FinBERT-FOMC: Fine-Tuned FinBERT Model with Sentiment Focus Method for Enhancing Sentiment Analysis of FOMC Minutes

Ziwei Chen*
ziwei.chen@student.unisg.ch
University of St.Gallen
St.Gallen, Switzerland

Sandro Gössi*
sandroantoniours.goessi@student.unisg.ch
University of St.Gallen
St.Gallen, Switzerland

Wonseong Kim
wonseongkim@korea.ac.kr
University of St.Gallen
St.Gallen, Switzerland
Korea University
Seoul, South Korea

Bernhard Bermeitinger
bernhard.bermeitinger@unisg.ch
University of St.Gallen
St.Gallen, Switzerland

Siegfried Handschuh
siegfried.handschuh@unisg.ch
University of St.Gallen
St.Gallen, Switzerland

ABSTRACT

In this research project, we used the financial texts published by the Federal Open Market Committee (FOMC), known as the FOMC Minutes, for sentiment analysis. The pre-trained FinBERT model, a state-of-the-art transformer-based model trained for NLP tasks in finance, was utilized for that. The focus of this research has been on improving the predictive performance of complex financial sentences, as our problem analysis has shown that such sentences pose a significant challenge to existing models. To accomplish this objective the original FinBERT model was fine-tuned for domain-specific sentiment analysis. A strategy, referred to as Sentiment Focus (SF) was utilized to reduce the complexity of sentences, making them more amenable to accurate sentiment predictions.

To evaluate the efficacy of our method, we curated a manually labeled test dataset comprising 1,375 entries. The results demonstrated an overall improvement of 5 % in accuracy when using SF-enhanced fine-tuned FinBERT over the original FinBERT model. In cases of complex sentences containing conjunctions like *but*, *while*, and *though* with contradicting sentiments, our fine-tuned model outperformed the original FinBERT by a margin of 17.4 %.

CCS CONCEPTS

• **Computing methodologies → Natural language processing**; Supervised learning by classification; Neural networks; • **Applied computing** → *Economics*.

KEYWORDS

Natural Language Processing, Economics, Financial Economics, Sentiment Analysis, FinBERT, FOMC Minutes, Sentiment Focus

*Both authors contributed equally to this research.

ICAIF '23, November 27–29, 2023, Brooklyn, NY, USA

ACM ISBN 979-8-4007-0240-2/23/11.
https://doi.org/10.1145/3604237.3626843

ACM Reference Format:
Ziwei Chen, Sandro Gössi, Wonseong Kim, Bernhard Bermeitinger, and Siegfried Handschuh. 2023. FinBERT-FOMC: Fine-Tuned FinBERT Model with Sentiment Focus Method for Enhancing Sentiment Analysis of FOMC Minutes. In *4th ACM International Conference on AI in Finance (ICAIF '23), November 27–29, 2023, Brooklyn, NY, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3604237.3626843

1 THE FEDERAL RESERVE

The Federal Reserve, or *Fed* for short, is the name of the central bank of the United States of America. The Fed consists of a central governmental agency in Washington, D.C., and twelve regional Federal Reserve Banks located in major cities throughout the United States. This central banking system was established in 1913 to provide the country with a safe, flexible, and stable monetary and financial system[1].

The Federal Reserve's responsibilities include conducting national monetary policy, supervising and regulating banks, maintaining the stability of the U.S. dollar, and providing financial services to depository institutions, the U.S. government, and foreign public entities.

With the evolution of monetary policy, the Federal Reserve has implemented policy tools that address concerns as well as publicly communicate information about the state of the economy.

1.1 The Federal Open Market Committee (FOMC)

The Federal Open Market Committee (FOMC) is a branch of the Federal Reserve that is responsible for making key decisions about interest rates and the growth of the United States' money supply.

It is arguably the most important part of the Federal Reserve System because it is in charge of formulating and implementing monetary policy, which has wide-ranging impacts on the national and global economy [12].

The FOMC consists of twelve members, including the seven members of the Board of Governors of the Federal Reserve System, the president of the Federal Reserve Bank of New York, and four of

[1] https://www.federalreserve.gov/aboutthefed/structure-federal-reserve-system.htm

the remaining eleven Reserve Bank presidents, who serve one-year terms on a rotating basis.

1.2 FOMC Meetings, Minutes, and Statements

The FOMC holds eight regularly scheduled meetings during the year and additional meetings as needed to discuss the outlook for the U.S. economy and make adequate policy decisions, whilst concurrently acknowledging and navigating the complexities and exigencies of the international economic landscape. The decisions taken during these meetings can influence the size and rate of growth of the money supply, which in turn affects interest rates[2].

Following these meetings, a statement is released to announce the decisions made regarding monetary policy which provides a summary of the committee's current assessment of economic conditions and projections and the rationale behind any policy changes.

The minutes of the FOMC meetings are a more detailed record of the released post-meeting statement. They provide a full account of policy discussions, including differing views and analysis of the economic conditions that influenced the votes of the members. These minutes are released three weeks after each meeting. The objective of this natural language processing research was to predict the sentiment embedded within financial texts utilizing the FOMC minutes.

1.3 Motivation

These tools of the Federal Reserve—the meetings, statements, and minutes—are important because they provide insight into the thinking of one of the most influential economic policy-making bodies in the world. For example, since the US Federal Reserve began raising the federal funds rate in March 2022, almost all asset classes have performed poorly [9].

They are closely watched by economists, traders, and investors because they can provide clues about the future direction of monetary policy and thus about future economic conditions [12]. There have been well-documented cases of world-renowned hedge funds using sentiment analysis to capture monetary policy decisions in fractions of a second [2]. Also, large Banks such as Morgan Stanley have patented AI-powered tools that help them predict monetary policy actions by analyzing the Federal Reserve's communications [7].

Modifications in monetary policy, like Fed policy rate changes, significantly affect financial markets because they are closely interconnected. Understanding the FOMC's sentiment can be very important for decision-making in many different sectors of the economy [8].

The FOMC Minutes are chosen as financial text data because they present a compelling problem statement with substantial relevance to both economic research and practical financial management. The FOMC minutes are unique in that they are a rich source of qualitative information reflecting the Federal Reserve's comprehensive thoughts on the U.S. economy [8].

The text-based format of the FOMC Minutes provides a valuable opportunity to apply natural language processing techniques facilitating a deeper exploration of sentiment analysis models.

[2]https://www.federalreserve.gov/monetarypolicy/fomccalendars.htm

Despite the complexity of the linguistic nuances often found in FOMC minutes, we were motivated by the potential insights that can be gained from transforming and quantifying these textual data into actionable economic sentiment indicators.

1.4 Task Description

The task of this project was to train a model that can accurately determine, based on the analysis of a range of sentences extracted from the FOMC Minutes, whether the overall sentiment of the financial economy expressed by the Federal Reserve is positive or negative.

This is a typical classification problem, with the class labels being *positive*, *neutral*, or *negative*. To address such sentiment analysis problems, there exist a variety of natural language processing and machine learning algorithms that can be employed, including Naive Bayes, logistic regression, support vector machines, deep learning models such as LSTMs/RNNs [4], and transformer-based models like BERT [3], GPT [6], and others. These models can capture and quantify the sentiment conveyed through the nuances of language in texts.

Our ultimate goal is to choose and use an existing pre-trained model and improve its performance and effective prediction capabilities for sentiment analysis for domain-specific financial texts such as FOMC Minutes through transfer learning and the power of fine-tuning.

1.5 Research Objectives

In this research project, we define our objectives in four key areas of benefit where sentiment analysis of the FOMC minutes can be valuable in providing nuanced understanding and insights into economic trends and monetary policy. The identified areas of focus are:

1.5.1 Understanding Market Situations. Sentiment analysis on the FOMC minutes can provide valuable insights into market expectations. The sentiment expressed by the Federal Reserve reflects its assessment of economic conditions and the potential direction of future monetary policy. By analyzing this sentiment, one can gain a deeper understanding of the market's anticipations about future economic trends and policy shifts, which can influence investment decisions, portfolio management, and risk assessments.

1.5.2 Predicting Market Reactions. Financial markets often react strongly to the information contained in the FOMC minutes. Positive or negative sentiment can drive market optimism or pessimism, affecting asset prices, volatility, and trading volumes. Through sentiment analysis, it is possible to gain a predictive edge regarding market reactions.

1.5.3 Assessing Economic Outlook. The sentiment expressed in the FOMC minutes is a reflection of the Federal Reserve's view on the economic outlook of the U.S. economy. It covers various facets including growth prospects, inflation, labor market conditions, and potential risks to the economic forecast. By performing sentiment analysis on these aspects, we can quantify the Federal Reserve's economic outlook, capturing nuanced views that can enhance traditional economic forecasts.

1.5.4 *Understanding Policy Implications.*
The Federal Reserve's views and sentiments expressed in the FOMC minutes have far-reaching policy implications. They often hint at the future direction of monetary policy, including decisions about interest rates and the management of the money supply. Understanding this sentiment can shed light on the potential impact of such policy changes on various sectors of the economy, such as impacts on borrowing costs, investment decisions, and spending behavior.

1.6 Data Set

We have used the Federal Open Market Committee (FOMC) Minutes as our primary and only data source. The dataset spans from January 2006 to February 2023, providing a comprehensive longitudinal perspective on the sentiment and policy shifts of the Federal Reserve during this time period.

Existing research on sentiment analysis in FOMC Minutes has typically utilized word-level analysis, even when dealing with longer paragraphs. However, this research project seeks to explore the benefits of sentence-level sentiment analysis in order to gain a more contextual understanding of the FOMC language data.

Kim et al. consists of 32,330 sentences, which offers a comprehensive collection of financial texts for sentiment analysis. The entries within the dataset were initially categorized by an author, necessitating a subsequent phase of manual labeling. This task was further enriched by the incorporation of labels from two additional researchers. This initial phase facilitated the establishment of a foundational set of categorized data, which was subsequently employed as a test set to assess our models, post-training, within the context of sentiment analysis.

1.7 Target Evaluation

As outlined in the task description, the performance of our sentiment analysis model on the FOMC Minutes will be evaluated using the accuracy metric for three classes: *positive*, *negative*, and *neutral* sentiments. Accuracy measures the proportion of true results (both true positives, true negatives, or true neutral) divided by the total number of sentences classified. It is an effective performance measurement tool for binary and multi-class classification problems. The range of accuracy is from 0 to 1, with a higher accuracy value indicating superior model performance. In our case, a high accuracy score would signify our model's strong ability to correctly classify the sentiment of sentences from the FOMC Minutes.

1.8 Baseline

Establishing a baseline for this project presents a unique scenario as our original dataset was originally unlabeled and required manual categorization. So there exists no conventional benchmark for a direct comparison.

However, after the initial labeling, we used the original FinBERT model as a sentiment analysis classifier to provide a preliminary performance measure. In this way, we obtained a baseline accuracy of 0.84 derived from the hand-crafted and manually labeled test dataset.

This figure will serve as our reference point for model improvement and comparison throughout our research. The comprehensive methodology used to achieve this initial accuracy is discussed in section 2.

As our dataset is novel in its formation, it is essential to note that its individuality contributes to the novelty and significance of our research.

2 METHODOLOGY

2.1 Related Work

In the initial stage of our research, we embarked on a comprehensive study of related work in the field, specifically those employing the FOMC Minutes. By reading academic papers and seminal works, we aimed to understand the current state of knowledge in the field, the methodological approaches commonly adopted, and the challenges faced by other researchers. We also focused on identifying gaps in the existing literature that our research could address. This helped to inform our methodology and also provided a theoretical grounding for our research, aligning our work within the broader context of sentiment analysis in financial economics.

This research project is based on the research of Kim et al., which highlights the challenges and limitations of using current NLP techniques to analyze FOMC communications, as well as the potential for improving language models and exploring alternative approaches.

In Kim et al., they employed advanced language modeling techniques such as VADER and FinBERT, as well as an experimental test with GPT-4 on the FOMC Minutes dataset, and analyzed the results. The results from this study show that FinBERT outperforms other techniques, especially in accurately predicting negative sentiment. Nonetheless, the study suggests that further improvements are necessary to optimize the performance of both FinBERT and GPT-4 in analyzing the sentiment of FOMC Minutes.

These findings formed the basis for our research project, which aimed to focus on improving the accuracy of sentiment analysis for complex financial texts and exploring alternative approaches to provide more comprehensive economic insights.

2.2 Model Selection

2.2.1 *Reasoning.*
Our analysis of related work has shown that the unique characteristics of the FOMC Minutes, such as its complex sentence structures and lower emotional content, present challenges for analysis by current NLP techniques and models.

The paper by Kim et al. demonstrates that FinBERT is superior to other methods in extracting sentiment from FOMC minutes, while GPT-4's performance is comparatively weaker, despite being a cutting-edge technique. Their findings revealed that GPT-4 struggles to accurately discern sentiment in intricate sentences, with many being classified as *neutral* despite potentially being *positive* or *negative*. According to the study, FinBERT also appears to outperform VADER in sentiment analysis of FOMC financial text data.

These considerations and findings were why FinBERT was chosen for this research.

2.2.2 *FinBERT.*
FinBERT is an open-source pre-trained NLP model based on the BERT architecture to analyze the sentiment of financial texts.

It is therefore a variant of the BERT model [3]. BERT models are designed to understand the context of words in a text by analyzing the words that come before and after it and have caused a revolution in the world of NLP by providing superior results on many NLP tasks, such as question answering, text generation, sentence classification, and many more compared to other methods.

According to Araci, general-purpose models such as BERT are not effective enough for domain-specific tasks because of the specialized language used in the financial context. Pre-trained language models can help with this problem because they require fewer labeled examples and they can be further trained on domain-specific corpora. That's why many domain-specific models have emerged using BERT as the base and are being used for a variety of use cases.

FinBERT is built by further training the BERT language model in the finance domain. It uses a large financial corpus, including news articles, corporate reports, and regulatory filings, thereby fine-tuning it for financial sentiment classification. The additional training corpus is a set of 1.8M Reuters news articles called TRC2 which were published by Reuters between 2008 and 2010 and Financial PhraseBank, which consists of 4,845 English sentences selected randomly from financial news found in the LexisNexis database [1]. These sentences then were annotated by 16 people with backgrounds in finance and business. In summary, it is trained on the following three financial communication corpora with a total corpora size of 4.9B tokens [11].

- Corporate Reports 10-K & 10-Q: 2.5B tokens
- Earnings Call Transcripts: 1.3B tokens
- Analyst Reports: 1.1B tokens

The model is designed to capture the nuances and complexities of financial language and terminology, making it suitable for various financial text analysis tasks. Consequently, FinBERT can be a powerful tool for sentiment analysis, document classification, named entity recognition, and other NLP tasks within the financial domain. It outperforms almost all other NLP techniques for financial sentiment analysis. The sentiment score of the FinBERT model is normalized to a range of -1 to 1.

2.3 Data Preparation

In the initial phase of our research, we focused on curating and segmenting the FOMC Minutes data. The data was accessed from the Federal Reserve website by navigating to *News & Events*, selecting *Press Releases*, and specifying the press type as *Monetary Policy*, where the *Minutes of the Federal Open Market Committee* can be found. Our primary goal was to convert large bodies of text into individual sentences for later analysis.

2.4 Run FinBERT Using Raw Data Set

Our dataset, which was preprocessed into individual sentences, was fed to the FinBERT model [1]. For each sentence, we utilized the model to perform text classification, outputting a sentiment label and corresponding confidence score. The label represented the sentiment of the sentence, while the score quantified the model's confidence in its prediction. These results were stored alongside the original sentences in a structured format. This iterative process was carried out for all sentences in our dataset. The total runtime was around two hours for 32,330 sentences, averaging a processing speed of about five sentences per second.

Table 1: Examples of Misclassification Errors on Complex Sentences

Example Sentence	Label	Score
"Aggregate hours fell slightly in December owing to a decrease in the workweek, but they rose over the fourth quarter as a whole."	*negative*	0.9988
"Real spending in the high-tech sector declined, although real outlays for computing equipment posted their first gain in a year."	*positive*	0.9994

2.5 Re-Label Data Set

In order to assess the performance of the FinBERT predictions, we independently and manually appraised the sentiment of each sentence, assigning our own sentiment labels manually based on our expert economic judgment. By placing our labels adjacent to the FinBERT predictions, we facilitated a straightforward comparison between the predicted and actual sentiments.

This exercise allowed us to discern the instances where the FinBERT model accurately captured the sentiment of a sentence, and more importantly, identify the cases where the model's predictions proved to be false and deviated from our manual assessment. This side-by-side comparison was crucial for evaluating the performance of the FinBERT model in correctly classifying sentiment in our specific dataset and helped us analyze where the state-of-the-art model fell short in predicting the sentiment of the FOMC Minutes.

2.6 Problem Analysis

We conducted a comprehensive review to uncover the limitations of the FinBERT model in accurately capturing sentiment in our financial texts and developed strategies for potential improvements. Our analysis has shown that FinBERT exhibits exceptional performance in processing straightforward financial statements. However, we found that the classification accuracy of the model decreased significantly with increasing sentence complexity, particularly for sentences containing disjunctive or contrasting conjunctions such as *although*, *while*, and *but*. These sentences often encapsulate nuanced sentiments that seem to pose a challenge for the FinBERT model and can lead to misclassification. Two example sentences that illustrate this misclassification problem are as shown in table 1.

The sentiment focus of the first sentence should be on *rose* in the part after *but*, therefore the correct classification would be *positive*; the sentiment focus of the second sentence should be on *declined* in the first part, and the part of the sentence containing *although* should be ignored, the correct classification should be *negative*. This problem analysis provides a solid foundation for developing improvement plans and model enhancement strategies.

Table 2: Example of the Preprocessing Steps

Example sentence	Label	Score
(Before) "While light vehicle sales had slowed in the fall, consumer spending outside the auto sector appeared to have remained vigorous."	*negative*	0.9847
(After) "consumer spending outside the auto sector appeared to have remained vigorous."	*positive*	0.8261

2.7 Find and Apply Solution

Our findings highlight the need for a more sophisticated interpretation of linguistic subtleties by the model and suggest the incorporation of more advanced preprocessing steps, linguistic features, or training strategies that can help FinBERT better understand these complex sentence structures and thus improve its overall performance.

2.7.1 Remove Comma. This preprocessing step aims for sentences characterized by the presence of disjunctive or contrasting conjunctions. From the target sentence, it selectively removes commas that were followed by the main verb or conjunction. This selective removal of the comma aimed to simplify sentence structure and mitigate instances of model misinterpretation, especially in cases where the conjunction introduces contrasting sentiments or alternatives. This first preprocessing step, therefore, played a pivotal role in reducing sentence complexity and enhancing model accuracy.

2.7.2 Sentiment Focus. As a second step in our text simplification process, we employed *sentiment focus* to shift the emphasis of a sentence based on certain keywords such as *but, although, though*, and *while*. This step changed the focus of sentences by retaining the sentence segments that were most likely to contain the main sentiment. Our observation was that in complex sentences, the main sentiment is often encapsulated in the section of the sentence that follows these contrasting conjunctions. By adjusting the sentence focus on this part of the sentence, we ensured that the sentiment analysis was conducted on the most sentiment-relevant portion of the text, thereby increasing the accuracy of FinBERT's predictions.

These two preprocessing steps combined, provide a more sophisticated approach to processing complex sentences, aimed at simplifying complex sentences and reducing misinterpretation difficulties, potentially enhancing the overall performance of sentiment analysis using the FinBERT model.

2.7.3 Preprocessing Example. In the example in table 2, we can see that when the original sentence contains the contrasting conjunctions *while*, FinBERT is influenced by the word *slowed* in the sentence and takes this as the main sentiment, which causes the real sentiment focus *vigorous* to be ignored and thus makes the incorrect classification *negative*.

After *SF*, the sentence part containing *while* was removed, and FinBERT succeeded in giving the correct classification of positive. For all contrasting conjunctions, *SF* will remove the non-sentiment focus to ensure correct classification.

2.8 Create Complex Sentence Data Set

After preprocessing all 32,330 FOMC Minutes sentences in our dataset, only 3,535 records were changed by our preprocessing method. Therefore we can conclude, that only 3,535 out of 32,330 sentences were complex sentences and the rest were simple sentences.

For our ongoing analysis, we created a new dataset solely comprised of complex sentences. These sentences were specifically those for which our preprocessing had identified and shifted the sentiment focus and modified the sentence, indicating the presence of more intricate linguistic structures. This resulted in a new specialized dataset containing 3,535 complex sentences, thus isolating a subset of our data that could be particularly challenging for sentiment analysis.

2.9 FinBERT Fine-Tuning

We applied FinBERT to a complex sentence dataset containing both simplified and focus-shifted sentences and stored the new sentiments predicted by FinBERT as new labels. This dataset of 3,535 complex sentences containing the new labels then constitutes the training set for fine-tuning FinBERT, where traditional FinBERT seemed to underperform. For fine-tuning FinBERT, we followed a similar approach and method as Wang proposed.

We loaded our saved training data and split it into training, validation, and testing sets in an 8:1:1 ratio. During this data preparation step, the labels *neutral*, *positive*, and *negative* were encoded into the numerical values $(0, 1, 2)$.

The model was trained for five epochs with a learning rate of 2×10^{-5}. We used the accuracy metric for model evaluation and set the model to save the best model parameters based on this metric.

2.10 Model Evaluation and Comparison

Finally, we used our manually labeled test dataset for the crucial task of evaluating the performance of our fine-tuned FinBERT model. This additional dataset from section 1.6 consists of 1,375 individual entries, each of which were manually selected and assigned the appropriate sentiment label. As an important step in the validation process, it allowed us to compare the model's predictions against the actual labels, which allowed us to determine the effectiveness of the fine-tuning process and sentiment focus technique. This selection of data served as the basis for our *ground truth* against which we evaluated the performance of our models. In total, we evaluated three methods on this manually labeled test set:

1. **FinBERT**: Accuracy of the original FinBERT model
2. **Text Simplification & FinBERT**: Accuracy after applying the text simplification pre-processing steps (remove comma and Sentiment Focus) before using FinBERT
3. **Fine-Tuned FinBERT**: Accuracy of the fine-tuned FinBERT model

3 RESULTS

3.1 Accuracy: Manually Labeled Test Set

First, we evaluated and compared the performance of the original FinBERT model and the fine-tuned model on the manually

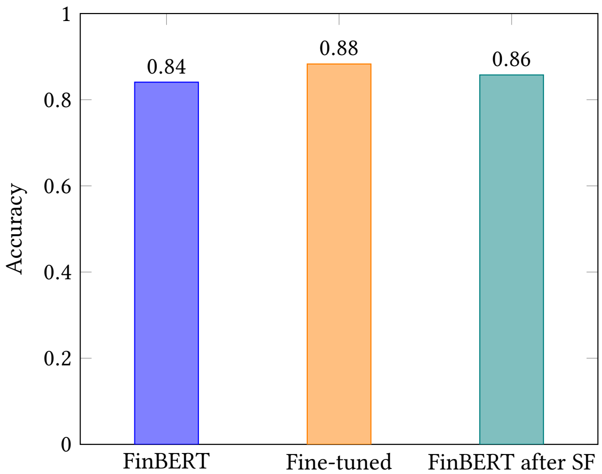

Figure 1: Overall: FinBERT vs. Fine-tuned FinBERT vs. FinBERT after SF

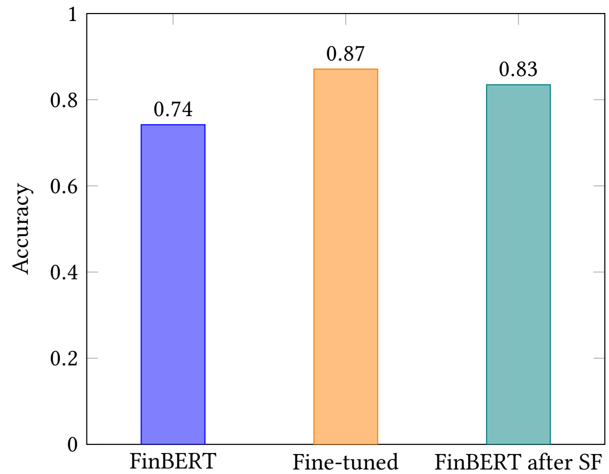

Figure 2: Complex only: FinBERT vs. Fine-tuned FinBERT vs. FinBERT after SF

labeled test dataset. Figure 1 shows the improvement in the performance of the fine-tuned FinBERT model compared to the original FinBERT model. Before any manipulation, the original FinBERT model achieved an accuracy of 0.8407 on the manually labeled test dataset. However, after simplifying the sentences in the dataset using the predefined pre-processing steps from sections 2.7.1 and 2.7.2 which simplified the text inputs and helped to set the focus on the sentence's sentiment-bearing components, the accuracy improved to 0.8575. More notably, the fine-tuned model's performance surpassed both, with an accuracy score on the original sentences of approximately 0.8829. Both these findings validate our hypothesis that Sentiment Focus or a model fine-tuned on a dataset of complex financial sentences can perform better at sentiment analysis tasks.

In evaluating the results of the model, it is important to emphasize that the 5.02 % performance improvement of our fine-tuned model over the original FinBERT may appear marginal at first glance. However, this modest increase in performance has a critical context that reinforces its significance.

Our test data set and the original data set consisted predominantly of simple sentences, whereas the original FinBERT already demonstrated considerable proficiency in its sentiment classification abilities. Thus, this minor improvement is primarily attributable to the model's ability to better handle more complex sentences, which was the main challenge for the original FinBERT. Consequently, the success of the fine-tuned model and the Sentiment Focus method in these challenging cases underscores its improved robustness and capacity to more accurately classify sentiments across a variety of sentences.

3.2 Accuracy: Complex Sentences Sub Testset

After our first performance comparison, we sought to compare the performance of the original FinBERT model, the fine-tuned model, and FinBERT after applying sentiment focus (SF) on the subset of complex sentences in our test dataset. These complex sentences, as identified in the problem analysis section, pose a significant challenge to sentiment analysis with the original FinBERT model.

In the comparison in fig. 2, we observe a substantial improvement in accuracy of 17.40 % with the fine-tuned model, reaching 0.8710 accuracy, compared to the original FinBERT's 0.7419. The original FinBERT model, in combination with the sentiment focus pre-processing step, also showed improved performance, achieving an accuracy of 0.8347. These results demonstrate the effectiveness of the Sentiment Focus solution and our fine-tuning process, particularly for complex sentences, which presented the primary area of difficulty for the original FinBERT model. The enhanced performance of FinBERT after SF underscores the significance of simplifying sentences to improve sentiment classification, even without additional model fine-tuning.

3.3 Accuracy: Different Pronouns

In addition, to get more in-depth insights, we also performed a more granular analysis of the sentiment classification results on the test data. We specifically examined sentences that contain the words *but*, *while*, and *though* to get a better understanding of where our model improved. As we identified in the problem analysis, these words often introduce complexity into a sentence by indicating contrast or presenting multiple sentiments in the same sentence. Our analysis showed notable differences between the performance of the original FinBERT model and our fine-tuned model concerning sentence complexity. The fine-tuned model consistently outperformed the original FinBERT in accurately classifying the sentiments of such sentences. For instance, in sentences containing *but*, the fine-tuned model achieved an accuracy of 0.8614, compared to FinBERT's 0.6733. Similarly, for sentences with *while*, the fine-tuned model's accuracy was 1, matching FinBERT's 0.8750. In sentences with *though*, the fine-tuned model attained an accuracy of 0.9038, which was significantly higher than FinBERT's 0.7308. These findings underline the improved capability of our fine-tuned model in handling

Table 3: Accuracy for Sentences with Different Conjunctions

Conjunction	FinBERT	Fine-tuned
but	0.6733	0.8614
while	0.8750	1.0000
though	0.7308	0.9038

sentence complexity, which makes it a valuable tool for nuanced sentiment analysis necessary in these complex financial sentences.

4 DISCUSSION

4.1 Discussion of Results

This study set out to enhance the performance of the original FinBERT model in classifying financial sentences, with a focus on solving the problems that come with more challenging complex sentences.

The primary findings suggest that our proposed method of simplifying sentences through the sentiment focus technique before feeding them into the FinBERT model significantly improves its performance in sentiment analysis, especially for complex sentences containing conflicting and contradicting sentiments.

One notable observation is the model's increased performance on sentences containing conjunctions such as *but*, *while*, and *though*, which our problem analysis showed, posed a significant challenge to existing models.

In general, our fine-tuned model and the prior use of only the sentiment-focus technique outperformed the original FinBERT in accurately categorizing the sentiments in these sentences. This improvement provides evidence that the SF sentence simplification method helps the model better handle complexity and ambiguity, which are common in such financial texts as published by the FOMC.

The comparison between our fine-tuned model and the original FinBERT model demonstrated an overall performance improvement of 4.2 %. While this might seem minimal, it becomes significant when considering the nature of our manually labeled test data set, which consisted primarily of simple sentences. The impact of our method becomes more prominent when applied only to complex sentences, which typically pose a bigger challenge to sentiment classification and were the focus of our research, as identified in our problem analysis.

4.2 Implications of the Research

The results of this study have far-reaching implications for natural language processing in financial contexts.

Financial texts are inherently complex due to a variety of factors. They incorporate nuanced language and financial jargon, contain multiple layers of meaning, and express uncertainty, subtlety, or ambiguity. These complexities pose challenges to NLP models that attempt to accurately understand and interpret financial texts. Methods such as ours that can improve a model's ability to understand such complex language are critical and provide new strategies for addressing these challenges.

Furthermore, the study sheds light on the importance of customization in NLP models and highlights the importance of continuous fine-tuning and testing to improve model performance. It should be clear that while pre-trained models like FinBERT provide a solid foundation for text analysis, there is often room for task-specific improvements. By refining these models to better suit our specific tasks and data, we can realize their full potential, which can lead to more accurate and reliable results.

Moreover, the impact of our method goes beyond the immediate task of sentiment analysis. The problem of understanding complex sentences is common in various NLP applications in finance, such as information extraction, document summarization, and forecasting. By improving a model's ability to handle complex sentences, our method could potentially enhance the performance of these other applications as well.

4.3 Limitations of the Study

Although our study has made significant efforts to improve sentiment analysis in FOMC Minutes' sentences, we must acknowledge its limitations.

Firstly, the absence of a large, pre-labeled dataset for training and testing presented a challenge. We manually labeled a small data set for evaluation purposes but this approach could not ensure the completeness and diversity that a large, machine- or professionally expert-labeled dataset might provide.

Secondly, while the process of manual labeling was valuable for gaining context-specific understanding and insights, it also introduced the potential for human error or bias. Even with the utmost care, subjective judgment or inadvertent mistakes could have affected the quality of the labels.

In addition, we identified only a subset of instances where FinBERT made incorrect predictions. A more beneficial approach might have been to collect a complete list of all cases that were predicted incorrectly in a separate dataset for more detailed examination, problem analysis, and model refinement.

Furthermore, our study was limited in its scope of model testing. We used and fine-tuned FinBERT, but other models available in the NLP landscape were not evaluated. Therefore, our findings and methods might generalize to the performance of other models for the same or similar tasks.

As another limitation of our study, we must acknowledge the specificity of our data set. Our analysis was based only on the sentences of the FOMC Minutes, which are written in a particular style and structure. This specificity might limit the generalization of our results to all types of financial documents, which can significantly vary in terms of language use, length, structure, and complexity.

Moreover, our study addressed only one specific problem we found in sentiment analysis in financial texts, namely complexity and conflicting and contradicting sentiments after certain conjunctions. Other issues inherent to financial documents, such as financial lingo, remained unaddressed.

While the sentence simplification approach proved effective in this study, it might not be universally applicable to all types of financial texts. Different financial documents have their unique linguistic characteristics, and further research is needed to test the generalization of our method.

4.4 Future Work

The recognition of these limitations also offers opportunities for future research to expand upon our work. By addressing these constraints, we believe that continued research in this field of sentiment analysis for finance can yield even more promising results. Our research has built on the earlier groundwork by Kim et al. [5] and laid the foundation for several promising avenues of further research in the future.

4.4.1 Use of Novel Models: BloombergGPT. While our research has provided valuable insights into the potential of incorporating sentiment focus and fine-tuning the FinBERT model, which is a state-of-the-art domain-specific NLP model for finance, it also highlights the potential for exploring other advanced and novel model methods for sentiment analysis. Cutting-edge NLP models such as *GPT-4* and *BloombergGPT* in combination with different prompting techniques could be evaluated for their performance on similar tasks and might even produce better results, allowing a broader understanding of the strengths and weaknesses of different approaches to financial sentiment analysis.

4.4.2 Address Additional Problems: Financial Lingo. Secondly, our study only addressed one specific issue—the problem of complexity in sentences featuring conjunctions and contradicting sentiments. Future research could delve deeper into the various challenges present in financial texts, such as a better understanding of financial language and jargon, where our problem analysis showed that FinBERT also had difficulties. Scientists could, for example, fine-tune FinBERT with a large amount of financial corpus from the FOMC Minutes or similar financial texts that contain such expert language. We are convinced that a follow-up analysis with a more comprehensive and complete data set of FOMC Minutes' sentences would reveal more issues that could be addressed. Addressing these additional problems could significantly enhance the performance of NLP models like FinBERT in financial and task-specific contexts, providing more nuanced and accurate sentiment analysis capabilities.

4.4.3 Create Complete Data Set. Finally, future research could involve the professional development of a comprehensive and accurately labeled FOMC dataset. Our study was limited by the availability of a labeled data set. The manual labeling process was tedious work and extremely time- and energy-consuming. The creation of a larger, more robust dataset with all 32K or even more specifically selected individual sentences accurately labeled, could provide a much stronger foundation for training and testing a variety of different NLP models and methods. This could also involve research into exploring more objective and reliable processes and methods for dataset labeling. A more representative data set would not only enhance the reliability of individual studies but could also facilitate more comprehensive comparisons between different NLP models and techniques, as well as an overall faster research process.

5 CONCLUSION

This study examined the potential of refining FinBERT, a state-of-the-art NLP model, for sentiment analysis of FOMC Minutes. We have fine-tuned the *FinBERT-FOMC* model, a language model based on enhanced sentiment analysis of FOMC meeting minutes. It is more accurate than the original FinBERT for more complex financial sentences.

Our method demonstrated promising results in enhancing FinBERT's sentiment classification performance through fine-tuning and text simplification techniques such as sentiment focus. Despite the original FinBERT's commendable performance, our fine-tuned model was able to capture nuanced sentiments in complex sentences more accurately, achieving a notable increase in performance. So, our strategy of simplifying complex sentences by focusing on the core sentiment-bearing parts of the phrases has proven effective.

Although our research is limited by factors such as dataset size and manual labeling, and only targets a specific problem in a specific type of financial text, the results seem to be promising. Even though the implications of this study extend beyond FOMC Minutes to a broader financial context, it is important to highlight the importance of continuous testing, fine-tuning, and innovation of NLP models to better handle the complexity of the financial language.

Looking ahead, there are numerous opportunities for future research, including testing other advanced models for sentiment classification in financial texts, and developing more comprehensive and reliable labeled datasets.

The journey to perfecting sentiment analysis in financial texts, especially FOMC Minutes, is long. Still, this study marks a step forward in the right direction and opens the door for potential improvements in other financial text analysis tasks to provide better insights and drive the enhancement of decision-making processes in this society-critical field.

REFERENCES

[1] Dogu Araci. 2019. FinBERT: Financial Sentiment Analysis with Pre-trained Language Models. https://doi.org/10.48550/arXiv.1908.10063 arXiv:1908.10063 [cs]
[2] Arthur Böök. 2019. Sentiment Analysis on FOMC Statements. https://www.linkedin.com/pulse/sentiment-analysis-fomc-statements-arthur-b%C3%B6%C3%B6k/
[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (Minneapolis, Minnesota, 2019-06). Association for Computational Linguistics, Minneapolis, Minnesota, 4171–4186. https://doi.org/10.18653/v1/N19-1423
[4] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Computation* 9, 8 (11 1997), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735 arXiv:https://direct.mit.edu/neco/article-pdf/9/8/1735/813796/neco.1997.9.8.1735.pdf
[5] Wonseong Kim, Jan Frederic Spörer, and Siegfried Handschuh. 2023. Analyzing FOMC Minutes: Accuracy and Constraints of Language Models. arXiv:2304.10164 [cs] https://arxiv.org/abs/2304.10164
[6] Alec Radford and Karthik Narasimhan. 2018. Improving Language Understanding by Generative Pre-Training. https://api.semanticscholar.org/CorpusID:49313245
[7] Morgan Stanley. 2023. Our Patented Fed Sentiment Indicator. https://www.morganstanley.com/articles/mnlpfeds-sentiment-index-federal-reserve
[8] Raul Cruz Tadle. 2022. FOMC Minutes Sentiments and Their Impact on Financial Markets. *Journal of Economics and Business* 118 (2022), 106021. https://doi.org/10.1016/j.jeconbus.2021.106021
[9] Tomokuni Higano, Shuxin Yang, and Akio Sashida. 2023. Machine Learning and FOMC Statements: What's the Sentiment? https://blogs.cfainstitute.org/investor/2023/01/18/machine-learning-and-fomc-statements-whats-the-sentiment/
[10] Yifei Wang. 2023. Aspect-Based Sentiment Analysis in Document – FOMC Meeting Minutes on Economic Projection. https://doi.org/10.48550/arXiv.2108.04080 arXiv:2108.04080 [cs]
[11] Yi Yang, Mark Christopher Siy UY, and Allen Huang. 2020. FinBERT: A Pretrained Language Model for Financial Communications. https://doi.org/10.48550/arXiv.2006.08097 arXiv:2006.08097 [cs]
[12] Keane Ong Yang Wei, Andrew Walker, Yan Qi Ho, Shah Dhruv, and Jacob Pang Jiarong. 2022. NLP on FOMC Meetings. https://medium.com/@nusfintech.ml/nlp-on-fomc-meetings-50b48c447fe1