of one model is smaller than the lower bound of another model's confidence interval and there is no overlap in the confidence intervals. In that regard, we proceed to calculate the 90%, 95%, and 99% confidence intervals of the models' performance as seen in table II. It can be clearly seen that even at a 99% confidence interval we do not have an overlap and the BERT model shows a better result.

D. Paired Observations

To confirm our analysis that the BERT model has a better performance than the SVM model, we perform another test called the paired observations. In this case, we calculate the confidence interval at a certain significance level for the difference in performance of the models. If the confidence interval does not contain 0 in its range, we can conclude that one model is better than the other. We calculated the confidence interval after calculating the mean, variance, and standard deviation and used the following formula to calculate the confidence interval, where $\bar{d}$ represents the mean of the difference between the performance of the models, s_d is the standard deviation of the differences, and $t_{\frac{\alpha}{2}}$ is the value of the t-distribution with $n-1$ degree of freedom (number of pairs = 10).

TABLE II
APPROXIMATE VISUAL TEST RESULTS WITH 90%, 95%, AND 99% CONFIDENCE INTERVALS

Metric	**BERT**	**SVM**
Sample Mean	0.4522941	0.3731151
Variance	0.000193259	6.48778E-05
Standard Deviation	0.013901752	0.008054673
90% Confidence Interval		
Lower Bound	0.452182072	0.373077492
Upper Bound	0.452406128	0.373152708
95% Confidence Interval		
Lower Bound	0.452155851	0.373068689
Upper Bound	0.452432349	0.373161511
99% Confidence Interval		
Lower Bound	0.45209549	0.373048426
Upper Bound	0.45249271	0.373181774

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i) \tag{12}$$

$$s_d^2 = \frac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n-1} \tag{13}$$

$$s_d = \sqrt{\frac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n-1}} \tag{14}$$

$$CI = \bar{d} \pm t_{\frac{\alpha}{2}}\left(\frac{s_d}{\sqrt{n}}\right) \tag{15}$$

We calculated the confidence interval with confidence of 90%, 95%, and 99% as seen in table III and none of them had zero in the interval. This concludes that the BERT model has a significantly better performance than the SVM model for this defined task.

E. Result Summary

The overall summary of the performance analysis is presented in the table IV. It can be clearly seen that the ANOVA test states that the performance of the two models (BERT and SVM) are significantly different, thus rejecting the null hypothesis and accepting the alternative hypothesis. The Approximate Visual Test and Paired Observations come to the same conclusion that the BERT model is better than the two models. Thus, by performing these analyses, we can be sure that the models used for the task defined for this study are significantly different and the BERT model outperforms the SVM model in predicting the themes of literary quotes.

TABLE III
PAIRED OBSERVATIONS WITH 90%, 95%, 99% CONFIDENCE INTERVALS

Metric	Value
Sample Mean	0.079179
Variance	0.0003532
Standard Deviation	0.018793627
90% Confidence Interval, 9 df	
Lower Bound	0.078974256
Upper Bound	0.079383744
95% Confidence Interval, 9df	
Lower Bound	0.078926336
Upper Bound	0.079431664
99% Confidence Interval, 9df	
Lower Bound	0.07881602
Upper Bound	0.07954198

TABLE IV
SUMMARY OF STATISTICAL TESTS AND DECISIONS

Tests	ANOVA	Approximate Visual Test	Paired Observation Test
Result	Statistically Different	BERT model is better	BERT model is better

VI. CONCLUSION

The research demonstrated that the BERT model significantly outperforms the SVM model in the thematic sentiment analysis of the Goodreads Quotes dataset. Despite the computational expense of the BERT model, its superior ability to generalize to new and unseen data makes it a more viable option compared to the SVM model, which struggles with generalization. The insights gained from this research extend beyond thematic sentiment analysis, indicating promising applications for advanced NLP models in various other NLP tasks. The findings underscore the potential of using sophisticated models like BERT for a deeper and more nuanced understanding of text, paving the way for future innovations in the field of NLP. This study contributes to the growing body of knowledge in NLP and sets a precedent for further explorations into the effective application of advanced NLP techniques in diverse linguistic contexts.

ACKNOWLEDGMENT

The "Goodreads Quotes" Dataset has been collected from Kaggle which was scrapped and compiled into a CSV file by the original creator of the Dataset.

REFERENCES

[1] K. Lybarger, N. J. Dobbins, R. Long, A. Singh, P. Wedgeworth, Ö. Uzuner, and M. Yetisgen, "Leveraging natural language processing to augment structured social determinants of health data in the electronic health record," *Journal of the American Medical Informatics Association*, vol. 30, no. 8, pp. 1389–1397, 2023.

[2] Z. Chen, W. Chen, C. Smiley, S. Shah, I. Borova, D. Langdon, R. Moussa, M. Beane, T.-H. Huang, B. Routledge, et al., "Finqa: A dataset of numerical reasoning over financial data," *arXiv preprint arXiv:2109.00122*, 2021.

[3] T. Lalwani, S. Bhalotia, A. Pal, V. Rathod, and S. Bisen, "Implementation of a Chatbot System using AI and NLP," *International Journal of Innovative Research in Computer Science & Technology (IJIRCST) Volume-6, Issue-3*, 2018.

[4] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 4, e1253, 2018.

[5] W. S. Noble,"What is a support vector machine?," *Nature Biotechnology*, vol. 24, no. 12, pp. 1565–1567, 2006.

[6] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[7] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[8] H. Saif, Y. He, M. Fernandez, and H. Alani, "Contextual semantics for sentiment analysis of Twitter," *Information Processing & Management*, vol. 52, no. 1, pp. 5–19, 2016.

[9] S. Malviya, A. K. Tiwari, R. Srivastava, and V. Tiwari, "Machine learning techniques for sentiment analysis: A review," *SAMRIDDHI: A Journal of Physical Sciences, Engineering and Technology*, vol. 12, no. 02, pp. 72–78, 2020.

[10] B. Agarwal and N. Mittal, "Machine learning approach for sentiment analysis," tin *Prominent Feature Extraction for Sentiment Analysis*, pp. 21–45, Springer, 2016.

[11] A. Yadav and D. K. Vishwakarma, "Sentiment analysis using deep learning architectures: A review," *Artificial Intelligence Review*, vol. 53, no. 6, pp. 4335–4385, 2020.

[12] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 4, e1253, 2018.

[13] T. Rahman, J. Nwokeji, R. Matovu, S. Frezza, H. Sugnanam, and A. Pisolkar, "Analyzing Competences in Software Testing: Combining Thematic Analysis with Natural Language Processing (NLP)," in *2021 IEEE Frontiers in Education Conference (FIE)*, pp. 1–9, 2021.

[14] E. Fitkov-Norris and N. Kocheva, "Are we there yet? Thematic analysis, NLP, and machine learning for research," in *ECRM 2023 22nd European Conference on Research Methods in Business and Management*, 2023.

[15] A. Kamr, "Goodreads Quotes [Data set]," Kaggle, 2023. [Online]. Available: https://www.kaggle.com/datasets/abdokamr/good-reads-quotes

[16] F. Colas and P. Brazdil, "Comparison of SVM and some older classification algorithms in text classification tasks," in *IFIP International Conference on Artificial Intelligence in Theory and Practice*, pp. 169–178, 2006.

[17] M. Goudjil, M. Koudil, M. Bedda, and N. Ghoggali, "A novel active learning method using SVM for text classification," *International Journal of Automation and Computing*, vol. 15, pp. 290–298, 2018.

[18] I. Tenney, D. Das, and E. Pavlick, "BERT rediscovers the classical NLP pipeline," *arXiv preprint arXiv:1905.05950*, 2019.

[19] M. V. Koroteev, "BERT: a review of applications in natural language processing and understanding," *arXiv preprint arXiv:2103.11943*, 2021.

[20] "Kaggle [Online]," Available: https://www.kaggle.com/.

[21] "Anaconda [Online]," Available: https://www.anaconda.com/.

[22] "Python 3.9.0 Release [Online]," Available: https://www.python.org/downloads/release/python-390/.

[23] "PyTorch Documentation [Online]," Available: https://pytorch.org/docs/stable/index.html.

[24] "Hugging Face's BERT Base Uncased Model [Online]," Available: https://huggingface.co/bert-base-uncased.

[25] S. Qaiser and R. Ali, "Text mining: use of TF-IDF to examine the relevance of words to documents," *International Journal of Computer Applications*, vol. 181, no. 1, pp. 25–29, 2018.