Machine Language Advancements in Identification of Insider Trading

[1]Ms. Saumya Joshi, [2]Ms. Radhika Baid, [3]Mr. Samarth Jain, [4]Mr. Divyank Gupta

[1,2,3,4]*Department of Computer Science,MPSTME, NMIMS University*
Mumbai, India

Abstract: As Information Technology has progressed over time, issues like noise and multiple data sources have made it a tedious assignment to identify insider trading. A trading is known as illegal insider trading when people privy to the information which is not available to the general public indulge in trading based on this information. They may be people working in the company who have access to confidential information or their relatives, friends or others who have been tipped off about the information. Illegal insider trading might have been going on for decades, but these were going unnoticed in most of the cases. With the development of Machine Learning, it has become possible to design software that has made it easy to assess the data and identify illegal insider trading. In this review paper, we aim to compare the various models that researchers and experts have proposed over time and scale down to the most accurate ones in terms of variety in industry and information.

Keywords: Illegal Insider Trading; Legal Insider Trading; Artificial Intelligence, Machine Learning; Extreme Gradient Boost; SEBI; Decision Tree; Dense Neural Network; Random Forest Classifier.

I. INTRODUCTION

The securities market is an important component which has contributed significantly to the financial market. Illegal insider trading occurs frequently due to asymmetric information, inadequate structures of corporate governance, and poor regulatory mechanisms. [1] According to SEBI regulations, a "insider" is a connected individual or someone who has access to UPSI. That is unpublished price sensitive information. Insider trading in the stock market is the trading based on non-public information (not accessible to the general public). It can be both legal and illegal. Proper guidelines must be followed in order to conduct legal insider trading.[2] The following trading is based on non-public (private, leaked, or tip) information (such as a new product launch, quarterly financial condition, acquisition or merger proposal) before the information is made public. A linked person could be anyone who has had a relationship with the corporation during the six months preceding the insider trade. The source for the leaking of confidential company information could be a director of the company, an employee, a close family, a banker, a lawyer, or even a representative of the stock exchanges, trustees or employees who are an integral part of an asset management company who interacted with the company. [2, 3]

Insider trading hurts the integrity of capital markets. In stock markets, the playing field is leveled by the symmetric information,because it enables investors to compare how they interpret and analyze events. Insiders do, however, have an unfair advantage over regular investors when trading on UPSI.. [4]

Experts have curated an array of indicators to identify insider trading that will aid companies and organizations in combating the issues that follow such practices. Generally, insider trading stocks have distinguishing non-zero excess returns. In order to investigate the effect of insider trade on price, studies like Tavakoli et al.(2012) and Chang and Suk (1998) have worked with abnormal returns, whereas Aktas et al. (2008) have used daily average transaction prices. To monitor trades by insiders, These trades must be declared using Form 4 [5], which is mandated by the Securities and Exchange Commission of the United States (SEC). It is difficult to find unlawful trading among the numerous reported trades.

A. HISTORY AND EVOLUTION OF INSIDER TRADING

The earliest known instances of insider trading in India date to the 1940s, when government bodies like the Thomas Group of 1948 were established. This committee examined, among other things, US laws governing short swing profits under Section 16 of the Securities Exchange Act of 1934. [6] Following that, Sections 307 and 308 of the Companies Act, 1956 were added with provisions relating to insider trading, requiring shareholding statements from company directors and managers. [7]

The lack of adequate enforcement measures under the Companies Act of 1956 led to urge the creation of a separate statute addressing insider trading by the Sachar Committee, the Patel Committee and the Abid Hussain Committee during the years followed.[8, 9]

B. Insider Trading in India - Today

India was early to recognise the damage that insider trading may do to public shareholders' rights, Indian corporate governance, and the financial markets. The Thomas Committee was established in 1948 and assessed several international approaches for limiting insider trading, including the Securities Exchange Act of 1934, as the first actual attempt to regulate insider trading. Sections 307 and 308 of the Companies Act of 1956 were added as per the Thomas Committee's recommendation. [10] This modification made some disclosures by directors and management required, although it had little impact on the goal of curbing insider trading.According to the SEBI Act (Insider Trading) Regulations, "insiders" are forbidden from trading in exchange-listed securities on behalf of others using unpublished price sensitive information, communicating such information unless necessary for business operations, or giving advice to others using such information. [Figure 1] shows a visual representation of the registered investors in the Indian Subcontinent.

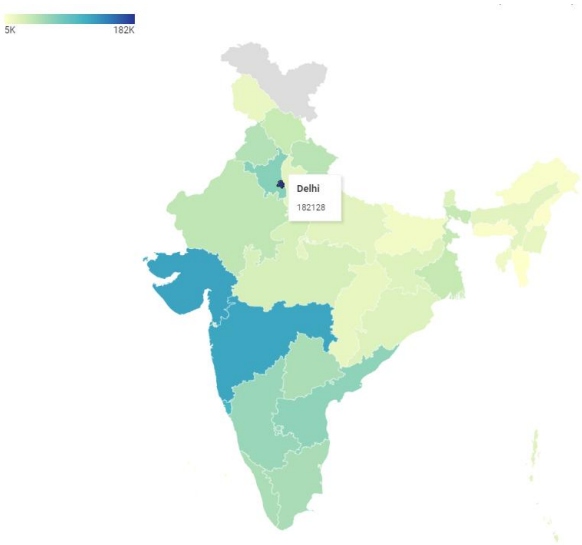

Figure 1- Registered investors and highest probability of insider trading Statewise

Trading or agreeing to trade as a principal or agent is referred to as "dealing in securities." [11] Tippees, or others who have received information from the insider, are not held accountable. In Figure 2, one can observe that in fiscal year 2017, approximately 15 cases were successfully completed, which was a negative rate from the previous fiscal year. [12]

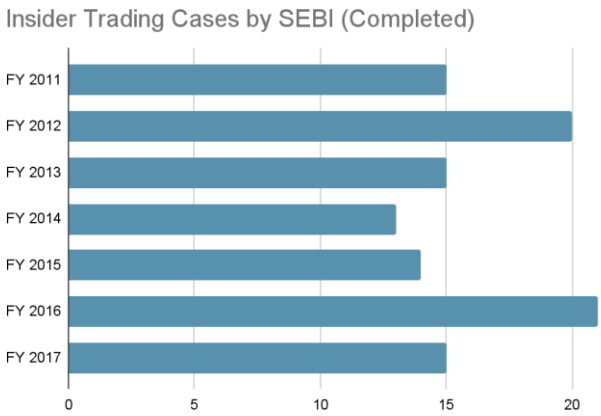

Figure 1 - Number of Insider Trading Cases completed by SEBI in India

II. IDENTIFICATION METHODS

Previously, preliminary detection of insider trading behaviors in different securities markets typically relied on artificial means, such as tracking the security account transactions of company managers and their families in China, as regulators may have believed that these individuals were the most likely to commit insider trading illegally and that their transaction data was therefore considered sufficient. However, given the state of the Chinese security markets today, those conventional methods turned out to be utterly ineffective. [13, 14]

Many academic scholars have discovered underlying relationships between insider trading activities and other associated variables, including stock price volatility and abnormal return. None of them proved to be reliable enough to produce results that could be put into action. The industry changed later as a result of the development of artificial intelligence and machine learning. [15]

A. Machine Learning Advancements

With the development of machine learning (ML) approaches over the last few decades, numerous academics have applied ML-based methods for automatic classification in a variety of application sectors. [16] In Machine Learning, a computer learns algorithms using training data to carry out particular tasks for performance improvement. ML is used to develop models that automatically detect various hazardous behaviors in the domain of insider threats. Since machine learning constantly learns from the data and updates the algorithms, its detection performance may be more precise and consistent. [17] In reality, we will examine the many ML models that different institutions have used to avoid insider trading in the sections that follow.

B. Literature Survey

- A classification technique called random forest produces many decision trees during training and outputs the class that corresponds to the average of the classes of the individual trees.. They are often used as "black box" models in business because they yield analytical predictions over a broad range of data with little configuration required.[18].
- A decision tree is a flowchart-like structure where each inner node represents a "test" for an attribute, each branch represents the result of the test, and each leaf node represents the result of the test[19].
- A set of parameters proportionate to the number of variables (features/predictors) in the training problem is needed for Naive-Bayes classifiers, which are extremely scalable.20].

- In the current meaning, an artificial neural network made up of artificial neurons or nodes is referred to as a neural network since it is a network or circuit of biological neurons. Artificial neural networks (ANNs) mimic biological neuronal connections as weights between nodes[21].
- The gradient boosting solution XGBoost pushes the limits of processing power for boosted tree algorithms. It is scalable and incredibly accurate. It was developed primarily to improve the efficiency and performance of machine learning models. [22]
- Support vector machines (SVMs) are helpful for categorizing text and hypertext because they significantly reduce the requirement for classified training periods in both the typical inductive and transductive contexts.Some strategies for shallow semantic parsing are primarily based totally on assist vector machines. The SVM set of rules has been broadly implemented withinside the organic and different sciences [23].
- The term "multi-objective optimization" refers to the process of determining the best values for the solutions to several desired goals. The MOO is chosen because it allows for optimization without the use of complex equations, which simplifies the issue.[24]
- The Goldberg-introduced idea of non-dominance is combined with the genetic algorithm (GA) in the non-dominated sorting genetic algorithm II (NSGA-II), an evolutionary algorithm. It has numerous variations due to the use of various crossover, evolution, and mutation operators. In many applications, the original NSGA-II has been used extensively to address multi-objective optimization issues.[25]
- The k-nearest neighbor (k-NN) algorithm is a nonparametric supervised learning method where An object is categorized based on the multiple votes of its neighbors, and the class with the highest frequency among its k closest neighbors where k is a positive integer and typically small is chosen to classify the object.[26].

C. Models Implemented in Identification

1.1 In the study cited [27], the researchers at Department of Computer Science, University of Texas at Dallas compared the performance of one-class SVM (Support Vector Machine) and two-class SVM to detect illegal trading. This was done after using two methods, namely updating the classifier ensemble and testing algorithm. One class SVM eventually performs better due to the benefit of learning a classification function using a relatively small training set.

1.2 A Deep Learning based illegal insider trading detection and prediction technique using the LSTM (long short-term memory network) was implemented. It is an assortment of recurrent neural networks (RNN) that can establish long-term dependencies, particularly in situations involving sequence prediction. This study employed our proposed algorithm [28] ANOMALOUS to find anomalous patterns in data originating from illegal insider trading, and it assisted in forecasting stock volume for a selected business. But there were a lot of false positives as a result.There is a scope of future research on this paper using different algorithms that are more effective with time series (e.g., GAM, ARIMA)

1.3 In this study, researchers combined extreme value theory with the Nearest Neighbor Dynamic Time Warping (NN DTW) pattern recognition algorithm. In this validation sample, the NN DTW model we develop successfully identifies 90% of all suspected illicit insider trades, significantly outperforming alternative approaches. Surpassing the best ARMA (1,1) Autoregressive Moving Average model detection method which detects about 60% of alleged illicit insider trades. Additionally, a 30% average false positive alert rate is achieved by the NN DTW model. [29]

1.4 Chen and He [Figure 3] suggested and created the tree-based boosting machine learning method known as XGboost. It is an enhanced technique built upon Friedman's gradient boosting decision. The XGboost can build boosted trees quickly and precisely, and it can use parallel tree boosting to quickly and accurately solve both classification and regression problems. [30]

1.5 Applying Random Forest (RF) is also an approach used to identify insider trading, which was proposed in the future scope of work cited [30] The Random Forest (RF) approach was then utilized, which is a tree-based method for classification or regression in which an individual tree from a set of decision trees votes to make an output decision. [31]

1.6 The work mentioned [32] created a combined technique using XGboost and NSGA-II for identifying insider trading cases, in which the XGboost was used for sample classification while the initial parameters were improved by the NSGA-II algorithm. Using features, the suggested approach XGboost-NSGA-II identified insider trading samples.

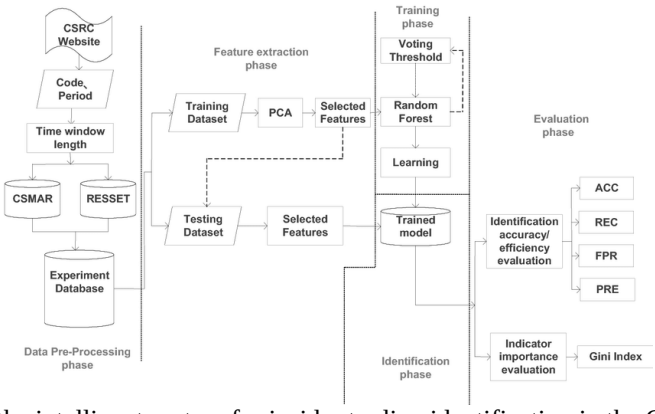

Figure 3 - Structure of the intelligent system for insider trading identification in the Chinese security market

D. Determinants - Feature Engineering

Several signs to help in the identification of insider trading are shed light on after carefully studying key researches. Stocks with insider trading often have excess returns that are notably nonzero. Insider trading results in wider bid-ask spreads for stocks and less liquidity on the securities market. Therefore, stock performance makes it easier for researchers to spot insider trading. Researchers that looked at insider trading in relation to corporate governance systems discovered that the extent of shareholders' rights is connected with the probability of insider trading. The probability of insider trading increases with increased shareholder rights.

In the paper cited [32], scholars worked with a dataset to realize that some indicators act as an alert more often than a confirmation. An increase in prices in a stock does not always signal insider trading. Even when combined with enormous volumes and other trading aberrations, it certainly raises suspicions, but insider trading cannot be proven. The instances in this dataset marked as positive should be considered "suspicious" or "worth reviewing" even though they have not been proven to be instances of insider trading.

III. RESULTS

A. Observations and Future Scope

Following a thorough review, the Multitask Deep Neural Network Model cited [30] is found to be the most accurate and promising in terms of the future scope of insider trading prevention, and the proposed model is carefully compared with regression, SVM, deep neural network (DNN), RF, and XGBoost models.

Multi-task learning (MTL) [Figure 4] is a subset of artificial intelligence in which a shared model is trained on many tasks at the same time. Such techniques provide benefits such as increased data efficiency, reduced overfitting via shared representations, and rapid learning via auxiliary information. However, there is room for improvement in the aforesaid approach.

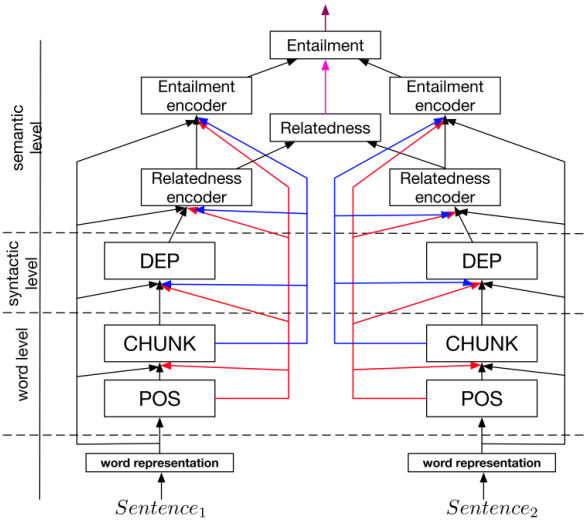

Figure 4 - Schematic Diagram of a Deep Neural Network in Deep Learning

Typical deep learning models use neural networks with a large number of nodes and several layers, resulting in a huge number of parameters that must be evaluated. This necessitates a large amount of data, which may or may not be present in every insider trading inquiry. As a result, we believe that the integrated XGBoost and NSGA-II cited [] will be a good alternative to the computational complexity that comes with MTL, because it can provide more robust and accurate results and is less prone to overfitting when compared to traditional machine learning techniques such as SVM, ANN, Adaboost, and RF. [Figure 5] It improves the initial indicators over simple XGBoost.

Window Length	ANN (%)	SVM (%)	Adaboost (%)	RF (%)	XGboost (%)	XGboost-GA (%)	XGboost-NSGA-II (%)
30-days	48.89	27.03	25	24.32	18.75	19.35	13.51
60-days	20	25.93	25	23.08	17.39	17.86	18.18
90-days	33.33	30.30	23.26	29.03	22.58	23.08	17.24
Average	34.07	27.75	24.42	25.48	19.57	20.10	16.31

Figure [5] - False positive rate results of insider trading identification in the time window of 30-, 60- and 90-days.

B. Conclusion

These findings should be beneficial for regulators and investors who are looking to understand insightful investors' trading activity, either to track down lawbreakers or to alter their own perceptions of the worth of a company's publicly listed stocks. It should also educate academics on how to structure future research that aims to elucidate indicators of informed trading as input on a related research topic or evidence of informed trading in situations that are yet to be identified.

REFERENCES

1. Li G, Li Z, Wang Z, Zhang K. Identification of Insider Trading in the Securities Market Based on Multi-task Deep Neural Network. Comput Intell Neurosci. 2022 Apr 22;2022:4874516. doi: 10.1155/2022/4874516. PMID: 35498190; PMCID: PMC9054416.
2. Securities and Exchange Board of India (Prohibition of Insider Trading) (Second Amendment) Regulations, 2021. (references)
3. Legal & Secretarial, " Code of Conduct for prohibition of Insider trading and Code for Fair Disclosure"
4. Sadbhav Infrastructure Project Limited, 16th Annual Report 2021-22

5. Form ISR-4, SEBI, dated January 25, 2022
6. Katarki, Suneeth. “India: ‘Mens Rea’ In Insider Trading – A ‘Sine Qua Non’?” Mondaq, 3 June 2015, www.mondaq.com/india/x/401724/Securities/Requirement Of Mens Rea As A Criterion for Penalizing Insider Trading In India.
7. Das, Sonakshi. “The Know-All of Insider Trading – Decades of Corruptive Prevention.” Academike, 15 Jan. 2015, www.lawctopus.com/academike/know-insider-trading-decades-corruptive-prevention/#_edn8.
8. Dhyey Vyas, Pooja Dhruve, “Insider Trading under SEBI(prohibition of insider trading) Regulation Act”
9. As defined under section 2(13), of Companies Act, 1956.
10. V.R. Maheswara Rao. Paleti, Dr. Sapna Bansal and Prof. Anu Solanki
11. “CRITICAL ANALYSIS ON ILLEGAL INSIDER TRADING AND ITS IMPACT ACTIONS TO CURB INSIDER TRADING IN INDIA”
12. SECURITIES CONTRACTS (REGULATION) ACT, 1956
13. Lee Biggerstaff a , David Cicerob , M. Babajide Wintokic, “Insider Trading Patterns”
14. Audemard, Gilles & Bellart, Steve & Bounia, Louenas & Koriche, Frederic & Lagniez, Jean-Marie & Marquis, Pierre. (2021). Trading Complexity for Sparsity in Random Forest Explanations.
15. Steve Donoho, “Early Detection of Insider Trading in Option Markets”
16. Quinlan, J.R., C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, 1993.
17. Deshpande, Rucha & Lunkad, Divyank & Kunjir, Madhura & Ingle, Samiksha. (2022). Stock Price Prediction and Analysis using Machine Learning Techniques. International Journal for Research in Applied Science and Engineering Technology. 10. 2092-2098. 10.22214/ijraset.2022.42747.
    RAMESHBABU, V & VIJAYAKUMARAN, C & PRABHAKAR, P.. (2022). Machine Learning. 10.47715/JPC.B.82.2022.9789391303372.
18. Breiman L (2001). "Random Forests". *Machine Learning*. **45** (1): 5–32. Bibcode:2001MachL..45....5B. doi:10.1023/A:1010933404324.
19. Kamiński, B.; Jakubczyk, M.; Szufel, P. (2017). "A framework for sensitivity analysis of decision trees". *Central European Journal of Operations Research*. **26** (1): 135–159. doi:10.1007/s10100-017-0479-6. PMC 5767274. PMID 29375266.
20. Zhang, Harry. *The Optimality of Naive Bayes* (PDF). FLAIRS2004 conference
21. Sanborn, Sophia & Shewmake, Christian & Olshausen, Bruno & Hillar, Christopher. (2022). Bispectral Neural Networks. 10.48550/arXiv.2209.03416.
22. Jones, Katelinh & Ong, Yuya & Zhou, Yi & Baracaldo, Nathalie. (2022). Federated XGBoost on Sample-Wise Non-IID Data. 10.48550/arXiv.2209.01340.
23. Platt, John C. (1998). *Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines* (PDF). NIPS. Archived (PDF) from the original on 2015-07-02.
24. Nyoman Gunantara | Qingsong Ai (Reviewing editor) (2018) A review of multi-objective optimization: Methods and its applications, Cogent Engineering, 5:1, DOI: 10.1080/23311916.2018.1502242
25. Doerr, Benjamin & Qu, Zhongdi. (2022). A First Runtime Analysis of the NSGA-II on a Multimodal Problem. 10.1007/978-3-031-14721-0_28.
26. Altman, Naomi S. (1992). "An introduction to kernel and nearest-neighbor nonparametric regression" (PDF). *The American Statistician*. **46** (3): 175–185. doi:10.1080/00031305.1992.10475879. hdl:1813/31637
27. Supervised Learning for Insider Threat Detection Using Stream Mining by Pallabi Parveen, Zackary R Weger, Bhavani Thuraisingham, Kevin Hamlen and Latifur Khan, Department of Computer Science, University of Texas at Dallas
28. A Deep Learning Based Illegal Insider-Trading Detection and Prediction Technique in Stock Market by Sheikh Rabiul Islam, Computer Science, Tennessee Technological University, Cookeville, TN, U.S.A
29. A Machine Learning Attack on Illegal Trading∗, Robert James† Henry Leung‡ Artem Prokhorov§, 1 April 2022
30. Identification of Insider Trading in the Securities Market Based on Multi-task Deep Neural Network, Guofeng Li , Zuojuan Li , Zheji Wang , and Ke Zhang
31. Shuxiang Xu; Ming Zhang ,"Data mining - an adaptive neural network model for financial analysis",Information Technology and Applications, 2005. ICITA 2005. Third International Conference on Volume 1, Issue , 4-7 July 2005 Page(s): 336 - 340 vol.1
32. Glasser J., Lindauer B. Bridging the Gap: A Pragmatic Approach to Generating Insider Threat Data; Proceedings of the 2013 IEEE Security and Privacy Workshops; San Francisco, CA, USA. 23–24 May 2013; New York, NY, USA: Institute of Electrical and Electronics Engineers (IEEE); 2013. pp. 98–104.