International Journal of Combined Research & Development (IJCRD)
eISSN:2321-225X;pISSN:2321-2241 Volume: 13; Issue: 7; July- 2024

PREDICTING FRAUD IN FINANCIAL PAYMENT SERVICES USING MACHINE LEARNING TECHNIQUES

Mary Anitha T
Associate Professor
Department of Master of Computer Applications
The Oxford College of Engineering
mary.anitha.charlton@gmail.com

S Sanjana
PG Student
Department of Master of Computer Applications
The Oxford College of Engineering
sanjanasuresh68@gmail.com

ABSTRACT

Financial companies need to be proactive in identifying transaction risks in order to improve customer satisfaction and losses. In order to predict financial transaction legitimacy in an efficient and effective manner, we compare and contrast different machine learning techniques. The study utilized are the subsequent methodologies: Deep learning, logistic regression, decision tree, bagging, MLP, Ada Boost, K neighbours, MLP Repressor, Random forest, Gaussian NB, Complement NB, and Bagging classifiers are some examples of machine learning techniques. The dataset came from the Kaggle depository. 6362620 rows and ten columns make up this data. When applied to an uneven dataset, the Random Forest Classifier excelled. Accuracy, precession, recall, and F1-score were all 99.97%, 99.96%, and 99.97%, respectively.

Keywords : *Financial Transactions Suspicious Activity, Chargeback Fraud, ML Algorithms, Account Takeover, Transaction Velocity and Unusual Spending Patterns.*

I. INTRODUCTION

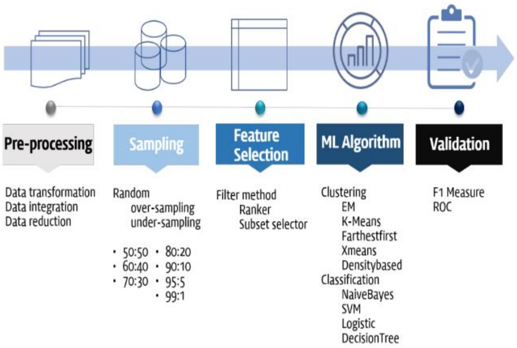
Fraud detectors and fraudulent transactions have long been associated with each other. Fraudulent actions are becoming increasingly common, especially with the rapid digitization of society, and they typically result in significant financial losses. Transaction fraud cost the economy $28 billion in 2019, over $30 billion in 2020, and much more this year (twice as much as in 2020). Globally, transaction fraud is predicted to continue increasing and reach $34 billion by 2022. Therefore, automated fraud detection systems may be required by financial institutions in order for financial transactions to be accurately recognized and validated. These fraud detection techniques are designed to recognize anomalous transaction behaviour from a large dataset and subsequently apply them for detection. Machine learning is an artificial intelligence (AI) application that gives systems the capacity to automatically learn from experience and get better at it without needing to be explicitly designed. The goal of machine learning is to create computer programs that can access data and utilize it to learn for themselves. In order to seek for patterns in data and make better decisions in the future based on the examples we provide, learning starts with observations or data, such as examples, firsthand experience, or instruction.The main goal is to enable computers to learn on their own, devoid of human guidance or involvement, and adapt their behaviour accordingly.

When it comes to artificial intelligence, deep learning is a subset of machine learning that uses networks that can learn on their own from unlabelled or inconsistent data. The process of creating face detection using profile photos and identifying whether they are authentic or fraudulent is called deep learning. Fraud transaction detection and classification have shown machine learning to be highly profitable. A fraud classifier can also be trained and validated using a large number of transaction records. Transactional fraud analysis methods are constantly evolving, even while supervised learning has proven to be incredibly effective in identifying fraudulent transactions.

II. LITERATURE REVIEW

A major worry today is the increasing likelihood of financial fraud in a world where wireless communications are essential for transmitting enormous amounts of data while guarding against interference. A special kind of artificial intelligence designed specifically for analyzing financial transaction data in real time is the ResNeXt-embedded Gated Recurrent Unit (GRU) model (RXT). Our artificial intelligence method is a methodical approach driven by the need to tackle the growing issue of financial fraud, which presents significant hazards to both clients and financial institutions. To mitigate data imbalance, we use the SMOTE to start the process with artificial intelligence data input and preparation. ResNet (EARN) and autoencoders are combined in feature extraction, an artificial intelligence ensemble technique, to uncover important data patterns. Feature engineering, on the other hand, improves the discriminative power of the model.

The RXT model, which has been optimized using hyperparameters using the Jaya optimization method (RXT-J), is the central component of our artificial intelligence classification challenge. Three real-world financial transaction datasets are used to thoroughly test our artificial intelligence model. It routinely outperforms state-of-the-art algorithms by a significant margin of 10% to 18% on a variety of assessment parameters, all while retaining an exceptional computational efficiency. With the potential to improve security and maximize efficiency in financial transactions, this ground breaking artificial intelligence research is a major step forward in the ongoing fight against financial crime. Our artificial intelligence work seeks to improve security, data availability, stability, and dependability against cyberwarfare attacks in the banking industry as a protection against interference with wireless communications.



III. EXISTING SYSTEM

In order to deceive users of financial statements, such as investors or creditors, financial statement fraud is defined as the deliberate misrepresentation of a company's or enterprise's financial outcomes. This is achieved through the omission of information or the intentional misstatement of amounts in the financial statements. Organizations frequently strive to enhance their financial reports in order to draw in investors. The company's primary goal in this regard is to conceal its debt and present falsely high profits.

IV. PROPOSED SYSTEM

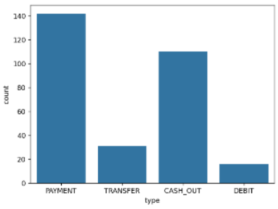
Financial statements are regarded as fundamental records that assist in illustrating a company's financial situation. Financial statements constitute the basis of the decision-making process for creditors, shareholders, investors and other consumers of important accounting information. As a result of the development of data mining tools, more sophisticated methods of knowledge discovery are now required to extract information from data that was not known before Supervised and unsupervised financial statement fraud prediction classes are based on machine learning techniques.

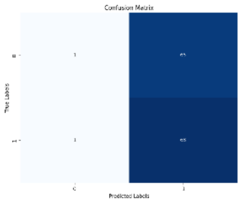
V. IMPLEMENTATION

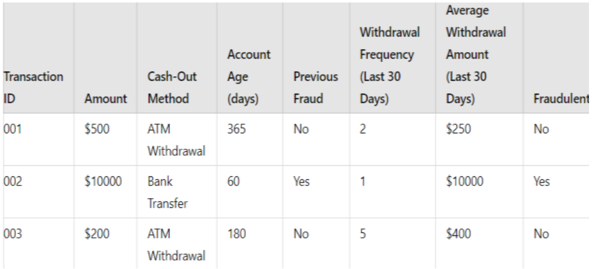
The implementation begins by collecting a comprehensive dataset of the historical transaction. This data may include both legitimate and fraudulent transaction with detailed features such as transaction amount merchant details location and user behaviour patterns. And the next is,it cleans the data set to address missing values, outliers and inconsistencies. Normalise or scale features to ensure that they are on a comparable scale which helps them in improving the performance of the machine learning models. Develop relevant features that can highlight potential fraud that might include transaction frequency average transaction amounts and deviations from typical spending patterns. Creating new features based on historical data can enhance models ability to detect anomalies. You can choose appropriate machine learning algorithms based on the problem complexity and data characteristics which are commonly used that include logistic regression decision trees, random forest etc. Later split the data into training and validation sets. Train the model using the training set and evaluate its performance on the validation set using metrics such as precision, recall and F1 score. This helps to ensure that the model can effectively distinguish between legitimate and fraudulent transaction. Integrate the trained model into the financial payment processing system for real time fraud detection. Ensure that the model can handle the real time data and provide alerts for only suspicious transactions. Continuously monitor the model’s performance in the real time environment. Full regularly update the model with new data to adapt the emerging fraud patterns and refine the accuracy overtime.

VI. RESULT

Highly imbalanced datasets are a common problem in fraud detection. We can demonstrate that our proposed approaches can detect fraud transactions with very high accuracy and low false positives. In fraud detection, there is frequently a trade-off between accurately identifying fraudulent samples and accurately classifying a large number of non-fraud samples. Every digital payment company must frequently make a business/design decision on this. Our decision-making process can be further enhanced by building user-specific models, which are based on the user's prior transactional behaviour. We think all these can be very useful in enhancing our quality of classification on this dataset. Recall and accuracy will always have to be traded off since there is no such thing as a flawless model. The organizations and its goals will help us determine which course of action are optimal in any given circumstances. Overall, we get to observe that almost all our proposed approaches seem to detect fraud transactions with high accuracy and low false positives.







Transaction ID	Amount	Cash-Out Method	Account Age (days)	Previous Fraud	Withdrawal Frequency (Last 30 Days)	Average Withdrawal Amount (Last 30 Days)	Fraudulent
001	$500	ATM Withdrawal	365	No	2	$250	No
002	$10000	Bank Transfer	60	Yes	1	$10000	Yes
003	$200	ATM Withdrawal	180	No	5	$400	No

VII. CONCLUSION

Equilibrium and unbalanced datasets can both produce good prediction results. With the ability to identify over 99.50% of fraudulent transactions while also avoiding labelling any non-fraudulent transactions as fraudulent, Bagging Classifier, Decision Tree Classifier, and Random Forest Classifier produced the best results. Recall and precision are always trade-offs in models because there is never a perfect one. The decision on which strategy works best in a given scenario rests with the organization and its goals.

VIII. FUTURE WORK

By using categorical information connected to accounts and users in transaction datasets, we may further enhance our methods through the application of algorithms such as decision trees. A time series may also be derived from a transaction dataset. Using algorithms like CNN, we may take use of this feature to create time series-based models. Currently, we train our models by treating the complete collection of transactions as a whole. Our decision-making process may be further enhanced by building user-specific models, which are based on the user's prior transactional behaviour. We think that these can be quite useful in enhancing our quality of categorization on this dataset.

XI. REFERENCES

[1]Al-Araj, R. S. A., et al. (2020). "Classification of Animal Species Using Neural Network." International Journal of Academic Engineering Research (IJAER) 4(10): 23-31.

[2]Al-Atrash, Y. E., et al. (2020). "Modeling Cognitive Development of the Balance Scale Task Using ANN." International Journal of Academic Information Systems Research (IJAISR) 4(9): 74-81.

[3]Alghoul, A., et al. (2018). "Email Classification Using Artificial Neural Network." International Journal of

Academic Engineering Research (IJAER) 2(11): 8-14.

[4]Al-Kahlout, M. M., et al. (2020). "Neural Network Approach to Predict Forest Fires using Meteorological Data." International Journal of Academic Engineering Research (IJAER) 4(9): 68-72.

[5]Alkronz, E. S., et al. (2019). "Prediction of Whether Mushroom is Edible or Poisonous Using Back-propagation Neural Network." International Journal of Academic and Applied Research (IJAAR) 3(2): 1-8.

[6]Al-Madhoun, O. S. E.-D., et al. (2020). "Low Birth Weight Prediction Using JNN." International Journal of Academic Health and Medical Research (IJAHMR) 4(11): 8-14.

[7]Al-Massri, R., et al. (2018). "Classification Prediction of SBRCTs Cancers Using Artificial Neural Network." International Journal of Academic Engineering Research (IJAER) 2(11): 1-7.

[8]Al-Mobayed, A. A., et al. (2020). "Artificial Neural Network for Predicting Car Performance Using JNN." International Journal of Engineering and Information Systems (IJEAIS) 4(9): 139-145.

[9]Al-Mubayyed, O. M., et al. (2019). "Predicting Overall Car Performance Using Artificial Neural Network." International Journal of Academic and Applied Research (IJAAR) 3(1):

[10] Alshawwa, I. A., et al. (2020). "Analyzing Types of Cherry Using Deep Learning."