RJPSS Oct.24-Mar.25 Vol. L No.1 ISSN: (P)0258-1701 (e)2454-3403 Impact Factor: 8.867
https://doi.org/10.31995/rjpss.2025.v50i01.028

Machine Learning Approaches for Fraud Detection in the Stock Market

Ujjwal Sharma
Mca Amity University,
Noida
Email: ujju210@gmail.com

Abstract

Fraud detection in the stock market is an ongoing challenge due to the increasing complexity of financial transactions and the sophistication of fraudulent activities. Traditional methods, such as rule-based systems and statistical approaches, often struggle to keep up with the dynamic nature of fraud. Recently, machine learning (ML) techniques have proven to be highly effective in detecting fraudulent behavior by analyzing large volumes of financial data to uncover anomalies and suspicious actions. This paper examines various ML methods, such as supervised learning techniques like logistic regression, decision trees, and support vector machines, as well as unsupervised learning approaches like clustering and anomaly detection. Additionally, ensemble methods such as XGBoost and random forests have shown promise in improving fraud detection accuracy. Deep learning models, including recurrent neural networks (RNNs) and graph-based models, are also gaining attention for their ability to detect complex fraudulent trading patterns. The paper also addresses challenges faced by these methods, including issues of data imbalance, interpretability, and the need for real-time detection. Finally, it suggests possible future research directions. By incorporating machine learning into stock market monitoring, financial institutions can improve the efficiency of fraud detection and reduce risks, ultimately supporting market integrity and fostering investor confidence.

Reference to this paper should be made as follows:

Received: 05.03.2025
Approved: 28.03.2025

Ujjwal Sharma

Machine Learning Approaches for Fraud Detection in the Stock Market

Article No.28
RJPSS Oct.-Mar. 2025,
Vol. L No. 1,
pp. 207-220

Similarity Check - 4%

Online available at:
https://anubooks.com/journal-volume/rjpss-2025-vol-l-no1-mar

https://doi.org/10.31995/rjpss.2025.v50i01.028

Introduction

The stock market plays a vital role in the global economy by enabling capital formation and providing investment opportunities. However, fraudulent activities such as insider trading, pump-and-dump schemes, and market manipulation present significant risks to both investors and financial institutions. Detecting and preventing fraud in the stock market is a complex task due to the large volume of transactions, fast-paced trading, and the constantly evolving methods used by fraudsters. Traditional fraud detection systems, which rely on predefined rules and manual monitoring, often struggle to keep up with emerging fraudulent patterns, leading to financial losses and diminished trust in the market.

In recent years, machine learning (ML) has emerged as a powerful tool in stock market fraud detection, owing to its ability to process large datasets, uncover hidden patterns, and adapt to new forms of fraudulent activity. ML techniques, including supervised learning, unsupervised learning, and deep learning, enable automated identification of anomalies in trading data, thereby improving the accuracy and efficiency of fraud detection systems. Supervised learning models, such as decision trees, support vector machines (SVM), and neural networks, classify transactions as fraudulent or legitimate based on historical fraud data. Unsupervised learning methods, like clustering and anomaly detection, excel at identifying new and unknown fraud patterns in real time. More advanced approaches, including graph-based learning and deep learning, enhance fraud detection by analyzing complex trading relationships and high-dimensional data.

Despite these advancements, several challenges persist in implementing ML-based fraud detection systems. One key issue is data imbalance, where fraudulent transactions are much rarer than legitimate ones, impacting the performance of models. Furthermore, the interpretability of ML models remains critical for financial regulators and stakeholders to trust And adopt these technologies. The incorporation of real-time detection capabilities and strong data security measures is also essential to ensure the effectiveness of these systems.

This paper examines various machine learning techniques employed in stock market fraud detection, discussing their applications, advantages, and challenges. By harnessing the power of machine learning, financial institutions and regulatory bodies can enhance fraud detection, protect investors, and ensure the integrity of the stock market.

https://doi.org/10.31995/rjpss.2025.v50i01.028

Figure 1

OBJECTIVE

The main goal of using machine learning approaches for fraud detection in the stock market is to create smart, efficient, and adaptive systems that can identify and prevent fraudulent activities in stock trading.

The specific objectives are as follows:

1. **Automated Fraud Detection –** Use machine learning models to automatically identify fraudulent activities in the stock market, such as insider trading, market manipulation, and abnormal trading behaviors.
2. **Real-Time Monitoring –** Develop AI-driven systems that can analyze vast amounts of financial data in real-time, allowing for the immediate identification of suspicious activities.
3. **Pattern Recognition and Anomaly Detection–** Apply both supervised and unsupervised learning techniques to uncover hidden fraud patterns and detect unusual trading behaviors that may indicate fraud.
4. **Improving Accuracy and Reducing False Positives –** Refine fraud detection models to reduce the number of false positives (incorrect fraud alerts), while ensuring that genuine fraudulent activities are not missed.
5. **Adaptive Learning to Counter Changing Fraud Strategies –** Utilize reinforcement learning and self-learning AI models to adapt to new fraud tactics and continuously enhance detection effectiveness.
6. **Compliance with Regulatory Standards –** Ensure that ML-based fraud detection systems meet the requirements of financial regulations and compliance standards, such as those set by the SEC and anti-money laundering (AML) laws.

7. **Privacy and Security in Fraud Detection –** Implement privacy-preserving methods like federated learning and blockchain-based AI to secure sensitive stock market data while improving fraud detection capabilities.
8. **Promoting Market Transparency and Investor Confidence –** Strengthen investor trust by deploying ML-powered fraud detection systems that foster a transparent and fair stock trading environment.

INTEGRATION OF ML IN STOCK MARKET

The use of machine learning (ML) in detecting fraud in the stock market has transformed how financial institutions identify and prevent fraudulent activities in real-time. By leveraging large volumes of trading data, ML techniques can uncover suspicious patterns, spot anomalies, and predict potential fraud.

The integration of ML into stock market fraud detection includes several critical components:

1. Data Collection and Preprocessing To train ML models effectively, a variety of data sources are needed, such as:

- Historical stock market data (including trades, prices, and volumes).
- Financial reports and investor actions.
- News articles, social media trends, and sentiment analysis.
- Regulatory filings and insider transaction reports.

Integration: Advanced ETL (Extract, Transform, Load) processes are used to clean and prepare raw data before feeding it into ML models.

2. Fraud Detection Models and Techniques Different ML techniques are used in fraud detection systems for the stock market:

3. Supervised Learning Models: These models are trained on labeled historical data, including fraud cases, to classify transactions. Models like logistic regression, random forests, and XGBoost are used to identify fraudulent transactions based on historical trading patterns. Deep learning models (LSTM, CNNs) are used to analyze stock data over time and detect unusual patterns.

4. Unsupervised Learning Models: When fraud data is scarce, unsupervised learning can be applied. Techniques like clustering (e.g., K- Means, DBSCAN) group similar behaviors and flag outliers, while autoencoders and isolation forests are used to detect anomalies.

- **Reinforcement Learning (RL):** This allows models to continuously improve by learning from new fraud cases, making fraud detection more adaptive over time.

https://doi.org/10.31995/rjpss.2025.v50i01.028

5. Real-time monitoring and Anomaly Detection ML can be integrated into real-time fraud detection systems to analyze trading activity as it happens.

Integration: Streaming frameworks like Kafka or TensorFlow Serving are used to process large amounts of real-time data, enabling the detection of fraud such as spoofing, wash trading, or pump-and-dump schemes.

6. Natural Language Processing (NLP) for Market Manipulation Detection NLP techniques are used to analyze textual data from various sources:

- Financial news and reports to identify misleading information that may affect stock prices.
- Social media platforms like Twitter, Reddit, or StockTwits to detect sentiment-driven market manipulation.

Integration: Sentiment analysis (using models like BERT or GPT) helps detect fraudulent stock hype, and fake news detection models identify misleading financial reports.

7. Blockchain and AI Integration for Secure Fraud Detection Combining blockchain technology with ML enhances fraud detection by providing transparent and unchangeable transaction records.

Integration: AI-powered smart contracts can automatically detect suspicious stock trades, and decentralized fraud detection networks using federated learning allow for fraud detection across different exchanges while maintaining data privacy.

8. Regulatory Compliance and AI-based auditing ML models support regulatory bodies in ensuring compliance by detecting fraudulent activities within the stock market.

Integration: AI-driven monitoring systems automate the process of fraud investigation to meet compliance standards set by organizations like the SEC or FINRA. Explainable AI (XAI) is used to make fraud detection models understandable for audits and legal purposes.

9. Predictive Analysis and Risk Assessment ML helps in predicting potential fraud risks by analyzing past market trends And forecasting future fraudulent activity.

Integration: Risk scoring models powered by AI assign risk levels to stock transactions, and time-series forecasting techniques (like ARIMA or LSTMs) predict fluctuations in stock prices that may indicate fraud.

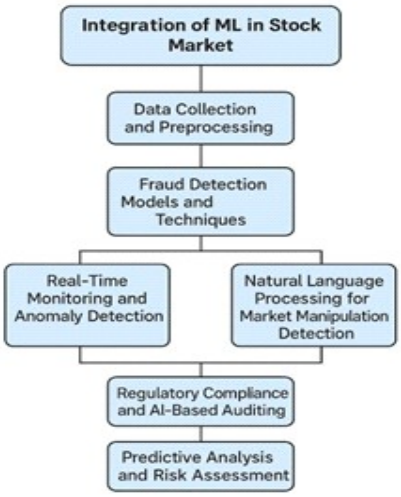

Figure 2

APPLICATIONS

Machine learning (ML) has transformed fraud detection in the stock market by enabling the automated, accurate, and real-time identification of suspicious trading activities. The use of ML techniques in detecting fraud has significantly enhanced financial security and regulatory compliance across various areas. Below are some key applications of ML in stock market fraud detection:

1. **Insider Trading Detection** Insider trading involves individuals using non-public information for personal financial gain. ML models analyze historical trading data, price trends, and market news sentiment to detect irregular transactions that may suggest insider trading. Supervised learning models like logistic regression and decision trees, along with deep learning techniques, are particularly effective in identifying anomalies linked to insider trading activities.

2. **Anomaly Detection in Trading Patterns** Fraudsters may manipulate stock prices through unusual trading behaviors. ML algorithms, particularly unsupervised learning methods such as clustering (e.g., K-means, DBSCAN) and anomaly detection, are used to spot suspicious trading patterns, such as irregular trading volumes or unexpected price changes. These models continuously adapt to new data, helping to identify evolving fraud tactics.

3. **Pump-and-Dump Scheme Detection** Pump-and-dump schemes involve

https://doi.org/10.31995/rjpss.2025.v50i01.028

artificially inflating stock prices with false information, only for fraudsters to sell off shares at a profit. Natural Language Processing (NLP) techniques are used to analyze financial news, social media discussions, and online forums to spot misleading stock promotions. When combined with time-series analysis, ML models can detect sudden, unjustified spikes in stock prices that are characteristic of pump-and-dump schemes.

4. **High-Frequency Trading (HFT) Fraud Detection** In high-frequency trading, fraudsters may engage in practices like spoofing and layering, where fake orders are placed to manipulate market prices. ML models, especially reinforcement learning and deep learning, analyze high-frequency trading data in real-time to detect these deceptive practices, enabling swift regulatory intervention.

5. **Predictive Risk Assessment** ML models can assess the likelihood of fraud by evaluating the behavior of traders, financial transactions, and stock price movements. Ensemble methods, such as XGBoost and Random Forest, combine multiple models to improve the accuracy of fraud risk predictions. These models help financial institutions develop risk scores for individual transactions and traders.

6. **Fraudulent Account Detection** Fraudsters often create multiple fake accounts to manipulate stock prices. Graph-based machine learning techniques analyze the relationships between accounts, transactions, and stock movements to detect fraudulent networks of accounts. This approach helps uncover coordinated fraud efforts and mitigate potential risks.

7. **Real-Time Fraud Prevention and Compliance Monitoring** Real-time fraud detection is crucial for stock exchanges and financial institutions. ML-driven surveillance systems monitor transactions continuously and flag suspicious activities in real-time. These systems support regulators in enforcing financial laws and preventing fraudulent activities before they cause significant harm.

SOLUTIONS

Fraud detection in the stock market demands robust, scalable, and adaptive solutions that can keep up with the ever-evolving tactics used by fraudsters. Machine learning (ML) provides a powerful framework for identifying fraudulent activities by processing large volumes of data, detecting hidden patterns, and enabling real-time analysis. Here are several key ML-driven solutions that can significantly improve fraud detection in the stock market:

1. **Supervised Learning for Fraud Classification** Supervised learning models are trained on labeled datasets containing both fraudulent and legitimate transactions, allowing them to classify new trades accurately.

- **Solution**: Algorithms like Logistic Regression, Decision Trees, Random Forest, Support Vector Machines (SVM), and XGBoost can be used to classify stock trades as either fraudulent or legitimate.
- **Benefit**: These models offer high accuracy in fraud detection when sufficient labeled data is available.

2. **Unsupervised Learning for Anomaly Detection** Fraud patterns can often be unknown or continuously changing, making traditional supervised learning less effective. Unsupervised learning helps detect unusual behaviors without the need for prior labels.

- **Solution**: Clustering algorithms (e.g., K- Means, DBSCAN), Isolation Forests, and Autoencoders can be used to identify abnormal trading activities such as irregular trade volumes or unexpected price changes.
- **Benefit**: These methods can detect new and previously unseen fraud patterns without requiring labeled data.

3. **Deep Learning for Pattern Recognition** Deep learning models excel at analyzing complex stock market data, enabling them to detect sophisticated fraudulent activities.

- **Solution**: Convolutional Neural Networks (CNNs) can be used for pattern recognition, while Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks are effective at spotting trends over time that may signal fraudulent activity.
- **Benefit**: These models offer more accurate detection of advanced fraud schemes like insider trading and market manipulation.

4. **Graph-Based Machine Learning for Network Analysis** Fraudulent activities often involve multiple coordinated accounts or traders. Graph-based ML models can uncover hidden connections between these traders.

- **Solution**: Graph Neural Networks (GNNs) and PageRank algorithms can analyze relationships between accounts to detect fraud rings or networks of colluding traders.
- **Benefit**: These methods help expose organized financial crimes, such as coordinated pump-and-dump schemes.

5. **Natural Language Processing (NLP) for Fake News & Sentiment Analysis** Fraud can be driven by misinformation in news reports and social media, which influences stock prices. NLP models are used to detect Misleading information that could manipulate the market.

- **Solution**: NLP techniques, including BERT, GPT, and LSTM-based models, can analyze financial news, reports, and social media discussions

https://doi.org/10.31995/rjpss.2025.v50i01.028

to detect fraudulent promotions or market manipulation.

- **Benefit**: These tools help identify false stock promotions and prevent misinformation from affecting market integrity.

6. **Reinforcement Learning for Adaptive Fraud Prevention** Fraud patterns are constantly evolving, making adaptive learning essential for real-time fraud prevention.

- **Solution**: Reinforcement Learning (RL) models allow AI agents to continuously learn from market trends and fraud incidents, improving their detection strategies over time.
- **Benefit**: These models adapt to new fraud patterns, enhancing the accuracy of fraud detection over the long term.

7. **Hybrid Models for Improved Accuracy** Combining multiple ML techniques can further enhance fraud detection efficiency.

- **Solution**: Hybrid models, such as XGBoost combined with LSTM for high-dimensional data analysis, or CNNs integrated with anomaly detection models, can provide a more comprehensive fraud detection approach.
- **Benefit**: These models balance accuracy, adaptability, and real-time detection capabilities, ensuring a more effective fraud detection system.

8. **Real-Time Fraud Detection with AI-Powered Monitoring Systems** Given the speed of financial markets, real-time fraud detection is critical.

- **Solution**: Implementing streaming ML frameworks like Apache Kafka and TensorFlow Serving allows for real-time processing of stock market data, flagging suspicious transactions as they occur.
- **Benefit**: Immediate fraud detection minimizes financial losses and ensures compliance with regulatory requirements.

EXAMPLE OF APPLYING ML IN THE STOCK MARKET

a) Creating Dataset by importing libraries:

Trade_ID	Trader_ID	Stock_Symbol	Trade_Volume
1001	T5708	MSFT	32065
1002	T4601	MSFT	44129
1003	T9863	GOOGLE	33605
1004	T4459	AAPL	7300

Table 1

Price_Per_Share	Trade_Time	Order_Type
2210.63	9:06 PM	Stop Loss
2974.14	15:25 PM	Market
1234.37	10:09 AM	Stop Loss
1999.1	10:50 AM	Limit

Table 2

Market_Impact	Sentiment_Score	Insider_Trade
Very High	0.66	No
Low	-0.82	No
Low	0.79	No
Very High	0.75	No

Table 3

High_Frequency_Trading	Unusual_Price_Spike	Historical_Fraud_Flag	Fraud_Label
No	No	No	Legitimate
No	No	No	Legitimate
No	No	No	Legitimate
No	No	No	Legitimate

Table 4

b) Data Preprocessing:

```
label_encoders = {}
for col in df.select_dtypes(include=['object']).columns:
    label_encoders[col] = LabelEncoder()
    df[col] = label_encoders[col].fit_transform(df[col])
```

Figure 4

c) Splitting for training and testing with scaling numerical features:

```
[7] X = df.drop(columns=['Fraud_Label']) # Assuming 'Fraud' is the target column
    y = df['Trade_Time']

[8] X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

[9] scaler = StandardScaler()
    X_train = scaler.fit_transform(X_train)
    X_test = scaler.transform(X_test)
```

Figure 5

https://doi.org/10.31995/rjpss.2025.v50i01.028

d) Training a model by applying the algorithm and taking out a classification report:

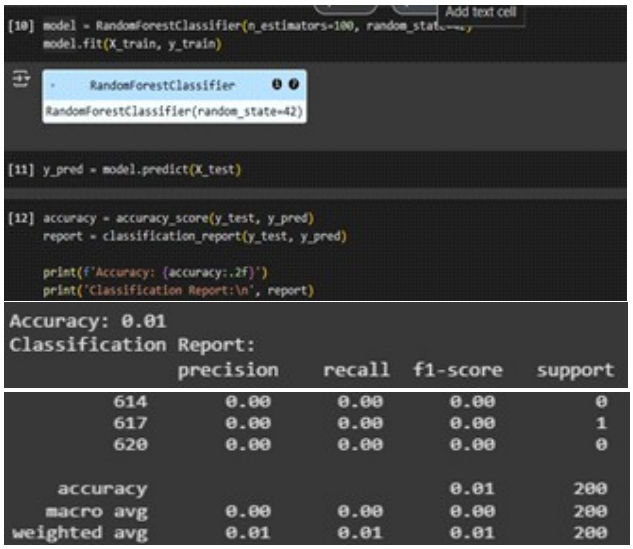

Figure 6

Despite the promising potential of machine learning (ML) for fraud detection in the stock market, several challenges continue to limit its effectiveness. These obstacles stem from the ever-changing nature of financial markets, the adaptability of fraudsters, and inherent limitations within ML models. The main challenges include:

1. **Data Imbalance** Fraudulent transactions are much rarer than legitimate ones, making it difficult for ML models to accurately learn to detect fraud.

- **Challenge**: Traditional ML models may become biased toward legitimate transactions, which affects their ability to detect fraud effectively.
- **Solution**: Techniques like oversampling, undersampling, and cost-sensitive learning can help balance the dataset and improve the model's ability to detect fraudulent activities.

2. **Evolving Fraud Tactics** Fraudsters are constantly changing their methods to avoid detection, making it difficult for ML models to keep up.

- **Challenge**: Models trained on past fraud data may not be equipped to detect emerging or unknown fraud strategies.
- **Solution**: Implementing adaptive learning models, such as reinforcement learning or online learning, can help the system adjust to evolving fraud tactics in real-time.

3. **Explainability and Interpretability** Many ML models, particularly deep learning models, operate as "black boxes," meaning their decision-making processes are not easily understood.

- **Challenge**: Financial institutions and regulatory bodies require transparent models that can explain how decisions are made.
- **Solution**: Explainable AI (XAI) methods, such as SHAP and LIME, can help provide insights into the model's decision-making process, improving transparency.

4. **Real-Time Fraud Detection Constraints** The stock market operates at very high speeds, requiring fraud detection systems to process large volumes of data in real-time.

- **Challenge**: Traditional ML models may struggle to keep up with the rapid pace of high-frequency trading (HFT) and large-scale data streams.
- **Solution**: Using streaming ML frameworks like Apache Kafka or TensorFlow Serving, along with edge computing, can enable real-time fraud detection by processing data faster.

5. **Feature Engineering Complexity** Fraud detection in the stock market involves a variety of variables, such as

trading volumes, prices, and individual trader behaviors, which makes selecting the right features a complex task.

- **Challenge**: Identifying the most relevant features for ML models requires in-depth domain knowledge.
- **Solution**: Automated feature selection tools, such as AutoML and deep feature synthesis, can assist in identifying useful features, improving the model's accuracy.

6. **False Positives and False Negatives** ML models can make errors by flagging legitimate trades as fraudulent (false positives) or failing to detect fraudulent transactions (false negatives).

- **Challenge**: A high false positive rate leads to unnecessary investigations, while false negatives allow fraud to go undetected.
- **Solution**: Optimizing models using techniques like precision-recall balancing, ensemble learning, and anomaly detection can reduce errors and improve model reliability.

7. **Privacy and Data Security Concerns** Stock market data is highly sensitive, and models need to comply with privacy laws and regulations.

- **Challenge**: Access to financial data for training ML models can be restricted due to privacy concerns.

- **Solution**: Privacy-preserving methods, such as federated learning and differential privacy, allow models to be trained without exposing sensitive information.

8. **Regulatory and Compliance Issues** Financial markets are heavily regulated, and ML models must align with legal and ethical standards.

- **Challenge**: Models must adhere to regulations such as GDPR, SEC guidelines, and financial fraud monitoring laws.
- **Solution**: Collaborating with regulators to develop AI models that meet legal Requirements and ensuring compliance can help address this challenge.

9. **Scalability and Computational Costs** Fraud detection models need to process vast amounts of data efficiently, which requires significant computational resources.

- **Challenge**: Large-scale ML models can be computationally expensive, increasing operational costs.
- **Solution**: Leveraging cloud-based AI platforms, distributed computing, and optimized algorithms can help improve efficiency and reduce costs.

Figure 6

References

1. **Aggarwal, C. C. (2017).** *Outlier Analysis.* Springer.
2. **Al-Jarrah, O. Y., Yoo, P. D., Muhaidat, S., Karagiannidis, G. K., & Taha, K. (2015).** "Efficient Machine Learning for Big Data: A Review." *Big Data Research, 2(3), Pg.* ***87-93****.*
3. **Fawcett, T., & Provost, F. (1997).** "Adaptive Fraud Detection." *Data Mining and Knowledge Discovery, 1(3), Pg.* ***291-316****.*
4. **Hochreiter, S., & Schmidhuber, J. (1997).** "Long Short-Term Memory." *Neural Computation, 9(8), Pg.* ***1735-1780****.*
5. **Li, J., Kong, X., & Yu, P. S. (2018).** "Financial Fraud Detection with

Convolutional Neural Networks." *IEEE Transactions on Systems, Man, and Cybernetics, 48(7), Pg.* ***1207-1219.***

6. **Lin, Y., & Hsu, J. (2020).** "Real-Time Stock Market Fraud Detection Using Machine Learning and Blockchain." *Journal of Financial Data Science, 2(4), Pg.* ***45-58.***
7. **Schumaker, R. P., & Chen, H. (2009).** "Textual Analysis of Stock Market Prediction Using Financial News." *ACM Transactions on Information Systems, 27(2), Pg.* ***1-19.***
8. **Venkatesh, A., & Ravi, V. (2019).** "Deep Learning Approaches for Fraud Detection in the Stock Market." *Expert Systems with Applications, 131, Pg.* ***285-295.***
9. **West, J., & Bhattacharya, M. (2016).** "Intelligent Financial Fraud Detection: A Comprehensive Review." *Computers & Security, 57, Pg.* ***47-66.***
10. **Yang, J., & Liu, Y. (2021).** "Stock Market Fraud Detection Using Machine Learning: Challenges and Solutions." *Financial Computing Journal, 3(2), Pg.* ***89-105****.*