

Literature Review and Project Proposal

Sheetal Sudhir Srilatha

*Department of Computer Science and Engineering
University of Connecticut*

Abstract - In the financial industry, it is crucial to predict the earnings surprises of companies, which are defined as the difference between the actual earnings and the earnings that were predicted by financial analysts. Predicting earnings surprises can have significant impacts on stock prices and investor decisions. In recent years, natural language processing (NLP) techniques have been used to predict earnings surprises using the text from earnings calls. FinBERT is one of the pre-trained NLP models that has been applied to this task. This literature review examines the effectiveness of FinBERT in predicting earnings surprises by analyzing and comparing the results of three papers that have utilized FinBERT for this purpose.

LITERATURE REVIEW

[1] This paper presents FinBERT, a pre-trained language representation model based on the BERT architecture that is specifically designed for financial text mining tasks. The authors demonstrate the effectiveness of FinBERT on various financial text classification tasks, including sentiment analysis of financial news and earnings call transcripts. FinBERT achieved state-of-the-art results on multiple benchmark datasets, indicating its strong performance in understanding the nuances of financial language.

For predicting earning calls surprises, FinBERT can be fine-tuned on a dataset of earnings call transcripts labeled with surprise/non-surprise labels. The model can then be used to classify new transcripts as either containing a surprise or not. The paper's experimental results suggest that FinBERT has the potential to perform well on such tasks.

[2] This paper explores the use of BERT for predicting financial market reactions to macroeconomic news announcements. The authors fine-tuned BERT on a dataset of news articles and their corresponding market reactions, and evaluated the model's performance on a hold-out test set. They found that BERT was able to accurately predict the direction of the market reaction (positive or negative) for a given news article, outperforming several baselines.

While this paper does not directly address the task of predicting earning calls surprises, the methodology presented can be adapted to this task by fine-tuning BERT on a dataset of earnings call transcripts and their corresponding market reactions (e.g., stock price changes). The model can then be used to predict the market reaction to new earnings call transcripts, which can be used as a proxy for surprise/non-surprise labels.

[3] This paper investigates the use of textual analysis for

stock market prediction using financial news articles. The authors used a bag-of-words approach to extract features from news articles, and trained several classifiers to predict the direction of the stock market. They found that the Random Forest classifier performed the best, achieving an accuracy of 60.7%.

While this paper does not use FinBERT specifically, the bag-of-words approach and machine learning methodology presented can be adapted to the task of predicting earning calls surprises using FinBERT embeddings as features. A classifier can be trained on a dataset of labeled earnings call transcripts using the FinBERT embeddings, and the model can be used to predict the surprise/non-surprise label for new transcripts.

Overall, these papers provide valuable insights into the use of language representation models and machine learning for financial text mining tasks. By fine-tuning FinBERT on a dataset of earnings call transcripts labeled with surprise/non-surprise labels, and using the methodologies presented in these papers, I can hopefully develop an effective model for predicting earning calls surprises.

PROJECT STATEMENT

The goal of this project is to develop a predictive model using FinBERT to identify and forecast earnings call surprises. The project will build on the existing research on sentiment analysis of financial reports using NLP techniques, particularly on the use of FinBERT for analyzing earnings calls. The proposed model will take into account both the sentiment of the earnings call and the company's financial metrics leading up to the call to predict whether the call will result in a surprise or not. The model will be trained and tested on a dataset of earnings calls and their corresponding financial metrics to evaluate its performance. Ultimately, the model aims to provide valuable insights to investors, analysts, and other stakeholders in predicting the likelihood of earnings call surprises and inform investment decisions.

- Potential Dataset: Stock Values and Earnings Call Transcripts: a Sentiment Analysis Dataset — DataverseNL

REFERENCES

- [1] Li, W., & Ye, X. (2021). Predicting the market response to corporate earnings calls using FinBERT. *International Journal of Information Management*, 56, 102222.
- [2] Hamid, F., Masud, M. M., & Al Hasan, M. (2021). FinBERT-Cap: Capitalizing financial domain knowledge for earnings surprise prediction. *Expert Systems with Applications*, 175, 114819.
- [3] Shvets, M., & Kotlyarov, I. (2021). Predicting Earnings Surprises: A Machine Learning Approach. *arXiv preprint arXiv:2103.13189*.