Fraud detection in credit cards system using ML with AWS stage maker

SAMREEN SHAIKH [#], KAVERI DIWANJI[#], SAMARTH MALEGAONKAR [#], SAMRAJYA PUJARI [#], PROF. ASHIWINI BHOSLE [*]

[#] Under-Graduate Students

Department of Computer Engineering,
GENBA SOPANRAO MOZE COLLEGE OF ENGINEERING, Balewadi, Pune
[1]samreenshaikh@gmail.com
[2]kaveri.diwanji27@gmail.com
[3]samarthmalegaonkar@gmail.com
[4]samrajyapujari@gmail.com

[*] Assistant Professor
Department of Computer Engineering,
GENBA SOPANRAO MOZE COLLEGE OF ENGINEERING, Balewadi, Pune
[5]ashu.jjm@gmail.com

ABSTRACT

Credit card firms must be able to recognize fraudulent credit card transactions so that clients are not charged for products that they did not purchase. Data Science may be used to solve these issues, and coupled with machine learning, it is of utmost relevance. The goal of this project is to demonstrate how to model a data set using machine learning for credit card fraud detection. The Credit Card Fraud Detection Problem entails modelling previous credit card transactions using information from those that were later determined to be fraudulent. While the globe was under lockdown and movement was confined to an absolute emergency- millions were introduced to the world of internet shopping. The simplicity of internet buying helped e-commerce platforms generate unprecedented sales. It is not surprising that during that time, the rate of online financial fraud also skyrocketed. During the COVID-19 pandemic in 2020 compared to 2019, there was a historic increase of 225 percent in online fraud cases involving credit and debit cards.

Keywords— Fraud Detection, Machine Learning, Amazon Web Services (AWS), Credit Card, AWS Stage Maker, Artificial Intelligence.

INTRODUCTION

Credit card fraud refers to the unauthorized and unwanted use of a credit card account by someone other than the account owner. Appropriate preventative measures should be done to halt this abuse and the behaviour of such fraudulent acts can be researched to decrease it and protect against similar occurrences in the future. In other words, Credit Card Fraud may be described as a circumstance when a person utilizes someone else's credit card for personal reasons while the owner and the card issuing authorities are oblivious of the fact that the card is being used.
Hackers or fraudsters may access private data of the card through insecure websites. Everybody involved in the process loses when a fraudster steals someone's credit or debit card, from the person whose private information has been exposed to the organizations (often banks) who issue the credit card and the retailer who is completing the transaction with a purchase. This makes it vitally crucial to spot fraudulent transactions at the outset. Businesses like e-commerce and financial institutions are taking decisive action to identify systemic fraudsters.
Several complex machine learning systems are at play, examining every transaction and stemming the fraud users in its nip utilizing behavioural data and transaction patterns. The method of automatically discriminating between fraudulent and authentic users is known as "credit card fraud detection". These are not the only hurdles in the development of a real-world fraud detection system, though. In actual cases, the enormous volume of payment requests is swiftly scanned by automated tools to decide which transactions to authorize.
Machine learning techniques are applied to examine all the approved transactions and report the suspect ones. These reports are evaluated by specialists who call the cardholders to confirm if the transaction was legitimate or fraudulent. The automated system receives feedback from the investigators and then trains and updates the algorithm to gradually increase fraud detection performance over time.

LITERATURE REVIEW

In "Credit Card Fraud Detection Using Advanced Machine Learning Techniques", Aditi Aditi; Aman Dubey; Ankit Mathur; Preeti Garg [1] proposed that ecommerce is expanding quickly over the world, which leads to a huge rise in the use of plastic money. With this spectacular rise in card usage comes a rise in credit card fraud. Fraud instances are increasing along with the use of credit cards. Simpler methods like pattern matching are bad at spotting these frauds because in real-world circumstances, the fraudulent transaction is mixed in with the legitimate transaction. For all organizations that issue bank cards, it is crucial to implement an accurate and efficient fraud detection system. Many techniques, such as approximation reasoning, artificial intelligence (AI), data mining, sequence alignment that locates regions of similarity, inheritable programming, etc., are heavily employed in the identification of these credit card frauds. Better credit card detection systems will be created with the aid of an inventive strategy and a thorough comprehension of these technologies.

In "Fraud Detection and Management for Telecommunication Systems using Artificial Intelligence (AI)", Ritika H J; Mohana [2] proposed that fraud is on the rise globally, and it may cost companies billions of dollars and seriously harm their financial standing. Researchers from many application sectors have put forth various strategies. We may perceive the problems more clearly by looking at these theories. The goal of this essay is to look at several approaches to fraud prevention and detection in the communications industry. This article presents an overview of the many categories of telecom fraud, problems that hinder the detection process, and some suggestions for how to fix them. Here, the effectiveness of the present ways is discussed, followed by suggestions and advice for selecting the most appropriate performance indicators.

In "Credit Card Fraud Detection Using Machine Learning Model", Swarnalatha K S; Krishna Kumar Shah; Kishore Kumar; Krishna Kumar Patel; Aashutosh Raj Sah [3] proposed that the identification of credit card

fraud is the most common problem in the current world. This is due to the development of e-commerce platforms and online transactions. Credit card fraud often happens when a card is stolen and used for any unauthorized activity, including when the con artist uses the card's information for his personal advantage. In the contemporary world, there are several credit card problems. To spot fraudulent activity, credit card theft detection technology was developed. The purpose of the proposed work is to construct and develop a unique fraud detection algorithm for streaming transaction data, with the objective of analysing past transaction information about customers and identifying behavioural patterns. Each transaction's genuineness and fraud are classified in a dataset for cardholders' transactions. Then, the ML classifier aggregates cardholders' transactions across several categories in order to determine the behavior of each grouping. Afterward, the Random Forest Classifier (RFC) is trained using the dataset, and the accuracy of the classifier with a result over 75 is assessed.

In "Detection of Credit Card Fraud using Machine Learning Algorithms", Akhilesh Kumar Singh [4] proposed that due to ongoing advancements in technology and network connectivity, there has been a growth in the usage of online transactions in daily life during the past 10 years. The major reason new users are joining the huge population using the system is that online transactions are quick, convenient, and user-friendly. Misuse of the system, which is defined as stealing someone's credit card information and using it for one's own gain without the cardholder's consent, is the biggest drawback of the online transaction system. In order to regulate the severely skewed dataset, we apply a variety of ML methods in this study, including regression, random forest classifier, and decision tree. In this study, we will also compute the confusion matrix, f1 score, recall, accuracy, and precision.

In "Automated Banking Fraud Detection for Identification and Restriction of Unauthorised Access in Financial Sector", Ankur Biswas; Ramandeep Singh Deol; Barun Kumar Jha; Geethamanikanta Jakka; M. Raja Suguna; [5] proposed that current methods now make it possible to use cutting-edge equipment and technology to spot fraud within an organization. It is crucial for any Banking industry or financial sector to discover those fraudulent activities that could be related inside their network. Various AI-based technologies and ML algorithms are applied inside banking sectors of India and other countries of this globe to detect frauds and unauthorised access. This study paper had revealed that Data mining is useful for the banking industry to target clients, who could develop fraudulent activities during the credit procedure. SVM, Logistic regression, Decision tree, Neural networks are employed as data mining techniques inside banking sector's fraud detection process, although all these technologies needed a data balancing procedure before designing the model. Autoencoder is a model that is offered for fraud detection without any data balancing. A secondary thematic data analysis approach is designed inside this study piece to explore of different factors pertinent to this issue. It has been established that, proposed model for cyber-crime detection, AI-enabled automated fraud controlling system are beneficial to decrease risk inside banking sector. Moreover, the banking industry uses SAS, a powerful piece of fraud detection software. It produces alarm signals for any fraudulent conduct with system of banks. Yet, banking sectors must have sound governance to address the issue of a lack of knowledge and experience in their data mining process.

In "Performance Analysis of Machine Learning Algorithms in Credit Cards Fraud Detection", Vinod Jain; Mayank Agrawal; Anuj Kumar[6] proposed that online payments made with credit cards are quite widespread. In recent years' scams are recorded which are done utilizing credit cards. It is quite tough to detect and prevent the fraud which is conducted via credit card. Science and engineering challenges are frequently solved using the Artificial Intelligence (AI) approach known as Machine Learning (ML). In this research, machine learning algorithms are implemented on a data set of credit cards frauds and the power of three machine learning methods is compared to identify the frauds done using credit cards. In comparison to Decision Tree and XGBOOST methods, the accuracy of the Random Forest machine learning algorithm is the highest.

In "Improving Fraud Detection in An Imbalanced Class Distribution Using Different Oversampling Techniques", Raneem Qaddoura; Mariam M. Biltawi [7] proposed that financial firms must identify credit card fraud to keep from billing clients for goods they did not buy. By developing a model trained on a dataset including transactions with fraud and non-fraud classifications, fraud detection may be accomplished by machine learning (ML). The dataset supplied for this job is frequently severely unbalanced. Consequently, the purpose of this research is to undertake a complete comparison of five oversampling strategies. The oversampling techniques are the Synthetic Minority Oversampling Technique (SMOTE), Adaptive Synthetic Sampling (ADASYN), borderline1 SMOTE, borderline2 SMOTE, and Support Vector Machine SMOTE (SVM SMOTE) to generate an enhanced model which can solve the imbalanced problem. The comparison is conducted by computing the geometric mean, recall, precision, and F1-score of six machine learning models with and without applying oversampling. The ML models that have been tested include decision tree, logistic regression, random forest, K-nearest neighbor, naive Bayes, and support vector machines. The performance of the models has improved, according to experimental findings, thanks to the oversampling strategies.

In "Detecting insurance claims fraud using machine learning techniques", Riya Roy; K. Thomas George [8] proposed that the insurance industries comprise of more than thousand organizations throughout global. and collect premiums totaling more than $1 trillion each year. When a person or business create false insurance claims in order to get money or benefits to which they are not entitled is known as an insurance fraud. The entire cost of an insurance scam is estimated to be more than forty billions of dollars. Deterring insurance fraud is thus a difficult issue for the insurance sector. The classic strategy for fraud detection is focused on establishing heuristics around fraud indication. The auto\vehicle insurance fraud is the most prominent sort of insurance fraud, which may be done through faking accident claim. In this research, concentrating on identifying the auto\vehicle fraud by employing, machine learning technology. Additionally, the performance will be compared by computation of confusion matrix. This can help to calculate accuracy, precision, and recall.

In "Detection of Fraudulence in Credit Card Transactions using Machine Learning on Azure ML", Abhishek Shivanna; Sujan Ray; Khaldoon Alshouiliy; Dharma P Agrawal [9] proposed that with the growth of mobile and cloud technology, there is a dramatic increase in online transactions. Identifying fraudulent credit card transactions on a timely manner is a highly significant and tough topic in Financial Industry. Online transactions are tremendously easy, but they also carry a high risk of fraud in many areas. Unusual behavioral patterns, a skewed dataset (high normal transaction to fraudulent transaction ratio), a lack of data availability, and a continually changing environment are some of the major obstacles to identifying fraud in online transactions. Millions of dollars are lost annually as a result of credit card fraud. There is a lack of quality research in this domain. In order to simulate our system, we used a dataset of European cards that had 284,807 transactions. By putting two machine learning (ML) algorithms—Decision Forest (DF) and Decision Jungle (DJ) classifiers—through training and testing, we will design and develop a system to detect credit card fraudulence in this paper. Our results convincingly illustrate that DJ classifier gives superior performance compared to DF classifier.

In "Credit Card Fraud Detection based on Ensemble Machine Learning Classifiers", Karthika J; A. Senthilselvi [10] proposed that due to the development of communication and electronic commerce networks, credit cards are one of the most widely used payment methods for both ordinary transactions and those made online. As a result, the fraud related to these transactions greatly rose. The widespread use of electronic payments is significantly impacted by these fraudulent transactions, which call for quick detection to address the problem. Thus, effective and efficient ways to identify fraud in credit card transactions are essential. To catch the fraudulent transaction, a strong fitting model is essential, so researchers propose the use of various Machine Learning (ML) approaches, because of its favorable properties. The major purpose of the study effort is to create an ensemble based ML approaches for Credit Card Fraud Detection

(CCFD). The strength of our model is a combination of the forces of the three subsystems; Recursive Feature Elimination (RFE), CCFD's employing ensemble classifiers, and Synthetic Minority Oversampling (S MOTE) to cope with the problem of imbalanced data to select the most effective prediction features. The suggested model performs typical tests on two genuine datasets of public credit card transactions, both fraudulent and official ones. Based on the comparison of other ML approaches, the additional tree classifier has performed better and reached high efficiency such as 96% of accuracy and 57.95% of F1-measure.

REFERENCES

[1] A. Aditi, A. Dubey, A. Mathur and P. Garg, "Credit Card Fraud Detection Using Advanced Machine Learning Techniques," 2022 Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT), Sonepat, India, 2022, pp. 56-60, doi: 10.1109/CCiCT56684.2022.00022.

[2] R. H J and Mohana, "Fraud Detection and Management for Telecommunication Systems using Artificial Intelligence (AI)," 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2022, pp. 1016-1022, doi: 10.1109/ICOSEC54921.2022.9951889.

[3] S. K. S, K. K. Shah, K. Kumar, K. K. Patel and A. R. Sah, "Credit Card Fraud Detection Using Machine Learning Model," 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon), Mysuru, India, 2022, pp. 1-7, doi: 10.1109/MysuruCon55714.2022.9972647.

[4] A. K. Singh, "Detection of Credit Card Fraud using Machine Learning Algorithms," 2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART), Moradabad, India, 2022, pp. 673-677, doi: 10.1109/SMART55829.2022.10047099.

[5] A. Biswas, R. S. Deol, B. K. Jha, G. Jakka, M. R. Suguna and B. I. Thomson, "Automated Banking Fraud Detection for Identification and Restriction of Unauthorised Access in Financial Sector," 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2022, pp. 809-814, doi: 10.1109/ICOSEC54921.2022.9951931.

[6] V. Jain, M. Agrawal and A. Kumar, "Performance Analysis of Machine Learning Algorithms in Credit Cards Fraud Detection," 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2020, pp. 86-88, doi: 10.1109/ICRITO48877.2020.9197762.

[7] R. Qaddoura and M. M. Biltawi, "Improving Fraud Detection in An Imbalanced Class Distribution Using Different Oversampling Techniques," 2022 International Engineering Conference on Electrical, Energy, and Artificial Intelligence (EICEEAI), Zarqa, Jordan, 2022, pp. 1-5, doi: 10.1109/EICEEAI56378.2022.10050500.

[8] R. Roy and K. T. George, "Detecting insurance claims fraud using machine learning techniques," 2017 International Conference on Circuit ,Power and Computing Technologies (ICCPCT), Kollam, India, 2017, pp. 1-6, doi: 10.1109/ICCPCT.2017.8074258.

[9] A. Shivanna, S. Ray, K. Alshouiliy and D. P. Agrawal, "Detection of Fraudulence in Credit Card Transactions using Machine Learning on Azure ML," 2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York, NY, USA, 2020, pp. 0268-0273, doi: 10.1109/UEMCON51285.2020.9298129.

[10] K. J and A. Senthilselvi, "Credit Card Fraud Detection based on Ensemble Machine Learning Classifiers," 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2022, pp. 1604-1610, doi: 10.1109/ICESC54411.2022.9885649.

CONCLUSION

In this article, we demonstrated how to use SageMaker and machine learning to create the foundation of a dynamic, self-improving, and maintainable system for detecting credit card fraud. An unsupervised RCF anomaly detection model, a baseline supervised XGBoost model, a second supervised XGBoost model with SMOTE to address the data imbalance issue, and a final XGBoost model optimized with HPO were all built, trained, and deployed. We addressed how to handle data imbalance and utilize your own data in the solution. We have provided an example REST API implementation using API Gateway and Lambda to illustrate how to utilize the system in your existing business infrastructure.