Comp 550 Final Project

Livesh Armoogum (260967119), Tevin Choi (260795752), Ahreum Lee (261054106)

Abstract

This research project investigates the performance of fine-tuned BERT models in comparison to classical machine learning models, including Support Vector Machines (SVM) and Random Forest, in the domain of financial sentiment analysis, with the expectation of convergent performance levels between two types of models: linear and non-linear. The study aims to uncover the strengths and limitations of each approach, considering factors such as computational efficiency, interpretability, and domain-specificity. The research contributes insights into selecting sentiment analysis models for financial applications and explores the trade-offs between state-of-the-art models and classical methodologies.

1 Introduction

The financial landscape is heavily influenced by market sentiment, making sentiment analysis a crucial component for informed decision-making. With the advent of advanced natural language processing (NLP) techniques, such as BERT (Bidirectional Encoder Representations from Transformers), there is a growing interest in understanding how these state-of-the-art models compare to traditional machine learning methods in the realm of financial sentiment analysis.

This research embarks on an exploration of sentiment analysis models by focusing on two paradigms: fine-tuned BERT models, known for their ability to capture complex linguistic patterns, and classical machine learning models, specifically Support Vector Machines (SVM) and Random Forest. The choice between these approaches is not merely an academic exercise but holds practical implications for industries reliant on timely and accurate sentiment predictions, such as finance.

The central research question driving this investigation is: **Do non-linear fine-tuned BERT models perform better compared to classical machine learning models in the context of financial sentiment analysis, and what factors influence the choice of model for predicting financial sentiment?** By addressing this question, we aim to provide valuable insights into the practical considerations that guide the selection of sentiment analysis models in financial applications.

2 Related Work

Previous research in sentiment analysis has explored various methodologies, ranging from traditional machine learning models to more advanced deep learning approaches. While the sentiment analysis domain has witnessed substantial progress, the intersection of sentiment analysis and financial markets remains a challenging and critical area. Here, we discuss two notable papers that contribute to our understanding of sentiment analysis in a financial context.

1. Title: "BERT for Stock Market Sentiment Analysis"

Authors: M. Gomes de Sousa, K. Sakiyama, L. S. Rodrigues et al.
Summary: The research conducted by Matheus Gomes de Sousa and his team focuses on addressing the challenges posed by the dynamic nature of the stock market and the need for rapid decision-making in response to breaking news. They propose leveraging Bidirectional Encoder Representations from Transformers (BERT) for sentiment analysis of financial news articles to provide timely and relevant information for stock market decision-making.

In their methodology, the authors pre-train the BERT model on a large general-domain document dataset through a self-learning task. The model is then fine-tuned on a manually labeled dataset of 582 stock news articles, categorized as positive, neutral, or negative sentiment. The fine-tuned BERT model achieves a notable F-score of 72.5/

The study goes beyond sentiment analysis and explores the practical application of the obtained model in predicting subsequent movements of the Dow Jones Industrial (DJI) Index. By conducting experiments, the authors demonstrate how the output of their model provides valuable insights for predicting market trends.

2. Title: "Stock Price Prediction using Sentiment Analysis and Deep Learning for Indian Markets"

Authors: N. Darapaneni, A. R. Paduri, H. Sharma, M. Manjrekar, N. Hindlekar et al.
Summary: The research conducted by Darapaneni et al. focuses on advancing stock market prediction through the integration of sentiment analysis and deep learning techniques, specifically applied to the Indian market context. The authors acknowledge the historical importance of stock market prediction and highlight the transformative impact of computing and machine learning on the pace and possibilities within this domain.

The study employs two distinct models to forecast future stock movements for selected Indian stocks, namely Reliance, HDFC Bank, TCS, and SBI. The first model utilizes Long Short-Term Memory (LSTM), leveraging historical stock prices as independent variables. In the second model, a Random Forest approach is employed, incorporating sentiment analysis data obtained through an Intensity Analyzer. Notably, macroeconomic parameters such as Gold and Oil prices, USD exchange rates, and Indian Government Securities yields are introduced to enhance the accuracy of the Random Forest Model.

The research stands out by combining traditional financial indicators with sentiment analysis, recognizing the impact of market sentiments on stock prices. The inclusion of macroeconomic factors further extends the scope of prediction models, showcasing a comprehensive approach.

Distinguishing Features of Our Work

While the aforementioned studies contribute significantly to the broader field of sentiment analysis in financial contexts, our work distinguishes itself in the following ways:

- **Comparative Focus:** Unlike studies that predominantly explore one type of model, our research systematically compares fine-tuned BERT models with classical machine learning models, providing a nuanced understanding of their relative performance in financial sentiment analysis.
- **Comprehensive Evaluation:** Our study incorporates a comprehensive evaluation of sentiment analysis models, addressing not only accuracy but also practical considerations that guide the selection of models in real-world financial scenarios.

By building upon the existing literature and focusing on these distinguishing features, our research aims to contribute valuable insights to the evolving field of sentiment analysis in financial domains.

3 Methodology

The code for all models are on our *colab file*. In the sentiment analysis of news headlines, we performed data preprocessing before training any of the models by tokenizing, removing stopwords, and applying TF-IDF vectorization.

3.1 Dataset

For the sentiment analysis experiments, we utilized the "Text Classification Financial News" dataset obtained from Kaggle. This dataset comprises a collection of financial news articles labeled with sentiment categories, making it suitable for our task. The dataset contains [specify the number of samples] samples, which were divided into training and testing sets.

- **Source:** Kaggle
- **Number of Classes:** 3 (Negative, Neutral, Positive)
- **Number of samples:** 4846
- **Train/Test Split:** 80/20

3.2 Fine-Tuned BERT Model

We performed fine-tuning on a pre-trained BERT model for sentiment analysis. The BERT model was tokenized using the BERT tokenizer from the Hugging Face Transformers library. To optimize its performance for our specific sentiment analysis task, we conducted hyperparameter tuning experiments, with a focus on the learning rate and batch size.

- **Pre-Trained Model:** bert-base-uncased

- **Tokenizer:** BERT Tokenizer (from Hugging Face Transformers)

- **Hyperparameter Tuning:**

  – Learning Rate: Explored different values, 2e-5, 1e-4, 1e-3, 1e-2.
  – Batch Size: Experimented with batch sizes such as 16, 32, and 64.

3.3 Non-Linear Models

Multi-Layer Perceptron (MLP)
For the MLP-based sentiment analysis model, we implemented a custom MLP class using the PyTorch library. The architecture allows flexibility in the number of hidden layers. Three hyperparameter tuning experiments were conducted to optimize the model.

- **Hyperparameter Tuning:**

  – Learning Rate: Explored learning rates of 0.0001, 0.001, 0.01, and 0.1.
  – Number of Hidden Layers: Investigated different numbers of hidden layers, including 1, 2, and 3.
  – Number of Epochs: Varied the number of training epochs, considering values such as 5, 10, and 15.

Random Forest
In addition to deep learning models, we explored the effectiveness of traditional machine learning using the Random Forest algorithm from the scikit-learn library.
We performed hyperparameter tuning on $n\_estimators$, which corresponds to the number of trees in the Random Forest, with a wide range of values 25, 50, 100, 200, 300, 400, 500, 600, 700.

3.4 Linear Models

For the linear models, we trained a Multinomial Naive Bayes classifier on TF-IDF vectors, fine-tuning the Laplace smoothing parameter (alpha) for optimal performance, using gridSearch. The Support Vector Machine (SVM) model was trained using the same TF-IDF vectors, with experimentation on kernel functions and regularization parameter (C). The evaluation of models was measured through the calculation of accuracy.

4 Results

In general, all models performed well with accuracy higher than 0.70. The fine-tuning steps led to higher accuracy in all models except for the SVM, thus all measures below are of the tuned models, except for the SVM, for which we grabbed the accuracy value of the default settings.

Models	**Accuracy (%)**
Fine-tuned BERT	82.58
SVM	75.13
Random Forest	73.71
MLP	72.47
Naive Bayes	70.03

Table 1: Impact of different models on Test Accuracy

Fine-Tuned BERT:
The fine-tuned BERT model demonstrated the highest accuracy among all models, achieving an accuracy of 82.58%. The BERT model's capacity to capture intricate relationships within the financial news texts contributed to its superior performance.
Support Vector Machine (SVM):
The SVM model exhibited a competitive accuracy of 75.13%. SVMs are known for their effectiveness in high-dimensional spaces, and their performance in sentiment analysis aligns with this reputation.
Random Forest:
The Random Forest model achieved an accuracy of 73.71%. While Random Forests are robust and versatile, they slightly trailed behind the fine-tuned BERT and SVM models in this specific sentiment analysis task.
Multi-Layer Perceptron (MLP):
The MLP model, implemented using PyTorch, demonstrated an accuracy of 72.47%. Despite its simplicity compared to deep transformer models, the MLP showcased competitive performance in sentiment classification.
Naive Bayes:
The Naive Bayes model achieved an accuracy of 70.03%. While Naive Bayes models are known for their simplicity and efficiency, their performance in sentiment analysis might be more pronounced in specific contexts.

5 Discussion and conclusion

The sentiment analysis experiments conducted on the Kaggle financial news dataset provided valuable insights into the performance of various machine learning models. The findings shed light on the strengths and weaknesses of each model, offering a basis for informed decision-making in sentiment analysis tasks.

Fine-Tuned BERT's Dominance: The standout performer in our experiments was the fine-tuned BERT model, achieving an accuracy of 82.58%. This result aligns with the model's reputation for capturing nuanced language patterns, especially in the context of financial news sentiment analysis.

SVM as a Strong Competitor: The Support Vector Machine (SVM) demonstrated robust performance with an accuracy of 75.13%. SVMs, known for their effectiveness in high-dimensional spaces, proved to be a strong competitor in this sentiment analysis task.

Competitive Alternatives: Random Forest, MLP, and Naive Bayes, while trailing slightly behind the top performers, showcased competitive accuracy values (73.71%, 72.47%, and 70.03%, respectively). These models offer viable alternatives, with their relative advantages depending on the specific requirements of the sentiment analysis application.

Our initial hypothesis, which posited that fine-tuning a pre-trained BERT model would lead to superior sentiment analysis performance, was verified by the experiments. The fine-tuned BERT model outperformed other models, emphasizing the efficacy of leveraging deep contextual embeddings for sentiment analysis in financial news.

Considering the experimental results and observations, several factors influence the choice of model for predicting financial sentiment:

- **Data Complexity:** In scenarios where financial news exhibits complex language patterns, deep learning models like BERT may be preferred for their ability to capture intricate relationships.

- **Interpretability:** Traditional machine learning models such as SVMs and Random Forests, while slightly trailing in accuracy, offer interpretability and transparency, crucial factors in financial decision-making.

- **Resource Constraints:** The choice of model may be influenced by resource constraints. Lighter models like Naive Bayes or simpler MLPs could be more suitable in resource-limited environments.

5.1 Conclusion

The sentiment analysis experiments performed on the Kaggle financial news dataset revealed BERT to be the best-performing model, followed by linear SVM, emphasizing its strength in high-dimensional spaces. Our findings highlight the importance of considering factors such as data complexity, interpretability, and resource constraints when choosing a sentiment analysis model, recognizing that the optimal choice may vary based on the specific requirements of the financial sentiment analysis task.

5.2 Limitations

BERT's success can be attributed to its ability to capture intricate linguistic relationships and contextual nuances, making it well-suited for sentiment analysis tasks requiring deep semantic understanding. On the other hand, SVM's robustness across different datasets makes it also a strong performer, especially in scenarios where the decision boundary is relatively linear. However, the effectiveness of these models is contingent on the task's complexity and the dataset's characteristics, and there are trade-offs to their performance. While BERT and SVM excelled in this particular sentiment analysis task, there are scenarios where they might not be the optimal choices. For instance, in tasks with limited computational resources, BERT's demands may become prohibitive and thus it is better to stick with a simpler model. Similarly, when faced with highly nonlinear relationships or tasks requiring an understanding of complex linguistic patterns, it may be better to use a non-linear model instead of the linear SVM.

In the dynamic landscape of sentiment analysis, particularly evident in social media contexts, results can deviate notably from conventional datasets. Social media posts introduce distinctive challenges, featuring character limitations, informal language, and a profusion of emojis and hashtags. Navigating these intricacies demands specialized preprocessing techniques. As we delve into the nuances of sentiment analysis, it becomes apparent that addressing these challenges is essential for ensuring accurate analyses, especially when transitioning from more structured datasets.

6 Statement of contributions

The success of this project can be attributed to the collaborative efforts of all team members—Livesh, Tevin, and Ahreum.

Each member contributed to one model; key contributions of: Livesh to BERT and Random Forest, Tevin to MLP, and Ahreum to the linear models. All team members actively participated in the preparation of the project report.

7 Figures

	Sentiment	Text	Encoded_Sentiment
0	neutral	According to Gran , the company has no plans t...	1
1	neutral	Technopolis plans to develop in stages an area...	1
2	negative	The international electronic industry company ...	0
3	positive	With the new production plant the company woul...	2
4	positive	According to the company 's updated strategy f...	2

Figure 1: Dataset view

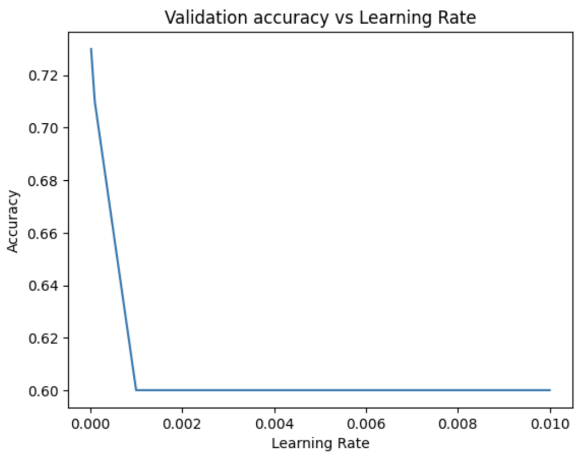

Figure 2: Validation Accuracy vs Learning Rate for BERT

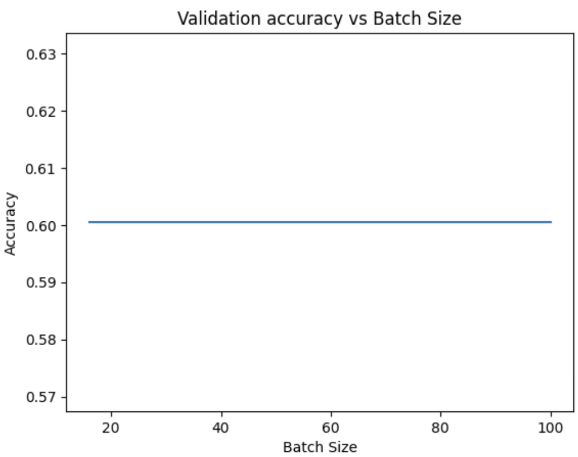

Figure 3: Validation Accuracy vs Batch Size for BERT

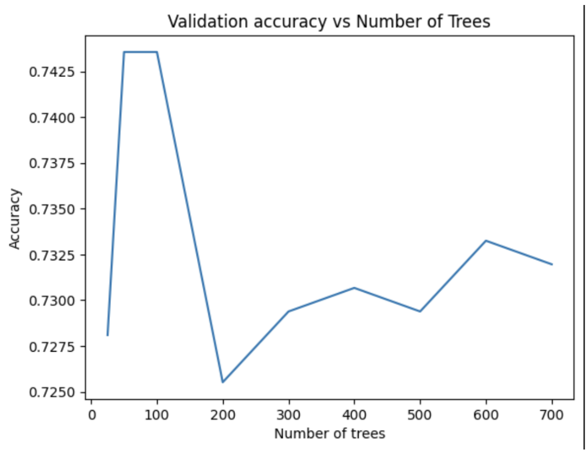

Figure 4: Validation Accuracy vs Number of Trees for Random Forest

References

1. **BERT Model:**
   Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805.
   https://arxiv.org/abs/1810.04805

2. **Multilayer Perceptron (MLP):**
   Sklearn MLP Documentation.
   https://scikit-learn.org/stable/modules/neural_networks_supervised.html

3. **Random Forest:**
   Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.
   https://link.springer.com/article/10.1023/A:1010933404324

4. **Support Vector Machine (SVM):**
   Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine learning, 20(3), 273-297.
   https://link.springer.com/article/10.1007/BF00994018

5. **Naive Bayes:**
   Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to Information Retrieval. Cambridge University Press.
   https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf

6. **BERT for Stock Market Sentiment Analysis:**
   M. Gomes de Sousa, K. Sakiyama, L. S. Rodrigues et al. FACOM/UFMS.
   https://www.researchgate.net/publication/339286476_BERT_for_Stock_Market_Sentiment_Analysis

7. **Stock Price Prediction using Sentiment Analysis and Deep Learning for Indian Markets**
   N. Darapaneni, A. R. Paduri, H. Sharma, M. Manjrekar, N. Hindlekar et al. Northwestern University/Great Learning.
   https://www.researchgate.net/publication/359918053_Stock_Price_Prediction_using_Sentiment_Analysis_and_Deep_Learning_for_Indian_Markets