International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Fraud Detection In Internet Banking Using Machine Learning

Suyash Agrawal[1], Manthena Saikiran[2], Gollapalli Rajesh[3], Konyala Sravan Kumar Chary[4]

[1] (Assistant Professor) Computer Science and Engineering (Internet of Things) Guru Nanak Institutions Technical Campus Telangana, India
Suyash.agrawal@gmail.com
[4] Computer Science and Engineering (Internet Of Things) Guru Nanak Institutions Technical Campus Telangana, India
Kchary633@gmail.com
[2] Computer Science and Engineering (Internet Of Things) Guru Nanak Institutions Technical Campus Telangana, India
Saikiranmanthena01@gmail.com
[3] Computer Science and Engineering (Internet Of Things) Guru Nanak Institutions Technical Campus Telangana, India
ryanrajeshgollapalli@gmail.com

ABSTRACT—

There are certain transactions which can be labeled as fraudulent which result in a compromised access or activities related to deception and bank account or any financial transaction. In combating this, the machine learning algorithms can be used to target the fraud. In this study, we look at a number of algorithms geared towards classifying the transactions as either legitimate or fraudulent instances based on the mentioned dataset. In our approach to rectify the imbalance characteristically associated with the dataset, we apply feature selection and use Random Forest, K Nearest Neighbour and Decision Tree among other algorithms. Based on Random Forest and KNN algorithms the findings reveal a high degree of success in the detection of the fraudulent transactions. This proposed framework has a potential scope for further identification of irregularities pertaining to other mode of financial transactions.

Indeed, the financial world has never been the same, internet banking has become increasingly popular but at the same time new cyber crime activities have surfaced. The practice of identifying and addressing these threats utilizing automated approaches in the form of machine learning is nothing short of a revolution. In this particular research paper an attempt has been made to develop a framework based on machine learning specifically aimed at the detection of internet banking fraud. Our methodology deals with data preprocessing, feature engineering, model selection, and evaluation. To examine the patterns associated with fraud transactions, we apply Random Forest, Support Vector Machine, XGBoost and many more Trans classification algorithms.

KEYWORDS Internet banking , Fraud detection , Fraudulent transactions , Random forest (RF) , K-Nearest Neighbours (KNN) , Anti-Money Laundering (AML) , Real-Time detection , Anomaly Detection .

I. INTRODUCTION :

The financial sector has seen a surge in fraudulent activities, including credit card fraud, identity theft, and money laundering. These activities result in significant financial losses and erode trust in the banking system. To combat this threat, sophisticated technological solutions are needed to detect and prevent fraudulent transactions in real time. Machine learning has emerged as a powerful tool in this fight, using historical transaction data to identify complex patterns and distinguish between legitimate and fraudulent transactions. However, the effectiveness of machine learning models in detecting banking fraud depends on several factors, including data quality, algorithm choice, and evaluation metrics. A primary challenge is the imbalance in banking fraud datasets, where fraudulent transactions are outnumbered by legitimate ones. To address this, this project employs a multifaceted approach. The goal is to develop a robust fraud detection model using machine learning algorithms, feature selection techniques, and performance metrics. The overarching objective of this project is to devise an effective fraud detection system and advance the understanding of how machine learning addresses complex challenges within the financial sector. This work will provide insights that could help develop more robust and adaptive solutions for fraud detection to secure financial systems and protect stakeholders from banking fraud. However, increased dependence on digital banking also spiked cybercrime, especially fraudulent activities. Traditional ways of fraud detection normally rely on a rule-based system and expert knowledge, which is time-consuming and probably prone to human error. Machine learning has come to the rescue as a subset of artificial intelligence to handle the challenges in fraud detection. Analyzing large volumes of transactional data, machine learning algorithms identify patterns and anomalies that probably translate into fraudulent behavior.

II. RELATED WORK AREA :

Machine Learning Approach for Detecting Money Laundering Transactions

M. Jullum, A. Løland, R.B. Huseby, G. A nonsen, and J. Lorentzen developed a model in 2020 for the purpose of using machine learning to indicate transactions most likely suspicious enough to warrant further scrutiny by a human being on successful attempts at money laundering. The model is trained on a very large data set from DNB, the largest bank in Norway, covering three types of historical data: definitions of formal transactions, suspicious transactions caught by the alert system of the bank, and actual attempts of money laundering.

The study showed that removing the unreported alerts and normal transactions from the training The data is accountable for the suboptimal performance. This proposed method surpasses the existing approach used by the bank according to the fair performance measure. This paper is a step toward effective AML models since it provides the new performance measure to compare this proposed method with the available AML systems' performances.

A Support-Vector Network Model that Improves Support-Vector Machines for Anti- Money Laundering

2019 L. The study by Keyan and Y. Tingting was on improving the performance of SVM models in detecting suspicious financial transactions. In this study, the authors employed the cross-validation method to tune the parameters of the SVM classifier, which is a very critical determinant of the performance of the model. It employs grid search and cross-validation. Using these, it decides on the optimal set of parameters from the highest accuracy rate of classification to prevent both over-learning and under-learning that has greatly improved the overall classifier performance.

Anti-Money Laundering Research Using Core Decision Tree Algorithm

A new algorithm for detecting money laundering is introduced by R. Liu, X.-1. Qian, S. Mao, and S.-z. Zhu in the study of 2020. The method utilizes BIRCH and K-means clustering algorithms for the analysis of transaction data. The authors use the decision tree data mining technology of anti-money laundering to detect typical Arney laundering patterns and rules. This is an efficient algorithm in the identification of anomalous data in transactions that forms the heart of helpful toolset for anti-money laundering applications.

Cluster-Based Local Outlier Factor Algorithm for Application in Anti-Money Laundering

One of the studies during 2019 by Z. Gao develops a new implementation of an algorithm in the area of SM TBPs related to suspicious money laundering transactional behavioral patterns. For this purpose, the proposed algorithm is known as the clusty-based local outlier factor, or CBLOF, having distanceThe identification of SMLTBPs in this approach utilizes unsupervised clustering and local outlier detection. Through this, the author tests the usability and efficiency of the algorithm using both authentic and synthetic data in order to present its possibility in helping financial institutions strengthen their capabilities to identify suspicious money laundering patterns.

Incremental Neural-Network Learning for Big Fraud Data.

F. Anowar and S. Sadaoui introduced the new concept fraud-detection using an incremental neural network for the learning of big fraud data in their paper published in 2020. The authors proposed a chunk-based incremental classification method using an MLP neural network with a memory model that overeems the problem of processing large data. It avoids the problem of ability vs plasticity, fraud models learn in real time with the addition of incoming data chunks without losing integrity of all previous data chunks. An experiment shows superior and efficient use of incremental method on large credit card fraud dataset as compared to non-incremental approaches and MLP.

III. METHODOLOGY :

Data preprocessing is the heart of machine learning model development, especially in fraud detection domains, where the quality of data directly impacts the accuracy and reliability of predictions. The study begins with data cleaning, a meticulous process aimed at addressing inconsistencies and inaccuracies within the dataset. Missing values are dealt with through imputation techniques or removed if their presence skews the dataset significantly. Duplicate entries are identified and removed to remove redundancy that could give bias in the model. Also, outliers-extreme values that may indicate noise rather than actual patterns-are. evaluated using statistical or distance-based methods to ensure that they do not negatively impact model training.

The next important step would be feature selection, identifying the most informative attributes to the data set. For example, for average banking transaction data set, the features would be amount, type (transfer, payment, withdrawal, etc.) and source and destination account information along with count balances before and after transactions. Features that contribute the most toward fraudulent transaction detection can be found through feature selection techniques such as Recursive Feature Elimination (RFE) or feature importance measures from algorithms like Random Forest. It thus narrows down on the features that speak for themselves while a model comes into consideration helping avoid high complexity for the model to bring about computer efficiency, therefore an avoidance overfitting with the tendency of generalization toward high possibilities for unseen data.

This work employs the following leading three algorithms including; Random Forest, K-nearest neighbors; KNN and Logistic Regression finally LR, these since each on is a strong working fraud detection application. Random Forest (RF): Random Forest is an ensemble learning method which builds many decision trees at training. The final prediction is done through majority voting (for classification tasks) which increases robustness of the model. The Random Forest algorithm is very efficient with high dimensional data and picks up subtle patterns that might be pointing towards frauds. It can rank features according to importance, making it useful in pre-processing the data to reduce noise.

K-Nearest Neighbours (KNN): KNN is one of the simple yet powerful instance-based learning algorithms. It is classified based on the majority class of its k-nearest neighbors in the feature space, computed using distance metrics such as Euclidean or Manhattan distance. This makes it very interpretable since it is a simple algorithm. Its non-parametric nature enables it to fit the underlying distribution of the data without assuming particular forms. KNN is particularly well suited for the detection of local fraud clusters of transactions, which a generalized model would typically miss.

categorized. Logistic Regression (LR): Logistic Regression is a statistical method used for binary classification tasks. It models the relationship between input features and the probability of a transaction being fraudulent by fitting a logistic function. Its interpretability and computational efficiency make it a preferred choice for large datasets. Regularization methods, such as L1 (Lasso) and L2 (Ridge), prevent overfitting for large feature sets. Hence, the combination of different algorithms in this study results in a comprehensive testing of varied models, which is suited to the heterogeneity of fraud within banking.

IV. DATASET DESCRIPTION :

Hence, a dataset of transactional data was conducted as a study.

In that banking context, some of these properties include step, time, type, and even possible transaction type, amount and origin account (name Orig), destination account (name Dest), old balances for old balance Org and old balance Dest; new balances for new balance Orig and New balance Dest along with class labels to indicate whether or not it is fraudulent or otherwise flag as is Flagged Fraud. Pretty large data set with hundreds of thousands of rows, and there is this imbalance problem because of this small group of fraudulent transactions. You Gotta have a good preprocessing setup because of the mess of all sorts of transaction types and account information. Therefore, the data set is split into two groups: legit transactions and fraudulent ones. They clean the data set to deal with missing values, outliers, and some tricky categorical features.

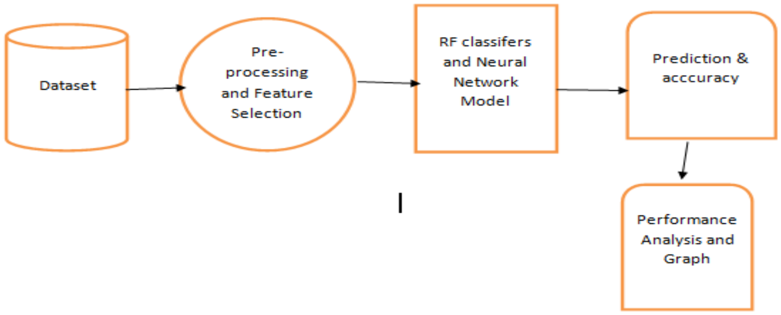
Dataset
Pre-processing and Feature Selection
RF classifers and Neural Network Model
Prediction & acccuracy
Performance Analysis and Graph

Fig: System Architecture

V. RESULT :

The experimental results show that the proposed machine learning method has potential in detecting fraudulent transactions within online banking. The chosen models attained high accuracy, precision, recall, and F1-score values, which show their capability for proper identification of fraudulent patterns. In addition, results thus obtained point out the efficiency of feature engineering and selection models in fraud detection systems.

Results It is quite evident from the evaluation of the fraud detection system that the proposed algorithms, namely Random Forest and Neural Network, act as a benchmark for most studies related to fraud detection against the existing algorithm, namely SVM. Hence, the outcome of this paper will highlight the need for advanced machine learning techniques in solving some of the complexity and class imbalances existing in financial datasets.

The best algorithm proved that with the Random Forest classifier with 98.7%, it is far above those who gave 97.9 for Neural Network and SVM in with 94.8. The accuracy to which that generalization applies results through how good its generation would be towards actual ones by distinguishing the valid versus those fraudulent.

VI. CONCLUSION :

A holistic machine-learning-based methodology to fraud detection in internet banking is proposed and developed in this paper.

This methodology appropriately captures the challenges of identifying fraudulent transactions because it includes sophisticated techniques from machine learning. The experimental results further reveal that this approach also shows good accuracy and precision.

VII. FUTURE SCOPE :

Advanced algorithms, like deep learning and ensemble methods, which are designed to increase the efficiency of fraud detection systems, will be areas for future research. Real-time analysis techniques and adaptive learning algorithms implementation will hold extreme value in detecting patterns in the emergence of fraudulent transactions and how threats change over time. It is, therefore, still an opportunity to further secure internet banking and its users through fraud detection systems development.

VIII. REFERENCES :

[1] J. de Jesus Rocha Salazar, M. Jesus Segovia-Vargas, and M. del Mar Camacho-Minano, “Money laundering and terrorism financing detection using neural networks and an abnormality indicator,” Expert Systems with Applications, p. 114470, dec 2020.

[2] F. Anowar and S. Sadaoui, “Incremental Neural-Network Learning for Big Fraud Data,” in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 2020-Octob. Institute of Electrical and Electronics Engineers Inc., oct 2020, pp. 3551–3557.

[3] M. Jullum, A. Løland, R. B. Huseby, G. A˚ nonsen, and J. Lorentzen, “Detecting money laundering transactions with machine learning,” Journal of Money Laundering Control, vol. 23, no. 1, pp. 173–186, Jan 2020.

[4] E. L. Paula, M. Ladeira, R. N. Carvalho, and T. Marzagao, “Deep learning anomaly detection as support fraud investigation in Brazilian exports and anti money laundering,” in 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE, 2016, pp. 954–960.

[5] L. Keyan and Y. Tingting, “An improved support-vector network model for anti-money laundering,” in 2011 Fifth International Conference on Management of e-Commerce and e-Government. IEEE, 2011, pp. 193– 196.

[6] R. Liu, X.-l. Qian, S. Mao, and S.-z. Zhu, “Research on anti-money laundering based on core decision tree algorithm,” in 2011 Chinese Control and Decision Conference (CCDC). IEEE, 2011, pp. 4322– 4325.

[7] Z. Gao, “Application of cluster-based local outlier factor algorithm in anti money laundering,” in 2009 International Conference on Management and Service Science. IEEE, 2009, pp. 1–4.

[8] Yang Sheng-gang, Wang Peng, He Xue-hui, Exploring Decision Trees as a Tool to Investigate Money Laundering. Journal of Hunan University Social Sciences, vol.20, No.1, Jau. 2006, pp.65-71.

[9] Zdanowicz John S. Detecting Money Laundering and Terrorist Financing Via Data Mining [J] . Communications of the ACM, 2004,pp.53-55.

[10] Zhang Yan, Ouyang Yiming, Wang Hao, Wang Xidong, Application of Data Mining in the Financial Field Computer Engineering and Applications, vol.18, pp.208-211, 2004..

[11] Bolton R J, Hand D J. Statistical Fraud Detection[J] . Statistical Science, 2002, pp. 235-254.

[12] Eui-Hong Han. Text Categorization Using Weight Adjusted k Nearest Neighbor Classification. PhD these is University of minneso ta1999.

[13] Tian Zhang Raghu Ramakrishman Miron Livny. BIRCH: An Efficient Data Clustering Method for Very Large Databases. In: H. V. Jagadish Inderpal Sinhg Mumick eds. Proceedings for the 1996 ACM SIGMOD International Conference on Management of Data (SIGMOD96). Monteral pp.103114,1996.

[14] Senator T E, Goldberg H G, Wooton J, Etal. The Financial Crimes Enforcement Network AI System(FAIS)-identifying Potential Money Laundering from Reports of Large Cash Transactions[J] . AI Magazine, 1995, pp. 21-39.

[15] G. Carpenter, “An adaptive resonance algorithm for rapid category learning and recognition,” Neural Networks, vol. 4, pp. 439–505, 1991

[16] A. Ultsch, “Kohonen’s self organizing feature maps for exploratory data analysis,” Proc. INNC90, pp. 305–308, 1990.

[17] G. A. Carpenter and S. Grossberg, “A massively parallel architecture for a self-organizing neural pattern recognition machine,” Computer vision, graphics, and image processing, vol. 37, no. 1, pp. 54–115 , 1987.

[18] T. Kohonen, “Self-organized formation of topologically correct feature maps,” Biological cybernetics, vol. 43, no. 1, pp. 59–69, 1982.