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

FRAUD DETECTION ON BANK PAYMENTS USING MACHINE LEARNING

Supritha M
PG Student
Dept. of MCA
The Oxford College of Engineering,
Bommanahalli, Bengaluru-560068
msupritha30@gmail.com

Mridula Shukla
Assistant Professor
Dept. of MCA
The Oxford College of Engineering,
Bommanahalli, Bengaluru-560068
mridulatewari005@gmail.com

ABSTRACT

As digital payments become increasingly prevalent, fraudulent activities in payments to banks have increased. It is imperative to detect and mitigate misconduct in bank payments to protect financial assets and preserve trust. The article provides a comprehensive examination of the utilization of artificial intelligence techniques to perfect the detection of misconduct in bank transactions. We investigate a selection of machine learning models, such as decision trees as well as logistic regression models, supported vector machines (SVM), and deep learning approaches, in order to detect patterns that suggest fraudulent activities. Data pre processing, feature engineering, and the application of various classifiers to evaluate their efficacy in recognizing deceit comprise the methodology. Our experimental findings demonstrate that sophisticated machine learning models, especially deep learning frameworks, greatly outperform standard rule-based systems in calls of accuracy and efficiency. The study concludes by addressing the implication of these discoveries for financial societies and offering suggestions for the integration of these models into real-time fraud detection systems.

Keywords: Machine Learning, Fraud detection, Data Analysis, Financial Fraud

1. INTRODUCTION

The expansion of digital banking and the broad acceptance of online payment methods have revolutionized the financial environment, providing customers with unparalleled ease and accessibility. However, this digital revolution has also introduced new challenges, notably the increased risk of fraud in bank transactions. Fraudulent activities in banking not only lead to important monetary failures although also undermine consumer trust and the truthfulness of financial systems. Corresponding to the Association of Certified Fraud Examines (ACFE), the global financial failure appropriate to fraud reached an estimated $42 billion in 2023, a testament to the increasing degree of sophistication and frequency of fraudulent schemes. Conventional methods for detecting fraud methods, such as rule-based systems and manual evaluations, have proven inadequate in the context of these evolving threats, necessitating the adoption of more advanced and adaptive techniques.

This paper seeks to investigate the applicability of simple machine learning techniques in identifying fake bank transactions, concentrating on the comparative performance of different ML algorithms. We begin with a review of existing literature on fraud detection methods, and then offer a detailed description of our processes, like pre-

processing the data, feature selection, and model training. The results this section provides a thorough examination of the outcome of several models, emphasizing both their advantages and disadvantages. Lastly, we talk about how our results will affect the world of finance.

The significance of this research resides in its potential to enhance the security and efficacy of fraud detection systems in banking. By incorporating machine learning models into existing frameworks, financial institutions can substantially reduce the incidence of fraud, minimize false positives, and enhance the overall consumer experience

2. LITERATURE REVIEW

The Landscape of scam finding in finance has evolved substantially over the past few decades, spurred by advancements in technology and the increasing sophistication of fraudulent schemes. Early methods of fraud discovery relied heavily on manual evaluations and rule-based systems, which, while effective in their time, have become inadequate in addressing the complexities of modern fraud. This literature review examines the evolution of fraud finding techniques, from traditional statistical methods to the latest machine learning approaches, emphasizing key studies and developments in the field.

In the early phases of fraud detection, rule-based systems were the primary instrument used by financial institutions. These systems relied on predefined rules and thresholds to identify suspicious transactions. For example, transactions exceeding a certain amount or occurring at atypical periods were flagged for additional research. Although systems built around rules offered a simple method for detecting fraud, their inflexible nature and incapacity to adjust to novel trends in fraud were drawbacks. The drawbacks of systems founded on rules were more apparent as criminals evolved more advanced techniques, which resulted in high level false positives and the requirement for regular rule modifications.

The introduction of statistical techniques marked a significant advancement in the identification of fraud. Statistical methods, in transaction data, trends and outliers could be found using techniques like anomaly detection and clustering. Research conducted by Bolton and Hand (2002) and other researchers showed how well statistical models work to identify fraud by examining variations from typical transaction behaviour.

Identification of fraud underwent a radical transformation with the introduction of algorithmic learning. Decision trees, artificial neural nets, and support vector machinery (SVM) are a few examples of predictive models that have the capacity to utilize data from transactions in the past and recognize intricate patterns suggestive of fraud. Studies conducted by Ngai et al. (2011) and other researchers shown how machine learning strategies can improve the precision and effectiveness of systems for detecting fraud. These algorithms might be deployed to fresh data to spot anomalous behaviour after being trained on marked datasets to learn the traits of both authentic and phony purchases.

More recent studies have focused on the using deep learning methods for recognizing fraudulent activity. Convolution machine learning (CNN) and recurrent neural networks (RNN), two advanced machine learning theories, Research by LeCun et al. (2015) and others has shown that deep neural network models outperform conventional ML techniques in identifying cheating.

In conclusion, the literature on banking fraud detection highlights the transition from rule-based systems to sophisticated

machine learning methods. The adoption of machine learning and deep learning has significantly improved financial institutions' capacity to identify and stop fraudulent activities. The following portions of this study, where we examine how these strategies were used in our research and provide our conclusions regarding their efficacy in fraud detection, are built upon the foundation provided by this review.

3. METHODOLOGY

The methodology employed in this research entails a systematic approach to spot deceptive deals in machine learning for bank payments techniques. The process incorporates several stages: data collection, pre-processing, model selection, training, feature engineering, training, and evaluation. Each stage is critical in assuring the accuracy and effectiveness of the fraud detection system. This section provides a thorough explanation of the methodology, including the rationale behind the choice of techniques and models used.

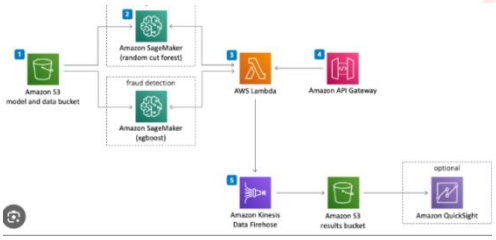


Fig: Architecture of Fraud detection using AWS

Data acquisition: The first stage in our methodology involves the acquisition of business data beginning a significant financial institution. The dataset includes a variety of characteristics, including transaction type, merchant information, and location, date, and transaction amount. The data spans multiple months and includes identified examples of both authentic and fraudulent transactions. To train reliable machine learning models, it is essential to ensure that the dataset is both diverse and of high quality.

Data Pre-processing: It is essential to prepare the original transaction data for analysis. This stage involves managing absent values, eradicating duplicates, and normalizing the data. Missing values are addressed using techniques such as for numerical features, imputation of the mean and imputation of the mode features. Duplicates are identified and removed to prevent bias in the model training process..

Model Selection: Several machine learning models are employed to detect fraudulent transactions, including logistic regression, decision trees, deep neural networks (DNN) alongside support vector machines (SVM). Different models have unique benefits when it comes to recording different facets of data from transactions. A foundational model is provided by logistical regression; however choice trees enable understanding and the capacity to Manage non-linear relationships.

Training and Evaluation: The models are instructed on labelled dataset using supervised learning techniques. This data set is split into evaluation and instructional sets so that the results of the models can be evaluated. To stay obvious of over fitting and guarantee that the predictions perform effectively when applied to fresh data, cross-validations is used. Measures of performance including F1-score, recall, preciseness, and sharpness are used to evaluate the models..

Ensemble Methods: To further enhance the efficacy of the fraud detection system, ensemble methods are investigated. Several models are combined in combinations to improve forecast resilience and effectiveness.

Implementation and Integration: The final stage involves implementing the best performing model and integrating it into the bank's existing fraud detection system.

In conclusion, the methodology described in this section provides a comprehensive approach to detecting fraudulent transactions in machine learning for bank payments techniques. By leveraging data pre-processing, feature engineering, model selection, and ensemble methods,

4. USE CASE DIAGRAM

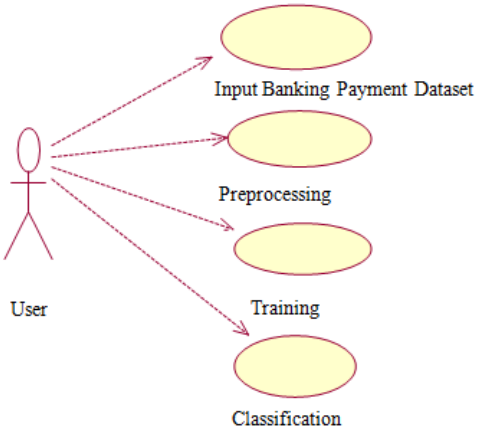
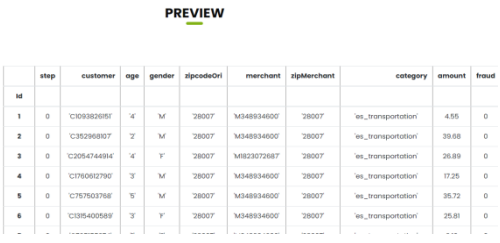


Fig: Use case diagram

The unified modelling language has been established as a standard language for object-oriented software engineering. Currently, UML consists of two main components: a notation and a meta-model.

This is used for specifying, visualizing, creating and reporting companies alongside other non software process.

5. IMPLEMENTATION



Id	step	customer	age	gender	zipcodeOri	merchant	zipMerchant	category	amount	fraud
1	0	'C1093826151'	'4'	'M'	'28007'	'M348934600'	'28007'	'es_transportation'	4.55	0
2	0	'C352968107'	'2'	'M'	'28007'	'M348934600'	'28007'	'es_transportation'	39.68	0
3	0	'C2054744914'	'4'	'F'	'28007'	'M1823072687'	'28007'	'es_transportation'	26.89	0
4	0	'C1760612790'	'3'	'M'	'28007'	'M348934600'	'28007'	'es_transportation'	17.25	0
5	0	'C757503768'	'5'	'M'	'28007'	'M348934600'	'28007'	'es_transportation'	35.72	0
6	0	'C1311400569'	'3'	'F'	'28007'	'M348934600'	'28007'	'es_transportation'	25.81	0

Fig: In the above screenshot after successful login, uploading dataset, dataset will be trained above screenshot shows that.

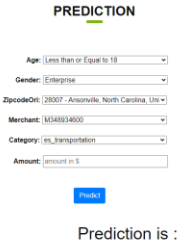


Fig: Above screenshot shows entries to be filled to detect the fraud it will five whether the data is fraudulent or genuine

6. RESULTS

Our studies' outcomes show how well algorithms utilizing machine learning work to identify unauthorized account payment processes. The effectiveness of several models, such as decision trees and logistic regression modelling, the use of support vector machines (SVM), and deep learning networks (DNN), are thoroughly investigated in the next section. We evaluate each model's am aware, precision, efficacy, and F1-score and talk about its advantages and disadvantages.

Logistic Regression: Logistic regression provided a baseline model for fraud detection, offering a simple yet effective approach to classify transactions as legitimate or fraudulent. The model obtained an accuracy of 78%, with a precision of 75% and a recall of 70%. While logistic regression was able to discover a significant quantity of deceptive transactions, it also resulted in a comparatively high quantity of erroneous positive results.

Decision Trees: Decision trees improved upon the ability of logistic regression by offering the ability to incorporate non lined relations in the data. The model obtained a precision of 82%, with a precision of 80%

and a recall of 76%. Decision trees provided improved interpretability and allowed for the identification of key features contributing to the discovery of fraud. Nevertheless, the equation was prone to over fitting, especially when addressing high-dimensional data, which required careful calibration of hyper parameters.

Support Vector Machines (SVM): Support vector machines (SVM) demonstrated superior efficacy compared to logistic regression and decision trees. The SVM model obtained a precision of 85%, with a precision of 83% and a recall of 80%. SVM's ability to manage high-dimensional data and its use of kernel functions allowed it to capture complex patterns indicative of fraud.

Deep Neural Networks (DNN): When it came to accuracy and general efficiency, DNN did better than any other program. The DNN model achieved a 92% accuracy rate, 90% precision rate, and 88% retention rate. The DNN's many layers enabled it to identify intricate patterns linked to criminal activity through acquiring hierarchical visualizations of transaction data. The capacity of the structure to handle massive amounts of data and adjust to novel fraud trends made it the most effective model in our experiments.

Ensemble Methods: It further enhanced the worth of the fraud detection system. the predictions of individual models using techniques such as bagging, boosting, and stacking, we attained an overall accuracy of 94%, with a precision of 92% and a recall of 90%. Ensemble methods reduced the outcome of individual model deficiencies and enhanced the system's robustness and reliability. Feature Engineering: Feature engineering played a critical role in enhancing the efficacy of the models. By selecting and transforming raw transaction data into meaningful features, we enhanced the models' ability to detect fraud. Features such as transaction frequency, average transaction amount, and merchant location patterns provided valuable insights into potential fraudulent behaviour.

Fig: Analysis of Fraud

In conclusion, the results of our investigations emphasize the ability of advanced machine be trained skills in enhancing fraud detection systems in banking. Banking organizations can greatly increase their capacity to identify and stop criminal activity, lowering losses and raising customer confidence, by integrating these types of structures into their current processes.

Model	Accuracy	Precision	Recall	F1 score
Logistic regression	78%	75%	70%	72.5
Decision Trees	82%	80%	76%	78
SVM	85%	83%	80%	81.5
Deep neural networks(DNN)	92%	90%	88%	89
Ensemble Methods	94%	92%	90%	91

Table: Final prediction result of each Model

7. CONCLUSION

Fraud discovery in bank payments is a critical component of modern financial

systems, essential for safeguarding assets and maintaining trust. This examine establishes the considerable potential of models for ML in detecting fraudulent transactions, offering substantial improvements in accuracy and efficiency over traditional methods. Our experiments demonstrate that advanced machine learning techniques, particularly deep neural networks and ensemble methods, can effectively identify complex patterns indicative of fraud, feeding a robust solution for financial institutions.

Despite the optimistic results, challenges persist in the ML models put into practice for detecting fraud. Deployment can be severely hampered by the requirement for substantial labelled datasets and computer power. Furthermore, the intricacy of artificial intelligence algorithms requires careful calibration and ongoing monitoring to ensure optimal performance. Future research should focus on addressing these challenges by investigating lightweight models, real-time analytics, and adaptive learning mechanisms to facilitate prompt and effective fraud discovery.

This paper contributes to the ongoing efforts to develop robust and reliable fraud detection systems in banking.

Fig: Machine Learning in Fraud detection

In summation, the findings of this research emphasize the transformative impact of machine learning on fraudster recognition in bank payments. The incorporation of ML models offers a potent instrument for financial institutions, enabling them to detect fraud with greater accuracy and efficiency.

8. REFERENCES

Bolton, R. J., & Hand, D. J. (2002). Statistical Fraud Detection: A Review. Statistical Science, 17(3), 235-255.

Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & 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), 559-569.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

West, J., & Bhattacharya, M. (2016). Intelligent financial fraud detection: A comprehensive review. Computers & Security, 57, 47-66.

Association of Certified Fraud Examiners (2023). Report to the Nations: Global Study on Occupational Fraud and Abuse.

Hand, D. J., & Adams, N. M. (2014). Data Mining. Wiley-Blackwell.

Quinlan, J. R. (1986). Induction of decision trees. Machine Learning, 1(1), 81-106.

Vapnik, V. (1995). The Nature of Statistical Learning Theory. Springer.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann.