Leveraging Large Language Models and Deep Learning for Detecting Illegal Insider Trading

Anoop Adusumilli
Department of Computer Science
Rutgers University, Camden
Camden, USA
anoopadu200@gmail.com

Sheikh Rabiul Islam
Department of Computer Science
Rutgers University, Camden
Camden, USA
sheikh.islam@rutgers.edu

Iman Dehzangi
Department of Computer Science
Rutgers University, Camden
Camden, USA
i.dehzangi@rutgers.edu

June Kim
Department of Accounting
Rutgers University, Camden, USA
jk1529@camden.rutgers.edu

2024 IEEE International Conference on Big Data (BigData) | 979-8-3503-6248-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/BigData62323.2024.10825461

***Abstract*— Illegal insider trading undermines the integrity of financial markets by exploiting non-public information for personal gain, posing risks to market fairness and investor trust. Addressing this issue is crucial to ensure transparency and maintain investor confidence. Estimates indicate that illegal insider trading occurs in about 20% of merger and acquisition events and 5% of quarterly earnings announcements. Our research introduces a deep learning-based approach that combines text classification, time series forecasting, anomaly detection and explainable AI to detect and predict potential illegal insider trading. By analyzing litigation related press releases and incorporating market and social media data, we develop a method to identify suspicious trading patterns. Our findings demonstrate that the proposed approach is promising in detecting potential instances of illegal insider trading, uncovering subtle patterns and anomalies. These insights offer financial regulators and institutions an enhanced surveillance mechanism, contributing to ongoing efforts to safeguard market fairness and reliability.**

Keywords—Large Language Models, Insider Trading, Timeseries Forecasting, Transformers, Explainable AI

I. Introduction

Insider trading can either be legal or illegal, based on the specific conditions surrounding the execution of the trades. Legal insider trading occurs when corporate insiders, such as executives or employees, buy or sell stock in their own companies in compliance with regulatory requirements. These trades are reported to agencies like the Securities and Exchange Commission (SEC) and made available to the public. On the other hand, illegal insider trading happens when individuals trade stocks based on non-public, material information, giving them an unfair advantage over other investors. This practice undermines market fairness and erodes investor confidence, as it distorts stock prices and creates an uneven playing field. Illegal insider trading is a pervasive issue that affects a significant number of merger and acquisition events, earnings reports, and other market-shifting announcements [1]. Ensuring a level playing field for all investors is essential for maintaining the stability and reliability of these markets. The covert nature of this activity makes it difficult to detect, especially as traditional systems struggle to identify the complex, nonlinear patterns in financial data that indicate potential misconduct.

Traditional methods for identifying illegal insider trading often fall short in capturing these subtle patterns, necessitating the development of more sophisticated approaches. Recent advancements in deep learning offer new avenues for addressing these challenges, particularly through the analysis of large volumes of structured and unstructured data.

Our study introduces a deep learning-based approach designed to detect and predict illegal insider trading with greater accuracy and transparency. A key innovation in our work lies in the integration of diverse data sources—litigation releases from regulatory bodies, stock volume, options volume, and social media data—into a comprehensive framework. By incorporating sentiment data from Twitter alongside stock and options data, we enrich the analysis, allowing for a more comprehensive analysis of market dynamics and potential insider trading activities. To the best of our knowledge, this is the first study to synergize these varied data sources with advanced deep learning models, including TimesNet [2] and transformer-based models for time series forecasting, in the context of insider trading detection. Furthermore, the addition of explainable AI techniques such as SHAP and LIME offers a valuable contribution, enhancing the interpretability and transparency of our models. This is important as it helps analysts gain a deeper understanding of the decision-making processes behind the predictions. In the following sections, we first present the literature review, discussing previous works on the detection of illegal insider trading. We then outline our methodology, describe the dataset, and explain the data processing techniques applied. Finally, we detail the experiments conducted, analyze the results, and offer insights and recommendations for future work.

II. Background

Illegal insider trading detection has made substantial progress with the application of machine learning techniques. Islam et al. [3] introduced a deep learning approach using LSTM RNNs for forecasting stock volumes and tree-based visualizations for data interpretation. This method represents a significant advance in proactively identifying insider trading patterns. Our work builds on this by incorporating state-of-the-art models like BERT for text classification, TimesNet and Transformer models for time series forecasting, along with options volume data and social media data.

Lauar et al. [4] employed the XGBoost algorithm to predict whether impactful news events would be disclosed within the following days. Their model utilized features such as the number of trades, largest trade size, total volume traded, and asset returns. The authors aimed to help regulators focus their limited resources on the most promising cases. Their models showed significant improvements over dummy classifiers, including stratified classifiers, which mimic the class distribution in the training data, and uniform classifiers, which make random predictions with equal probability for each class.

Tallboys [5] tackled market manipulation detection using an LSTM-based method and TadGAN (Generative Adversarial Network). They analyzed real-world datasets where market manipulation was suspected, using LSTM to identify contextual and local anomalies. The dynamic thresholding approach worked particularly well for detecting anomalies in time series data, while TadGAN was less effective due to parameter tuning challenges. ARIMA, a statistical model, showed faster results in some cases but struggled with detecting contextual anomalies. The study concluded that deep learning techniques like LSTM, when combined with statistical models, could provide a more effective solution for anomaly detection in financial markets.

Chen et al. [6] applied logistic regression to detect insider trading by modeling the probability of illegal activity based on features such as historical price movements, trading volumes, and other market indicators. Their model was trained on labeled datasets where instances of insider trading were identified, allowing it to detect patterns associated with suspicious trading activity. However, one limitation of their approach is that logistic regression assumes a linear relationship between the features and the probability of insider trading. This assumption can miss more complex, nonlinear patterns in the data, which are often present in real-world financial markets. Additionally, their model was not designed to adapt to changes in market dynamics over time, which can be crucial in detecting evolving insider trading strategies.

Seth et al. [7] introduced a multi-stage framework combining deep neural networks (DNN) with consensus models and statistical techniques to detect illegal insider trading events. The first stage used a DNN to generate initial predictions, which were then refined by a consensus model—a method that aggregates predictions from multiple models to reach a final decision, thereby reducing errors from individual models. A distance-based scoring system ranked the predictions to improve accuracy. Although this approach effectively combined deep learning with traditional statistical methods, it was constrained by its reliance on predefined features and fixed scoring mechanisms. This rigidity could limit the model's ability to adapt to evolving market conditions or detect more complex patterns in insider trading behavior over time.

James et al. [8] introduced a novel surveillance model combining the Near-Nearest Neighbor Dynamic Time Warping (NN DTW) algorithm with Extreme Value Theory (EVT). This model effectively identified abnormal trading sequences in broker-specific high-frequency order book data. NN DTW allowed for the detection of misaligned trading patterns by adjusting the time series alignment to capture subtle irregularities in trading behavior. EVT was applied to the upper tail of anomaly scores to set dynamic thresholds for distinguishing between normal and illegal trades. This approach achieved a good success rate in detecting suspected insider trading without requiring historical examples of fraud, while maintaining a lower false positive rate compared to traditional models. However, the model's reliance on high-frequency, broker-specific data limits its generalizability to contexts where such data is unavailable, and the sensitivity of EVT thresholds may pose challenges during periods of extreme market volatility.

Rizvi et al. [9] proposed an unsupervised model based on Kernel Principal Component Analysis (KPCA) to detect insider trading. KPCA non-linearly transformed stock price and trade data into higher dimensions, where trades were clustered into normal and suspicious categories using multi-dimensional kernel density estimation (MKDE). By automating parameter selection and removing the need for labeled data, the model enhanced detection rates while reducing false positives. Compared to other methods such as PCA, GMM, and OCSVM, the KPCA-MKDE approach demonstrated superior performance, especially in identifying subtle and complex manipulative trading patterns. However, a limitation of this approach is its sensitivity to the choice of kernel and bandwidth parameters in KPCA, which can impact detection accuracy, particularly in the presence of noisy or high-dimensional data.

Liu et al. [10] developed a support vector machine (SVM) model combined with Borderline Synthetic Minority Oversampling Technique (SMOTE) to detect stock market manipulation using China Securities Regulatory Commission punishment cases. They found that SVM outperformed logistic regression, and Borderline SMOTE significantly improved detection of manipulated stocks. Incorporating market sentiment indicators from analyst reports, financial news, and social media further enhanced accuracy. The study demonstrated the potential of machine learning for regulatory monitoring of market manipulation.

Zhang et al. [11] introduced TEANet, a Transformer Encoder-based Attention Network that integrates social media data from Twitter with stock prices to predict stock movements. The model addressed the challenge of temporal dependence in financial data while effectively merging diverse data sources. TEANet's attention mechanisms enabled it to outperform both traditional models and other Transformer-based methods. This approach parallels our work, as both studies utilize Transformer models to predict stock-related outcomes—stock price movements in their case—stock volume in our case, and —demonstrating the value of integrating diverse data sources like social media for financial analysis.

In summary, while the existing literature demonstrates significant advancements in using machine learning for insider trading detection, a gap remains in fully integrating heterogeneous data sources such as social media sentiment and options volume. Many approaches focus on singular data types or rely on less sophisticated models, limiting their adaptability in complex and evolving market conditions. Our research aims to fill this gap by employing cutting-edge deep learning models, including transformer-based architectures and TimesNet, to analyze a comprehensive dataset of litigation releases, social media data, and transaction volumes. This integration enhances the accuracy and robustness interpretability of insider trading detection.

III. METHODOLOGY

This section presents the flowchart (Fig. 1) depicting the methodology, followed by descriptions of the three main stages: the litigation classifier, timeseries forecasting, and anomaly detection.

A. Litigiation Classifier

The initial phase involved collecting litigation-related press releases from multiple sources, including the U.S. Securities and Exchange Commission (SEC) [12], Federal Bureau of Investigation (FBI) [13], and the US District Court

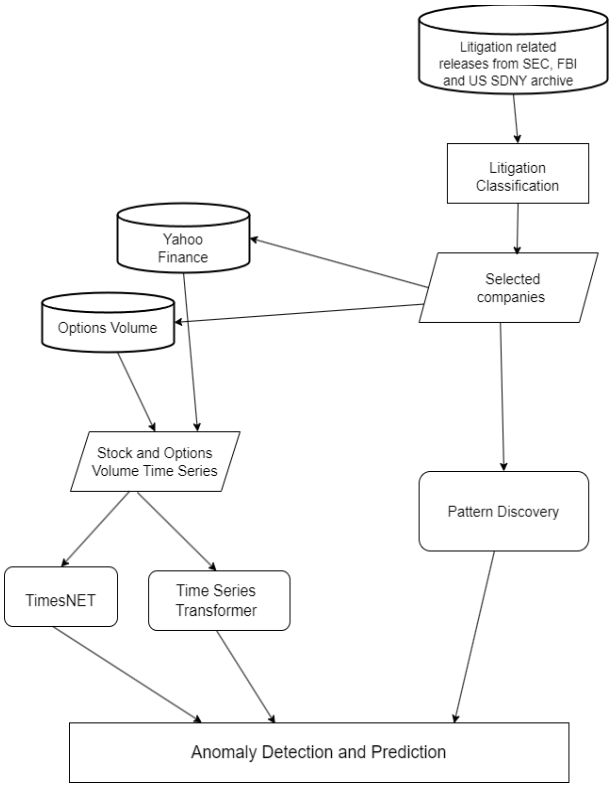

Fig. 1. Flowchart of the proposes approach

for the Southern District of New York (SDNY) [14]. This comprehensive dataset improved the model's ability to accurately determine whether a case involved insider trading, allowing it to better differentiate the key factors that distinguish insider trading cases from other types of financial misconduct. We employed transformer models, such as BERT (Bidirectional Encoder Representations from Transformers) and ELECTRA (Efficiently Learning an Encoder that Classifies Token Replacements Accurately), to classify these litigation releases as either 'insider' or 'non-insider' events.

Transformer models are particularly effective for text classification tasks due to their ability to process entire sequences of words simultaneously using self-attention mechanisms. This allows the model to better understand the relationships and context within the text. Explainable AI techniques such as SHAP (SHapley Additive exPlanations) [15] and LIME (Local Interpretable Model-agnostic Explanations) [16] were employed to provide insight into the model's decision-making process. These tools helped us determine how different features of the press releases, such as timing, individuals involved, and the nature of disclosed information, contributed to classifying a case as related to insider trading.

B. Timeseries Forecasting

In this phase, we selected companies involved in prominent cases of illegal insider trading (e.g., American Semiconductor Corporation, Palo Alto Networks). We then collected day-wise historical stock volume and options volume data for these companies and forecasted future volumes. For this task, we used TimesNet, a state-of-the-art deep learning model designed for general time series analysis, as well as a Transformer-based model for long-range dependencies.

TimesNet employs a novel approach that transforms 1D time series data into multiple 2D representations based on discovered periodicities, making it particularly effective for capturing short to medium-range temporal patterns in time series data. The model's core component, TimesBlock, can capture both intraperiod and interperiod variations simultaneously. Intraperiod variations represent short-term fluctuations within a single period (e.g., daily changes within a week), while interperiod variations capture longer-term trends across multiple periods (e.g., week-to-week or month-to-month patterns). This ability to capture multi-scale temporal dynamics is crucial for understanding complex patterns in stock and options volume data.

The process begins with Fast Fourier Transform (FFT) to identify significant frequencies in the time series, which are then used to reshape the 1D series into multiple 2D tensors. For example, consider a 1D time series of daily stock volumes over a trading year (approximately 252 days). TimesNet might discover weekly and monthly periodicities in this data. It would then reshape the 1D series into multiple 2D representations - one could be a 5x51 matrix (with some padding) representing weekly patterns across the year, and another could be a 21x12 matrix capturing monthly trends. This allows the model to simultaneously analyze short-term weekly fluctuations and longer-term monthly patterns in stock volumes. The model processes these 2D tensors using parameter-efficient inception blocks with multi-scale 2D kernels. This approach allows TimesNet to detect patterns more efficiently than traditional 1D approaches, particularly for short to medium-term dependencies. In the reshaped tensors, columns typically represent intraperiod variations, while rows capture interperiod variations, enabling a comprehensive analysis of temporal patterns at different scales.

One of TimesNet's key advantages is its ability to handle variable-length input sequences flexibly, which is particularly beneficial when analyzing historical data of varying lengths for different companies. By considering time series as sets of 2D tensors, TimesNet can identify local and global patterns within its effective range, as well as relationships between different time points and periods. This enhanced representational power for short to medium-range dependencies, combined with its adaptive period discovery and multi-scale analysis, enables TimesNet to make accurate and reliable predictions for near to medium-term forecasting tasks in stock and options volume data.

For capturing long-range dependencies, we employed HuggingFace's Time Series Transformer model [17], a variant of the Transformer model specifically designed for time series data. Unlike traditional time series models or recurrent neural networks that may struggle with long-term dependencies, Transformers can directly attend to any part of the input sequence, regardless of the distance between time steps.

The key to Transformers' success in handling long-range dependencies lies in their self-attention mechanism [18]. This mechanism computes pairwise similarities between all time steps, allowing the model to capture complex temporal patterns and interactions across the entire time series. This is especially valuable in financial time series, where long-term trends and cyclical patterns can significantly influence future behavior.

The HuggingFace Time Series Transformer model adapts the vanilla transformer architecture for probabilistic time series forecasting. Its encoder-decoder structure processes historical data to generate a distribution of possible future outcomes, rather than point estimates. This probabilistic approach is particularly valuable in the volatile stock and options markets, offering a nuanced view of potential future scenarios. By providing a range of outcomes with associated probabilities, the model accounts for the inherent uncertainties in financial time series. This enables more robust decision-making based on a spectrum of potential future states, rather than relying on a single deterministic forecast.

Similar to the experiment by [3], our approach involved forecasting stock and options volumes for a predefined number of days, determined by the selected context length. We applied TimesNet and HuggingFace's Time Series Transformer model for multi-variate time series forecasting on both stock and options volume data using the following three methods:

- Predicting a future window of transaction volume based on the previous window (TimesNet).
- Predicting the transaction volume of the next day based on the previous window (TimesNet).
- Predicting a future window of transaction volume based on all previous transaction data (Time Series Transformer).

The output of our forecasting models includes: historical transaction volumes, predicted volumes for the next window (based on prior data), predicted next-day volumes, and predicted volumes for a future window based on all historical data.

C. Anomaly Detection

In the final stage of our process, we employed a modified version of the ANOMALOUS algorithm, which we call ANOMALOUS-DTW (Fig. 2), adapted from the original work by Islam et al. [3]. Unlike the original algorithm which used Normalized Cross-Correlation (NCC), we implemented Dynamic Time Warping (DTW) to handle the complex and non-linear variations often present in financial data.

DTW calculates the optimal alignment between two time series by allowing elastic shifting of the time axis. It measures similarity by finding the path through a distance matrix that minimizes the total distance between the aligned elements of the sequences. This flexibility makes DTW better suited for financial time series, where patterns such as sudden spikes or drops in transaction volume may occur at different time intervals but still be indicative of similar trading behavior.

Like NCC, DTW retains several important benefits while offering additional advantages. It can handle variable length sequences, allowing comparison of time series that occur over different spans, which is crucial when analyzing trading patterns of varying duration. DTW also maintains scale invariance when implemented with a normalization step, enabling comparison of patterns regardless of their absolute scale - a key feature when examining trading volumes across different stocks or time periods. Similar to NCC, DTW exhibits shift invariance, detecting similar patterns even when they occur at different points in a trading cycle. However, DTW goes beyond these shared capabilities by excelling in capturing non-linear alignments in sequences. This provides a more robust comparison of transaction volumes across various time lags, which is particularly important in financial markets where trading behaviors are often inconsistent and fluctuate due to external market conditions. By allowing time series to stretch or compress, DTW improves the accuracy of anomaly detection, even when patterns are misaligned in time - a capability that NCC lacks.

The ANOMALOUS-DTW algorithm analyzes both predicted and historical transaction volumes to identify potential illegal insider trading patterns. The key parameters of our approach include:

- Companies (C): Selected companies involved in illegal insider trading cases.
- Methods (M): The different forecasting methods used to predict future transaction volumes, including predictions based on previous windows or the entire historical data.
- Windows (W): Time segments, such as 50 consecutive days, used to compare predicted and historical volumes.
- Patterns (P): Predefined illegal trading patterns used as benchmarks for comparison.

Algorithm 1 ANOMALOUS (C, M, W, P, DATA)

```
for each company c of C do
    for each method m in M do
        for each window w in W do
            R_1 ← NCC(c, m, w, 0, 0)
            for each pattern p in P do
                for day d = 1 to size(w) + overlap − size(p) do
                    R_n ← DTW(c, m, w, p, d)
                end for
            end for
        end for
    end for
end for
return R_1 ∪ R_2 ∪ ... ∪ R_n
```

Fig. 2. ANOMALOUS Algorithm

By applying DTW within this framework, we leverage its advantages to enhance anomaly detection in financial time series data. The algorithm's ability to handle non-linear alignments and variable-length sequences is particularly valuable when analyzing trading patterns across different time windows and companies. It's worth noting that while DTW offers these advantages, it does come with increased computational complexity compared to NCC. However, for the unpredictable and complex nature of financial time series data, the improved accuracy in anomaly detection justifies this trade-off.

IV. DATASET

This section outlines the datasets utilized in our study, including how they were collected, processed, and prepared for analysis.

A. Litigation-Related Data

To build a comprehensive dataset of litigation-related press releases, we developed custom web crawlers using Python's BeautifulSoup library for HTML parsing and the requests library for HTTP session handling. These crawlers extracted press releases from the websites of the Securities and Exchange Commission (SEC) [12], the Federal Bureau of

Investigation (FBI) [13], and the US District Court for the Southern District of New York (SDNY) [14].

Our dataset from the SEC archive spans from 1995 to 2023 and contains a total of 10,326 press releases (Fig. 3), of which approximately 15% involved illegal insider trading charges. Specifically, 782 press releases had the keyword "insider" in the title, while 1,514 contained the term in the body. We found 1,609 press releases that featured the term in either the title or body, and 687 press releases contained the keyword in both. This process ensured a robust dataset for training our models to detect patterns specific to insider trading.

2021	SEC Charges Former Investment Adviser for Making Unsuitable Investments, Misleading Clients, and Misappropriating Client Funds	The Securities and Exchange Commission announced charges against Jacob C. Glick of Scottsdale, Arizona, a former investment adviser representative associated with Advanced Practice Advisors, LLC (APA), for repeatedly defrauding and breaching his fiduciary duty to advisory clients. The SECs complaint alleges that from mid-2016 through mid-2018, Glick defrauded his advisory clients in three different ways. First, Glick allegedly placed the majority of his clients, many of whom had moderate or conservative risk tolerances, in unsuitable and risky investments that resulted in substantial losses. In addition, Glick failed to disclose the risks involved in these investments to clients. As also alleged in the complaint, after APA told Glick to liquidate the risky ...

Fig. 3. A random litigation release

We initially labeled each press release as “Non-Insider” or “Possible Insider” based on the presence of the keyword 'insider' in the title or body. Press releases with this keyword were labeled as ‘Possible Insider’, while the rest were classified as ‘Non-Insider’. To refine this classification, we employed semantic analysis using two large language models: OpenAI GPT-4 Turbo and Anthropic Claude 3 Opus. These models analyzed the 'Possible Insider' cases, categorizing them as either 'Insider' or 'Non-Insider'. Cases with concordant 'Insider' classifications from both models were confirmed as insider trading events, while those with concordant 'Non-Insider' classifications were added to our 'Non-Insider' dataset. This dual-verification process using state-of-the-art LLMs ensured a reliable and accurate classification dataset for the subsequent analysis. The remaining ambiguous cases were manually reviewed and labelled.

B. Transaction Volumes and Related Data

We collected historical daily stock volume data for ten companies involved in known illegal insider trading cases. The selected companies include Palo Alto Networks, British Petroleum, GTx Inc., Ubiquiti, American Semiconductor Corporation, Spectrum Pharmaceuticals, Allscripts Healthcare Solutions Inc., Sangamo Biosciences, Evercore, and PetMed Express. Alphabet Inc. was included as a control group, as there were no reported illegal insider trading incidents.

Additionally, we gathered daily options volume data from OptionMetrics, part of the WRDS (Wharton Research Data Services) [19] database. WRDS is widely recognized as a leading provider of historical options and implied volatility data, commonly used by researchers and market practitioners.

To complement the stock and options volume data, we built a Python-based web scraper to collect tweets containing the selected companies’ tickers. We focused on a period of about two and a half years before and after the period when the accused made transactions based on inside information, collecting all available tweets within this time span. These tweets were analyzed using the pyFin-Sentiment model [20], which assigns sentiment probabilities to categorize tweets as ‘Positive’, ‘Negative’, or ‘Neutral’ (Fig. 4).

tweet_text	positive	neutral	negative
$bp for $nxy chatter making rounds---sc	0.418	0.327	0.2555
UK citizens should trade their Notes witl	0.408	0.318	0.2743
I look forward to the day British Fund an	0.49	0.186	0.3241
If you were buying tomorrow, what woul	0.493	0.31	0.1967
$BP goes ex-dividend tonight and yields	0.282	0.58	0.1376
Price targets: $BP 400.00 USD per share	0.553	0.265	0.1817
Royal Dutch Shell's dividend yield now a	0.176	0.188	0.6361
I'm sorry but $BP at 35.00 USD makes fc	0.398	0.313	0.2888

Fig. 4. Random tweets with sentiment scores

The sentiment scores for each tweet were averaged daily and integrated into the historical stock and options volume dataset as additional features (Fig. 5). For days where no tweets were found, we filled in the missing values with 0s for each sentiment category. This inclusion of social media sentiment data adds a more nuanced understanding of how market perception and public discourse might influence trading volumes.

1291	1/3/2011	13735400	2011-01-03	0.6908287868272569	0.1388978414268998	0.1702733717458432
1292	1/4/2011	20036400	2011-01-04	0.314189658410086	0.4180207089869888	0.267789632602925
1293	1/5/2011	11881600	2011-01-05	0.2310610271967426	0.3271040604848664	0.4418349123183908
1294	1/6/2011	12556800	2011-01-06	0.4514572228215049	0.332243953959419	0.2162988232190759
1295	1/7/2011	8037500	2011-01-07	0.1842171338526789	0.5606591709166037	0.2551236952307175
1296	1/10/2011	11165000	2011-01-10	0.4693194288100724	0.2848766394061262	0.2458039317838014
1297	1/10/2011	11165000	2011-01-10	0.4450668428889793	0.3442635048266107	0.2106696522844098
1298	1/11/2011	9228400	2011-01-11	0.4095172108001822	0.3805314132468081	0.2099513759530094
1299	1/12/2011	9980500	2011-01-12	0.5678378652755243	0.1731415070941739	0.2590206276303016
1300	1/13/2011	13854800	2011-01-13	0.2913403504913899	0.5927234758877716	0.1159361736208383
1301	1/14/2011	35199100	2011-01-14	0.3405622595917699	0.3545605089539699	0.30487723145426
1302	1/18/2011	15798500	2011-01-18	0.4507950234794408	0.3450171743728688	0.2041878021476903
1303	1/18/2011	15798500	2011-01-18	0.4961045158777803	0.3020636873415868	0.2018317967806326
1304	1/19/2011	10347900	2011-01-19	0.3182284185304297	0.4535417675365001	0.22822981393307
1305	1/20/2011	11543100	2011-01-20	0.230566656369708	0.4900696269392045	0.2793637166910873
1306	1/21/2011	9596900	2011-01-21	0.6129457794049651	0.0699304826344007	0.3171237379606342

Fig. 5. Stock/options volume dataset

The pyFin-Sentiment model, a logistic regression framework developed for financial tweet analysis, outperforms traditional sentiment analysis models due to its focus on market-specific vocabulary and sentiment nuances. Trained on 10,000 manually labeled finance-related tweets, it accurately interprets financial jargon, abbreviations, and implicit sentiments common in investor discussions. The model recognizes terms like "bull" and "ATH" as positive, and "dump" and "short" as negative. It employs sub-word tokenization to capture variations in word spellings used for emphasis in social media posts (e.g., "bullllish"). This domain-specific approach enables pyFin-Sentiment to achieve superior accuracy in classifying financial tweets as bullish, bearish, or neutral, outperforming even BERT-based models while requiring significantly less computational resources for training and inference.

V. EXPERIMENTS AND RESULTS

This section discusses the experiments conducted and the results obtained while evaluating the performance of our models in detecting and predicting illegal insider trading.

A. Litigation Classifier

For classification, we utilized various BERT-based transformer models including FinBERT [21] – a model pre-trained on financial documents. To tailor these pre-trained models for our specific task, we engaged in a fine-tuning process on our labeled dataset. Fine-tuning, a form of transfer

learning, involves taking pre-trained models, which have already been trained on large corpora of relevant data, and further training them on a smaller, specific dataset. In our case, this meant adjusting the models' existing knowledge base to better align with the nuances of insider trading related press releases. This process modifies the internal weights of the model, making it more adept at identifying and categorizing the unique features and patterns present in our dataset. By fine-tuning these models, we leveraged their extensive pre-trained knowledge in recognizing and classifying the cases in our dataset.

Our findings demonstrate that transformer models excelled in classifying the litigation releases as either "Insider" or "Non-Insider." FinBERT performed slightly better than other models, such as BERT [22], convBERT [23], RoBERTa [24], and ELECTRA[25], achieving an accuracy of 0.99, precision of 0.99, recall of 0.94, and an F1 score of 0.97. The Matthews Correlation Coefficient (MCC) was 0.96, and the Area Under the Precision-Recall Curve (AUC-PR) was 0.98, suggesting highly reliable predictions. Table 1 provides a comparison of these models' performance metrics.

TABLE I. CLASSIFICATION METRICS

Model	Accuracy	Precision	Recall	F1-Score	MCC	AUC-PR
FinBERT	0.99	0.99	0.94	0.97	0.96	0.98
BERT	0.99	0.98	0.94	0.96	0.95	0.97
convBERT	0.99	0.98	0.94	0.96	0.96	0.97
RoBERTa	0.99	0.97	0.95	0.96	0.95	0.98
ELECTRA	0.99	0.97	0.95	0.96	0.95	0.98

The confusion matrix for FinBERT (Fig. 6) highlights its precision, with very few false positives and negatives. Other models, such as BERT, convBERT, ELECTRA, and RoBERTa, also showed similarly strong results.

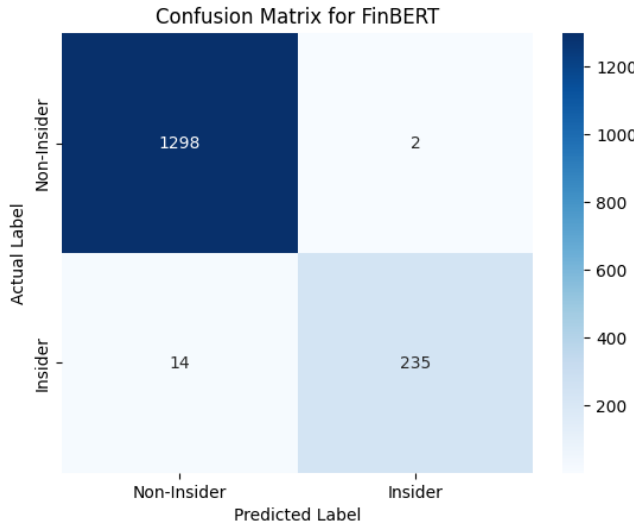

Fig. 6. Confusion matrix for FinBERT

From these findings, we understand that FinBERT's financial lexicon and context awareness give it a slight edge in identifying insider trading indicators within legal documents. The high performance across all evaluation metrics confirms the robustness of transformer models for our text classification task.

To interpret our model's decision making process and enhance transparency, we integrated the Explainable AI techniques, SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations). SHAP values provided a detailed breakdown of feature contributions to the model's predictions, offering quantifiable insights into how specific words or phrases impacted the classification.

Below is a SHAP visualization (Fig. 7) that shows how features influence classification outcomes for 'Insider' and 'Non-Insider' predictions. Darker shades indicate stronger impact. For insider predictions, phrases like "prior to the announcement" and "illegal profits" carry significant weight, while non-insider predictions show less contextually relevant terms have subdued influence.

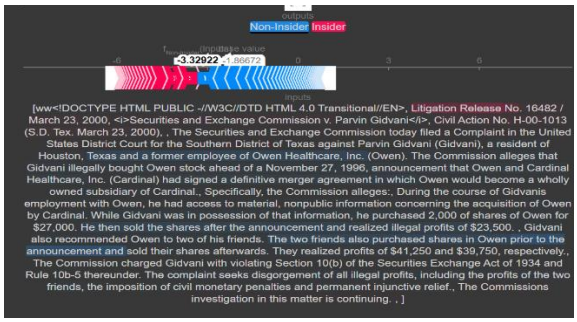

Fig. 7. SHAP Illustration 1

LIME complements SHAP by providing interpretable explanations for individual predictions. It generates local surrogate models to approximate the black-box model's predictions, helping assess reliability and accuracy case-by-case. The LIME visualizations (Fig. 8) below presents class probabilities and highlights influential keywords, quantifying their contribution to the model's decision.

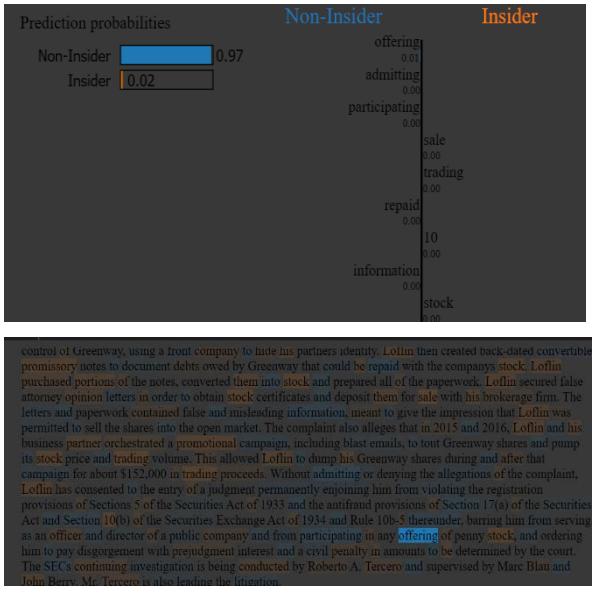

Fig. 8. LIME Illustration 1

B. *Timeseries Forecasting*

In this section, we examine the forecasting performance for both stock and options volume data using advanced deep

learning models. Specifically, we explore the TimesNet and Transformer models for time series forecasting. The task is two-fold: (1) predict future stock and options volumes based on previous windows of data (using TimesNET), and (2) assess the predictive capability when utilizing the entire historical dataset (using the Time Series Transformer model). Each approach is evaluated through the Mean Absolute Scaled Error (MASE) and Symmetric Mean Absolute Percentage Error (sMAPE).

1) Predicting the Next Window of Transaction Volume Based on the Previous Window(TimesNET):

We first use the TimesNet model to predict future stock volumes by relying on a window of 50 consecutive days from the historical stock volume data. TimesNet's unique ability to transform 1D time series data into 2D representations provides a more nuanced understanding of periodicities in the data, such as weekly or monthly trading patterns.

The model was tested across all selected companies, and performance was evaluated using MASE and sMAPE. On average, the results were as follows:

- **MASE:** 0.827
- **sMAPE:** 0.468

These results suggest a 17.3% improvement over naive forecasting methods, though the sMAPE score indicates that there is room for improvement in predictive accuracy.

Figure 9 illustrates the actual versus predicted stock volumes for American Semiconductor Corporation, where the orange line represents the model's forecast, and the blue line shows the actual values. The initial overlap indicates the historical context window used for prediction.

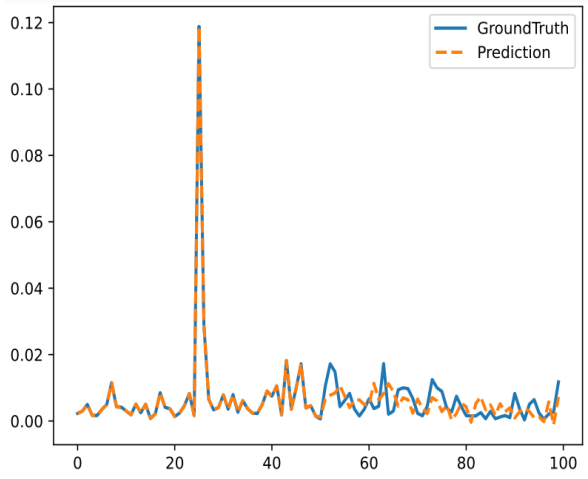

Fig. 9. Normalized forecasted stock volume using the window method

The TimesNet model was then applied to predicting future options volume based on a 50-day window of historical data.

- **MASE:** 0.840
- **sMAPE:** 0.860

Though the model demonstrated a 16% improvement over naive methods, the higher sMAPE score for options volume suggests that predictions in this domain are more volatile, with less consistency in performance compared to stock volume predictions. Figure 10 shows the predicted and actual options volumes for one of the selected companies.

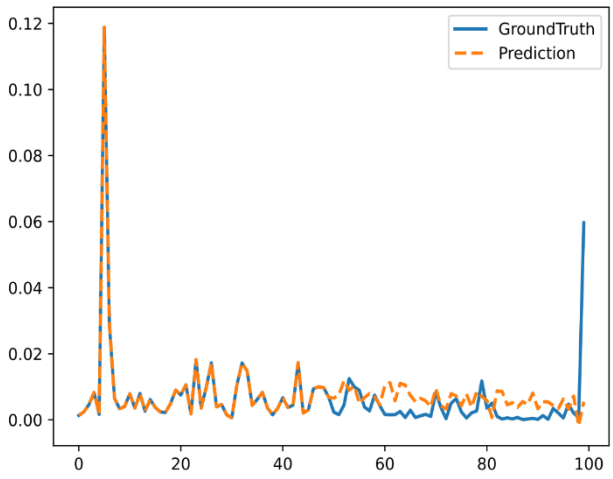

Fig. 10. Normalized forecasted options volume using the window method

2) Predicting the Next Day of Transaction Volume Based on the Previous Window (TimesNET):

For short-term predictions, we forecast the stock volume for the next day using the previous 50 days as input. This day-ahead forecasting is intended to capture shorter-term fluctuations in the stock market.

Results for this method showed:

- **MASE:** 0.792
- **sMAPE:** 0.397

This model exhibited around 20.8% better predictive accuracy than naive methods. As shown in Figure 11, the forecasted values (orange) align closely with the actual values (blue), showcasing the model's capability to predict day-ahead stock volumes with a reasonable degree of accuracy.

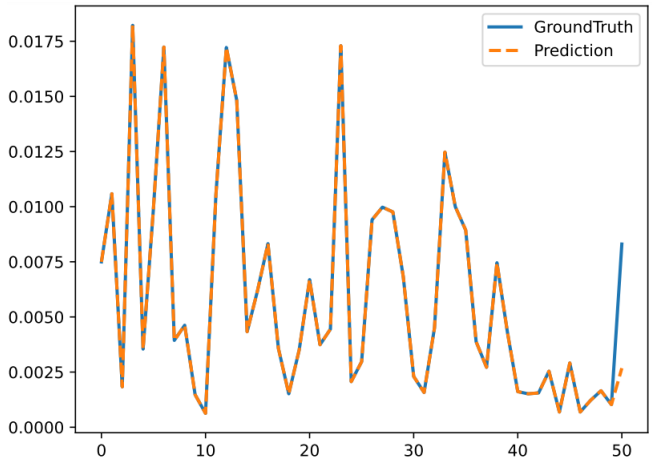

Fig. 11. Normalized forecasted stock volume using the day ahead method

Short-term, day-ahead options volume forecasting yielded the following results:

- **MASE:** 0.801
- **sMAPE:** 0.785

The model improved predictions by approximately 20% compared to naive methods. Similar to stock volume

forecasts, this approach performed well in capturing short-term trends, as visualized in Figure 12.

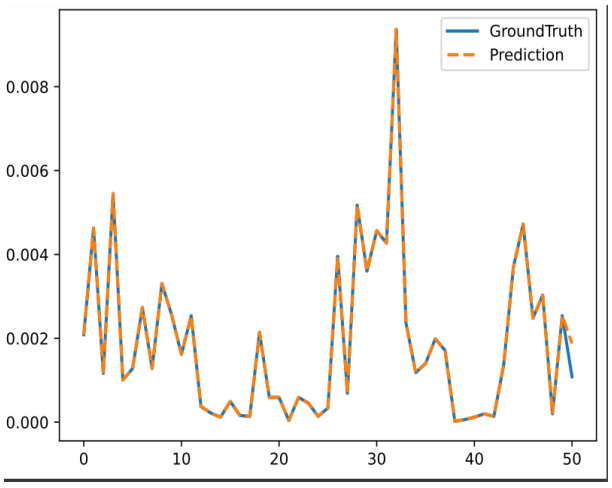

Fig. 12. Normalized forecasted options volume using the day ahead method

3) Entire History-Based Transaction Volume Forecasting (using Transformer Model):

We employed the Time Series Transformer model for long-range forecasting by utilizing the entire historical data available for each company. This method excels at capturing long-term dependencies and is more flexible than traditional models, particularly in handling variable-length sequences.

By leveraging the full span of historical stock volume data, the Transformer model aimed to predict future trends more comprehensively. The results were as follows:

- **MASE:** 0.879
- **sMAPE:** 0.495

While the performance was slightly lower than that of day-ahead forecasting with TimesNet, the Transformer model still provided reasonable predictions. Figure 13 illustrates the probabilistic forecasting approach of the Transformer model, where the orange line represents the median prediction, and the shaded area indicates the uncertainty (±1 standard deviation) around the prediction.

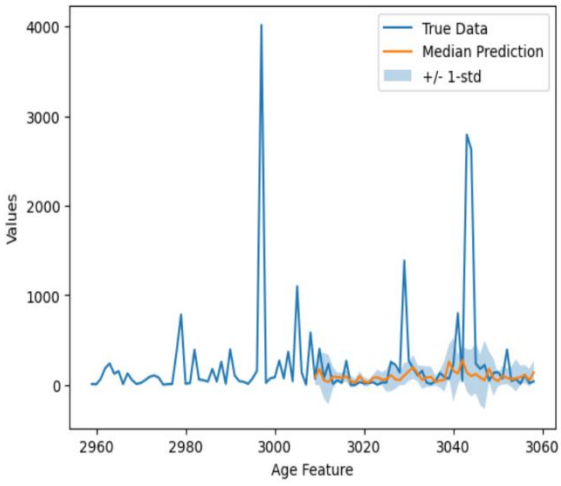

Figure 13: Normalized forecasted stock volume using entire history method (Transformer model)

When applied to options volume data, the Transformer model showed the following results:

- **MASE:** 0.847
- **sMAPE:** 0.836

Though the results were not as robust as for stock volume, the model still demonstrated a 16.3% improvement over naive methods. Figure 14 visualizes the model's probabilistic predictions for options volume, with uncertainty levels indicated by the shaded region.

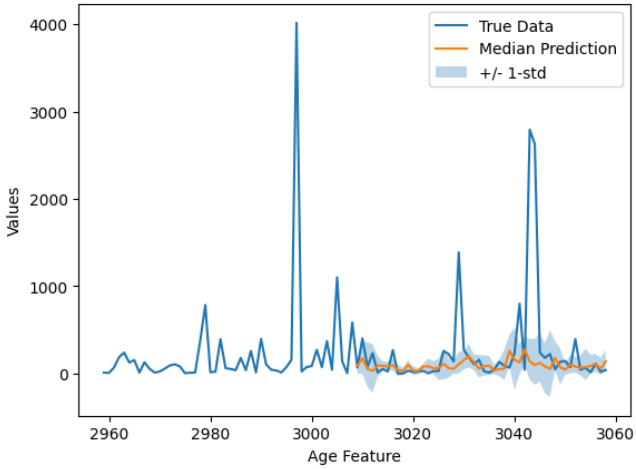

Figure 14: Normalized forecasted options volume using entire history method (Transformer model)

C. Anomaly Detection

For anomaly detection, we utilized the ANOMALOUS-DTW algorithm, applying Dynamic Time Warping (DTW) to compare predicted stock volume time series with established patterns of illegal insider trading. DTW is a technique that measures the similarity between two temporal sequences which may vary in speed or length. The DTW distance is non-negative, with 0 indicating identical sequences and larger values indicating greater dissimilarity.

In this context, the 'minimum DTW distance' refers to the lowest observed distance between the predicted stock volume and the known patterns of insider trading during our analysis. A lower DTW distance indicates a stronger similarity between the predicted and established patterns, suggesting a higher likelihood that the detected anomaly could be related to illegal insider trading.

1) Day Ahead Forecasting:

For the day-ahead method, we observed a minimum DTW score of 1.77 for window 38, pattern 3, day 6 (fig. 15). This high score indicates a strong correlation between the predicted stock volume and the insider trading pattern on that specific day, suggesting an anomaly potentially related to insider trading activities.

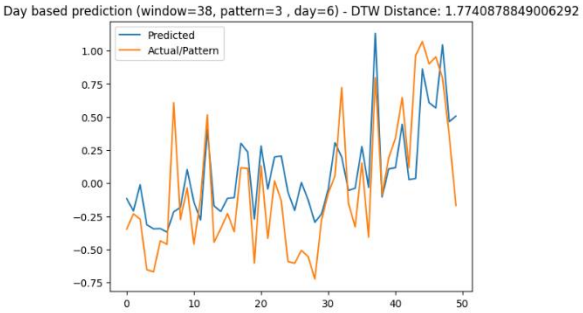

Figure 15: DTW Alignment of True and Predicted Signals for the Day-Ahead Method

2) Window Based Forecasting:

In the window-based forecasting method, the minimum DTW score observed was 2.16 for window 6, pattern 5, day 10 (fig. 16). This strong correlation also highlights a possible period of suspicious trading behavior that aligns with insider trading patterns.

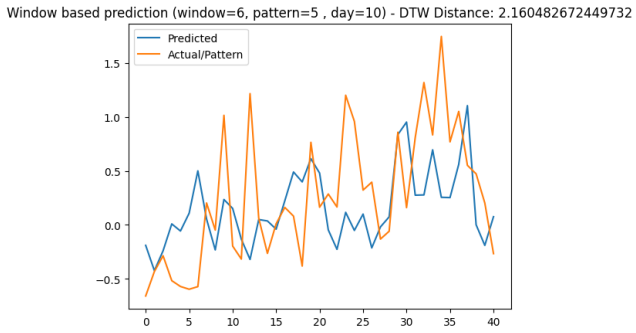

Figure 16: DTW Alignment of True and Predicted Signals for the Day-Ahead Method

3) Entire History Based Forecasting:

Using the entire history-based approach, we detected a minimum DTW score of 2.32 for window 4, pattern 3, day 11 (fig. 17). This method similarly identified a period where the trading patterns closely matched those of illegal insider trading.

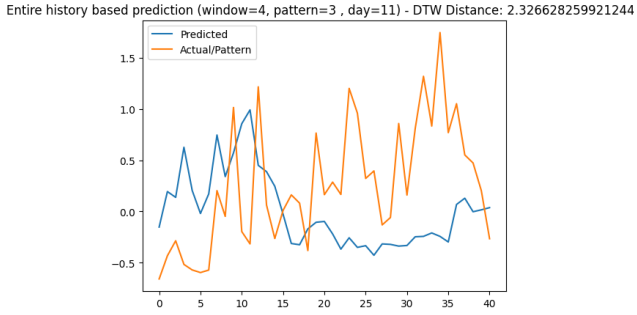

Figure 17: DTW Alignment of True and Predicted Signals for the Day-Ahead Method

To determine the actual day corresponding to the lowest value of DTW distance, we use the following formulated equation derived from our observations of the data:

$$day = window\ size + (w * window\ size) + d$$

In this equation:

- ***window size*** refers to the number of consecutive days used in each analysis window (in our experiment, this is set to 50 days).
- ***w*** is the window number which indicates the sequence of the window in the analysis.
- ***d*** is the specific day within that window where the maximum NCR value is observed.
- $day = 50+(38\times50)+13 = 1963$

This calculation tells us that the minimum DTW value of 1.774 corresponds to day 1963 in the time series data. This day is identified as having a high correlation with known patterns of insider trading.

To assess the effectiveness of these anomalous patterns as indicators, we extended our experiment to Alphabet Inc, for which there have been no reported cases of illegal insider trading to date. Our findings show that when using a DTW distance threshold of 2.2 or lower, no anomalous patterns were detected for Alphabet Inc. This suggests that our model can effectively distinguish between normal trading patterns and those potentially indicative of illegal insider trading activities.

Table 2 summarizes the hit counts per pattern across the selected companies, where each pattern represents a specific sequence of trading volumes indicative of insider trading. The table shows the percentage of companies where each pattern was detected ('% Comps'), the total number of hits (signals with DTW distance less than the threshold of 3), and how these hits are distributed across the three methods.

TABLE II. HIT COUNTS PER PATTERN

Pattern	Total Hits	Day	Window	History
1	290	189	48	53
2	602	423	96	83
3	1318	837	267	214
4	77	37	17	23
5	816	579	138	99
6	442	336	57	49
7	130	92	21	17
8	525	384	73	68
9	31	17	11	3
10	381	246	64	71

However, this approach is not without its limitations. While DTW is more flexible than NCC in handling non-linear relationships and time warps in the data, it can be computationally expensive for large datasets. DTW's flexibility might make it harder to interpret the alignment and match results compared to NCC, which provides a more straightforward measure of similarity between time series.

The method's effectiveness is heavily dependent on the quality and representativeness of the established insider trading patterns used for comparison. The reliance on historical patterns may limit the detection of novel manipulation strategies. While our approach can identify suspicious patterns, it cannot definitively prove the occurrence of illegal insider trading. False positives may occur due to legitimate market events that coincidentally resemble insider trading patterns. Conversely, sophisticated actors might develop strategies to evade detection by avoiding known patterns. These limitations underscore the need for

complementary analytical methods, continuous refinement of the model, and human expertise in interpreting the results.

VI. CONCLUSION

Our study presents a significant advancement in the proactive detection and prediction of illegal insider trading through the analysis of diverse data sources. The integration of LLMs, options data, and social media sentiment represents a novel contribution in this domain. Our BERT-based classifier demonstrated high accuracy in categorizing litigation releases, while the TimesNet model showed promising results in predicting stock and options volumes, particularly with the day-ahead forecasting method. A key addition in our work is the use of Dynamic Time Warping (DTW) for anomaly detection, which proved valuable in identifying potential insider trading patterns given the complex nature of financial markets. This approach detected anomalous patterns with high precision, indicating strong correlations with known insider trading patterns.

However, we acknowledge several limitations in our study. While our approach shows promise in detecting anomalous patterns, it heavily reliant on established insider trading patterns. It may also generate false positives, highlighting the need for further refinement of our detection thresholds. In addition, the computational resources required for multiple stages of our analysis were substantial, constraining our ability to perform extensive hyperparameter tuning and limiting the scale of our experiments.

Future work should focus on extending this analysis to a larger corpus of companies to provide more generalized findings. There is also a need to develop methods capable of identifying previously unprosecuted instances of insider trading, which could lead to more comprehensive detection systems. Improving the accuracy of the forecasting models, is an important area of focus to enhance the overall effectiveness of the system.

In conclusion, our work represents a significant step forward in the application of deep learning and time series analysis to the detection of illegal insider trading. By leveraging LLMs and modern deep learning models, we have developed a framework for financial regulators and institutions to enhance their surveillance mechanisms, contributing to the ongoing efforts to maintain the fairness and integrity of financial markets.

REFERENCES

[1] V. Patel and T. J. Putniņš, "How Much Insider Trading Happens in Stock Markets?," in *American Finance Association (AFA) Annual Meeting*, 2020.

[2] H. Wu, T. Hu, Y. Liu, H. Zhou, J. Wang, and M. Long, "Timesnet: Temporal 2d-variation modeling for general time series analysis," *arXiv preprint arXiv:2210.02186,* 2022.

[3] S. R. Islam, S. K. Ghafoor, and W. Eberle, "Mining illegal insider trading of stocks: A proactive approach," in *2018 IEEE international conference on big data (Big Data)*, 2018: IEEE, pp. 1397-1406.

[4] F. Lauar and C. Arbex Valle, "Detecting and predicting evidences of insider trading in the Brazilian market," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 2020: Springer, pp. 241-256.

[5] J. Tallboys, Y. Zhu, and S. Rajasegarar, "Identification of stock market manipulation with deep learning," in *International Conference on Advanced Data Mining and Applications*, 2022: Springer, pp. 408-420.

[6] F. Chen, K. Di, H. Tao, Y. Jiang, and P. Li, "Insider trading detection algorithm in industrial chain based on logistics time interval characteristics," in *International Conference on Parallel and Distributed Computing: Applications and Technologies*, 2023: Springer, pp. 118-129.

[7] T. Seth and V. Chaudhary, "A predictive analytics framework for insider trading events," in *2020 IEEE international conference on big data (Big Data)*, 2020: IEEE, pp. 218-225.

[8] R. James, H. Leung, and A. Prokhorov, "A machine learning attack on illegal trading," *Journal of Banking & Finance,* vol. 148, p. 106735, 2023.

[9] B. Rizvi, D. Attew, and M. Farid, "Unsupervised Manipulation Detection Scheme for Insider Trading," in *International Conference on Intelligent Systems Design and Applications*, 2022: Springer, pp. 244-257.

[10] Q. Liu, C. Wang, P. Zhang, and K. Zheng, "Detecting stock market manipulation via machine learning: evidence from China Securities Regulatory Commission punishment cases," *International Review of Financial Analysis,* vol. 78, p. 101887, 2021.

[11] Q. Zhang, C. Qin, Y. Zhang, F. Bao, C. Zhang, and P. Liu, "Transformer-based attention network for stock movement prediction," *Expert Systems with Applications,* vol. 202, p. 117239, 2022.

[12] SEC. https://www.sec.gov/litigation/litreleases (accessed.

[13] FBI. https://www.fbi.gov/investigate/white-collar-crime/news (accessed.

[14] US-SDNY. https://www.justice.gov/usao-sdny/pr (accessed.

[15] S. Lundberg, "A unified approach to interpreting model predictions," *arXiv preprint arXiv:1705.07874,* 2017.

[16] M. T. Ribeiro, S. Singh, and C. Guestrin, "" Why should i trust you?" Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016, pp. 1135-1144.

[17] "Hugging Face Time Series Transformer." https://huggingface.co/docs/transformers/en/model_doc/time_series_transformer (accessed.

[18] Q. Wen *et al.*, "Transformers in time series: A survey," *arXiv preprint arXiv:2202.07125,* 2022.

[19] UPenn.https://wrds-www.wharton.upenn.edu/pages/about/data-vendors/optionmetrics/ (accessed.

[20] M. Wilksch and O. Abramova, "PyFin-sentiment: Towards a machine-learning-based model for deriving sentiment from financial tweets," *International Journal of Information Management Data Insights,* vol. 3, no. 1, p. 100171, 2023.

[21] D. Araci, "FinBERT: Financial Sentiment Analysis with Pre-trained Language Models," *arXiv preprint arXiv:1908.10063,* 2019.

[22] J. D. M.-W. C. Kenton and L. K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of naacL-HLT*, 2019, vol. 1, p. 2.

[23] Z.-H. Jiang, W. Yu, D. Zhou, Y. Chen, J. Feng, and S. Yan, "Convbert: Improving bert with span-based dynamic convolution," *Advances in Neural Information Processing Systems,* vol. 33, pp. 12837-12848, 2020.

[24] Y. Liu, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692,* 2019.

[25] K. Clark, "Electra: Pre-training text encoders as discriminators rather than generators," *arXiv preprint arXiv:2003.10555,* 2020