Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 1 (2024)

Detecting Accounting Fraud in Publicly Traded Firms Using a Machine Learning Approach with Live Implementation

Siddharth Nanda[1*], Dr. Vinod Moreshwar Vaze[2]

[1*]Research Scholar, Department. of Computer Science and Engineering, Shri JJT University, Jhunjhunu, Rajasthan, India.

[2]Guide, Department of Computer Science and Engineering, Shri JJT University, Jhunjhunu, Rajasthan, India.

Abstract:

In this research, the authors use a machine learning-based approach to creating a state-of-the-art fraud prediction model. It is beneficial to construct models using a combination of domain knowledge and machine learning techniques. The paper provides a literature review and classification for the investigation of data mining's function in identifying instances of financial fraud. In this paper, researchers used some techniques for detecting accounting fraud, i.e., Random Forest, Genetic Algorithm, Decision Tree, and Convolutional Neural Network (CNN). A comprehensive literature research on Financial Fraud Detection (FFD) has not been conducted despite the reality that it is a rapidly developing field of critical relevance. Financial accounting fraud has increased, making financial accounting fraud detection (FAFD) an issue of major relevance in academia, research, and industry due to the current financial environment. Forensic accounting techniques are now used to identify instances of financial accounting fraud due to the lack of effectiveness of the company's business auditing system. Data mining methods are proving to be of considerable assistance in the identification of fraud in financial accounting due to the difficulties of forensic accounting in handling high data quantities and the complexity of financial data. As a result, it can be used for a wide variety of things, including machine learning, web development, and software testing. Performing a visual representation of genuine data and fraudulent data, which are indicated by 0 and 1.

Keywords: Financial Fraud Detection (FFD), Machine Learning, Random Forest, Genetic Algorithm, Decision Tree.

1. Introduction

Financial statement fraud occurs throughout the globe. Significant damage could be inflicted not only on the stakeholders of fraudulent organizations (like Enron and WorldCom) but also on the indirect stakeholders of numerous non-fraudulent enterprises if fraud is not recognized and stopped promptly. Consequently, it could be difficult to identify fraudulent activity in accounting. Furthermore, by the time it is discovered, significant harm has already been done. Therefore, regulators, auditors, and investors might all benefit greatly from more efficient and effective techniques for detecting corporate accounting fraud [1].

Accounting fraud is committed by creating fabricated financial accounting statements in which the numbers are changed by overstating resources, producing false profits and sales entries, misappropriating taxes, or understating liabilities, debts, costs, or losses. Accounting professionals describe fraudulent activity as the deliberate distortion of a company's financial position to make its accounting information look more favourable.Financial fraud is a growing issue in today's economy, and accounting professionals have long faced a crucial but difficult challenge in recognizing instances of fraud in the accounting system [2].

There is substantial evidence that book cooking accounting procedures are widely used around the globe to commit financial crimes, making internal audits of financial problems in firms an increasingly time-consuming and difficult task. Using traditional means of internal auditing to uncover financial statement fraud is a challenging task, and in certain cases, it could be impossible. First, auditors do not have sufficient expertise in recognizing the signs of accounting fraud. Second, most auditors don't have the knowledge and competence to identify and prevent fraud since the illegal manipulation of accounting information is so rare [3].

The detection of accounting frauds in publicly [4] listed organizations is one of the most exciting and demanding applications of Machine Learning (ML) in computational finance [5]. Publicly traded companies have a global issue with accounting fraud committed through insiders (i.e., management and controlling shareholders). Non-fraudulent

businesses might be harmed by fraud because they typically must compete with fraudsters for limited investor funds and consumer spending. The ambiguity of the financial market regarding the presence of frauds could also affect the proper operation of financial markets and economic development in a nation because of the knowledge asymmetry between business insiders and outside investors [6]

1.1 Machine Learning-based Fraud Detection

Fraudulent financial transactions are becoming more problematic in the Financial Services Industry (FSI). Flag, and analyse these actions to identify large financial resources are required. The amount and speed of fraudulent online transactions and the urgency with which they must be addressed have both increased in parallel with the rise in popularity of doing business online. ML could provide FSIs with the capacity to equip their fraud decision-makers with the ability to make educated judgments to prevent fraud before it damages the bottom line of the company and the entire brand. Financial fraud is currently very difficult to identify, and it takes a lot of time and effort in the form of manual forensic accounting effort. Numerous AI and ML techniques, such as decision forests or artificial neural networks, have been used in the fraud detection and auditing area to minimize the amount of human labour required and to advise financial auditors [7]. The challenge of financial fraud detection, however, remains in the fact that most real-world datasets suffer from a high-class imbalance. This is because ML algorithms perform best when given a large and equal number of observations for every class to be predicted [8]. Figure 1 depicts the detection of accounting fraud with ML.

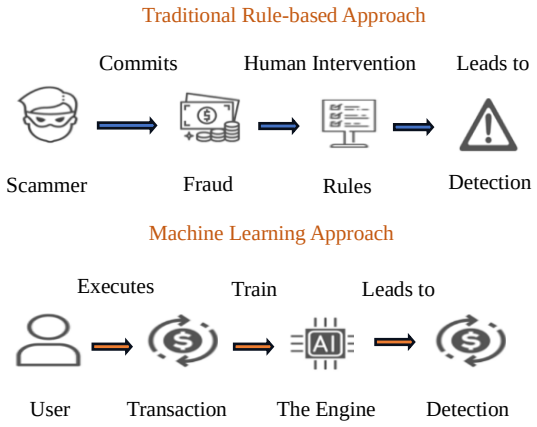

Fig 1: Detection of accounting fraud with ML [9]

2. Literature Review

The following study expands on detecting accounting fraud in publicly traded firms using an ML approach. Several researchers explained their discoveries, as seen below.

Huang et al., (2022) [10] stated that the development of emerging technology, such as data analytics and ML, is influencing the field of accounting. Audit and assurance practices are expected to change considerably because of these effects. There could be a transition into techniques, including audit sampling. Audit sampling is becoming more irrelevant in the age of big data because it is based on such a limited portion of the population. One feasible strategy is to use audit data analytics and ML to evaluate the population rather than a collection of transactions. The paper explains a method for combining audit data analytics and ML into comprehensive population testing and resolves the associated challenges.

Pranto et al., (2022) [11] suggested a blockchain and smart contract-based method for developing a powerful ML algorithm for detecting fraudulent purchases online. The suggested solution takes advantage of blockchain to secure data privacy. A smart contract implemented inside the network completely facilitates the system. Throughout eight simulations, the model maintained a testing accuracy of 98.93% and a Fbeta score (recall-biased f measure) of 98.22%. The results demonstrate that blockchain mining time is affected by both the amount of data and the difficulty level. There is a positive association between mining time and difficulty level for difficulties less than 5. Their system's block extraction time for levels 2 and 3 of difficulty is less than one second.

Yang Li et al., (2022) [12] evaluated a revolutionary deep-learning strategy to predict the future of stock prices. As a research study for the research, they would be using the S&P 500 Index. The processing ensemble deep learning model significantly outperforms the best existing prediction model using the same dataset by reducing the mean-squared error from 438.94 to 186.32, a 57.55% reduction, increasing the precision rate by 40%, recall by 50%, the F1-score by

44.78%, and the movement direction accuracy by 33.34%. The goal of this research is to explain the development model and demonstrate that ensemble deep learning technologies are superior to more conventional approaches in predicting future stock price patterns and assisting investors in making the most effective investment decisions.

Lokanan et al., (2022) [13]found an ML system for predicting fraud in the Canadian securities sector using data obtained from the Investment Industry Regulatory Organization of Canada (IIROC). The amount spent on the transaction and whether the offender was from a bank-owned investment business were the most significant predictors of fraud based on the standardized coefficients from four different ML models. Portfolios with recurrent losses should raise red flags for branch managers and regulators, who should investigate immediately. The results provide more clarity to the self-regulation of Canada's financial markets and are of relevance to authorities exploring creative and efficient fraud detection approaches.

Liao et al., (2022) [14] calculated a fraud detection model for publicly traded businesses using information obtained from the SEC's Accounting and Auditing Enforcement Releases (AAERs). Furthermore, a Nonlinear Activated Beetle Antennae Search (NABAS) method, a version of the meta-heuristic optimization technique Beetle Antennae Search (BAS), is used to solve this computational finance model. Initial steps include remaking the fraud detection model as a loss function minimization optimization problem, which is then solved using the NABAS algorithm. Finally, experimental results demonstrate that the NABAS algorithm improves comparable techniques in detecting fraud in publicly available datasets in terms of both accuracy and efficiency.

Celal Uslu et al., (2021) [15]stated that capital market instrument trade-based manipulation is the subject of this investigation. The data was collected from 22 cases of manipulation in Borsa Istanbul (BIST) between 2010 and 2015. The researchers present an ML strategy based on supervised ML classification models to identify trade-based manipulation in the regular data of manipulated stocks. The results of this research demonstrate that supervised ML algorithms are effective in identifying trade-based manipulations in trading networks using accuracy, sensitivity, and F1 score as measures of performance. The research shows that the suggested approach is 91% accurate, has 95% sensitivity, and 93% specificity when it is used to identify instances of market manipulation.

Zhang et al., (2021) [16] computed that the primary contribution is a system for detecting fraudulent transactions, which makes use of a deep learning architecture and an innovative feature engineering method based on Homogeneity-Oriented Behaviour Analysis (HOBA). They evaluate the efficacy of the suggested framework through a comparative analysis using a real-world dataset from one of the major commercial banks in China. Experiment results indicate that the suggested technique is an efficient and practicable method for identifying credit card fraud. If they compare their suggested technique to the benchmark methods, they observe that it can detect more fraudulent transactions while maintaining a false positive rate that is acceptable to everybody.

Eachempati et al., (2021) [17] illustrated that the study attempted to collect sentiments about transparency data so that its effects on financial assets can be evaluated. The study utilizes a prediction model based on deep neural networks to do sentiment analysis on data sets that are not consistent. They develop a sentiment simulation model of voluntary disclosures to determine whether managers can utilize market sentiment as a calculated input to improve market presentation through appropriately designing the tone and substance of voluntary disclosures without sacrificing quality or truthfulness. The Deep Neural Networks with the LSTM method is shown to be superior to both the Deep Neural Networks with RNN and the other basic ML classifiers when comparing the predicted accuracy of the NSE NIFTY50.

Sailusha et al., (2020) [18] evaluated that to focus on ML approaches. The Adaboost and random forest algorithms are used in this project. The parameters utilized to assess the performance of the two techniques are precision, recall, accuracy, and F1-score. The ROC curve is generated using the confusion matrix. The optimal method for detecting fraud is determined by comparing the Random Forest and AdaBoost algorithms and selecting the one with the highest accuracy, precision, recall, and F1 score. Random Forest performs better in terms of precision, recall, and F1-score when compared to the Adaboost algorithm. Therefore, they derive the conclusion that the Random Forest Method is more effective than the Adaboost algorithm in detecting credit card fraud.

Dhiman Sarma et al., (2020) [19] found that bank fraud is a federal offense that includes the deception of financial institutions to get monetary advantages. Consistently, fraudsters use avoidance strategies that allow them to get over anti-fraud security systems. In this study, authors present a method for detecting occasions of bank fraud by using a community detection algorithm to detect the types of patterns that must identify fraud. The web-based application developed to identify fraud was developed using an agile method. The program served as a centre between the financial institutions and their clients. The database was constructed and represented using the graph database Neo4j and accessed through the graph query language Cypher. The suggested technique identified all frauds presented during the experiment.

3. Background Study

This research looks at trade-based manipulations of capital market assets. A total of 22 instances of modification in Borsa Istanbul (BIST) were collected for the study between 2010 and 2015. To identify trade-based modification in the regular information of manipulated stocks, authors offer an ML strategy based on supervised ML classification models. This research shows that supervised ML algorithms are effective in identifying trade-based manipulations in trading networks using accuracy, sensitivity, and the F1 score as measures of performance [20].

4. Problem formulation

One of the most significant issues faced by most businesses, especially those operating in the banking, finance, retail, and online shopping industries is detecting fraudulent activity. A company's financial line, reputation, and capacity to recruit and keep customers are all negatively impacted by fraud. However, accounting fraud persists despite intensive regulations imposed by both internal and external auditors and several statutory reforms. The purpose of this research is to evaluate existing methods for detecting accounting fraud. The manipulation of financial records is a difficult and extensive social issue with a convoluted remedy. To make improvements to detection, it is required to have improved assistance, such as fraud detection models. Explanatory variables are vital to creating such models, but earlier financial fraud detection research has not standardized the selection method.

5. Research Objectives

- To create a state-of-the-art fraud prediction model using the ML approach.
- To identify fraud, one needs to know what sort of fraud one is looking for, what signs of fraud one should be on the lookout for, and in what areas fraud could be prevalent.
- To discover instances of financial accounting fraud caused by the ineffectiveness of the company's corporate auditing system by using Forensic accounting methods.

6. Techniques Used

In this section there are three techniques used in the research methodology of detecting accounting fraud, i.e., Random Forest, Genetic Algorithm, and Decision tree.

6.1Random Forest

This is one of the ensemble approaches that is exclusively utilized to increase the success and accuracy of ML algorithms in artificial intelligence. A random forest technique could also aid in identifying the relevant independent variables and allowing the system to choose functionality. The information obtained from the trees is used to create the most accurate projections possible. A single decision tree can only reach one decision and has only a select few groups from which to choose, but a forest of decision trees could guarantee a more precise result since they consider a bigger number of categories and options. In addition to the benefits of injecting randomization into the model, this technique also offers the added benefit of picking the best feature from a pool of characteristics chosen at random [21]. Figure 2 demonstrates a classification and regression decision tree for dependent variables.

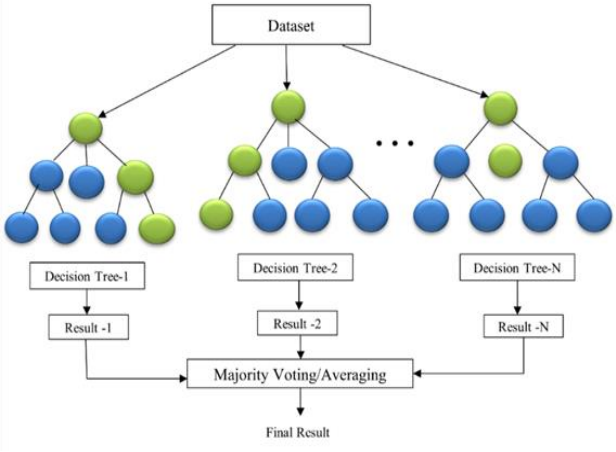

Fig 2: The architecture of Random Forest

6.2Genetic algorithms

Genetic algorithms are a subset of the class of algorithms known as evolutionary algorithms. Evolution strategies (ES), Genetic Programming (GP), Genetic Algorithms (GA), and Evolutionary Programming (EP) are all sorts of algorithms based on natural evolution. The algorithms of this class are all based on a group of people. Evolving algorithms have been used to resolve a broad variety of management challenges involving inventory management and distribution schedules. Figure 3 depicts the process of the Genetic Algorithm.

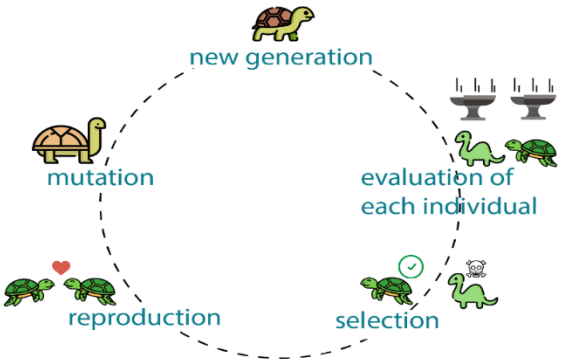

Fig 3: Process of Genetic Algorithm [22]

The discipline of materials science and manufacturing can benefit greatly from the usage of genetic algorithms. Reproduction, crossover, and mutation are all examples of genetic operators in a genetic algorithm. For a given problem class, it is worthwhile to experiment with factors,

for example, the crossover probability, mutation probability, and population size [23].

6.3Decision Tree

In ML, a decision tree structure is a predictive model that adds observable information about a phenomenon to predictions about the phenomenon's goal value. Decision tree learning is an ML approach for generating a decision tree from data, and it is one of the most familiar "data mining techniques". Each node, which corresponds to a variable, and each arc, which corresponds to a child, represents a potential value for that variable. The projected value of the target variable is represented by a "leaf node, with the values of the variables denoted by the route from the tree root to that leaf node". The leaves of a tree symbolize clusters, while the branches indicate seasonal combinations of characteristics that result in clusters. Dividing a resource set into subgroups centered on a characteristic value test might be used to learn a tree. In each subfolder created by the separation, this procedure is performed recursively. When the partition is no longer advantageous or when one class can be utilized for all the samples in the subclass, the return process is completed. Figure 4 shows the decision tree [24].

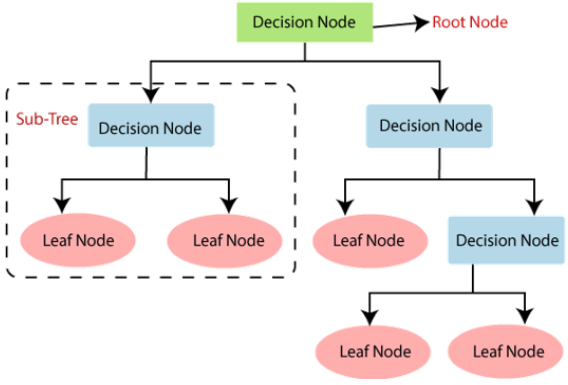

Fig 4: Decision Tree

7. Proposed Methodology

In the proposed methodology initially training dataset is taken as an input. After data pre-processing occurs, it divides the data into two variables: financial variables and linguistic variables. These variables generate the sequence pattern. If the pattern has any variation, then it updates the model; otherwise continues with the same model. The same procedure

occurs in the testing module to check the accuracy of the model. The proposed methodology is shown in figure 5 and their steps are discussed below.

1. Data Collection

In the proposed methodology, the analysis of the dataset will be done using the Python tool. For the analysis, there is a requirement for primary data and secondary data.

a) Primary dataset: The first step in every research effort is to analyze the data collected, known as primary data analysis. Primary data analysis refers to the steps used to make sense of raw data to conclude the study's assumptions and questions.

b) Secondary dataset: The term "secondary data analysis" refers to the process of analyzing information that has already been obtained by someone else. By using secondary analysis, scholars could save time and resources while still researching under-represented communities using massive datasets.

2. *Data Pre-processing*

In data mining, data pre-processing is the process of cleaning and organizing data in preparation for further analysis. A dataset is compiled, and then the information is pre-processed and cleaned. Numerous techniques exist for cleaning and organizing raw data before analysis. These features are not informative enough to be included in the dataset. The instrument for various types of occurrences in the subsequent stage. A system detects and discards information that does not belong in the dataset. The data can't be utilized without first being parsed to extract certain features. Typically, the data would go through several pre-processing steps before the characteristics are selected and extracted.

3. Financial Variables

Financial indicators were selected for the ability to identify monetary irregularities that could point to statement fraud. Financial statement fraud detects different forms, such as the intentional or negligent underreporting of revenue, assets, expenses, or debts. The chosen financial variables must include all aspects of a company's financial execution.

4. Linguistic Variables

The "Management Discussion and Analysis (MD&A)" section of annual reports in terms of linguistic aspectssince it provides investors with superior qualitative data about a company's performance and developments from the perspective of the management.

5. Sequencing

Each fraudulent transaction shares several transactional features. There is no correlation between the sequence of these trade features and the transaction's meaning, but different feature combinations would have different impacts on the model after the convolution process. As a result, the model has a layer for the sequencing of features.

Each transaction feature is available globally, ensuring that all possible feature configurations can be implemented. Therefore, the nodes between the first input layer and the last input layer are entirely connected, but the link weight changes with each iteration.

6. Convolutional Neural Network (CNN)

A CNN is a "Deep Learning system" that can take an image as input, value distinct features and objects in the picture, and distinguish between them. The Feature Sequencing Layer is part of the structure of a CNN network. The network structure of this model is distinct from that of the CNN model currently in use in that it has a feature sequencing layer. The network was set up so that it could be used to network transaction data and so that information about online transactions could be found quickly. The architecture consists of four alternating scattering convolutional layers, two pooling convolutional layers, a fully connected layer, and a feature sequencing layer.

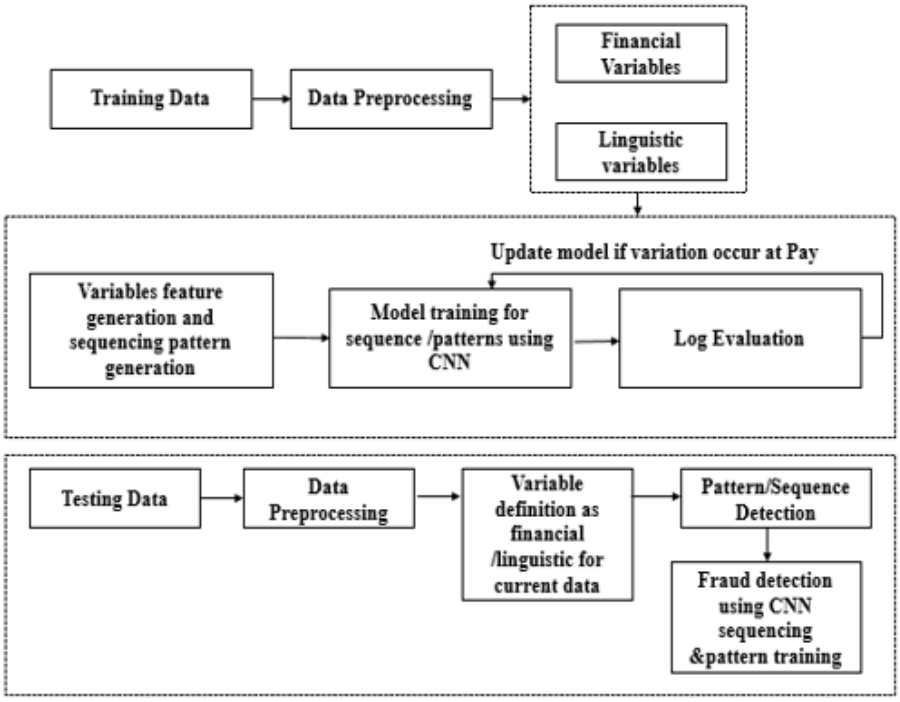

Fig 5: Proposed Methodology

8. Implementation and Results

This section of the research details the implementation carried out using the suggested technique, and the implementation tools are provided below.

8.1 Dataset Collection

In this section, the authors used the synthetic financial dataset which datasets are sparse in the field of finance, especially in the relatively new field of mobile money transactions. Financial datasets are crucial to many academics, including those doing fraud detection studies. The fact that monetary transactions are often conducted in secret contributes to the lack of openly accessible statistics. As a solution to this kind of issue, researchers provide a synthetic dataset developed with the help of a simulator called PaySim. The private dataset is aggregated and used by PaySim to create a synthetic dataset into which harmful behavior is inserted to test the efficacy of fraud detection algorithms.PaySim is a mobile money transaction simulator that models transactions using data taken from actual financial logs of a mobile money service operating in an African nation over one month. Mobile financial services are presently available in more than 14 countries throughout the globe, and the initial logs were given by the firm responsible for providing these services. This artificial dataset is specially designed for Kaggle and is scaled down to 1/4 the size of the actual dataset.

8.2 Tools Used

In this research, the authors used the Python tool to obtain the results. In recent years, Python has risen to become one of the most popular programming languages used all over the globe. It has applications ranging from ML to the construction of websites and the testing of software. Both developers and non-developers could utilize it. The following are the results implemented to support the proposed work discussed in given below.

8.3 Performance Analysis

8.3.1 Representation of Fraudulent data and Genuine Data

Performing a visual representation of the data in step one of the results. Figure 6 shows the value for genuine data to be 284315, while fraudulent data has a value of 492. The value 0 represents the protected data, which is referred to as Genuine data, while the number 1 represents fraudulent data.

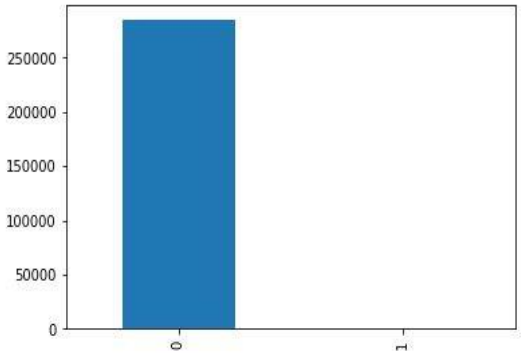

Fig 6: Visual representation of 0 and 1

8.3.2 Components Analysis of 0 and 1

The main component analysis for this result is carried out in the manner shown in Figure 7. The entire objective provided by principal component analysis is visualization, and this visualization depends on the classes being completely out of balance with one another. After that, the courses were subdivided into the 0 class and the 1 class.

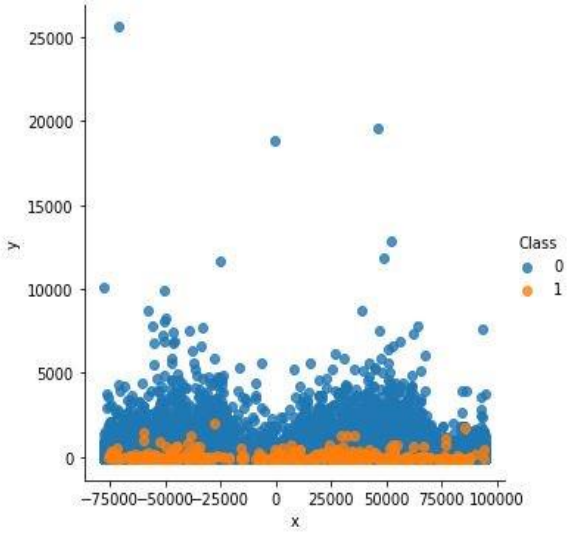

Fig 7: Visualization of component analysis.

8.3.3 Transactions of Genuine data and Fake data

In this result of the paper, the authors would discuss the number of transactions for both fake and genuine data, which is depicted in Figure 8, following the time and variations.

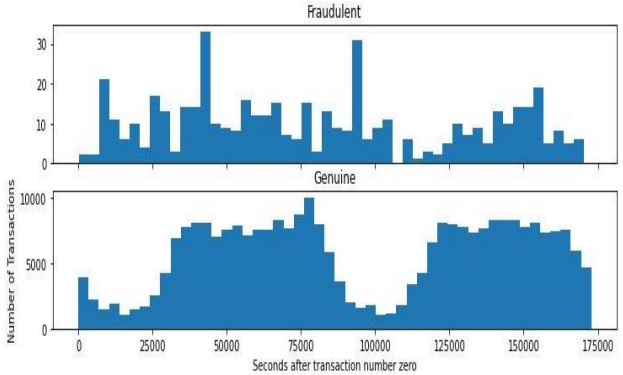

Fig 7: Genuine data and Fake data

8.3.4 Correlation between time and variance

In this result, the time and variances were used to evaluate the correlation between the features and the likelihood of the transaction being fraudulent on the unbalanced dataset. The imbalance in the dataset, as shown, prevents a proper visualization of the correlations that would otherwise exist. This is because of the significant imbalance between the classes, which influences the correlation matrix. First, let's make sure our classes are balanced, and then the author will examine the correlation matrix. Figure 9 depicts the correlation matrix.

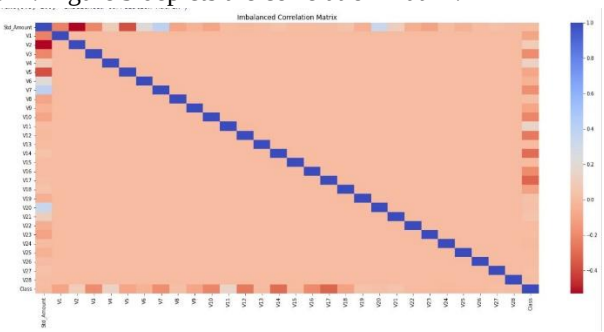

Fig 8: Correlation Matrix.

8.3.5 Classification of the Model

In the section that tests from the matrix, both the model's accuracy and its loss have been examined. In addition, the matrix can be observed below. Because the figure demonstrates very clearly that the model has the maximum accuracy and that its value has reached 0.97, which is the number that is closest to 1, this demonstrates that our model is successful and verified. Figure 10 shows the accuracy matrix.

	precision	recall	f1-score	support
0	0.50	1.00	0.67	99
1	0.00	0.00	0.00	98
accuracy			0.50	197
macro avg	0.25	0.50	0.33	197
weighted avg	0.25	0.50	0.34	197

Fig 9: Accuracy Matrix.

8.3.6 Training and Validation Accuracy of Model

The model achieves a training accuracy of 0.97 and a validation accuracy of 0.90. The accompanying graph illustrates the model's accuracy trends on both the training and validation datasets as a function of the epoch. Notably, the accuracy stabilizes around 20 epochs, indicating limited improvement beyond this point.

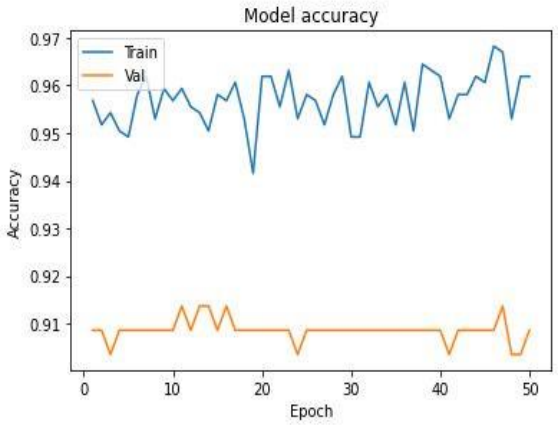

Fig 10: Training and Validation accuracy of the model.

8.3.7 Training and Validation Loss of Model

The training loss of the model is 0.11, while the validation loss is 0.22. The graph itself appears to show the model's accuracy on the training and validation datasets over time, plotted as a function of the epoch. In this case, the model's loss appears to have stabilized after around 20 epochs, with no significant improvement beyond that point.

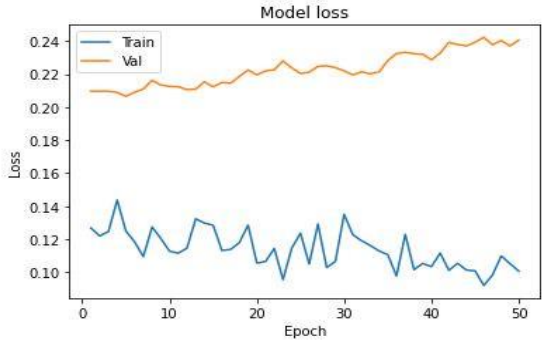

Fig 11: Training and Validation Loss of the model

9. Conclusion and Future Scope

It is extremely difficult to identify fraud instances in the accounting system. Accounting research should focus on the early detection of corporate accounting fraud to limit damages. The purpose of this research was to develop a better method for determining the probability of financial statement fraud. The approach includes financial and non-financial risk categories and a hybrid evaluation technique. In this research, authors explored the difficult issue of imbalanced learning with massive data, which develops in the domains of Data Mining and Machine Learning. This method utilizes state-of-the-art blockchain technology to establish a platform for gradually and cooperatively training fraud detection ML models while respecting participating organizations' security. The author of this study uses several methods, including a decision tree, a random forest, a genetic algorithm, and others, to identify instances of fraud. In the results, using the Python tool, the author presents a graphical representation of the difference between genuine data (specified as a 0) and fraudulent data (specified as a1), respectively. On the unbalanced dataset, the time and variances were utilized to assess the connection between the attributes and the probability of the transaction being fraudulent. In the future, it is possible to assess the effectiveness of the examined strategies by using data from other nations.Finally, studies in the future will examine to use of machine learning to identify offenders. Eventually, it would be useful to consider when the fraud was uncovered so that undetected frauds from previous years can be uncovered retrospectively.

10. Acknowledgement

In our research article, "Design and Optimization of Heat Transfer Fins for Enhanced System Performance," we extend our appreciation as the research behind it would not have been possible without the exceptional support of my supervisor,Dr. Arun Kumar, Associate professor. Department of Mechanical Engineering NITPatna. His enthusiasm, knowledge, and exacting attention to detail have been an inspirationandkept my work on trackfrom my first encounterto the final draft of this paper. The key contributors and our research facility of the Department of Mechanical Engineering NIT Patna facilitated its completion. There are no conflicts of interest among anyone. Our gratitude extends to the broader research community for their contributions, enriching our literature review, specifically in next-generation capability models.

References

[1]. Bao, Y., Ke, B., Li, B., Yu, Y.J. and Zhang, J., 2020. Detecting accounting fraud in publicly traded US firms using a machine learning approach. *Journal of Accounting Research*, *58*(1), pp.199-235.

[2]. Ngai, E.W., Hu, Y., Wong, Y.H., Chen, Y. and Sun, X., 2011. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision support systems*, *50*(3), pp.559-569.

[3]. Sharma, A. and Panigrahi, P.K., 2013. A review of financial accounting fraud detection based on data mining techniques. *arXiv preprint arXiv:1309.3944*.

[4]. Li, B. and Hoi, S.C., 2014. Online portfolio selection: A survey. *ACM Computing Surveys (CSUR)*, *46*(3), pp.1-36.

[5]. Abbasi, A., Albrecht, C., Vance, A. and Hansen, J., 2012. Metafraud: a meta-learning framework for detecting financial fraud. *Mis Quarterly*, pp.1293-1327.
[6]. Li, B., Yu, J., Zhang, J. and Ke, B., 2016, February. Detecting accounting frauds in publicly traded US firms: A machine learning approach. In *Asian Conference on Machine Learning* (pp. 173-188). PMLR.
[7]. Song, X.P., Hu, Z.H., Du, J.G. and Sheng, Z.H., 2014. Application of machine learning methods to risk assessment of financial statement fraud: Evidence from China. *Journal of Forecasting*, *33*(8), pp.611-626.
[8]. Hasanin, T. and Khoshgoftaar, T., 2018, July. The effects of random undersampling with simulated class imbalance for big data. In *2018 IEEE international conference on information reuse and integration (IRI)* (pp. 70-79). IEEE.
[9]. https://piexchange.medium.com/fraud-detection-with-machine-learning-a-use-case-6866e95ca982
[10]. Huang, F., No, W.G., Vasarhelyi, M.A. and Yan, Z., 2022. Audit data analytics, machine learning, and full population testing. *The Journal of Finance and Data Science*, *8*, pp.138-144.
[11]. Pranto, T.H., Hasib, K.T.A.M., Rahman, T., Haque, A.B., Islam, A.N. and Rahman, R.M., 2022. Blockchain and Machine Learning for Fraud Detection: A Privacy-Preserving and Adaptive Incentive Based Approach. *IEEE Access*, *10*, pp.87115-87134.
[12]. Li, Y. and Pan, Y., 2022. A novel ensemble deep learning model for stock prediction based on stock prices and news. *International Journal of Data Science and Analytics*, pp.1-11.
[13]. Lokanan, M.E. and Sharma, K., 2022. Fraud prediction using machine learning: The case of investment advisors in Canada. *Machine Learning with Applications*, *8*, p.100269.
[14]. Liao, B., Huang, Z., Cao, X. and Li, J., 2022. Adopting nonlinear activated beetle antennae search algorithm for fraud detection of public trading companies: a computational finance approach. *Mathematics*, *10*(13), p.2160.
[15]. Uslu, N.C. and Akal, F., 2022. A machine learning approach to detection of trade-based manipulations in Borsa Istanbul. *Computational Economics*, *60*(1), pp.25-45.
[16]. Zhang, X., Han, Y., Xu, W. and Wang, Q., 2021. HOBA: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture. *Information Sciences*, *557*, pp.302-316.
[17]. Eachempati, P., Srivastava, P.R., Kumar, A., Tan, K.H. and Gupta, S., 2021. Validating the impact of accounting disclosures on stock market: A deep neural network approach. *Technological Forecasting and Social Change*, *170*, p.120903.
[18]. Sailusha, R., Gnaneswar, V., Ramesh, R. and Rao, G.R., 2020, May. Credit card fraud detection using machine learning. In *2020 4th international conference on intelligent computing and control systems (ICICCS)* (pp. 1264-1270). IEEE.
[19]. Sarma, D., Alam, W., Saha, I., Alam, M.N., Alam, M.J. and Hossain, S., 2020, July. Bank fraud detection using community detection algorithm. In *2020 second international conference on inventive research in computing applications (ICIRCA)* (pp. 642-646). IEEE.
[20]. Uslu, N.C. and Akal, F., 2022. A machine learning approach to detection of trade-based manipulations in Borsa Istanbul. *Computational Economics*, *60*(1), pp.25-45.
[21]. Patil, S., Nemade, V. and Soni, P.K., 2018. Predictive modelling for credit card fraud detection using data analytics. *Procedia computer science*, *132*, pp.385-395.
[22]. Bhoskar, M.T., Kulkarni, M.O.K., Kulkarni, M.N.K., Patekar, M.S.L., Kakandikar, G.M. and Nandedkar, V.M., 2015. Genetic algorithm and its applications to mechanical engineering: A review. *Materials Today: Proceedings*, *2*(4-5), pp.2624-2630.
[23]. https://www.generativedesign.org/02-deeper-dive/02-04_genetic-algorithms/02-04-01_what-is-a-genetic-algorithm
[24]. Panhwar, M., Rajpar, S.P., Talpur, N., Baig, Q.A., Kumar, K. and Banglani, M.A., 2021. Information technology (IT) application and challenges faced by medical and dental undergraduate students. *Rawal Medical Journal*, *46*(2), pp.438-438.