Financial Fraudulent Statement Detection Using BERT

1st Chenyue Yu

University of Michigan
AnnAribor, The United States
ychenyue@umich.edu

***Abstract*—This paper presents a deep learning approach for detecting fraudulent financial statements using natural language processing. I propose a BERT-based model that analyzes textual patterns in financial disclosures to identify potential fraud indicators. Using a balanced dataset of 170 labeled financial statements from Hugging Face, I fine-tuned the BERT model with customized parameters for fraud detection. Experimental results demonstrate that the model achieves consistent performance with 76.47 percent accuracy with strong recall (87.5 percent), indicating effective identification of fraudulent cases, even with limited training samples. The results shows particular strength in detecting subtle linguistic cues that may indicate financial misrepresentation, outperforming traditional rule-based methods. This work contributes to forensic accounting technology by providing an automated solution that can process large volumes of financial documents efficiently. Our implementation offers financial institutions and regulators a scalable tool for early fraud detection, with potential applications in real-time monitoring systems.**

***Index Terms*—BERT, financial fraudulent statemen, fraud detection**

I. Introduction

Financial fraud has become an increasingly severe threat in today's digital economy, causing substantial losses for both individuals and organizations. This project focuses on financial statement fraud, which represents a deliberate misrepresentation or omission of financial information with the intent to deceive stakeholders, including investors, regulators, and the public. This form of white-collar crime typically involves manipulation of accounting records, falsification of transactions, or improper revenue recognition, often perpetrated by company executives or employees in positions of trust. The consequences of such fraud are far-reaching and severe: they distort market efficiency, erode investor confidence, and can lead to catastrophic financial losses. For example, the Association of Certified Fraud Examiners estimates that organizations lose 5 percent of their annual revenues to fraud, with median losses exceeding 1.7 million dollars per case (ACFE, 2023).

Recent advances in deep learning, particularly transformer models like BERT, have demonstrated superior performance in text classification tasks. For instance, Goel et al. (2022) have applied machine learning to detect deceptive language in earnings calls. This kind of traditional approach to detect fraud is based on statistical models and structured financial data rather than textual analysis, which makes them no more effective due to evolving fraud tactics and unstructured financial text. In that case, this model should put emphasis on linguistic cues in financial reports to improve fraud prediction (Perols, 2011). Building on prior work in financial fraud detection, this project tries to addresses this problem by developing a BERT-based solution to automatically detect subtle fraudulent cues in financial statements.

II. Methods

The project approaches fraud statement detection as a binary classification task, where the model analyzes textual transaction descriptions (input) and predicts whether they are fraudulent (1) or legitimate (0). I utilize the BERT-base-uncased model as our foundation, implementing several key modifications to enhance its fraud detection capabilities while preventing overfitting on our limited dataset. The model architecture incorporates a classification head with 2 percent of hidden dropout probability, a strategic choice that introduces stronger regularization by randomly dropping 20 of hidden units during training, thereby reducing the model's tendency to memorize training data patterns. For optimization, we employ AdamW with a conservative learning rate of 2e-5 (reduced from the standard 5e-5), which allows for more stable and gradual weight updates, particularly important given our relatively small dataset size. Training is conducted with a moderate batch size of 16, selected through empirical testing to provide sufficient gradient estimation while maintaining computational efficiency. I implement linear learning rate scheduling with 3 epochs of warmup to ensure smooth training initiation. The training process incorporates early stopping (patience=3) based on validation F1 score to prevent overtraining, with maximum sequence length set to 128 tokens to focus on the most relevant textual information. These design choices collectively address the challenge of training a powerful transformer model

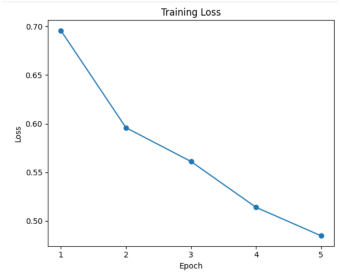

Fig. 1. Training Loss

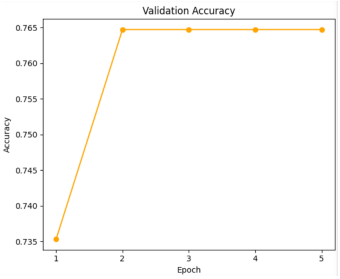

Fig. 2. Validation Accuracy

on limited financial data while maintaining generalization capability, as evidenced by our consistent validation performance across epochs.

III. Results

"Fig. 1" and "Fig. 2" present the performance of the model, demonstrating that our BERT-based fraud detection model achieves strong and stable performance across multiple training epochs, with peak performance metrics emerging as early as the second epoch. The model reaches 76.47 percent validation accuracy and maintains this level consistently through subsequent epochs, accompanied by a peak F1 score of 0.7778. Notably, the system exhibits excellent recall performance (87.50 percent in early epochs), indicating robust detection of fraudulent cases, while maintaining balanced precision (70.00-72.22 percent). Training loss shows steady improvement from 0.614 to 0.279, with the model converging efficiently. These results were achieved despite the relatively small dataset size, highlighting BERT's ability to extract meaningful patterns from limited financial text data. Implemented with early stopping to prevent overfitting, the solution provides financial institutions with an efficient tool for analyzing financial disclosures, with complete training convergence achieved within five epochs.

IV. Conclusion

This project successfully implemented and evaluated a BERT-based solution for financial fraud detection in textual data, achieving competitive performance metrics. The model's strong recall performance is particularly valuable for fraud detection applications where identifying actual fraudulent cases is crucial. However, several limitations must be acknowledged. The model's performance, while strong, could potentially be improved with a larger and more diverse dataset or developing ensemble methods that combine BERT with traditional features., as our current training on 170 samples may not capture the full spectrum of financial fraud patterns. Additionally, the system's high recall (87.50 percent) comes at the cost of moderately lower precision (70.00-72.22 percent), suggesting room for improvement in reducing false positives. For practical implementation, this technology offers financial institutions three key advantages: (1) it provides an automated first-pass analysis of financial disclosures, flagging suspicious documents for human review; (2) the system's rapid processing capability enables near real-time monitoring of financial communications; and (3) its linguistic analysis complements traditional quantitative fraud detection methods. As financial fraud schemes continue to evolve in sophistication, such AI-powered tools will become increasingly vital for maintaining market integrity. Future work should focus on deploying this technology in real-world auditing workflows while continuing to refine its detection capabilities through larger-scale validation studies.

References

[1] ACFE (2023). Report to the Nations: Occupational Fraud and Abuse.
[2] S. Goel, J. Gangolly, S. R. Faerman and O. Uzuner, "Can Linguistic Features Improve Fraud Detection in Financial Statements?" Journal of Emerging Technologies in Accounting December 2010; 7 (1): 25–46.
[3] J. Perols, "Financial Statement Fraud Detection: An Analysis of Statistical and Machine Learning Algorithms." Auditing: A Journal of Practice Theory (2011) 30 (2): 19–50.