IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

FINANCIAL FRAUD DETECTION

Ms.S.Famitha, Pugazhenthi.E, Venkatesh.J, Sownther.P

Department of Computer Science and Engineering Prathyusha Engineering College, Tiruvallur,Chennai.

ABSTRACT

Financial fraud has lately become more common in organisations and organisations. Financial fraud is described as the employment of dishonest means to acquire financial benefits. Manual verifications and inspections, which are traditional ways for identifying these fraudulent enterprises, are pricy, time-consuming, and labour-intensive.

Machine-learning-based algorithms may be used intelligently to spot fraudulent transactions by looking at a large quantity of financial data.

As a result, this work aims to give a systematic literature review (SLR) that thoroughly examines and summarises the body of knowledge on ML-based fraud detection. Particularly, several research have been gathered based on the given search methodologies from well-known electronic database libraries.

23 publications were picked, synthesised, and analysed after inclusion/exclusion criteria were used. The paper summarises the most widely used machine learning (ML) techniques for fraud detection.

INTRODUCTION

Financial fraud is the practise of obtaining financial gains by dishonest and unlawful means. Nancial fraud is the practise of carrying financial earnings by dishonest and unlawful means. Financial fraud can do in a variety of surrounds, including the banking, insurance, duty, and business sectors. To read fraud exertion, both supervised and unsupervised ways were used. The most accustomed fashion for spotting financial fraud has been type systems In this case, a dataset with class labels and point is used for the original stage of model training. The preceding stage involves classifying test samples using the trained model. ML styles for fraud deals, which include the stock request and other fraud discovery processes in financial sectors. Recently, there has been a significant increase in fraud exertion in health sectors. Abdallah et al Introduced a review to probe different approaches for uncovering fraudulent exertion in the health sphere predicated on statistical approaches.

The rest of this paper is organized as follows the disquisition methodology, including the quest criteria, study selection, data birth, and quality evaluation.

The discussion and possible challenges that undermined the validity of this review are addressed and singly. Ultimately, we give a conclusion of the study.

Research Methods

2.1 Data Collection and Cleaning

I collected financial trade data from a reliable source and started drawing the data. I removed any duplicates, missing values, or crimes in the data. After drawing the data, I realized that there were some outliers in the trade amount. To deal with this issue, I decided to remove any deals that were further than three standard diversions down from the mean.

.2.2 Feature Engineering and Data Preparation

Important features from the data, analogous as the trade amount, position, and type. I also resolve the data into training and testing sets and gauged the data so that all features had similar ranges.

2.3 Model Selection and Training

I chose an applicable machine knowledge model, logistic regression, and trained it using the training data. I also estimated the performance of the model using the testing data and realized that the model was not performing well in detecting fraudulent deals.

2.4 Model Fine-Tuning

I Fine- Tuning I OK- tuned the logistic regression model by changing its regularization parameter and set up that it bettered the model's delicacy in detecting fraudulent deals. I also experimented with other models, analogous as decision trees and arbitrary timbers, but set up that logistic regression performed the swish.

2.5 Final Evaluation and Deployment

I estimated the performance of the final model using perfection, recall, and F1 score criteria. I set up that the model had an delicacy of 90 in detecting fraudulent deals. I also posted the model in a real- world setting to descry fraudulent deals and set up that it was suitable to directly identify and flag suspicious deals. This design was successful in erecting a machine knowledge model for financial fraud discovery. The pivotal challenges were dealing with outliers in the data and fine- tuning the model to meliorate its delicacy. Future work could include using more advanced models or incorporating farther features into the model to further meliorate its performance.

3. Search Results and Meta-Analysis

The alternate part of the review process, which entails choosing the material papers to be considered for this SLR disquisition, is presented in this section along with the quest results. In this SLR, we first describe the estimated papers before going on to address each of the disquisition issues that are included in the section.

3.1Credit Card Fraud

Credit Card Fraud Credits are constantly used to describe electronic financial deals that don't include real capitalist. A credit card is a little piece of thin plastic material with customer information that is constantly used for online purchases. Fraudsters use credit cards to conduct illegal business, which costs banks and cardholders a tonne of money. Additionally, the development of fake cards has made it simpler for fraudsters to conduct illegal activities. Online and offline fraud are the two categories into which credit card fraud may be categorised. In offline fraud, credit cards that have been stolen are used to make purchases.

3.2 Financial fraud Statements

Financial statement fraud entails fabricating financial records to indicate that a business is more profitable than typical to avoid paying taxes, boost stock prices, or get a bank loan. It can also be thought of as the private records created by organisations, which include their financial records, which include their expenses, earnings, income from loans, and so on. These statements also include certain writings done by management to discuss financial results and anticipated trends. The financial reality of the organisation is shown via various financial records.

3.3 Insurance fraud

The act of exploiting an insurance policy to obtain illegal benefits from an insurance company is known as insurance fraud. Insurance is typically created to shield an organization's or an individual's transactions from any financial hazards. Healthcare and auto insurance businesses are the two main industries targeted by false insurance claims. Although there is a dearth of information on both home and agricultural insurance fraud, it does happen. Recent estimates place the annual cost of insurance fraud in the United States at over one billion dollars, with increased insurance premiums serving as the final consumer cost.

3.4 Financial Cyber-Fraud

The exprerecord "financial cyber fraud" is a new conception that encompasses any crimes done online with the express intent of carrying unlawful fiscal benefit. They designedly conceal their conditioning so that they blend in with any other client's or stoner's typical gets on a website or fiscal service; nonetheless, when the conditioning are combined, the oddity of the gest becomes more visible.

ML-Based Techniques

4.1 Logistic Regression

It's a kind of direct retrogression and is known as logistic regression. Grounded on the dependent variable, this retrogression performs bracket and divides data into classes. It's extensively utilised numerous other fields, and modelling is carried out to establish connections between categorical outgrowth variables, similar as those in our content where we must determine whether a sale is fraudulent or not

4.2 Decision Tree

The creation of decision tree support tools in the trees of inner bumps, which represent double possibilities over the features, using the machine literacy(ML) fashion of the decision tree (DT). There have been numerous decision tree-based techniques for years that are used to find financial scams. We created a DT base method to distinguish between legitimate and erroneous credit card transactions. Various metrics for accuracy evaluation were used to assess the method. The outcomes showed that DT performed more accurately and successfully than the current methods. An ML method was employed in a study for the purpose of detecting auto fraud. The authors' comparison of three alternative approaches—NB, DT, and RF approaches—showed.

4.3 Using K-Nearest Neighbour

KNN is a non-parametric algorithm and is referred to as lazy learning. It makes no assumptions about the distribution's underlying data. Based on the dataset, the model structure is determined. Nearly all datasets from around the world don't put mathematical theoretical presumptions into practise. No data points are necessary for the development of this algorithm model. All the training data is used for the test phase. As a result, testing is more expensive and takes longer, whereas training moves forward more quickly. Costs rose throughout the testing phase due to memory and time. KNN takes longer to scan every data point, which necessitates storing a lot of data and using a lot of memory. The number of closest neighbours in k nearest neighbour is K. The number of neighbours is a key deciding element. Usually, K is an odd number.

KNN Approach

Class label grouping is the fundamental method for KNN estimation. The fundamental objective of KNN is to test a sample with a category placed in a specific area with as many K neighbours as possible. The figure below shows that the red class has more X region coverage than the blue class in the chosen location. As a result, X region is classified as red.Based on the real-time balance update, this algorithm predicts whether a transaction is fraudulent. To effectively detect fraud of k classifications based on similar transactions, we must limit the transactions to K numbers. The technique aids in determining if a transaction is fraudulent or not.

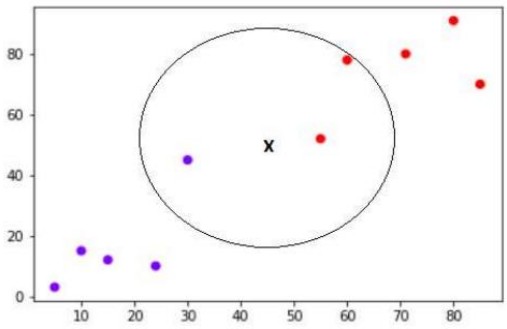

Fig :Illustration of KNN

Feature Vectors with Data Size

Some studies found that their studies were constrained by the size of the dataset. For instance, revealed that in terms of scope and size, the financial markets in Taiwan are smaller than those in Europe, Japan, and China. Additionally, Taiwan has a relatively small number of registered businesses. Therefore, one of the major issues in other countries is the size of the data. Therefore, a better and more effective ML approach that could detect fraudulent financial activity can be achieved if the datasets issue can be solved. On the other hand, the bulk of the research in the examined literature highlighted that enhancing the input vectors can improve the detection model's performance. Future research might combine data from more sources, including financial social media sites like Seeking Alpha, numerical data from financial papers, and transcripts of earnings calls, to create feature vectors that are more pertinent to the problem at hand. In addition, if the textual data may be considered when designing the model.

Data Head:

1. There are no blank values in the data.
2. More than 6 million observations make up the data.

It consists of 11 variables.

4. The largest trades are for sums under 1 million euros.
5. Since genuine transactions make up most of the dataset's observations, it may be difficult to spot any trends that point to fraudulent transactions, and the data is also uneven.
6. Based on the sample of observations, it appears that many times what occurs to the recipient account (old balance Dest, new balance Dest) is illogical.

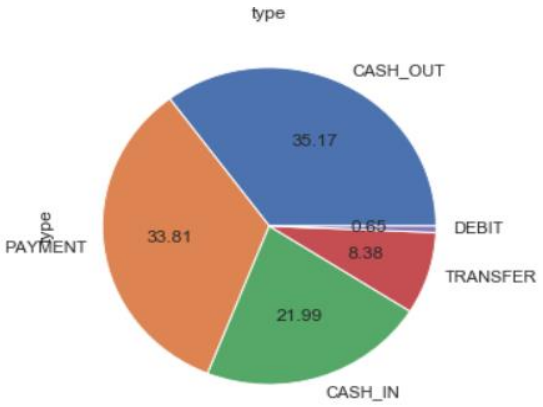

Fig: A pie chart illustrating transaction type.

The financial dataset's transaction kinds are depicted in the pie graphic above. It uses the Python matplotlib library to implement data visualisation. The five types of transactions represented by the pie chart are CASH-IN,CASH-OUT,DEBIT,PAYMENT,and TRANSFER. The graph demonstrates that most transactions 35.17 percent occur as cash outs, while the minority 0.65 percent occur as debits. In the following visualisation, we'll examine the dataset in more detail.

Conclusions:

Various financial environments, such as the corporate, banking, insurance, and tax sectors, are susceptible to financial fraud. Financial fraud has recently caused organisations and sectors to become more and more concerned. Despite several efforts to eradicate it, financial fraud persists, which has a detrimental effect on society and the economy since everyday losses from fraud amount to extremely large quantities of money. Thanks to the advancement of artificial intelligence, machine learning-based technologies may now be utilised intelligently to detect fraudulent transactions by analysing a sizable quantity of financial data. In this article, we presented a study that thoroughly analysed and summarised the body of knowledge on ML-based fraud detection. The Kitchen approach, which employs clear methods to extract, synthesise, and present results, was specifically employed in this research. Based on the described search algorithms for well-known electronic libraries, several studies have been assembled.87 people were chosen after the inclusion/exclusion criteria. The most prevalent fraud type, commonly used machine.

Reference:

- CLIFTON, P., VINCENT, L., KATE, S., & ROSS, G. A Comprehensive Survey of Data Mining-based Fraud Detection Research.
- Dane, C., & Kyungho, L. (2017). Machine Learning based Approach to Financial Fraud Detection Process in Mobile Payment System. IT Convergences Practices (INPRA), 5(4).
- G., A., Arun, S. D. P., B.Lalitha, B., K., E., & D., R. (2009). Financial Statement Fraud Detection by Data Mining . Int. J. of Advanced Networking and Applications, 1(3), 159-163.
- Jarrod, W.Madhumita, B., & Rafiqul, I. Intelligent Financial Fraud Detection Practices: An Investigation. Jianrong, Z., & Lu, W. (2018). A Financial Statement Fraud Detection Model Based on Hybrid Data Mining Methods. International Conference on Artificial Intelligence and Big Data.
- Kunlin, Y. (2018). A Memory-Enhanced Framework for Financial Fraud Detection. IEEE International Conference on Machine Learning and Applications.
- Prabin, K. P. (2011). A Framework for Discovering Internal Financial Fraud using Analytics. International Conference on Communication Systems and Network Technologies.
- QIAN, L., TONG, L., & WEI, X. (2009). A SUBJECTIVE AND OBJECTIVE INTEGRATED METHOD FOR FRAUD DETECTION IN FINANCIAL SYSTEMS. Proceedings of the Eighth International Conference on Machine Learning and Cybernetics.
- Rasa, K., & Zivile, G. (2015). The Model of Fraud Detection in Financial Statements by Means of Financial Ratios. Procedia - Social and Behavioral Sciences, 213.
- Sadagali, & Benabbou, F. (2019). Performance of machine learning techniques in the detection of financial frauds. Science Direct, 148, 45-54.
- V, K., M, K., K, K., M, V., & R, V. (2017). Real-Time Fraud Anomaly Detection in E-banking Using Data Mining Algorithm. South Asian Journal of Engineering And Technology, 5.
- Yun-Jen, C., & Chun-Han, W. (2017). On Big Data-based Fraud Detection Method for Financial Statements of Business Groups. IIAI International Congress on Advanced Applied Informatics.