Journal of Harbin Engineering University Vol 45 No. 9
ISSN: 1006-7043 September 2024

Leveraging Aiml Ops for Fraud Detection and Prevention in Fintech

[1]Pradeep Chintale, [2]Piyush Ranjan,[3]Nithin Reddy Desani,[4]Gopi Desaboyina, [5] Rajesh Kumar Malviya,

[1] Lead DevOps Engineer, Downingtown, PA, USA, chintale.pradeep@gmail.com
[2]AVP S/W, Barclays USA NJ , Piyushranjangc@gmail.com
[3] AWS, 2805 Pine branch Rd, leander, TX-78641, nithindesani@gmail.com
[4]SEI Investment Company, Phoenixville , Pennsylvania, USA, gopidesaboyina@gmail.com
[5]Individual Researcher, 14 Tall Meadow Court, Painted Post NY 14870, rajesh.malviya@gmail.com

Abstract

In this overall analysis it has provided information regarding the leveraging aiml ops for fraud detection and prevention in fintech. It has also provided information regarding how new improved innovations and technologies have started to prevent the fraud with specific software and tools.

1.0 Introduction

In the recent years it has been seen that there is a rapid up trend in the financial technology sector which is probably known as (fintech), the significant challenge which most probably ensures that the models would perform at its best and will ensure that the unseen data are crucial. This specific set of challenges are not new to fintech as several different fundamental aspects which reflect around the Machine Learning Known as (ML). One of the major concerns are regarding the amount of risk which is related to the overfitting of the models which are adjusted to the training data and that it can fail to address the new set of data.

2.0 Detailed discussion on SVM and KNN

Several authors such as shamsolmoali and zareapoor scientifically conducted a research where it was found that the fraud detection and the emptying of the several different analytical techniques provides significant information such as the SVM and KNN bagging ensemble classifier. The research paper also highlighted the key challenges of the specific fraud detection which was detected in the research and the security breach in the real time data. It was also noticed that there is a particular scarcity primarily due to the specific financial institutions which are like the banks which are safeguarding the basic set of data and its privacy concerns, which is enforcing towards researchers as the researchers actually rely upon

the simulated databases (Hein *et al*.2022). It also provides information regarding the issue of data imbalances which provides the data of fraudulent transactions which are typically around a percentage of 2% of the total. The rest of the 98% are actually known to be legitimate. Through a deep set of analysis it was found out that the experimental research is actually determined towards the traditional metrics which is specifically having higher accuracy and lower error rate which is actually inadequate for overall assessment of the model performance in the provided context.

3.0 Fraud Detection through Random Forest

The Random Forest method of recognition has proved to be a very reliable and accurate in detecting of fraudulent activities in financial services transactions. Through this learning method, learning innumerable decision trees is performed by randomly selecting subsets of training data and merging the votes of the predictions. The process of randomization prevents the model from being pumped and, instead, provides the foundation for generalization. A random forest classifier showed a better performance of 0. mathew correlation coefficient (MCC) rate. Categorization with 990% accuracy, obtained by deployment of Convolutional Neural Network and beating other approaches like Support Vector Machines and gradient-boosted trees. However, the combined application of SVM and AdaBoost produced the most accurate results and gave the model maximum Matthews Correlation Coefficient (100%) in a Southeast Asian database with biased distribution and less than 1% of fraudulent transactions. The main advantage

Random Forest has lies in its capacity to work with high-dimensional and difficult data (Riikkinen *et al*.2023). This makes it the perfect methodology for implementation in the financial sector, which involves several features. The approach of Random Forest to this situation involves using the output of several decision trees and then combining this outcome to identify variables that may have a link with fraud with accuracy. The researching has drawbacks too. They also warn users to take conservative precautions using the current model as not all categories of intangible assets have been included. The hybrid models that use together different machine learning methods deliver more accurate and reliable fraud detection outcomes than the method, that use only one algorithm. Overall, Random Forest algorithm in particular, in conjunction with ensemble techniques like AdaBoost, is a powerful tool in the sense that it is used in improving fraud detection and prevention in the Fintech industry, making it an effective way of safeguarding against the current financial fraud from occurring.

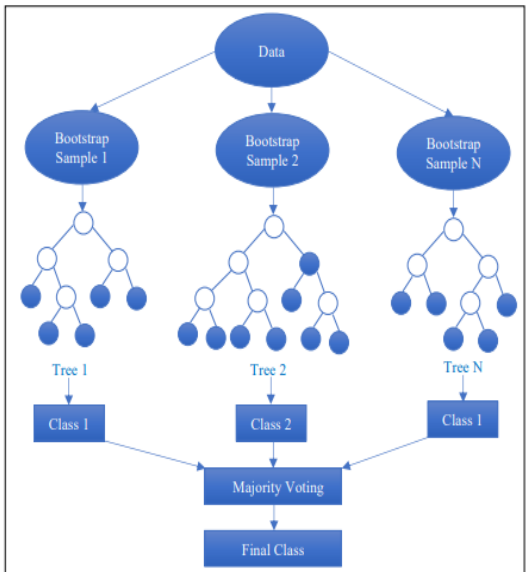

Figure 1: Random Forest

(Source: Riikkinen *et al*.2023)

4.0 Utilisation of the Convolutional Neural Networks

Convolutional Neural Network (CNN), which is a kind of machine learning strategy, is the initial stage to perform well in regards to effectively identifying suspicious activity coming from Fintech transactions. This analysis technique targets the conversion of financial data into a medium that would be able to support machine learning research, discovering potential fraud by searching into huge amounts of transactions. The approach initiates with collecting, categorizing and defining a set of Fintech transactions. Afterward, the transaction data is formatted in a way that matches the CNN parameter, either as one dimensional arrays or formally represented image arrays. The CNN architecture is built to deal with the regularity of transaction data, it might have the one-dimensional sequences of convolution image-based data representations which can be of two dimensions (Pattnaik *et al*.2024). The CNN model is trained on the dataset which is prepared with the methods of regularization and dropout being employed to stay away from an overfitting and to allow the model to generalize the new transactions which have not previously observed. There is the new validation set which is used for evaluating model's effectiveness and metrics such accuracy, recall and AUROC (AUC - Receiver Operating Characteristic Curve) are explained. Post training and data improvement, the CNN model can be linked in with a transaction or batch processing system that spots questionable transactions for indicative transaction investigations. It involves CNN's capabilities that are very good at recognizing patterns by which fraud is detected effectively in financial transactions thereby showing the strength of deep learning finance applications in security (Ajmani *et al*.2020). The CNN model utilized for fraud detection is coupled with issues such as selecting batch and real-time processing, and the maintenance of regulative compliance.

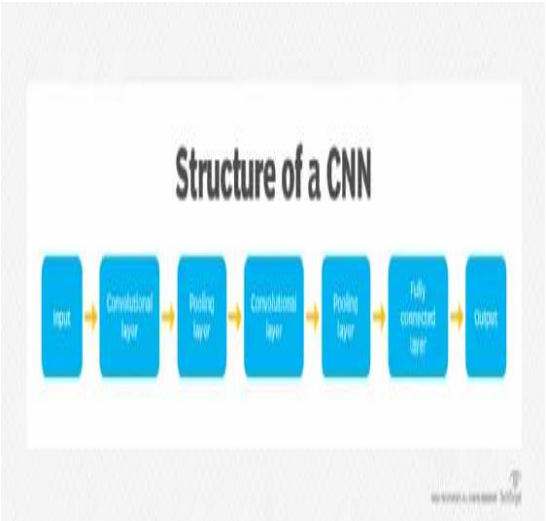

Figure 2: Structure of CNN

(Source: Ajmani *et al*.2020)

5.0 Using Anomaly for Fraud Detection

The strange event detection model is comprised of methods that enable it to detect minor anomalies which could signify evading transactions available. This approach consists in establishing deviations ("anomaly detection") as thresholds of fraudulent sidesteps. Each set, like transactional, balances, queries, and customer information changing are accounted for by a given score system of aggregation that highlights the similarities and progressively grade those exceeding the norm. Such anomalous observations can be detected by a blend of statistical as well as deep learning approaches, such as Z-score, Mahalanobis distance, and tailored encoders. Nevertheless, the Isolation Forest algorithm leads the pack in machine learning at this problem (Pang *et al*.2022). Isolation Forest methods is an outlier discovery algorithm which abandons random decision trees, showing the potential application of it in discovering anomalous patterns in financial transactions. The Isolation Forest algorithm is a base-by-base partitioning approach whose result is that each observation is excluded. Outliers are observed easily as it is a small amount of data and the value of attributes has a large difference from the other data. One major assumption in the technique is that an aberration would be more likely to be delay than the common case since it has a diversion from the sample mean. Anomaly detection in fraud detection consists of several steps, including accurate identification of the problem, enough knowledge of the regulatory and legal requirements, drawing of profiles showing usual user behavior, and the process of feature engineering to get the vital features. Models selection, data collection, preprocessing, training, assessment, and periodic feedforward are the main tasks to be considered in the construction process (Bangera *et al*.2023). Outliers detection systems, especially Isolation Forest algorithm, stand out among the tools for anomaly recognition and analysis of financial transactions, thus providing us with a great number to fight against the harm.

Figure 3: Credit Card Fraud Techniques

(Source: Bangera *et al*.2023)

6.0 Data Collection and Preprocessing

The whole machine learning model for fraud detection and money laundering monitoring is data-driven in a financial technology (FinTech) realm. Ensuring all data are of sufficient quality, and representative of our target rules, is a key factor for correctly trained models. Transaction information, user's data and past records, all of these are the primary factors for countering anti-money laundering and fraud. Some of the largest financial institutions in the world could acquire the data on which they want to make their decisions from third parties that collect and curate this type of information. Nevertheless, the model`s performance will be impossible without establishing automatic filtering algorithms, so that it does not operate inefficiently and without predjudice. The data cleaning is a vital point and needs diverse of the activities. First of all, the data should be cleaned by removing missing numbers as well as outliers' data and any noise or noises that are not accurate. Feature engineering for new data that may have been overlooked previously is performed to develop new features that may be more helpful or related to the model. Another type of pre-processing is data transformation that includes normalization or scaling the data which makes features of various units or sizes not having undue effect on the process of training. Based on the data properties inscription techniques like min-

max normalization, standardization, and logarithmic transformations will be used. Aside from these preparation acts, also the assignment of the data to the training, validation, and testing items is worth giving some thought. On the one hand, the training set is used for training the model (Aggarwal *et al*.2022). On the other hand, the validation set is reflected on the tuning of parameters and the estimate of the accuracy of the model during the process of training. The set of test data, which must be separated from the training data, is employed for the verification of the model's performance with respect to the samples with unknown attributes and which is lacking familiarity as opposed to the training data. Proper data collection and preprocessing are so important for fraud detection model development in Fintech as good data preparation ensures that models are trained on high-quality, representative data and can efficiently learn and achieve high validation accuracy on transferred cases.

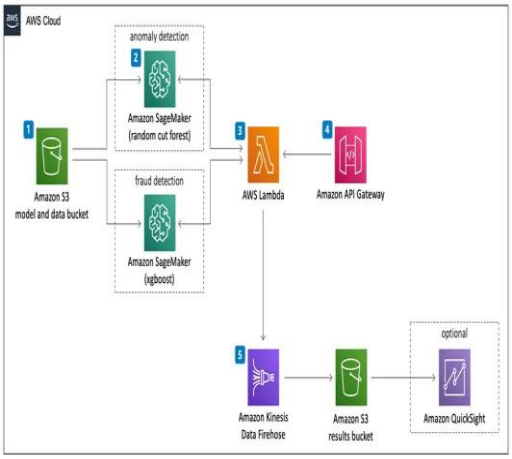

Figure 4: Machine Learning Architecture

(Source: Aggarwal *et al*.2022)

7.0 Unauthorised Transaction in fintech

The issue with unauthorized transactions in financial sector is that it can cause great sums of money to client's, account's, and banks accounts. AI and machine learning can aid in detecting and preventing fraud by closely inspecting customer expenditure and quickly-red flagging abnormalities. These models performa the historical transaction data and user behaviour patterns to build up a sample baseline for the regular expenses. With such an approach, models are capable of spotting deviations from regular transactions, which would be alerted as allegedly fraudulent activities. In particular, Mastercard's Decision Intelligence combine AI, the cardholder spending habits, behavioral baseline, and the current transaction, plus the time of the transaction is immediately used to establish if the new transaction is a fraud or not. If a particular attention is given to the kind of transactions which go against the usual purchasing habits, location, type of merchant, and other factors, the system could deny it during the process of its approval thus avoiding any losses. AI and machine learning algorithms may also flag identity theft symptoms like that of new passwords, invalid emails, and so on, by detecting irregular trends that are outreached from the account owner's normal behavior (Adedoyin *et al*.2024). The case when a consumer's data has been hacked or the system wants to send extra reminders for password change, the system can notify the consumer and adopt other sense of security measures, for example multi-factor authentication, to stop illegal access. AI-based fraud detection tools that implement highly advanced pattern recognition and anomaly detection functionality could be programed to get rid of the existing fraud methods and make illegal money transfers irreversible. This not only guards the customers' financial assets, but also allows the overall public system to maintain trust and raise the level of confidence.

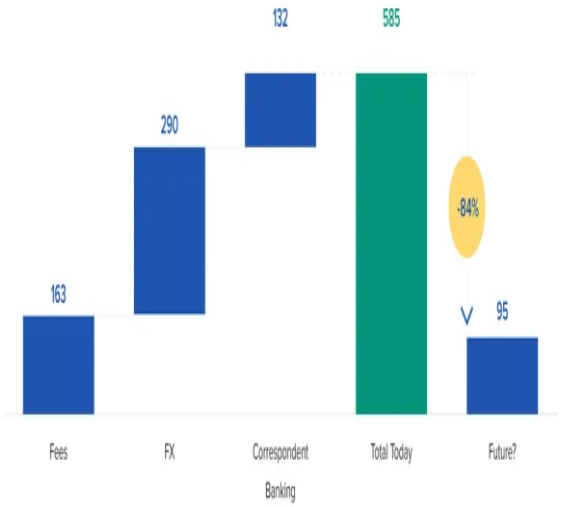

Figure 5: Unauthorised Transaction in Europe

(Source: Adedoyin *et al*.2024)

8.0 Identifying of Thefts

The financial sector is one sector where identity theft becomes serious when robbers' steal clients' personal details and then use their accounts for fraudulent purposes. AI and Machine learning

models can detect and prevent an act of identity theft which they can detect abnormally by observing the unusual behavior patterns different from the normal ones of an individual. These models work on the historical data to train the model and achieve a benchmark of activity for each customer, such as the transaction patterns, login activities, and account changes. Tracking of these acts by models may allow them to identify any anomalous behaviors likely tied to identity theft including replacing passwords, amending contact information, and transactions from remote locations or devices that do not usually perform these actions (Huang *et al*.2023). For instance, if an account of a customer is being reset with a new password or an information of a person posted on its account is being changed by someone abnormally, AI may notice that the activities cause alert for a possibility of identity theft. The system could then initiate a customer notification, request for additional security enhancements (multi-factor authentication) or a quick account freeze to prevent the unsuccessful logins. AI-controlled system of monitoring and tracking known criminals' modus operand may as well be used as a tool for detection of kinds of malicious identity theft activity. The models can detect and curtail these criminals' ways and tactics and thus stop identity theft at source. To sum up, AI as well as machine based learning procedures can be employed for enhancing the biometric identification systems like face recognition or speech.

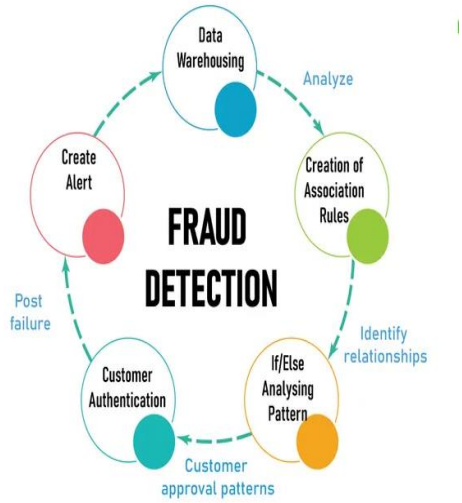

Figure 6: Identifying of Fraud
(Source: Huang *et al*.2023)

9.0 Preventing the fraudulent in Fintech

The AI and machine learning tools related to the fintech sector are currently being more and more implemented for the purpose of detecting different types of criminal activities in finance and substantiating the security measures. Current technologies become mighty tools to defend against the fraudulent cases, on which people and financial institutions might be protectionlessly damaged very seriously. It is one of the significant ways that an AI and machine learning is applied to the thwarting of frauds. These algorithms do analysis of the colossal amount of data which may include behavior of users, pattern of transactions and earlier frauds. They detect the tiny abnormalities from the general activity that signals fraudulent conduct. The method of working with the methods such as Isolation Forest algorithms has shown effectiveness in pinpointing suspicious behaviors in financial transactions which can be used to indeed take action immediately and to prevent the fraud. AI and machine learning algorithms may be pre-trained to spot different types of fraud, that involve holding bank accounts, identity theft, forgery of documents, as well as money laundering (Chanda *et al*.2023). These models are trained on data from previous account activities, including transactions, file uploads, and other suspicious behavior. Therefore, the models can highlight any accounts that are exhibiting abnormal or unusual trends for further investigation. On the other hand AI- supervised fraud detection system operate with a capability to monitor the normal patterns of the customer who in turn submits the abnormal habits to the system and as a end effect it results in prevention illegal access to the account. As well as these systems could trace the tendencies of identity theft like a change of anything about credentials or the contact information which immediately in turn leaded consumer to be notified. When evolving document fraud the AI and ML models can learn the patterns which are inherent to original signatures, ID documents, and other forms of paperwork, however in cases where even slight deviations from this pattern are observed, the human eye might miss them. By detecting information inconsistencies, these sensors affect a fraudster from getting away with cashing checks or getting a

loan using irrelevant documents. Moreover, this technology and approach can be of assistance with minimization of false positives, which may be an important concern for financial institutions.

Conclusion

AI and machine learning algorithms in obvious fraud detection and prevention financial technology have brought revolutionary change. Financial businesses are likely to do a good job at finding illegal transactions, suspicious patterns and identity threats with help from algorithms like Random Forests, Convolutional Neural Networks and Anomaly Detection Models. Besides, operation of such models requires proper data to be of the high-end, a good infrastructure, and persist monitoring and updating on a continuous basis. With the fast improvement of criminals' methods, the function of artificial intelligent and machine learning system in preventing the financial system to be hacked will play a meaningful role. This suggests that the field of studying AI and machine learning will become more efficient in detecting frauds.

Reference List

Journals

[1] Fritz-Morgenthal, S., Hein, B. and Papenbrock, J., 2022. Financial risk management and explainable, trustworthy, responsible AI. Frontiers in artificial intelligence, 5, p.779799.

[2] Riikkinen, M., 2023. Facilitating or Enabling Value Creation? Reconfiguring value creation in financial services.

[3] Kalyani, S. and Gupta, N., 2023. Is artificial intelligence and machine learning changing the ways of banking: a systematic literature review and meta analysis. Discover Artificial Intelligence, 3(1), p.41.

[4] Pattnaik, D., Ray, S. and Raman, R., 2024. Applications of artificial intelligence and machine learning in the financial services industry: A bibliometric review. Heliyon.

[5] Ajmani, P., Sharma, V., Sharma, S., Alkhayyat, A., Seetharaman, T. and Boulouard, Z., 2023, September. Impact of AI in Financial Technology-A Comprehensive Study and Analysis. In 2023 6th International Conference on Contemporary Computing and Informatics (IC3I) (Vol. 6, pp. 985-991). IEEE.

[6] Pang, G., Aggarwal, C., Shen, C. and Sebe, N., 2022. Editorial deep learning for anomaly detection. IEEE Transactions on Neural Networks and Learning Systems, 33(6), pp.2282-2286.

[7] Bangera, S. and Bhat, S., 2023. A Systematic Study of Application of Cognitive Intelligence in Mphasis–a Case Study. International Journal of Case Studies in Business, IT and Education (IJCSBE), 7(2), pp.360-380.

[8] Pang, G., Aggarwal, C., Shen, C. and Sebe, N., 2022. Deep learning for anomaly detection. Institute of Electrical and Electronics Engineers.

[9] Adedoyin, F.F. and Christiansen, B. eds., 2024. Generative AI and Multifactor Productivity in Business. IGI Global.

[10] Wan, H., Li, K., Huang, Y. and Zhang, L., 2023. Blockchain and financial e-services. In Springer Handbook of Automation (pp. 1371-1383). Cham: Springer International Publishing.

[11] Chanda, R. and Prabhu, S., 2023, May. Secured Framework for Banking Chatbots using AI, ML and NLP. In 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 60-65). IEEE.

[12] Nguyen, V.L., Lin, P.C., Cheng, B.C., Hwang, R.H. and Lin, Y.D., 2021. Security and privacy for 6G: A survey on prospective technologies and challenges. IEEE Communications Surveys & Tutorials, 23(4), pp.2384-2428.