IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 8, Issue 12, December 2020

DOI 10.17148/IJIREEICE.2020.81207

Real-Time Fraud Detection in Digital Payments Using Machine Learning and Big Data Analytics

Abhishek Dodda

Principal Product Manager, abhishek.dodda1@gmail.com, ORCID: 0009-0000-6728-945X

Abstract: Digital transactions are becoming faster and easier, providing numerous benefits to businesses and consumers alike. However, the increase in digital transactions is also fueling a rapid rise in digital payment fraud, an illicit phenomenon that attempts to exploit or bypass the control systems of online banking services. For banks and credit card companies, consumer protection against system failure is critical, and fast and real-time solutions are essential. Most traditional fraud detection methods rely on either approaches based on customer behavior, such as analysis of past transactions, or supervised learning based on domain-specific features, such as fraud rules and regulations. However, the rush to implement new services has left many banks and credit card companies ill-prepared to protect consumers from losses during online or mobile transactions. In this paper, we explore the use of novel machine learning techniques in the detection of online payment fraud. We explore the idea of deploying high-performing detections systems for real-time detection of payment fraud, while refining and learning local classifiers that can help even in situations that are very peculiar to a region. In addition, Big Data, if used in combination with proper validation processes, can facilitate rapid analysis of fraud schemes that require real-host interaction through the use of transaction fingerprinting techniques. This provides crucial information that could possibly lead to the unmasking of the true perpetrator. Furthermore, the use of actual user profiles can provide additional valuable data, but great care must be exercised to protect sensitive digital assets. Since it takes only seconds to commit digital payment fraud, business organizations must have not only the right technological framework, but also a detailed design and well-defined functional processes. By utilizing powerful detection systems, and also properly designing and implementing functional business processes, business organizations can implement a time-critical detection framework to efficiently handle the inherent unpredictability of this digital security threat.

Keywords: Real-time fraud detection, digital payments, machine learning, big data analytics, anomaly detection, transaction monitoring, behavioral analytics, supervised learning, unsupervised learning, feature engineering, predictive modeling, streaming data processing, fraud prevention, payment security, data pipelines, real-time scoring, model training, model deployment, decision trees, neural networks, support vector machines, clustering, risk scoring, false positive reduction, precision and recall, scalability, data lake, data ingestion, data enrichment, latency minimization, financial fraud, AI-driven security, fintech analytics.

I. INTRODUCTION

With the emergence of electronic devices and the Internet for online business, the financial services sector has moved into new forms that can provide users with ease of processing, computing, and accounting. E-commerce is currently the fastest-growing form of business. At present, many companies participate in e-commerce as well as small and medium-sized enterprises. Usually all of these companies will provide a website where users can purchase the desired items without having to go to the store. To support the e-commerce activities, the finance company provides various methods of payment systems that are very simple, easier, and time saving such as credit cards, electronic money, banking, etc. Today, online payment systems have become very popular and have sprung up that provide online payment transactions with the introduction of Debit cards, Cash cards, Credit cards and so.Since digital payment systems provide solutions to various transactions such as bill payment, ticket purchase, etc., the use of this method is increasing. However, the digital payment systems should also address security concerns related to sensitive information such as user passwords and authentication, customer databases, digital signatures, payment verification, etc.One of the biggest technical issues that all digital payment systems should consider is that transactions should be performed wisely and be protected in order to prevent fraud. E-commerce and Internet banking operations have become increasingly common; however, these services can be misused and abused by criminals. Nowadays, many Ecommerce companies and Internet banking organizations have been vulnerable to malicious fraudsters. Fraud may lead to substantial inappropriate profits along with major losses for the affected companies and predominately, the society at large. Fraud in the Digital Payment sector is the wrongful or criminal deception intended to result in financial or personal gain. People are being preyed upon, manipulated and robbed of hard-earned money after being enticed by "Too Good to Be True" Offer.As a result, businesses must adopt

IJIREEICE
ISSN (Online) 2321-2004
ISSN (Print) 2321-5526
International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering
Vol. 8, Issue 12, December 2020
DOI 10.17148/IJIREEICE.2020.81207

advanced security measures such as multi-factor authentication, encryption, and machine learning-driven fraud detection to mitigate these risks. Additionally, continuous monitoring and real-time transaction analysis are essential to swiftly identify and block fraudulent activities before they cause significant damage.

II. OVERVIEW OF DIGITAL PAYMENT SYSTEMS

The process of utilizing electronic methods, instead of traditional paper-based methods, to conduct a payment transaction is called digital (or electronic) payment. Digital payments facilitate the transfer of monetary values through electronic modes. In digital payment systems, users do not require to visit any physical location or output their payment in the form of cash or check. Instead, they can effect electronic payments any time from any location using a computer, smart device, or other electronic devices. The following advantages of digital payment systems from the perspective of a payer include: a) ease of payment, b) no expertise requirement, c) no time limit on payment, d) any location-enabled payment, e) payment via multiple modes, f) payment with less living cash, g) support of account held funds, and h) less time consumption. These advantages can facilitate faster payment processing as well as reduction of queues at merchant locations. From the merchant's perspective, digital payment systems involve less handling of cash, which can help in reducing the chances of cash-related frauds.

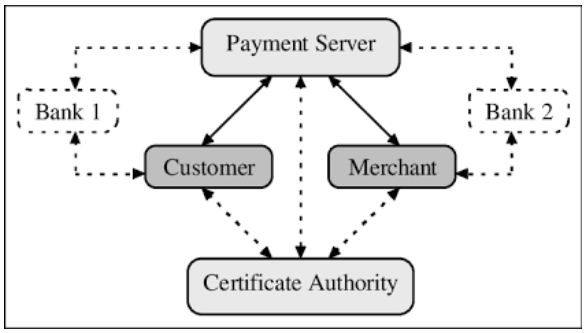

Fig 1 : Schematic diagram of a typical payment system

Digital payment exempts people from visiting banks and queues to pay utility bills, school fees, etc. On the other hand, merchants can get paid easily and quickly without requiring to count and manage large amounts of cash, which directly benefits business owners. Additionally, the use of payment gateway systems to pay online enables merchants from the e-commerce industry to accept money from customers, worldwide. Bank transfers, debit card, credit card, prepaid cards, e-wallets, cryptocurrencies, biometrics, UPI, etc. are some modes of digital payments. Adding to this, several banks and financial companies are developing innovative digital payment services to benefit their users.These advancements are streamlining transactions, enhancing user convenience, and fostering a more inclusive financial ecosystem. As a result, digital payment systems are not only driving economic growth but also encouraging financial inclusion for individuals in both developed and emerging markets.

III. FRAUD IN DIGITAL PAYMENTS

The digital payment landscape witnesses an incident of fraud approximately every 2 seconds globally, with approximately $28.58 billion lost to fraudulent transactions. The rapid advancement in related technology infrastructure, coupled with the internet becoming a ubiquitous channel for business and social collaboration, has given birth to one of the fastest-growing industries: digital payments, also known as e-payments or online payments. Digital payments allow the transfer of monetary value from one entity (payer) to another (payee) using electronic methods via the internet and, more recently, mobile devices.

Digital payments have undergone many transformations over the years, from the early days of electronic fund transfers to the renowned payment card systems to ever-evolving online banking facilities to the emergence of mobile wallets and near field communication payments. Digital payments are broadly categorized as card-based payments (credit and debit cards) and non-card-based payments (peer-to-peer transfers, mobile wallet payments, bill payments, NFTs, and cryptocurrency). It is forecasted that approximately 61% of global consumer retail purchases will be settled using e-payments by 2024, and several countries are taking every measure to promote a digital economy. Digital payments have been gaining increasing popularity among consumers due to the ever-growing need for real-time settlement, increasing mobile usage, availability of easy-to-use and contactless payment mechanisms, and the wide acceptance among merchants.

IJIREEICE
ISSN (Online) 2321-2004
ISSN (Print) 2321-5526
International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering
Vol. 8, Issue 12, December 2020
DOI 10.17148/IJIREEICE.2020.81207

3.1. Types of Fraud

Fraud in digital payments can be classified according to the different stages of payment process fraud affects. Merchant fraud occurs before the formation of the payment service agreement: a weak merchant intent creates the risk of many payment channels. User fraud happens before the fund transfer, as the bad user tries to abuse the payment channel without reimbursement. Client or cardholder fraud occurs after the payment service agreement is approved as the user manages to pay without the bank authorization. Gateway routing fraud disables the upper security levels of a payment gateway on a Merchant Based toward an acquirer that authorized an acceptable risk merchant classification. Acquirer fraud occurs when a weak acquirer intent creates the risk of many payment services on a Merchant Based. Interchange fraud occurs after the transaction has been executed as the bad user tries to abuse the payment service without a transfer from these schemes.

Merchant fraud is created by a malicious user that does not intend to supply the goods and services expected by the client. This can happen on any electronic transaction based on pre-payment; credit card payments are not unique for this type of fraud. It is not difficult to set up a web page that appears to operate a legitimate e-Commerce site and takes credit card purchases without delivering anything. Many merchants are present in the Internet trying to click and collect goods from Internet Provider Payment Processor Servers. These processes have been extremely penalized by phishing, done to point malicious users to legitimate looking merchant pages.

Fig 2 : Digital Payment Frauds

3.2. Impact of Fraud on Stakeholders

Fraud is a serious problem for stakeholders involved in electronic payments. First, fraud harms consumers as their financial assets are lost, and their card numbers are abused. But even more serious are the emotional consequences: many victims report feelings of anger, stress, and panic, and claim that a hard lesson was learned from being a fraud victim. Some victims will terminate their existing contracts with the payment service providers and switch to a company with better security measures, putting the loss during that period down to bad luck. In severe cases, dealers are also held partly liable. For merchants selling products via a payment provider, often the consumer receives a refund from the payment provider without any consequences, while the merchant has to bear the loss on the purchased goods. Merchants also bear the risk of false declines, meaning that a payment transaction was rejected by the payment provider, although the customer was just on the verge of committing fraud. This might put potential customers off and cost merchants additional sales.
Finally, fraud also harms payment service providers. Costs incurred in the period between fraud detection and transaction delay, cost of dealing with these companies either through email, telephone, or chat when they are not satisfied, and costs incurred through false declines are major business detractors. At a certain point, customer satisfaction levels will decrease, driving away merchants. This phenomenon is called negative Word of Mouth Development. Suffering from fraud, the level of risk appetite for increasing market shares is driven down. An inaccurate fraud detection system is costly, as well, when either a drastic increase in sales or large amounts of voided transactions occur. This is because the company will either have to employ more agents in the target area or will have to optimize its current resources. Therefore, traditional fraud detection systems are typically threshold-based, but identifying abnormal behavior through machine learning and big data approaches is sensitive and successful.

IV. MACHINE LEARNING IN FRAUD DETECTION

As the number of financial transactions continues to grow, fraud detection has become one of the most important issues in the field of finance. A wide range of artificial intelligence techniques are employed in this area. Machine learning encompasses a series of technologies that attempt to learn regularities from past data, evaluate these observations with new data, and detect these unusual or different cases. Machine learning applications in fraud detection can be classified

IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 8, Issue 12, December 2020

DOI 10.17148/IJIREEICE.2020.81207

according to the type of model used, the nature of the training data, and the engine used to create the model. Consequently, we can distinguish between supervised and unsupervised models.

Supervised learning models are often used to solve fraud detection problems. The most popular algorithms in this area include regressions, decision trees, artificial neural networks, support vector machines, k-nearest neighbor classifiers, Bayesian classifiers, and boosted classifiers. Supervised learning is accomplished through supervised training sequences involving labeled example transactions. This approach suffers from certain limitations. On the one hand, training on labeled transactions involves a great deal of manual effort. On the other hand, there is a lack of labeled training data, with labeled examples coming only from past fraud cases. These limitations extend to supervised models specifically designed to cope with data scarcity, such as semi-supervised and active-learning approaches, as well as to early warning models that are built using transactions made in real time prior to the fraud attempt. The use of labeled training data is necessary in order to detect known fraud schemes, such as money laundering and card theft, but is not sufficient to detect new fraud schemes. In other words, supervised learning techniques alone cannot provide adequate solutions for all fraud detection problems.

Consequently, unsupervised learning techniques train fraud detection models without having to extract labeled transactions. After building these models, the algorithms assign scores to transactions according to the extent to which these transactions differ from the normal behavior reflected in the models. Unsupervised learning is applied to a diversity of models, such as clustering, self-organizing maps, and outlier detection algorithms. Reinforcement learning also serves as the basis of certain models in the market.

Eqn 1 : Binary Classification Prediction

$$P(\text{Fraud}|x) = \frac{1}{1 + e^{-w^T x}}$$

- x: feature vector (e.g., amount, location, frequency)
- w: weight vector (learned during training)
- Output: probability of fraud (0–1)

4.1. Supervised Learning Techniques

Supervised learning is a conventional machine learning paradigm that assumes the prior availability of a fully labelled dataset, allowing the model to be trained using labelled examples. Classification and regression are the two types of supervised problems. The simplest supervised machine learning approach is to train a single classifier that receives all of the labelled examples, or a class-conditional model. Unfortunately, the current massive dataset sizes encountered in many practical applications do not allow the development of such simple models. Despite the wide availability of massive datasets with millions or billions of examples, the classification of rare events, like detecting fraud, is still a challenge and can be very costly in terms of both effort and resources.

Traditional machine learning models require biasing towards models capturing the minority class during training. In fraud detection, the classes are defined based on the outcome of the activity monitored; thus, a model trained using outcomes from only the majority class has no chance of learning those features associated with the minority class, which ultimately results in a high false negative rate. Additional issues with traditional machine learning include the problem of many redundant training attributes and the fact that some available attributes may not be relevant at the time of prediction. In addition, many traditional machine learning models also rely on assumptions of the distributions of the attributes and classes which may not hold true or known. These assumptions may lead to systematic errors or bias in the prediction of the outcome. To mitigate the above issues with traditional supervised learning, several approaches have been proposed for bias correction, such as over- and under-sampling, cost-sensitive learning, and ensemble methods. Ensemble methods learn multiple classifiers that can combine multiple weak classifiers to produce a more accurate classifier.

4.2. Unsupervised Learning Techniques

However, there are some important limitations using supervised learning methods: firstly, in the case of fraud detection, fraud samples are much smaller than genuine samples, and secondly, labeled data will not contain a sample of each unique kind of fraud. Consequently, it is so hard to identify adequately, most of the time, the existing classes of fraud. Unsupervised approaches work with unlabeled datasets, clustering the samples, and attempting to identify the fraud samples. Such techniques can be further divided into three approaches: clustering methods, distance/similarity-based methods, and validation of fraud fingerprints.

IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 8, Issue 12, December 2020

DOI 10.17148/IJIREEICE.2020.81207

Clustering methods are based on grouping similar transactions together. The basic idea is that since fraudulent and genuine transactions are normally very different, they will be in different clusters. However, the main issue with this approach is how to choose the appropriate clustering methods, meaning the appropriate number of clusters and the features used for clustering.

Distance/similarity-based methods work with the assumption that fraudulent instances have different or even unusual characteristics than other instances in the dataset. An uncommon row does not imply a fraud operation, but it has a high probability of being a false operation. Such unusual characteristics can be determined by measuring distance or similarity toward the normal transaction profile and flagging as probable fraud operation whose distance or similarity exceeds an established threshold.

4.3. Reinforcement Learning Applications

In recent years, Reinforcement Learning (RL) has emerged as a popular approach for solving complex sequential decision-making problems in Artificial Intelligence (AI). RL is conceptually motivated via the notion of trial-and-error learning from environmental feedback. There is significant present-day interest in RL, inspired by empirical successes in applying deep neural networks coupled with RL to problems such as playing video games and robotic control tasks. We first cover some AI results applied to domains such as computer games, card games, and web-advertising. Although not fraud-specific, these results implicitly suggest a role for RL in solving Fraud Detection (FD) problems via reusable RL modules.

Reinforcement Learning (RL) begins with the definition of a Markov Decision Process (MDP) model that specifies the sequential decision-making problem. An agent interacts with the problem environment by taking actions and receiving rewards based on observed states. A common environment model for RL studies is stochastic including time-varying properties, which may be time-homogeneous or time-varying. In this chapter, we will focus on time-homogeneous MDP problems with discrete states and actions with random reward sequences. Reinforcement learning builds on the MDP framework in two basic ways: in theory, via fundamental results and algorithms and, in practice, RL designs specialized usable modules that can be instantiated to solve specific MDP problems of interest. Reinforcement Learning (RL), a sub-field of Machine Learning (ML), concentrates on learning optimal sequential decision-making policies by optimizing rewards associated with each decision, with actions selected in sequence over time. Unlike Supervised Learning, Reinforcement Learning need not be provided with either label actions or a corrective signal indicating appropriate updating. Thus, the agent generally learns through trial and error, discovering only the rewards associated with the actions undertaken. The general concept of Reinforcement Learning is that of an agent interacting with the environment over time, learning how to choose actions to maximize a criterion that depends on the actions chosen.

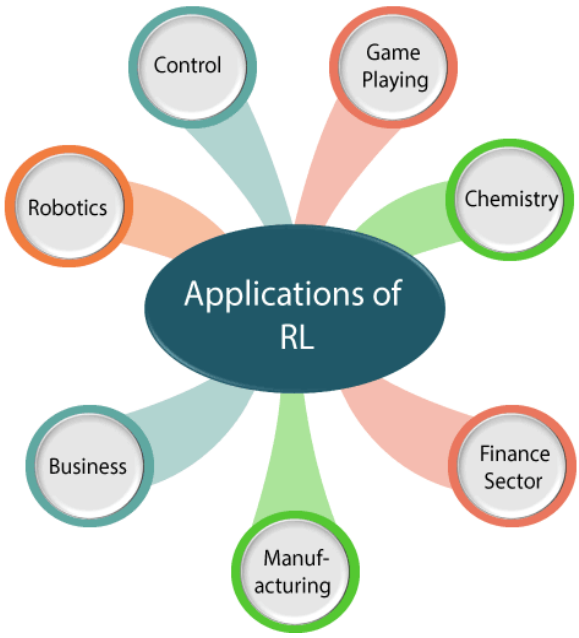

Fig 3 : Applications of Reinforcement Learning

IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 8, Issue 12, December 2020

DOI 10.17148/IJIREEICE.2020.81207

V. BIG DATA ANALYTICS IN FRAUD DETECTION

A large number of Internet users are now performing financial transactions through the use of smart mobile devices. The extremely rapid increase in electronic payment methods has led to an exponential increase in the possible methods for electronic fraud. This has created the need for a technological solution that is able to identify fraudulent transactions in real time and effectively detect irregular behavior. While the traditional technologies and security measures currently in place such as risk profiling, authentication, data encryption, intrusion detection systems, and public key exchange mechanisms have been able to slow the increase in fraud activity, the surging volumes of generated transactional and user data have made it increasingly difficult to use the traditional methods to flag potential violations based on the established rules.

Data generated by e-service application receives, captures, stores, and collects a substantial amount of structured and unstructured data. This data needs to be analyzed to extract significant patterns. The knowledge and information gained from this pattern is then used to create models for fraud detection. This is done either by utilizing the characteristics gained from user transactions or behavioral information of individuals or by collaborating with security experts to extract rules for the model. Machine Learning is a category of data analysis that recognizes regularities in data and enables prediction of new data. The big data ecosystem consists of a collection of tools that when applied in a proper way are capable of processing data in an effective way. Features leveraged from built models are used to train machine learning models on different types of classification. The rest of this section discusses in detail the sources, types, and processing techniques for data that modern fraud detection systems use. It also discusses real-time analytics in the area of fraud detection.

5.1. Data Sources and Types

Over the years, numerous frauds have surged involving large amounts of monetary funds. Hence, organizations such as credit card companies, banks, mobile wallet companies, etc., have started collecting large amounts of real-time transactional data in order to detect such fraudulent transactions. In order to use big data analytics for real-time fraud detection, it is important to have already existing large classified historical datasets that contain records of legitimate transactions and also of fraudulent transactions with valid labels. These records can then be used to model the big data algorithms for prediction purposes. Two sources for developing historical labeled datasets are the datasets available publicly as well as the datasets obtained from the organization involved in transactions. Likewise, other datasets include the past records of transactions of the credit cards especially related to charges which are disputed or otherwise declined. The organizations collect all types of transactional data while the customers deal with them such as bank transfers, withdrawals, deposits, credit card transactions, net banking transactions, ATM counterpart transactions, card transactions, etc.

The types of data used in case of transactional fraud detection are diverse and available in volumes. These data include the customer attributes such as name, relationship, address, age, occupation, email address, household income, number of years with the bank, employment type of the customer, etc. For the account, the data include the type of account, branch address, account balance, number of recent transactions on the account within a specific timeframe, known regular deposits, category of deposit accounts, deposits made, time taken, date of deposit, etc. For the involved transaction, the concerned data include the transaction fee charged, transaction size, transaction date and time, transaction description, channel type used for the transaction, transaction location, transaction type, type of transaction counter-party, status of transaction, etc. Moreover, the data also include the device and surrounding attributes such as IP address, type and number of the devices, geographical location where the transaction occurred, event logs of the surrounding device, etc. All this information captured from the extensive transactional data of customers is then used for modeling purposes in machine learning for such significant risks related to finance, security, etc., that the organizations deal with

IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 8, Issue 12, December 2020

DOI 10.17148/IJIREEICE.2020.81207

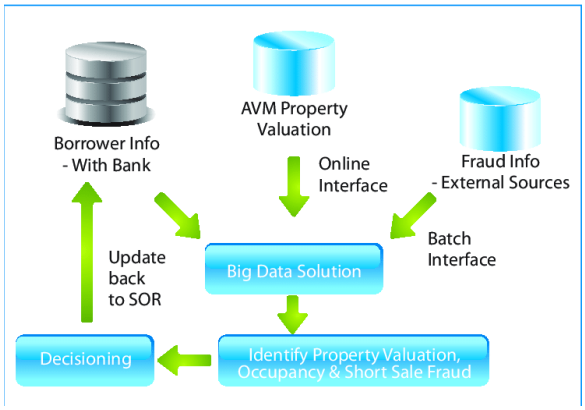

Fig 4 : Enable Fraud Detection using Big Data Solution

5.2. Data Processing Techniques

Data preprocessing is an important stage of data analytics. It generally consists of converting the raw data into a clean dataset. Data processing techniques include data cleaning, data transformation, data integration, and data reduction. There are multiple ways to achieve data cleaning, for example, by addressing duplicates, removing irrelevant features, filling in missing values, filtering noise from data, and fixing inconsistencies. In addition, the methods used to accomplish data transformation can include normalization, data type conversions, and attribute encoding techniques. Furthermore, to integrate different data sources, the ontology mapping approach can be achieved to track ontologies, map them to a common metadata using a controlled vocabulary, create an integrated ontology, and map them to a common metadata.

Moreover, to reduce the size of the dataset, data reduction can be utilized to compress the data, reduce the number of records, or reduce the dimensionality of the dataset. However, it's important to apply data reduction carefully because if the data is reduced too much, it could affect the results. This section provides a detailed description of the data preprocessing techniques applied to the network data used in the framework.

This includes the validation of the transactions in the main dataset and the modes of the transaction amount by day of week. The main dataset collected from the analytics system includes several redundant transactions. In this case, a day starts at midnight UTC but depending on the customer location, the merchant location, and the merchant acquiring processor location time zone differences can be applied the transaction might have different dates. Therefore, the network transactions during that day have to be grouped based on the day of the week of the transactions. After grouping the transactions, the most common amount of the transactions needs to be validated in a way to protect against the outliers.

5.3. Real-Time Data Analytics

The importance of data analytics in fraud detection is especially emphasized by real-time data analysis techniques. Consider that a reasonable compromise between the accuracy of a detection score and the available time for fraud detection systems to operate implies that, at the very beginning of a financial transaction, we need to assess the risk with a response time of less than 200 ms. If we compare this to the actual Transaction Time, we see that a small fraction can be made available to the fraud detection script, usually 1%-2% of the total Transaction Time, which would mean that our model should operate in under 5-10 ms. Although this does not seem much, this implies that each detection evaluation is performed on a very small set of features, severely compromising the accuracy of a detection score based on transaction-level variables. Moreover, replaying the entire transaction only for detection purposes cannot be always performed. This has led several stakeholders to propose and adopt a hybrid approach that considers a transaction on the fly by sending the relevant parameters to the fraud detection engine, while also considering historical data related to the subject, the service, and other known neural networks.

Solutions based on one-class classifiers are proposed to improve detection accuracy by assessing a transaction according to a small set of transaction variables supported by historical identity-related and behavioral information deployed in tree-based models. Class-based solutions that use multiple independent trusted networks are also suggested. However, these solutions are not suitable for real-time detection of financial fraud. As in the transport sector, where speeding cannot force traffic managers at entrances or exits of cities to continuously evaluate probe data picked for all vehicles, transport flow controllers could not allocate more than a small part of the time for mitigation purposes for any transport vehicle if excessively charged in charge of managing complex signals for the area, or nearby districts could not deploy conflicting manipulations for too much differentiated times.

IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 8, Issue 12, December 2020

DOI 10.17148/IJIREEICE.2020.81207

VI. INTEGRATION OF MACHINE LEARNING AND BIG DATA

It is widely recognized that definitive solutions to big-data challenges will be achieved only by developing and deploying intelligent big-data technologies. In fact, the future of big data science lies in research that focuses on improving the lack of intelligence of true 'big-data' processing tools. To accomplish this, we are currently in an Arms Race to capitalize in commercial products on 'big-data' toolkits that have little or no inherent intelligence. In this regard, providing a solution to the challenges that arise from hybridizing machine-learning methodologies with big-data toolkits is a critical issue both for the future of artificial intelligence and the future of big data science. For example, despite the fact that machine learning plays a critical role in the intelligent processing of data, integrated solutions that effectively combine machine learning with big-data methods are only at the initial stages of development. Embedding machine-learning capabilities into data-processing systems would empower users who may not possess extensive machine-learning expertise and enable them to take full advantage of intelligent capabilities. Magic may happen when ML is embedded/leveraged in Big Data technologies.

In data integration, operations are crucial for optimal integration algorithms. However, those algorithms require high computing power and, more importantly, dozens of parameters to be defined by the user, typically along the entire categorization process. In the case of very big data, an absurdly high number of parameters will need to be manually set up to boards' cluster partitions. Therefore, it may be far more efficient to perform data integration tasks at lower clustering levels and in the initial phases of ML mining tasks. After that, each group of similar data points may become a specific task; special domain experts could concentrate on solving problems related to special types of data points. Results could be combined after an access level has been defined, making hierarchical ML a type of 'Big Data' task.

Eqn 2 : Model Training Objective (Parallel Gradient Descent)

$$w^{(t+1)} = w^{(t)} - \eta \cdot \frac{1}{n} \sum_{i=1}^{n} \nabla \mathcal{L}(f(x_i; w^{(t)}), y_i)$$

- $\mathcal{L}$: loss function (e.g., cross-entropy, MSE)
- w: model parameters
- η: learning rate

6.1. Architectural Framework

Data analysis is not a single point task; it is a whole iterative process that starts with data collection and ends with data visualization. To go through this whole process successfully, there are some very important players that come into action such as: the input data that comes from external sources, the data architecture that answers the basic questions about the system workflow such as: if the input data stream is organized or if it is a real time flow of unstructured data or of structured data, the data storage that is used to organize and make sense of the data, the data processing in which the data is analyzed using different processing models in batch or stream, in memory or disk and by using distributed or centralized execution, cloud or edge storage, the data processing tools that consists of the different techniques and models used to analyze the data, and finally the data presentation tools and models such as data visualizations that let users exploit the data.

The research work proposed in this section presents a clear architectural framework for the integration of machine learning and big data with the different models clearly identified and presented for each processing stage. The proposed framework is of a general use and comes to cover different domains and applications with different requirements and constraints. Its architectural design is modular such that developers can build and exploit real-time big data machine learning applications for many areas such as cloud or edge computing, autonomous vehicles or smartphones. In this section, the proposed framework concepts and components are presented with some examples to help readers understand the framework and to encourage them to use and adapt it for their machine learning and big data applications. Other solutions for different application constraints are also proposed such as supervised or unsupervised learning, easy or hard data collection, edge or cloud users, and offline or online learning processes. These other solutions can build a library of reusable big data modules that developers can use to fast build their machine learning big data applications.

6.2. Challenges in Integration

The integration of Big Data and ML comes with numerous challenges. Some of the first challenges are due to different storage technologies. Integrating systems written in different languages presents additional challenges, as does the need for businesses to access and use analytics tools for preparing data for ML, building ML models, and deploying and managing them. Another challenge is what can be called the "Deep Learning Gap": deep learning is a new technology that leads to fundamentally improved results in many fields. Traditionally, deep learning has required experts to design

IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 8, Issue 12, December 2020

DOI 10.17148/IJIREEICE.2020.81207

specialized systems to be embedded in larger systems. It's not realistic for most companies to have a cadre of experts working full time on deep learning problems. The limited number of producers leads to a "Deep Learning Gap": a very long time between companies needing deep learning solutions and having them available. Without novel solutions to the DL Gap, many companies will simply not be able to deploy and manage deep learning.

A third challenge has to do with the rapid adoption of cloud solutions. The flexibility and low cost of cloud technologies are leading companies to move their infrastructure to the cloud. Companies that use a cloud-first strategy want technology tools that work on the cloud. In practice, however, most AI applications will find it easiest to deploy systems that integrate both public cloud and on-premises infrastructure. A delayed technology push from the on-premises world to the cloud are tools that allow companies to develop true end-to-end ML pipelines where any component could be used in the workflows, but overall processes could be executed seamlessly. They also want AI client tools that assist them with every step of the journey: preparing data for ML, building ML models, and deploying and managing them.

VII. CASE STUDIES OF FRAUD DETECTION SYSTEMS

This section focuses on the applications of fraud detection system in three different domain areas. These are, fraud detection in e-commerce, fraud detection in bank credit card, and fraud detection in mobile payment systems. We present these three applications below.

Phishing is increasingly becoming a big problem in e-commerce. Phishing is a crime in which a person gains access to sensitive information of others by masquerading as a source trusted by the visitors of the web site. Most people are likely to fall into the trap of a phishing web site because the phishing web site tries to recreate a similar environment to a real web site. Hence it is difficult to detect that the web site is a phishing web site.

Phishing simulation systems try to simulate the real world by generating a phishing web site in order to study the detection capabilities of different people, groups, and tools against fishing attacks. Therefore there is a considerable research interest in the detection of phishing web sites. Researchers have proposed models that use classifier algorithms to detect phishing web sites based on different features of the web site. Several features have also been proposed for incorporation into the classifiers which can be levied from different levels.

Credit and debit card fraud is a crime in which a person uses another person's credit or debit card illegally to remove money from that person's account. Activity of every card holder is tracked and recorded in order to detect fraud. Detection of abnormal patterns or violations of a model characterizing the normal transaction patterns is key to credit and debit card fraud detection.

Credit and debit card transactions are nearly impossible to be expected without any reliable prior knowledge about the models of an abnormal transaction. Hence the model of the normal transaction is often considered as the model of abnormal transactions. A normal behavior of every single credit and debit card holder is developed and is represented as a model. In order to update these models a real time system is required that monitors all transactions of each credit and debit card holder. The adaptation of the model's parameters of the credit and debit card holder is very important.

Mobile payment is a transaction where a customer pays for products or service fees using his/her mobile devices. Mobile device has gained more popularity due to its functionality of use, such as operating electronic mails, web browser access, navigation system, and navigation system. In a mobile payment system, customers have to register their account either with a mobile payment service provider, retailers, or banks. After remaining in the database, they are ready to conduct mobile payment transactions with merchants for purchasing goods or services.

A merchant contacts a payment service provider to issue a token for a customer and sends a purchase request via the mobile client. The provided token exists to authorize transaction processing without going back to the trusted third entity, signature generation. The role of this authentication is to complete transactions via an authorization routine, so that the merchant can act on behalf of the customer to conduct payment without compromising transaction confidentiality and privacy of the service providers.

IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 8, Issue 12, December 2020

DOI 10.17148/IJIREEICE.2020.81207

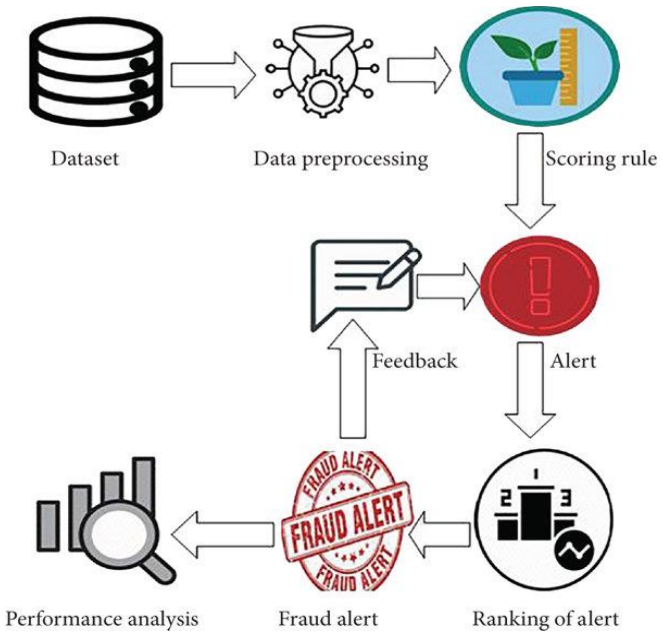

Fig 5 : Proposed system block diagram

7.1. Case Study 1: E-Commerce

Detecting fraud in digital payment systems used for e-commerce is indeed a big deal - people get annoyed because some arbitrary payment didn't go through and they have to use some alternative payment method or system. On the other hand, if payments are going through but somebody else is enjoying the products or services instead of the consumer, the service provider's resources are exhausted at their own cost. Thus, reliability of detection algorithms in either false acceptance or false rejection of fraudulent transactions is critical for the success of e-commerce businesses. On one hand, online store front needs to discourage people from committing fraud by having stringent rules applied to transactions going through. Obviously, it can't be unreasonable where people shopping online do not get fooled and bail out of using a particular system. On the other hand, people engaged in legitimate online shopping should not have a whoopsie when they see their credit card or bank transfer is not being approved for some recipient, especially if the transaction is a high frequently transacted or purchased product like digital music or videos. Maintaining this balance between the two is what is necessary to keep fraud to the minimal in e-commerce.

Many e-commerce online stores have now web interfaces enabled using machine learning techniques - mostly supervised learning - to extract features of transaction details, and trained on transaction sets of labeled data in terms of whether legitimate persons hold accounts there and made legitimate purchases there. Features that these algorithms examine include patterns in online time stamps, products and prices being bought or services being requested, Client IP address, and the predicted probability of the transactions being fraudulent based on user modeling. Algorithms then output probability scores that are interpreted by business rules applied on the data set to determine if the next incoming possible transaction is fraudulent or legitimate.

7.2. Case Study 2: Banking Sector

In our second case study, we focus on the banking sector. The traditional banking sector has experienced a major transformation due to the increasing popularity of online and mobile banking. Banks have become the primary target of both opportunistic criminals and highly sophisticated hacker organizations. Absolute safeguards that guarantee no financial fraud can take place have yet to be developed. Just as payment card and e-commerce providers are seriously concerned with financial fraud, so too are banks and the public at large. The possibility that consumers might suffer the loss of direct access to their bank accounts is a daunting prospect. Employing an experienced and professional fraud detection team to apply advanced analytical tools developed specifically for the financial sector is the single most effective way that banks can protect themselves.

IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 8, Issue 12, December 2020

DOI 10.17148/IJIREEICE.2020.81207

Eqn 3 : Customer Segmentation (K-means Clustering)

$$\mathbf{J} = \sum_{i=1}^{n} \sum_{j=1}^{k} \|x_i^{(j)} - c_j\|^2$$

- $x_i^{(j)}$: data point (customer) in cluster j
- c_j: centroid of cluster j
- n: number of data points (customers)
- k: number of clusters (customer segments)

Machine learning is increasingly being used as one of a suite of techniques for identifying cyber-attackers in relation to banking fraud. With the rise of fund misappropriation, card-not-present fraud, and identity theft; subjects are now more heavily reliant on cashless technology more than ever. Cybercriminals increasingly targeting cashless payment delivery, making a bank an attractive mark for a range of criminal activities. Furthermore, the rapid growth of the banking, finance and insurance services has resulted in increasing levels of competition. Financial services organizations are increasingly turning to machine-learning techniques used for detecting fraud in a bid to gain competitive advantage and to become leaders in fraud detection competency. While machine learning holds a key in enabling the efficiency and accuracy in the fraud detection process, the successful application of machine learning to the task hinges on the banking and financial services organizations using the right technique or techniques. The challenge for the industry is discerning the most effective combination of models, and their hyper-parameters, and having the technology capabilities and computing power to support real-time operations on a sustained basis. The sheer volume of transactions combined with the very low rate of fraud make extensive use of predictive analytics necessary in detecting banking fraud.

7.3. Case Study 3: Mobile Payments

Research on mobile payment fraud detection is fewer than credit card and online fraud detection. Fraud detection in mobile payments is slightly similar to credit card and online fraud detection. In addition, some of their methods are also similar. The mobile payment fraud detection approaches include: 1. Machine Learning. A multi-modal deep learning model for Mobile Online Banking (MOB) has been designed. The model has two levels. The first level is RNN-Encoder, which will extract multi-modality features from mobile online banking transaction data source. Then RNN-Encoder will transmit features to the second level for feature fusing with Fully Connected Network (FCN). All features are obtained by the fully connected layer. Furthermore, the IFCN will classify the features of actual transaction instances to identify each class; including Normal Transactions (NT) and Fraud Transactions (FT). A privacy-preserved multiple mobile phone collaborative design method for e-currency has been proposed. This approach can overcome the security challenges in e-money. Since there are multiple mobile clients, advanced secret sharing is utilized to secretly share the data needed for the fraud detection across multiple clients. Therefore, the e-money system provider can operate and protect sensitive information from fraudsters. The experimental results confirmed that the utilized methods are reliable and efficient. Unsupervised anomaly detection has also been relied upon.

A lightweight multi-dimensional fraudulent transaction detection framework for mobile payments has been designed. The detection is based on spatial-temporal sequence embeddings of mobile payments, which are vectors consistent with temporal and spatial zoning. The learned embeddings are combined with two classifiers for detection. For classification, a class-aware random forest classifier is utilized for fast and sensitive detection of large-scale and massive suspicious transactions. For detection confirmation, an extended XGBoost model is designed to be more sensitive to multi-category customers. The experimental results confirm that the proposed methods are reliable and efficient.

VIII. EVALUATION METRICS FOR FRAUD DETECTION MODELS

After training various classifiers and detecting fraudulent and legitimate transactions, quantitative evaluation would help us analyze the effectiveness of our models. Performance metrics assessing classification performance help determine the effectiveness of models; however, the crypto-image domain is a “Classification with Imbalanced Class Distribution” domain where the positive class is a tiny fraction of all transactions. So, the normal metrics, such as accuracy, precision, etc., would not be able to assess the models’ performance meaningfully. Precision and recall together help offer a more confident model evaluation than those isolated due to the fact that precision relies on a model to not misclassify legitimate transactions while recall requires the model to detect all frauds since both precision problems and recall problems are the common reasons of why classification fails.

IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 8, Issue 12, December 2020

DOI 10.17148/IJIREEICE.2020.81207

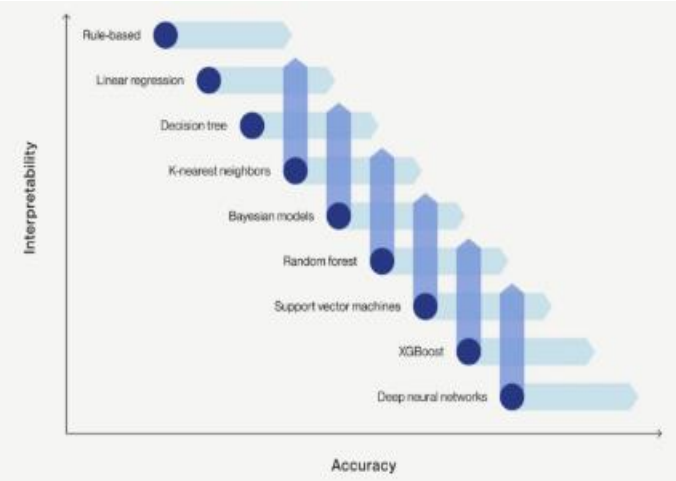

Fig : Data and Machine Learning in Financial Fraud Prevention

Precision and recall, ideally, want their values to be massive. Mathematically, precision is defined as follows:

Where TP is the count of True Positives, while FP is the count of False Positives. And recall is defined as follows:
Where FN is the count of False Negatives. In other words, precision reflects about false complaints and thus helping in avoiding negative impacts of the model; that is, warned but innocent/legitimate users are caused to suffer due to the warning while recall reflects about missed detections and thus helps avoid potential risk factors; that is, risk-loosing/exploiting users are caused to forsake possible damage and thus avoiding responsibility. That's why together precision and recall are the two most important model performance metrics of alerting fraud transactions in the fraud detection domain. Clearly, using only precision or recall alone won't be able to report the weaknesses of a model.

8.1. Precision and Recall

Precision and recall are the fundamental building blocks used in algorithms to evaluate the performance of binary classifiers, especially for imbalanced classes. This characteristic of precision is important in the case of fraud detection, as a high precision score would mean that a considerable number of verified frauds are indeed real frauds and are thus not false positives. More importantly, this would mean that innocent customers do not have to go through the hassle of disputing a transaction made by someone else, as their payment method would be involved in a fraud. However, this needs to come at a cost of missing out on a few fraudulent transactions. Accuracy cannot be relied on in this case as most of the predictions would be of the majority class, which is genuine transactions. Making sure that the transactions that are targeted as fraud are truly frauds is more important. On the other hand, recall would imply that we are covering a vast majority of the actual frauds. That is, a considerable proportion of the fraudulent transactions in the test set are detected as fraud. This results in a situation where a considerable percentage of the customers are not intercepted when committing a fraudulent act. The captured fraudulent transactions do not have a direct adverse effect on genuine customers. However, there is a hard requirement to satisfy. If predicting a transaction as fraud, that transaction should be returned as fraud, as any transaction that caused a major loss would make recommender systems useless and thus would discredit the bank. In that case, precision and recall are equally important. It may happen that on increasing recall, precision decreases and vice versa. In these cases, a weighted average score or a value called the F1 score is used.

8.2. F1 Score and ROC-AUC

While precision is a rough estimate of the percentage of correct frauds detected and the recall is a measure of correctness of the total frauds present in the data, the F1 score is the harmonic mean of precision and recall. When both precision and recall are high, the F1 score will be high. However, F1 score is usually a low score as it does not give additional weightage to precision or recall. The F1 score is most useful in cases of imbalanced class distribution or datasets having false negatives and false positives and the decision may take the lesser of the two. The F1 score is typically lower than the other scores as precision and recall bounds it. In order to increase the F1 score or maintain a high F1 score, both precision and recall should be increased. In marketing campaigns, in cases of higher false negatives, leads might miss the opportunity to convert. Similarly, if the false positives are higher, then the organization may lose money for fraud detection models. However, in real scenarios, the decision of what the threshold should be based on lies with the organization pursuing the study.

IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 8, Issue 12, December 2020

DOI 10.17148/IJIREEICE.2020.81207

Often, it may happen that we have both recall and precision pretty good including F1 score. But still, it does not show the actual insight into the model performance. For instance, AUC–ROC of 0.99 would mean that we have an extremely well-performing model when performing in test data. Now just to explain this, True positive rate (TPR), which is equal to sensitivity, is the ratio of correctly predicted positive samples and all actual positive samples. Whereas, the false positive rate (FPR) is the ratio of incorrectly predicted positive samples and all actual negative samples.

8.3. Cost-Benefit Analysis

The cost-benefit analysis (CBA) underpins many business models for fraud detection systems within digital payments. The key metric is the fraud recovery from false negatives over time, which can be used to assess the return on investment (ROI) of deploying mechanisms to manage fraud risk. In high-risk payments verticals or low-volume transactions, the fraud savings can be both large and expected in a short timespan. Conversely, low-risk payment activity with high-frequency will only provide small benefits over time.

In most use cases we have assessed, employing Machine Learning-centric approaches where we embed and deploy algorithms on downstream applications proves financially beneficial to the firm. We emphasize that the CBA will favor solutions offered only as a service if the provider can match the performance of company-owned models or processes. Although the rationale for detecting behaviors and actors who perpetrate frauds is generally undisputed, the industry has developed rules of thumb on when to apply what fraud mechanisms, and information transfer between industry members is opaque. As a practical matter, decision-makers must retrieve and compile costs and losses which banks within specific domains or industry verticals have experienced, as well as invoke a sanctioning policy for each specific type of fraud.
The observed flow and volume of transactions influence operating and functional costs. Overhead is influenced by manual processes. The latter can prove both unavoidable and long-term habitual if fraud cases require extraordinary attention, like cross-selling or up-selling cases where agents need to strengthen indirectly the primary focus on what is but a customer service attention. Generally speaking, both manual or semi-automatic processes with a high rate of known false negatives lead banks to commission automated risk scoring and user profiling algorithms.

IX. CHALLENGES IN REAL-TIME FRAUD DETECTION

Fraud detection systems can face a number of issues, including data privacy concerns, scalability issues, and high levels of false positives and false negatives. First, for liability reasons, digital payment processors may choose not to store transaction records, which would be a necessary step for training fraud detection algorithms. In such cases, third-party sources may be solicited to provide such sensitive information. However, since there is no guarantee of data collection accuracy or to what extent the data yield any useful representations, potential implications include loss of personal and sensitive data and creation of faults in data patterns. Due to these factors, it may be better to leave out a few transactions than to implement novel architectures while putting sensitive personal data at risk. Scalability is an important element of any ML model but is especially critical for predictions in time-sensitive areas like fraud detection. A payment network is designed to support a large volume of transactions occurring in a relatively small window of time. As the amount of data available to detect fraudulent transactions during this time continues to grow, the design of the detection models and their subsequent deployment face increasingly challenging requirements for scalability. In addition, unlike most other ML applications, fraud detection benefits greatly from utilizing large-scale models that have been trained with significant quantities of accurate data, such as those capable of representing rare-event prediction or transfer learning, which would be infeasible if traditional, centralized ML training methods were followed. There are other unique challenges that arise from how predictions for detecting fraudulent periods can result in extreme levels of false sensitivity or selectivity that can both be harmful. It is also hard to imagine how one could implement an ideal sampling strategy when the risk of being labeled as inadequate may be dire, meaning that punishment for failure will be much more severe than punishment for false alerts.

9.1. Data Privacy Concerns

One of the primary challenges in implementing machine learning and big data analytics solutions for real-time fraud detection in digital payments is data privacy. Keeping financial transaction data secure and private is very important. Users do not support easily sharing the sensitive, highly personal information contained in digital payment transactions with third parties. Privacy-preserving machine learning aims to protect users' information confidentiality without sacrificing the integrity and accuracy of the machine learning model or predictions. Preserving data privacy can be done during both data storage and execution time. Various techniques preserve private information at execution time, on-device machine learning, federated learning, and feature anonymization. Privacy-preserving machine learning is still a young and active research area. The solutions require trade-offs in the security level or accuracy of results versus computational time and complexity. A study identified that zero-knowledge proof techniques provide the highest level of security but require the longest computational time.

IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 8, Issue 12, December 2020

DOI 10.17148/IJIREEICE.2020.81207

The emerging legal framework enshrines the privacy and security rights of individuals for protecting their personal data, especially that of citizens. Various implications and challenges are posed to organizations or groups using machine learning or artificial intelligence by the order. The regulation promotes privacy through active design in machine learning technology and workflow, including compliance documentation, minimal data-set retention time, anonymization of private features, risk assessments, and public transparency regarding the machine learning technologies and methods employed. Balancing user data privacy with accuracy is particularly important in machine learning solutions, including machine learning of payments fraud. It is user data privacy, not model accuracy that users value most.

9.2. Scalability Issues

Fraud detection in the real-time transaction stream is a challenging and daunting activity due to the high volume and increased velocity of transactions. Over the last ten years, several measures have been adopted to significantly reduce the processing time even up to seconds; but still challenges remain. With such high transaction volumes, classification algorithms easily take several minutes and even hours to classify a single transaction. The situation worsens with the deployment of such models for detection of data streams where each transaction is unique, and unless the model is hosted all the time to classify the transaction as it is generated, the required results will not be obtained. Hence, we cannot afford the luxury of generating and buffering a batch of transactions, it needs to be processed in real time.

Time is indeed an important factor in any machine learning deployment in a fraud-detection scenario; but equally important is accuracy. False negatives might be acceptable to a certain extent, but false positives cause significant damage not just in terms of cost involved in reversing the transaction but also the loss of business due to distrust of the customers on the payment service provider. There is a lot of overhead involved in a false positive. Such a scenario will require the entire transaction stream to be classified post-mortem which will increase costs and also worsen customer experience during the entire process. Therefore, building models and algorithms that can work offline but will also support online processing is very desirable. At the same time, the fact that the incoming transaction is unique means that we do not have the luxury of waiting to detect several transactions at the same time. This poses a tremendous challenge to the fast and efficient design and service of the models used.

9.3. False Positives and Negatives

A wider transaction history can lead to more reliable detection of anomalous transactions, but first-time offenders might not always be picked up right away if their transaction patterns are similar to those of legitimate customers. This problem can be partially mitigated by examining not just the user but also the moved funds as well as their point of origin. If they have entered the digital payment network very recently, it is suspicious when large amounts flow from that address to somebody else's account, or when many small amounts pass through the exchange to different user accounts that look just like the target account simultaneously.

Unfortunately, adding too many levels of filtering might also risk tagging potential threats as false negatives. The classification of incoming transactions can therefore result in the following economic consequences. A false negative occurs when genuine fraud occurs, but the detection system fails to identify it. Such detection is a false negative or type two error, which means that real positives bias the false negatives but this leads to a false security for the consumer and a high loss for the provider. On the other hand, a false positive search warns about a problem, which is benign. This hinders consumers from using the identification technology. False positives influence the customer's experience negatively in all provided transactions. Credit card issuing companies have to bear in mind the costing conducts of false positives, which are dealt with differently by the establishment.

X. FUTURE TRENDS IN FRAUD DETECTION

The trends for the digital payment fraud detection technology are on the verge of numerous changes in the future because of digital transformation impact on economic, lifestyle, and business. However, at the same time, due to the increased functionalities of payment services, the ability to exploit the payment systems are coming to the criminals domain. Furthermore, the challenge of adaptive dynamic techniques applied to fraud detection systems is not properly addressed by the Banks and Payment Service Providers. Hence, it is essential that the researchers should suggest future directions in fraud detection in the area of emerging technologies like Artificial Intelligence, Machine Learning, Big Data and Block Chain and the implications on the changes brought by the Digital Transformation.

Various emerging technologies which have the power to mitigate innovative fraudulent practices are discussed. However, the mitigating power solely rests on the way these technologies are utilized. In the true spirit of Digital Transformation, organizations should adopt a parallel strategy of implementation of these technologies and also adoption of appropriate

IJIREEICE
ISSN (Online) 2321-2004
ISSN (Print) 2321-5526
International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering
Vol. 8, Issue 12, December 2020
DOI 10.17148/IJIREEICE.2020.81207

cultural changes at both the organization level as well as societal level. Hand-in-hand with the HUMINT or human intelligence track are the AI or Artificial Intelligence and Big Data Analytics track. The three tracks provide the criminal justice system a much broader protective wall against the newest in payment fraud schemes, and ensure swifter apprehension of those responsible, and a quicker road to justice for the victims. Researches and investigations could evolve more into being predictive and intelligent, rather than forensics centric focusing on after the fact analysis with a retrospective motion.

Future regulatory changes are expected in a few areas: adoption of ASIC and AI in the Criminal Justice System will have an impact on research and collaboration; more focus will be on the use of information security and fraud risk management standards; tightening of data privacy regulations; increased global collaboration; and upgrading payment fraud fraudster profiles. Adoption of these trends will help organizations to plan the digital transformation for their organizations, but more importantly use this plan to serve the society in a more productive way.

10.1. Emerging Technologies

Technology is a highly dynamic and fast-growing field. Emerging technologies such as Cloud Computing, Machine Learning, Big Data Analytics, and the Internet of Things (IoT) hold great promise for use in the financial sector for gaining a competitive edge over rivals. Thanks to various technological breakthroughs, the focus in the financial sector has shifted from transactional to interactive, with an emphasis on strengthening ties with consumers and offering customized services. These innovations have led to an increase in various instruments and channels, and the pace of transactions has become rapid. This guarantees efficient services and an enhanced consumer experience. However, modernization has also spread vulnerabilities. Financial institutions must invest heavily, not only on emerging technologies but also on protective measures to build and assure cybersecurity. A high level of trust in financial institutions as custodians of consumer funds is essential for the financial ecosystem to function smoothly.

Data is omnipresent and always growing, essentially due to the predicating systems. Big data refers to vast amounts of data assets that are difficult to capture, manage, and process within a predictable time frame, using traditional technologies. The challenges highlighted by big data in getting the businesses the most needed insights have necessitated the development of integrated system architectures or environments that can address the issue of data capture, creation, storage, management, sharing, transfer, analysis, and visualization of big data. The management involves operational processes and the policies implemented by all other technologies, and thus involves the use of both hardware, software, and services.

10.2. Regulatory Changes

As new services in financial sectors are being made available on a large scale, updates and changes to the regulations are also being made simultaneously. With the growing wealth and increasing power of mainly the developing countries, these new set of rules need to provide a high level of security and privacy for its users. The last decade has seen massive innovations and the establishment of new payment technologies worldwide. Financial institutions and organizations, in general, are forced to ensure the safety of the new financial services enabled through Fintech solutions. They will have to ensure that the effects of the problems such as the global financial crisis and the pandemic do not happen again. To ensure that the innovations are sustainable and do not affect potential users, the regulations need to be updated. It is important to strike a balance between the level of restriction and control that regulations impose on firms, so as not to stifle creativity and development of new and efficient solutions that help underpin economies, while ensuring that risks associated with the provision of these services are adequately addressed, so as to maintain integrity and stability of the financial system and broader economy.

The year 2020 was deemed the year of Cross-Border Payment regulations, which was centered around Enhanced Due Diligence. In 2022, the Payment Service Directive 2 was implemented. Its second version restricts the use of legacy security methods in the Account Authentication Service Provider and Strong Customer Authentication Service Provider. Recently, in October 2022, the Financial Conduct Authority and the Treasury published their strategy for the regulation of Crypto-Assets. The document aims to limit the use of Crypto-Assets for illicit activities and to protect users. With the regulations being updated and given a higher priority, we expect to see more centralized regulation such as stricter rules preventing the movement of illicit activities such as money laundering.

10.3. Consumer Awareness and Education

Consumer education is essential for the successful deployment of AI/ML tools in combating mobile payment fraud. As the pandemic has propelled contactless payments into the forefront, customers are more exposed than ever to threats such as account takeovers, credential stuffing, and social engineering. Payment solutions cannot eliminate the effects of these attacks by themselves; customers also need to be aware of the risks and access available security features in order to use

IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 8, Issue 12, December 2020

DOI 10.17148/IJIREEICE.2020.81207

them properly. As technology is an ever-evolving process, consumer education should also be on-going. Given the jump in new users in online payment systems, this becomes even more important for both tech-savvy and non-tech-savvy generations alike. Phishing email scams, for instance, affect virtually every internet user across the demographics, no matter their level of expertise. Progress regarding consumer education could be accomplished through greater outreach programs, better standardized user instruction techniques and policies, as well as collaborations with banking agencies and authorities.

Increases in cybercrime as a result of the pandemic have also contributed to ambition from international bodies for updated regulations related to privacy issues of e-payments and collaboration with machine learning models in detecting fraud patterns. Payments fraud losses are expected to climb significantly; this has triggered a rise in customer payment safety awareness, and agencies have expressed their intention to steer the implementation process of payment security solutions in the short term. Due to the potential risks involved for customers – including a significant loss of money or leaked sensitive information – it is crucial to develop collaborative security measures aiming to both protect user data while preventing fraud attempts – a feat made even more difficult as new mobile users flood the e-commerce market. In this situation, educating bank customers is fundamental in guiding them through the process of securely entering the world of digital payments and enhancing trust and confidence in the banking system. Payment security authorities and institutions are encouraged to issue clear messages to educate consumers on safe mobile payments in this sentimental context, as a simple mistake could compromise their financial stability.

XI. CONCLUSION

The digital payment systems are becoming popular day by day due to the advancement of technology, user-friendly features, and fast transaction qualities but on the other hand due to these factors digital payment systems are also become the top target of the fraudsters or attackers. Online transactions fraud or digital payment fraud loses billions of dollars annually and are increasing day by day. Traditional fraud detection systems cannot handle the huge transactions produced by the digital payment systems, thus some modern technologies are required to handle the huge transaction data as well as fraud detection and prediction. Machine learning and big data analytics techniques can easily manage the huge-size and high velocity transaction data and provide an efficient fraud detection system in a real or near-real-time manner. This research mainly focuses on the digital payment fraud detection sector and provides a clear view of using machine learning and big data analytics technologies to design and develop an efficient fraud detection system. It not only covers the theoretical aspects of digital payment fraud detection but also covers the practical aspects like different datasets, machine learning algorithms, big data analytics tools used for the digital payment fraud detection system design and development. In this research, a hybrid big data platform using big data analytics tools and machine learning algorithms is presented to develop the digital payment fraud detection system. Tools are used to design the hybrid platform for the data collection, storage, preprocessing, and analysis. Three different datasets are used to evaluate the performance of the machine learning algorithms, logistic regression, decision tree, random forest, extra tree, naive Bayes, k-nearest neighbors, support vector machine, artificial neural network, convolutional neural network, recurrent neural network models. Finally, detailed comparative analyses are performed to provide a clear understanding of which machine learning model performs better than others in the real-time prediction. The results show that the hybrid model of the convolutional neural network and recurrent neural networks with the transaction records provide the better fraud detection and prediction results.

REFERENCES

[1] Karthik Chava, "Machine Learning in Modern Healthcare: Leveraging Big Data for Early Disease Detection and Patient Monitoring", International Journal of Science and Research (IJSR), Volume 9 Issue 12, December 2020, pp. 1899-1910, https://www.ijsr.net/getabstract.php?paperid=SR201212164722, DOI: https://www.doi.org/10.21275/SR201212164722

[2] Data Engineering Architectures for Real-Time Quality Monitoring in Paint Production Lines. (2020). International Journal of Engineering and Computer Science, 9(12), 25289-25303. https://doi.org/10.18535/ijecs.v9i12.4587

[3] Vamsee Pamisetty. (2020). Optimizing Tax Compliance and Fraud Prevention through Intelligent Systems: The Role of Technology in Public Finance Innovation. International Journal on Recent and Innovation Trends in Computing and Communication, 8(12), 111–127. Retrieved from https://ijritcc.org/index.php/ijritcc/article/view/11582

[4] Mandala, V. (2019). Integrating AWS IoT and Kafka for Real-Time Engine Failure Prediction in Commercial Vehicles Using Machine Learning Techniques. International Journal of Science and Research (IJSR), 8(12), 2046-2050.

[5] Ghahramani, M., Qiao, Y., Zhou, M., O'Hagan, A., & Sweeney, J. (2020). AI-based modeling and data-driven evaluation for smart manufacturing processes. IEEE/CAA Journal of Automatica Sinica, 7(4), 1026–1037. https://doi.org/10.1109/JAS.2020.1003114