Deep Science Publishing
https://doi.org/10.70593/978-93-49910-78-2_2

DeepScience
Open Access Books

Chapter 2: Implementing artificial intelligence for real-time fraud detection and risk mitigation

2.1. Introduction

Organizations implementing and considering artificial intelligence (AI) for real-time fraud detection see both practical benefits and challenges. On the upside, AI can significantly speed up the process; modern behavior-based AI algorithms do not rely on past fraud patterns to help detect high risk transactions, as does rule-based and machine learning technology; and are more effective at reducing the volume of false-positive transactions. However, AI implementation and use is complex. Organizations must manage the large data sets and ever-changing features required to train the AI algorithms, constantly tune model parameters, and set thresholds; and understand that AI applications have some inherent limitations; real-time AI for fraud detection is a hot area of research and many approaches used today are yet to be deployed in real-time scenarios at scale in the wild. Over the past decade, AI has profoundly transformed how several high-stake activities are performed across a wide range of sectors, with many of the responsible organizations actually relying on AI for decision automation (Liu & Li, 2022; Wang, 2023; Hossain et al., 2025). The fields of banking and finance are some of those most heavily dependent on algorithmic decision-making; in these areas, harmful, financial gain-seeking acts such as money laundering, payment fraud, credit card and loan fraud, securities fraud, and anti-money laundering can lead to severe consequences for both business organizations and society as a whole. As such, fraud detection and risk mitigation are critical tasks for the financial industry. The importance of these tasks is matched only by the challenges that insiders face when tending to them. In the age of Big Data, financial organizations are overwhelmed by an endless flow of data from disparate sources that underlie every aspect of their business processes. Detecting fraudulent events in such huge amounts of data, in real time, is a challenging task,

particularly because fraudsters constantly adapt their actions to sidestep detection (Yazici, 2020; Yamini et al., 2023).

2.2. Understanding Fraud in Financial Systems

Fraud is pervasive in business, primarily by six different categories: computer fraud, employee, consumer, check fraud, financial report fraud, and trade secret fraud. Computer fraud is using a computer and/or a computer network to commit illegal activities. The Internet has been instrumental in creating a whole new venue for illegal activities, including child pornography, online gambling, identity theft, pirated CDs and software, online scams, and so on. Although some degree of disgruntlement is probably necessary in order to commit fraud, there are multiple factors associated with computer fraud. These include poor computer security, overexpansion of the organization, inadequate education of users on the potential threats associated with using the computer, and unauthorized access to computers by employees (especially formerly trusted employees).

Fraudulent acts committed by employees may involve the theft of an asset belonging to the employer, the theft of an asset in the employer's possession but not ownership, or a financial misrepresentation. Within these three categories fall a multitude of different unpaid leave, travelers' expenses, excessive theft of office supplies, payment to fictitious vendors, kickbacks, etc. Consumer fraud covers a wide variety of deceits perpetrated by consumers on businesses, and involve such acts as check-kiting, purchase order invoices, defaulted charge accounts, counterfeit credit cards, and identity theft, to name a few. Check fraud involves writing bad checks on assets when the writer had insufficient deposits in his/her/their account to cover the check. Frequently, stop/check orders are filed after the check is issued but prior to the check clearing the bank. Check fraud also embraces counterfeit and altered checks, and bad business checks written to an establishment by someone who would not ordinarily do business with the establishment. Banks estimate the losses from check fraud to be in the range of $3 billion each year.

2.2.1. Types of Fraud

More recently, we have witnessed a significant surge in the variety of technology-enabled frauds and how structures previously reserved for rogue states have now been weaponized to disrupt legitimate businesses and their customers. Although some methods remain popular, new varieties of old scams are emerging, and the new landscape is offering malicious actors fertile ground for innovative fraud techniques. We present an overview of the most common types of financial fraud, enabling us to prepare advanced analytics to detect and mitigate real-time fraud.

Credit and debit card fraud is not new. But recently, with the rollout of new EMV-compliant credit and debit cards in the U.S., the landscape has shifted considerably. EMV is a chip-based, secure payment authentication model designed to eliminate card-not-present fraud and card-dominated counter/service fraud. The shift to EMV has turned the U.S. into a much more attractive target for CNP fraud, which is now taking place at an unprecedented pace. In tandem, with the rise of total CNP fraud loss levels nearing $33 billion by 2023 and double-digit percentage year-over-year increases, we are witnessing a new and escalating identity theft service that focuses on retraining drives behind account takeover fraud techniques. ATO fraud refers to malicious actors acquiring enough intermediate financial credentials from externalized credential dumps to enable them to successfully access financial accounts, including credit cards, bank accounts, and credit bureaus.

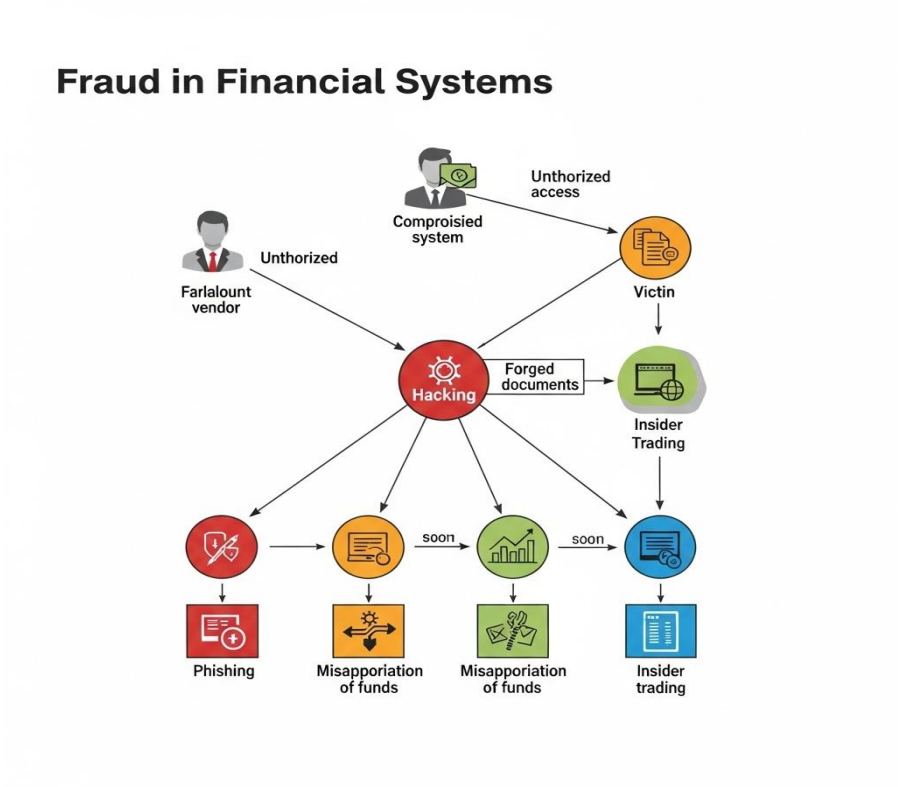

Fig 1: Concise diagram illustrating common methods of fraud within financial systems.

Routine credential stealing has now been boosted through a concentration of age old weak user and password schemes applied by many target companies who are forced to relax their onerous password rules because customers constantly forget their passwords and demand password resets. Hence restoring user password access now represents a

costly cycle of degradation for many target companies. The result of ATO fraud is now manifesting in the huge spikes in accounts linked to loans, credit, identity, and synthetic fraud.

2.2.2. Impact of Fraud on Businesses

Fraud is a major risk factor for organizations, particularly for financial services firms. In particular payment services, such as e-wallets, which offer digital payment systems, are often appealing targets for fraudsters. As these services grow in popularity and usage, criminal organizations are moving into this domain, forging identity documents, especially in digital identity verification, laundering money through the associated digital bank accounts and transferring it to physical currency. It is vital that firms have in place not only the right AI-based tools to detect these types of activities but also the trained specialists to interpret the information generated by the tools. Failure to detect such activity can expose a financial services firm to sanctions and fines by government regulators, as well as resulting in significant financial losses. When customers engage in highly suspicious behaviors, it could cause the concerned financial services firm to even terminate its business relationship with that customer. Such relationships are usually strong and mutually beneficial. Loss of such key client relationship could more than offset the revenue earned for enabling the suspicious transaction detected.

Various mitigative and detection methods for revealing and disclosing fraudulent transactions have been advanced and used. Financial service firms usually recognize fraud losses in the accounting period during which the loss is incurred, which is usually at the same time the transaction is approved, much earlier than the losses may be realized. However, estimating when realized losses occur in the case of fraudulent transactions involving a forged credit card is a much more problematic issue. Predicting the time and size of these exposures could help financial institutions better manage their overall liquidity risks.

2.3. Artificial Intelligence Overview

Artificial Intelligence is a deceptively simple term that includes numerous complex algorithms, techniques, and areas of research. Most broadly, it refers to the notion of building computer systems that are capable of performing tasks that would otherwise require human intelligence. This definition is broad enough to be applicable to areas such as expert systems development, natural language processing, and more, but also leaves questions unanswered regarding how one actually builds systems capable of AI. In more specific terms, Artificial Intelligence is an area of research focused mainly on the development of software systems capable of performing a particular set of functions with

minimal human assistance or intervention. The most common areas of task completion involve data analysis and interpretation, pattern recognition, and prediction capabilities.

What distinguishes such systems is the reliance on data-driven models that are able to learn over time. In this sense, AI models learn from the knowledge and experiences gained from the feedback on their previous actions in the domain and apply this knowledge to improve their future actions. The most rigorous and popular subset of Artificial Intelligence is machine learning, a discipline that draws from fields such as statistics, optimization, and others. The goal of machine learning is to use data to uncover patterns that govern the input-output mappings of a domain. These learned mappings are then used to make decisions regarding previously unseen data. Unlike predictive models based on regression analysis, for example, which rely on human specified functions to build the predictive model, the goal of machine learning is to automate this process, learning the functions underlying the input-output mappings themselves. What distinguishes machine learning from traditional methods of statistical modeling is, therefore, the fact that the models used are data-driven predictions that build a variety of functions, designed specifically for prediction accuracy rather than for any inherently statistical merit.

2.3.1. Definition and Key Concepts

Artificial intelligence has been a field in existence for decades, nevertheless the rapid ascent of its application to businesses and to our everyday lives has ignited a new flurry of developments, algorithms and advancements in many facets. Some define it simply as the replication of human intelligence, in other words the ability of a machine to perform any task that an intelligent being may perform. But persons can be said to be intelligent in many different facets, and in fact humans ensure that they do not rely predominantly on an artificial entity to perform any of the myriad of functions that humans possess. Traditionally, the term artificial intelligence was identified with tasks that required human-like characteristics or attributes of problem solving capability such as creativity and wisdom, or functions that were seen as highly complex. Despite differing definitions of intelligence among the general world population, AI has now encompassed not solely the replication of reasoning but now also visual processing. In this context, we adopt the following definition: Artificial Intelligence (AI) is the design and development of intelligent agents, which are computer programs that act intelligently. That is, they perceive their environment and take actions that maximize their chances of success. A more colloquial and common definition describes it as the science of making computers do things that require intelligence when done by humans.

2.3.2. Machine Learning vs. Traditional Methods

Real-time fraud detection systems are traditionally built using a combination of static business rules, pre-defined transactional thresholds, and expert rules that are hard-coded into the detection application. For example, a business user may want to block transactions that are over a certain dollar amount if they originate in a specified country. If a user executes more than 3 logins or failed password attempts within 30 minutes, flag the user or block user for 30 minutes. Machine learning adds a serious weapon to the fraud-detection arsenal by adding dynamic detection capabilities. Machine learning uses user or entity behavioral profiling for a broader view of transactions and not just the current transaction.

Machine learning techniques offer the following advantages over traditional systems. First, machine learning techniques can process very large volumes of data and are capable of very high throughput rates. As a result, they can be deployed in real time to make immediate approval/reject decisions on entering transactions. Traditional code-based systems may make those decisions, but they take longer, which can lead to frustrated customers. Second, ML techniques can analyze very complex relationships among features and user behavior and transaction data. Traditional scoring systems are limited by external scoring-determinant correlations that can only be updated and modified through frequent interactions by a business user with the IT team that built the detection system. Machine learning techniques, on the other hand, analyze internal relationships and correlations and update without outside intervention. Some may argue that these internal correlations are not always trustworthy, which is partly true. However, with the right training data, noise filtering of false fraud or false not-fraud transactions, and tuning, ML techniques can learn useful patterns from internal correlations. Third, ML techniques can adapt to changes in user behavioral profiles and quickly re-train on new data.

2.4. AI Techniques for Fraud Detection

Fraud detection is a critical component of risk management and mitigation in business. The available approaches for fraud detection fall into two categories: data-driven and process-driven techniques. Traditionally, the data-driven approaches use statistics and machine learning algorithms to detect anomalies or perform predictive modeling. The algorithms typically used are naive Bayes classification, logistic regression, decision tree, random forest, support vector machine, artificial neural network, etc. The prediction or anomaly model will learn from collected data, where the outputs are fraud or not fraud. The data used can be historical data or could be transactional data that are collected in real-time.

Approaches toward fraud detection have continuously evolved through a series of advancements over the years. A major evolution emerged with the increased computational power and availability of large data using new techniques such as big data analytics, real-time predictive analytics, and machine learning that use larger and more complex data sources. However, this new evolution still relies primarily on the traditional data-driven data and predictive-driven techniques. Fortunately, the arrival of artificial intelligence, energized through deep learning capability, has provided another new evolution in how we can detect and prevent fraud. Coupled with the availability of massive data, deep learning has been demonstrated to outperform existing machine learning techniques for visual and speech recognition and even for NLP tasks over a variety of text and language datasets.

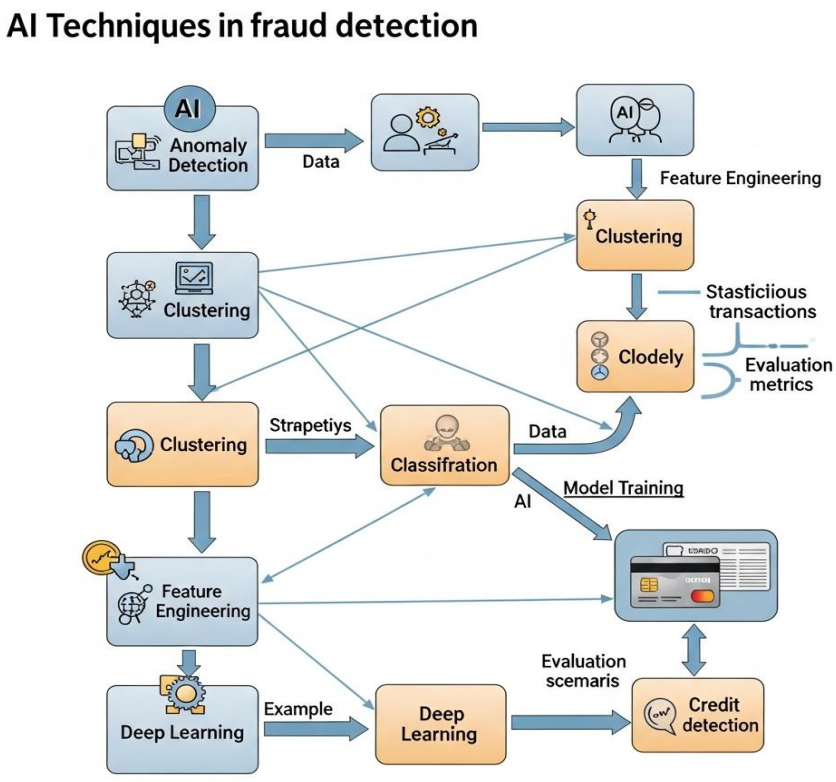

Fig 2: Diagram illustrating various AI techniques employed in fraud detection.

Deep learning techniques, operating on the top of the neural network, have been shown to be effective in handling large, complicated data and have excelled at supervised and unsupervised tasks in computer vision, speech recognition, and even natural language processing. It is only natural, then, for this technique, alongside AI-based NLP, to be

used to solve the problem. In fraud detection and risk mitigation activities, the use of AI-based high-tech analytics has two dimensions. The first one is anomaly detection and predictive analytics directly applying AI techniques to provide an accurate, smart input/output function model to identify fraudulent transactions based on the information provided. The second is AI-based intelligence assistant tools performed by the AI to assist human investigators and help the human intelligence discover hidden patterns or deep analysis explorations to point out dubious activities for further investigation.

2.4.1. Anomaly Detection

Anomaly detection in its widespread sense, is a concept that refers to a scenario where a newfound object, among a given set of data points, appears to behave differently from those of the rest of the data points. In other words, detecting the outliers or novelties is the primary goal of such a process, which is common in various application fields. Fraud and intrusion detection, fault detection, monitoring environmental disturbances, sensor network security, and detecting ecosystem disruptions in general are only some examples of anomaly detection applications.

While supervised anomaly detection can be approached using discriminative models, majority of proposals fall in the category of unsupervised anomaly detection algorithms. Under this easy-to-use scheme the set of data is unlabelled, thus no prior knowledge about what constitutes anomalies is needed, which is a desired characteristic in many real-world applications. In these latter cases, the data would mainly consist of normal instances, and anomalies would be rare. This attribute is typical for most of the relevant applications, from fraud detection to network intrusion and fault detection, because underlying data-generating distributions are unmodelled, times of normal behavior in the systems are usually far outweighed by times of unusual behavior, and were anomalies to be easily labelled, detection would not be hard. Negative and positive sets of labelled occurrences are however also available for some applications, though can be often misrepresented.

2.4.2. Predictive Analytics

Typically employed in relationship-intensive business areas such as retail banking, credit cards, insurance, or fraudulent acquisitions, predictive analytics can assist in determining the propensity for particular types of fraud. Predictive models are built for particular types of fraud or even models forecasting several types. So-called multievent models, which predict propensity for several claims simultaneously, are becoming more common. A small amount of data is usually available to develop the model in these particular types of areas. Small sample sizes impose severe data challenges that the

modeler needs to resolve; until very recently, the data obstacles in these situations would have precluded anything other than traditional predictive techniques — logistic regression using a judgment-sampled control or mirroring the data with the claims likely overrepresented — but a few groundbreaking advancements have made it possible to circumvent these data constraints and take advantage of the innovative advances in predictive technology pioneered by other fields. Fraud detection predictive models can have a striking impact. Since they target only a particular part of the claims universe — fraudulent claims — the return-on-investment ratios can be astronomical; in fact, just one identified fraud investigation can lead to recoveries that are 40 or 50 times the resource investment.

Much of the initial work in fraud detection was generated in North America 15 or so years ago by a small band of people. In fact, some of the initial system development was not done in insurance but in bank and credit card processing; most of that effort has been, understandably, proprietary in nature. Despite the maturity of the underlying predictive modeling technology, this pioneering work led to some of the first prototype commercial applications using advanced predictive detection. And although some more advanced predictive techniques began to penetrate the estimate long ago, there was a decision to stay with traditional technology until enough experience and good results had been published with more advanced techniques to justify developing business or models using these more advanced methods.

2.4.3. Natural Language Processing

NLP is often used to process online reviews in content customization, product comparison, fake review detection, and opinion mining, learning sentiment by detecting an issue and sentiment polarity associated with the issue. Fraud detection applications of NLP technology often use rule-based approaches to detect fraud features such as the text content of a message. Recently, NLP has made remarkable progress through the use of large-scale pre-trained transformer-based language models, which have become the main building block in many NLP applications and powered significant advances in a large number of sub-tasks in NLP. How can we take advantage of these pre-trained language models to tackle text-based fraud detection tasks? NLP applications often involve training a model with a relatively small number of labeled samples of specific downstream tasks. While previous deep learning methods may have learned a language model primarily built from language modeling; say how to create a model for the fraud detection task. One of the unique aspects of this fraud detection task lies in the fact that many existing preference manipulation strategies become the training sample, such as fake review detection, click fraud detection, fake news detection, and bot detection.

The key observation is that such a fraud detection task may be more appropriately framed as a "fringe case" detection problem spelled out a variant of rare-class detection. Such cases may comprise the long tail of a heavy-tailed distribution because only a small fraction of samples belonging to specific fringe classes violate the fraud rules. Such a data distribution is also commonly observed in many of the aforementioned fraud tasks as selecting samples that do not satisfy and the goal of the fraud detection task is to classify and remove fringe cases. Various strategies may be implemented to formulate the fraud detection tasks. In general, various downstream text-based fraud detection tasks, such as fake review detection, click fraud detection, and bot detection, need to collect the appropriate training set, annotation schema, and classifiers.

2.5. Data Requirements for AI Models

For the development of supervised machine learning models, labelled data sets are required where the observed risk events are labelled accordingly. Knowledge-based detection will likely require data from known fraud cases, as expert opinion suggests that fraud detection using value-based techniques is more of an inference process than a learning process. A few known case studies include the detection of fraudulent modification of diesel engine control software, the inspection of low cash flow of publicly listed corporations by their competitors, and the natural language analysis of filings of giant companies to know their risk propensity and the risk of misrepresentation due to manipulation, etc.

The real-time machine learning detection models can only be developed and successfully implemented when there are enough labelled data points true-positive as well as true-negative cases pertaining to the chosen time frequency like by minute, by hour, by days, etc. The magnitude of labelled data would be much higher for the latter time frequency compared to the former, given the data availability, and thus would need the least number of model iterations to build the foolproof models ready for retail operations. An action has to be taken on the machine learning predictions at least every now and then to avoid creating the orphan predictions, and to ensure that the model feedback loop remains functional.

Fraud detection datasets are confidential and, therefore, not publicly available, as compared to the more generic datasets like the data banks for the subfield of natural calamity detection. The publicly available datasets include the credit card fraud detection datasets from the banking sector, and the payment fraud detection datasets from the merchant store transactions. The transaction times are recorded, and care should be taken to manipulate the date and time transaction attributes properly if they need to be used in building the machine learning models, as misleading model predictions could occur due to wrong time zone considerations for the transactions concerned – for either zeroing out

the time zone attribute, or else using the time zone information as another predictor variable.

2.5.1. Data Collection and Sources

Target data for AI model development can come from multiple sources according to the requirements of the project and scope of the analysis. For fraud detection in credit, insurance, application or transaction data from documents, databases, lending system architecture, policies, and insurance guidelines would be needed. The size of the dataset also has a significant impact on the success of the model. More data generally implies better representation of various product factors, tenant controls, market conditions, underwriting and collections policies, product design, disbursement methods, the ability to contain different types of NBFCs, banking products, vendor credibility, market placed documents, input errors, product validation, and loan execution. In our experience, for models targeted towards predicting an event in a timeframe of 12-24 months, a minimum dataset containing at least ten times more than the number of positively labelled cases is a good starting point. For most commercial or industrial loans, that would imply a dataset containing minimally 100K-150K data for development, plus another 50K-100K for validation.

Data must also be consequently pulled in from multiple sources to add new bank statement metadata concepts as features which change over time like rent types or amounts, tax filing purposes, banking patterns over an applicant's lifestyle change, utilization behaviour pre- and post-disbursement, basic needs such as income and expense corroboration, availability of data in support of loan or product requirements, and seasonality which has an impact on repayments. Other sources are assessments of financial statements - profit-loss and balance sheet, and bank data changes, along with Public Credit Registry and other bureau reports to check for Tier II spending, among others. Validation of data from possible third-party tools must also be considered for social media checks and source verification, and for clean document uploads, data should be validated using relevant vendors and agencies.

2.5.2. Data Quality and Preprocessing

In real-world AI implementations, data may not always fulfill the requirements for machine learning. In many cases, raw data, especially if it comes from third parties, needs to go through various preprocessing steps before it can be used for exploratory data analysis, feature engineering, and training of the model. In AI – as in every other activity – you are only as good as your inputs. If your data is of low quality, your results will be equally poor at best.

However, low-quality data is not the whole story. Data cleaning is tedious work, especially when working on large amounts of uncurated data. You may end up spending sixty percent of your time cleaning data in a project, and institutions may go to great lengths to deploy tools and processes to ease the process of data quality monitoring. Institutions are often overwhelmed by the sheer number of manual data preparation tasks required to make data ready for high-quality analysis and subsequent machine learning pipelines. Common tasks include correcting data formats across many attributes, deduplication, trimming, parsing, data type estimates across all records, transforming records with incorrect lookup values, and joining records with any underlying domain data definitions that may have little semantic meaning in their original representation. Further, data is seldom in a single repository. So you may have to also deal with cross-dataset preparation tasks like finding candidate matches in disparate datasets and normalizing similar but different data records.

We have focused on tackling these pain points early on and seek to make data quality a transparent function of data collection. Our data validation framework needed to provide accurate feedback at all times during the data collection project, and facilitate speedy fixes to incorrect records across large populations. We needed our decisions to be rooted in data quality, as good-quality data leads to good insights. Underlying our desire for better data quality was a desire for data quality estimation and indexing that would be cheap enough to act upon feedback from the indices on every new record.

2.6. Model Development and Training

While this work is focused on deep learning for its superior performance in consecutive predictions for fraud data, it is worth mentioning that any machine learning model could theoretically be used, and one may also appreciate performing the model selection based on training results of multiple algorithms at first, followed by explanatory modeling to pick the best candidate, as was the common practice in the field. The candidate algorithms include logistic regression, ensemble models like random forests but particularly gradient-boosted trees, and simpler models such as support vector machine, neural network, or deep learning, whose performance may be particularly good if the scale of training data is sufficiently large. Predictably, simple models like logistic regression would be quicker to train and evaluate during experimentation, benefit from parameter tuning unless over-trained, tend to perform well in generalization, and be easier to interpret through deductive rule generation.

At the other end of the spectrum, a gradual release of work on deep learning and proper network setup in terms of design, costs, hardware, and optimization, possibly including additional randomized pretraining phase, has led to slower yet higher performance in practice, with suitable combination of multi-task learning to allow sharing of information

among tasks, or better transfer learning through fine-tuning individual networks for related via meta-learning, and some successful self-supervised learning for generalization. These suggestions have been used around work on other applicability like computer vision and natural language for their particular properties, and to various degrees related to available resources and inherent characteristics in overall shared tasks regarding tasks including stock price prediction.

2.6.1. Algorithm Selection

Due to the associated costs with fraud, including supplementary R&D costs, product shipping costs, and the loss of customers, e-commerce businesses have a major interest in minimizing the emergence of fraud. Thus, e-commerce businesses require efficient fraud detection solutions. By 'efficient', we refer to exactness and speed. The latter is paramount, as companies want the resolution of a fraud detection process to happen in real time. It does not suffice to accurately detect frauds in an undeterministic way. Thus, the model must also classify valid and fraudulent transactions while being adaptive within a short time period. In general, we can distinguish two different types of classification algorithms: deterministic models and ensemble models. The former use a probabilistic approach, while the latter combine different models to enhance both classification performance and minimize the time taken to identify a fraudulent transaction.

Along this path starts the model development and training process within the framework of the proposed real-time fraud detection and risk mitigation prediction insight tool. Regarding the selection of suitable algorithms, one could argue that the mainly used models would be suitable. However, the question is how such commonly used classification models could overcome the challenge of real-time fraud detection and risk mitigation prediction.

2.6.2. Training and Validation

Training is supervised learning aimed at minimizing the model's loss function. The training data consists of samples from either training, validation, or test data. The model's loss function is its training objective, which means that the model is trained by modifying its parameters to minimize the expected value of the loss function. The model makes a many-to-one mapping from a prediction space to labels, such as clustering data or predicting a class probability distribution. The training process generally consists of the following steps. The model adds noise to the input data or sample from the prediction space according to the sampling, prior, or conditional to create noisy versions of the data, such as automatically augmenting augmented data using dropout added and noise. The

model calculates the loss function for a batch of training samples without noise. The computation performs backpropagation and optimizes the training parameters or model, or optimizer algorithm for a mini-batch size of data.

Training strategies often require complex methods to scale machine-learning models to large datasets. The basic form of generalization in predictive modeling is to train a model using data from a training space and then evaluate it using validation data from the validation space for metrics. Storage space and time constraints for large datasets often make it impossible to apply all the techniques, such as scaling the models for large datasets, avoiding overfitting, and balancing the training, validation, and test space. These issues are even more important when fraud detection is divided into many tasks. The model may need solutions that act like data selectors in order to be trained by the training space. If all the sample training tasks were optimized together, the training space would become quite complex, which could make model training infeasible for a real-time stream of data.

2.7. Real-Time Processing and Implementation

The second main goal of this chapter is to present the real-time processing and implementation details necessary to realize the continuous distribution modeling and anomaly detection methods previously discussed. The system was originally developed, deployed, and validated as part of the biometric data collection from video authors. The original focus which comes in part as a result of the focus on social media videos allowed for the emphasis of real-time processing as it directly affects users submitting applications for video monetization as they are relying on a platform which promises payment for ad revenue-sharing at the time of video upload after their application for partner program is reviewed and approved. With the addition and intertwining of new emerging attack vectors which are predominantly exploitative as well as the introduction of additional data sources, the emphasis on near real-time processing remains, especially as additional designated resources to improve current implementations of fraud detection is a topic of continuous discussion within the tech industry and negatively affects the economic balance between platform and users.

A high-level architecture that implements the real-time processing element for anomaly detection is depicted below. The key processing component of the architecture is the anomaly detection server which continuously listens for new content to evaluate. The detection server requires continuous data sources of creative acts and external signals of legitimacy or news stories covering the hands that redistribute exploitable false information which would likely attempt to monetize, during early stages of a request for verification of monetization permissions and during the review and approval phase to

function effectively. For additional context on the data sources, a summary of the different data sources and configurations possible is provided.

2.7.1. System Architecture

The availability of major cloud providers to supply Lightning application services is transforming payment systems in financial service and e-commerce dynamic environments. In this context, real-time payment systems are assisted on instantaneous money transfers and receipts, promoting the emergence of new kinds of businesses. Accordingly, in this environment, people transfer different amounts of money to each other directly, including climates of trust and reputation, and there is no need for financial intermediaries, but they do not recognize the use of the payment system susceptible to being designed and implemented in a fraud-proof way. Fraud can occur, and it should be promptly detected and adequate penalties and sanctions impeached on users after the commitment. Transaction details can hardly allow a fraud detection model to categorize a transaction as normal or abnormal due to the lack of a previous reference.

The innovative proposal of this work is the implementation of real-time machine learning classifiers with risk mitigation techniques. This is performed through a system architecture capable of distinguishing dishonest transactions that are outside the current pattern of user behavior established with a known set of previous transactions. The pre and post-processing steps of the modeling procedure developed on unsupervised and supervised machine learning methods utilize the users' mobility data related to digital wallets. This architecture information offers a flexible solution that can be executed directly on Lightning payment systems and is applicable on distinct customers, minimizing costs and effort, through an off-line initial training phase. With the solution developed, and consequently, continuous learning systems. Although our proposed architecture has easier integration with Lightning compliant wallets, it can also be adapted to act as an intermediary between the Lightning network and other wallets, which do not comply to its protocol.

2.7.2. Integration with Existing Systems

Integration of a real-time AI-based fraud detection system with existing verification systems is one of the most critical issues for deploying such a solution. Often, older systems simply do not have the architecture or capabilities needed to handle external signals in real-time. After creating a backend machine that can do the required pattern scoring at appropriate speed, we needed to create a middleware service that could provide a minimal API for our real-time scoring services and receive thousands of calls each second. Performance testing was required with industry-leading solutions for low

latency middleware solutions from various vendors to choose the best one based on engineering effort, cost, and required performance. In addition, we also had to work very closely with several fraud risk business stakeholders in multiple parts of the company and get sign-off from the executives before deciding on the deployment strategy. Different countries had widely different requirements, and any changes driven by the AI system feedback on real-time verification thresholds had to be approved and signed off in advance.

In addition, there were also challenges with implementing the AI solution as additive to existing verification solutions in the backend. Aligning different databases and signals was critical to ensure the success of any adjustments made based on fraud and risk models. As a part of the solution, we decided to implement A/B testing of changes in decision thresholds after successful validation of model maturity and performance at the customer level. Once these thresholds are validated over different timeframes and transaction volumes, we will then be able to implement a solution where the real-time AI system directly makes tweaks to the verification business rule logic, creating a continuous feedback loop. This would enable the potential for a self-learning real-time feedback system that could help minimize risk and fraud without directly impacting the customer experience.

2.8. Risk Mitigation Strategies

The primary goal of risk assessment is the proposal of mitigation actions to reduce the likelihood of a significant event happening and the severity of the associated symptoms. Risk reduction is sometimes difficult to achieve, especially if it requires drastic changes on the company processes. For instance, avoid permitting more than a minimal amount of remote work. Remote environments are more difficult to monitor, making employees much less accountable for being honest, but forcing employees to commute every day may leave the company vulnerable to many additional physical security threats with unforeseen indirect costs. Alternatively, it would be less difficult to require that employees let certain tools be installed on their devices to monitor OS-level events, enabling the system to detect anomalies, such as a user account opening and executing a sensitive file at odd working hours on a holiday for a few moments, a sensitive file that he/she has not accessed in months or over a prolonged period seemingly downloading with no business need, a large number of sensitive files, which is unusual, or accessing those files using an unusual channel.

Reporting any strange behavior that they are observing from the system, trying to convince their executives of the risk and sensitivity of the area, and documenting any interactions they are having with suspicious customers or employees could help to decrease the likelihood of incidents occurring. A robust Incident Response and Action

Plan must be designed by the information security department, containing the detailed steps that those partnering units would take in case of a stimulation incident.

2.8.1. Identifying Vulnerabilities

Mapping out every point in a system where fraudsters can place their order within a proposed procedure is vital to understanding risk within that channel. Fraudulent actors dedicate considerable amounts of their resources to expand their reach and circumvent detection models. Identifying model weaknesses makes them extremely vulnerable and a lucrative target for threat actors. DoS, DDoS, and other amplifications cause destruction of otherwise mighty fraud tools. Alone these attacks deny the businesses from legitimate transaction processing and escalate client and partner anger, however, during a holiday season when transaction-processing throughput is maximal and the company makes their yearly profits, it becomes unforgivable and leaves a scar from which recovering will unlikely happen. What is even worse in terms of risk is that single-layered tasks are fragmentable, and bad actors can pick them one by one and create appropriate task scenarios in their control to fit their budgets. They can rent cheap but ineffective botnets to perform those attacks, one at a time in a distributed fashion. They lead to a lost connection on the business side and immediate loss of profits, yet during peak seasons, device fingerprints are refreshed and create a massive probability of false negative alerts.

Vulnerability mappings should understand which kind of accesses help create fraud force multipliers that are usually exploited: easy-to-guess variables, the incidence of arbitrary malware and other unrequested nurture, easy-to-spot open redirect rules, ties to other fraudulent zip code or geolocation locations, or referrers leading to clickaggregated, non-reconstructed sites. They should also identify specific layers or payment functionalities that are easiest to manipulate avoiding alerts or response activities. Bots, script injections, or recours-escalation into holistic task layers and just superficial combinations are the most effective. Only by understanding those access weaknesses for each active procedure, their activity frequency and drive, tactical response tools can be prioritized. Setting rewards and enabling disincentives at the appropriate entry points becomes possible for the various tasks so that fraud’s operating margins become thin enough as to turn the game unprofitable.

2.8.2. Building Response Protocols

A single approach is hardly ever sufficient to neutralize risk exposure. A natural extension, therefore, is creating different incident response protocols for different levels of fraud severity, modeled on the examples from other risk areas of the organization. In

this case, the business ramifications of a detected fraudulent behavior are an important parameter. For example, while a manual review of a suspected application fraud incident is often entirely justified for small loan amounts, it could be entirely inappropriate to do the same for a corporate loan. This is compounded by the fact that different types of transactions have different review and fraud recovery costs associated with them. For small auto loans, the cost of a dispute may be greater than the outstanding loan amount, while the costs of reviewing a multi-billion-dollar wire transfer are much greater.

Equally important is that real-time fraud detection is only one way of combating fraud. The second, and perhaps more potent way, is to act on the knowledge gained from having such a system in place. What is the point of creating a business advantage based on being able to detect fraudulent activity? It is wholly incomprehensible to then allow the criminals to take over and thrive in your organization so as to make your systems available for their activities. If you can detect card-not-present fraud while it is occurring, why not destroy the fraud ring by monitoring certain known trouble spots for suspicious activity? If you can detect application fraud in real-time, why not share information with the police regarding such applicants? If a bank holds an account in which unusual, and perhaps fraudulent, activity is occurring, why not encourage them through communication with the relevant authorities to conduct a little on-site surveillance?

2.9. Challenges in AI Implementation

The adoption of AI solutions for fraud detection can be a complex process. Financial institutions have to consider organizational readiness, talent shortage, infrastructure readiness, data policy, and so on before rolling out AI solutions for fraud detection. Implementation of AI in the complete operations of the organization goes through three to four cycles before complete efficiency in terms of cost and effort can be achieved. Since machine learning concept is based on the laws of statistical probability, false positives in the beginning stages of development are usually more in number. Budgetary controls need to be assigned to each deployment cycle depending on its impact.

With AI implementation, organizations have to deal with a new set of challenges. Companies have to deal with user data protection concerns while designing and using AI models. Apart from the ethical concerns in the usage of data, there are also regulatory requirements regarding data usage. Companies need to comply with the norms set up as a framework for AI developments or AI product utilization. Bias in AI models has a significant impact on the social judgment as well as on data-driven decisions. It leads to the reinforcement of social stereotypes and unintentionally creates inequalities. Models need to be designed in such a way that human bias during the design process gets neutralized.

With increased usage of third-party syndicates for script development and deployment, the issue of IP theft is increasing. Financial institutions need to develop in-house capabilities to avoid such challenges in intellectual transactions. Production systems need to be designed for continuous monitoring of model performance not only with the business algorithms but also for business matrices. AI models, if not monitored continuously, might lead to decision avoidance by the risk teams and other business stakeholders rather than usage of AI outputs and results for final decision making for fraud investigation and risk management.

2.9.1. Data Privacy Concerns

Data privacy protection is crucial for all organizations, but for finance organizations it is especially the case because of the amount of sensitive data they process and the importance of privacy in the finance sector. Financial fraud involves stealing someone else's money. The ways of discovering and preventing fraud in finance will basically involve accessing a person's most sensitive information, financial history, shopping journey, and other sensitive details. External parties audit these details in the name of financial security. While the finance organization will hold these details for security, the fact is that these details are being exposed by external parties.

In the finance sector, AI models are built on the basis of data related to customers' financial lives. Finance organizations monitor and analyze customers' financial transactions. Detecting fraud through AI gives rise to various concerns regarding data privacy. Organizations must ensure the safety of their data, and building a data-driven AI model for real-time fraud detection means that sensitive data is in the hands of a machine. Today, organizations outsource these operations by using a cloud platform involving multiple external companies. Sensitive data leave the organization and processes undergo multiple manual steps in order to render it usable. Cloud systems allow auditors to access this sensitive information and actions need to be taken just by using third-party systems, making the organization reluctant to support such a process. Organizations need to prioritize data privacy issues when deploying AI models to detect and prevent email fraud. It is also essential to respect the privacy of platform users while preventing the fraudulent activities of these users. This is truly a very complex balance to establish. What is most noteworthy about these processes in such a financially and socially relevant sector is how complex those developments are in an organizational space as tasked with password safety and prevention of illicit activities.

2.9.2. Regulatory Compliance

One of the key elements mitigating the option of using such AI technology within financial services is achieving regulatory sign-off. To be accepted within such risky environments, AI models require a strong level of explainability. In the past, adaptive methods have shown strong capabilities in a wide range of disciplines, from classifying image data to winning board game competitions against the best on the planet. However, their deterministic cousins have continued to dominate the regulated environments. This bias towards deterministic models is also shown by the fact some of the most elite current data scientist teams are using simple linear modellers (in their logistic form). The main reason for working with such models lies not in performance but their regulatory acceptance, extremely simple, completely linear.

In an environment where the cost of real-world errors is enormous (both in terms of loss to the company for false alerts and the customer experience), discriminatory processes threaten careers, and corporate failure threatens a multibillion-dollar industry, it is perhaps unsurprising that by far the largest proportion of large-scale AI/ML projects are focusing on internally facing problems. Such problems are generally lower in stakes as there is no end client; they would be working with internal data, often earlier data than would be internally available to face the regulatory hurdles typical of deploying externally facing AI projects within the financial vertical. Such activities permit the business to experiment with available technology to improve processes and reduce costs in a safe environment.

2.9.3. Bias in AI Models

Right from its inception, the focus of AI has always been on depicting human-like performance in problem-solving, and more recently, in machine-decision processes. Straddling a nexus of psychology, computer science, linguistics, neuroscience, and philosophy, AI is predicated on a sound understanding of reasoning, action, perception, communication, and learning in humans. However, over the last few decades, modelling complex human functions has become increasingly difficult, particularly due to the implicit assumptions made in modelling them. The success of current AI programs in computer vision, speech recognition, machine translation, and game play has diverted interest away from the challenge of creating human-like systems in their true sense. Contemporary AI models, often referred to as machine learning models, differ from conventional AI models by their ability to learn the relationships among variables in a problem domain directly from real-world data. The prediction performance of machine learning models often reaches — and sometimes surpasses — that of human experts when suitably trained on large amounts of high-quality data, in tasks such as prediction

problems in public policy, healthcare, physics, biology, economics, marketing, and finance.

In contrast to the cognitive understanding espoused by traditional AI models, machine learning models often function as ”black boxes,” rendering explanations for their predictions difficult and often impossible. AI/machine learning models used in sensitive domains, such as healthcare, the judiciary, finance, insurance, and the military, cannot be allowed a prediction performance monopoly, specifically because the modeling assumptions and intuition behind how these models learn to map predictor variables to response variables can greatly influence the predictions. For example, it is expected that AI models making crucial predictions for diagnosing tumorous growth employ the same facets that qualify such growth as tumorous for deriving probabilities rather than unrelated parameters that have little relationship with the actual data.

2.10. Case Studies

As the adoption of artificial intelligence technologies spreads in the financial and insurance industry, the amount of case studies reporting both successful and unsuccessful implementations has been increasing. We divide this section into two parts, the first being a collection of successful implementations, and in the second, we expose less fortunate attempts. This way, we give a broader perspective of the whole picture to better help organizations to adopt the best route to follow when implementing technology in its processes.

The healthcare sectors have long been susceptible to fraud due to costly services and little motivation to implement preventive strategies. However, those that choose to work with AI technologies can reduce fraud. A use case of using machine learning algorithms to better accurately detect medical sector falsifications in insurance is described. During the case study, it is explained how the technology would make the algorithms learn the pattern of the correct machine insurance, so those policies that were different from the pattern would be classified to the class of fraud cases, alleviating the manual task of the inspectors.

Continuing with the insurance sector, a use case of neural networks in claim rejection prediction is demonstrated. The implementation was a success and revealed interesting factors that directly influence the fact that these claims should be considered false. The tested architecture was a purely functional multilayer perceptron with supervised learning, backpropagation algorithm. By using AI technology for this task, employees were freed from this time-consuming task and it became possible for insurance companies to process a greater number of claims faster while still considering the needed investigation on claims with the highest chance of being fraudulent.

2.10.1. Successful Implementations

Artificial intelligence is proving to be a game changer in the world of financial fraud detection as seen through implementations made by start-ups and banks. Look through this list of pioneering companies in the field demonstrating best practices for real-time fraud detection and mitigation thanks to artificial intelligence.

Kount was founded in 2007 to focus on verification of online transactions for eCommerce mixture of four Artificial Intelligence methods: hidden Markov models, case-based reasoning, unsupervised clustering, and supervised classifiers, boosting decision trees in particular.

American Express being one of the strongest players in transaction card service market has implemented a diverse range of Artificial Intelligence systems based on historical fraud pattern data developed through the years. Transaction flagging systems, real-time customer transaction alert systems, and real-time merchant intelligence systems are just a few among the complete range of systems deployed.

BBVA has implemented a simpler use of Artificial Intelligence in the market that runs on rules. As BBVA is a multinational bank, filtering rules are used to filter out unlikely transaction matches according to transaction behavior during fraud detection and prevent account blocked problems for clients without risk. The system analyzes historical fraudulent transactions to detect patterns commonly present. The most popular services used are credit and consumer credit cards, personal loans for mortgaging, and other transactions during the Dark Web market.

Rakuten acquired a unique banking business model and launched Rakuten Bank in 2001 that has implemented a highly skilled artificial intelligence fraud detection businesses. The company operates all transactions and communications online, that gives both advantages and disadvantages.

2.10.2. Lessons Learned from Failures

Failures are an educational opportunity to establish policies and procedures to mitigate the risk of having a disaster within the organization. Implementing AI for higher revenue for your company can seem like the holy grail; however, failures are an essential part of the process. Even on the business level requiring an initial investment to set up a system, failures can be detrimental to the bottom line. Are you using analog algorithms only designed for previous standard detection techniques? Was the decision-making process insufficiently straightened? Does your algorithm have sufficient initial training or monitoring technology to avoid bias? Did you rely entirely on automated processing of your AI? These are some of the questions to begin asking the company when the initial

excitement turns into frustration due to system ineffectiveness. To avoid that disillusioned feeling, we want to underscore some general ideas that we collected from the mistakes of our predecessors. Since many people in the business world have had experiences, one could say the previous generation built an experience elbow that led them to develop correct operations. But the degree of chaos created in smaller companies can lead to their collapse in 6 or 18 months when there is a sustained failure. Having tools that let you take advantage of the discoveries of those who came before us can speed up the learning curve and has a high economic impact. Due to the chaotic nature of decision-making in smaller organizations in an area as complex as abandonment modeling can generate somewhat amateur trials that damage the development of the discipline.

2.11. Future Trends in AI for Fraud Detection

Emerging Technologies In the short term, we will see the growing use of machine learning to enhance methods that use predictive analytics to set and adapt fraud prediction thresholds more dynamically. The biggest challenge is prioritization: data science will help companies tell which transactions to focus on first, allowing them to maximize the impact of their fraud detection efforts. The future will also see the disaggregation of the fraud detection and risk mitigation areas in terms of organizational structure and technology architecture, flowing into a more active deterrent space, especially in cross-organizational settings. The rationale is simple: a complex payment and transaction ecosystem demands a more sophisticated risk approach than piecemeal transaction risk decisioning and IT architecture can offer. Moreover, this pivot to a more active deterrent space will allow organizations to positively influence consumers, making them active partners in the deterrent process, building on sentiment analysis to enhance personalization. In the medium-term timeline, we expect the deployment of deep reinforcement learning algorithms specifically designed to prevent risk actors from over-exploiting structural weaknesses in payment ecosystem design. 2.11.2. The Role of Blockchain In the medium- to longer-term horizon, advances in blockchain technologies may provide organizations, especially banks and payment service providers, with a risk mitigation tool that knocks down silos, sharing transaction responsibility across organizations while also making it safer to share data internally among risk actors. Blockchains can be used to lower costs, make settlement and clearing faster, and/or to heighten security. Organizations will also need to balance the risk-related costs of using blockchain-type technologies against the benefits that accrue by boosting consumer trust. Blockchain also has the potential to reduce instances per transaction for both the entity and the consumer while facilitating the establishment of a mutually trusted identity.

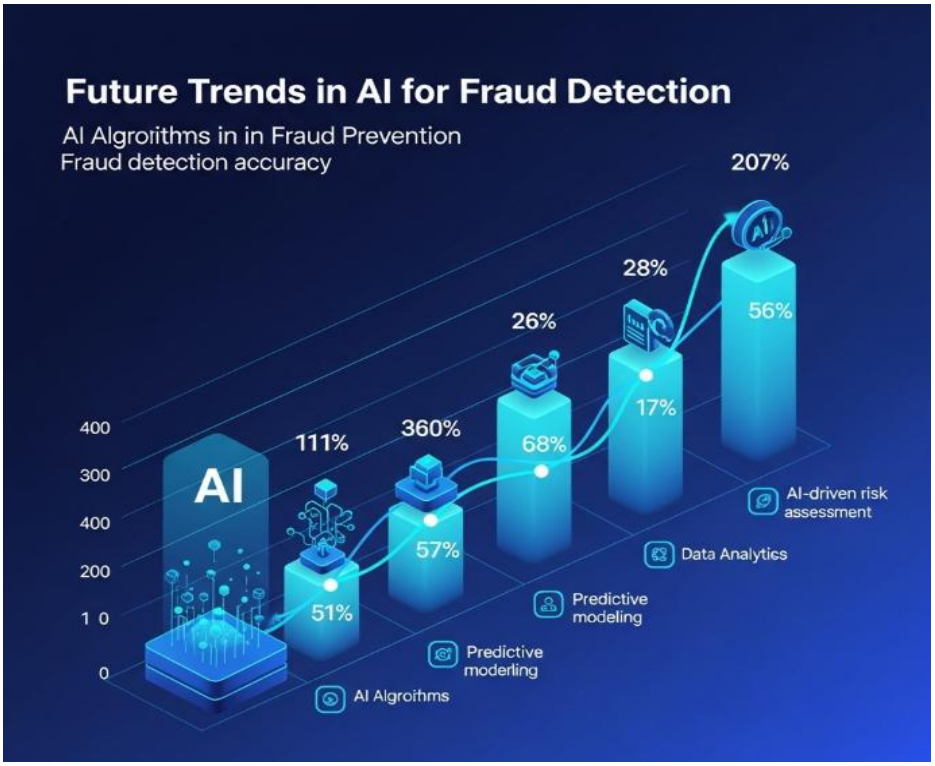

Fig : Graph illustrating future trends in AI for fraud detection.

2.11.1. Emerging Technologies

As we move forward, we can expect the development of more sophisticated AI-based tools and solutions that will be capable of carrying out more predictive and preventive analysis in order to help detect and prevent future fraud attempts. In the meantime, the development of new and exciting technologies in combination with the establishment of further global law and security regulations and standards will continue to minimize fraud opportunities. The emergence of new technologies such as distributed ledger and blockchain technology, cloud computing, the internet of things, ultrafast networks and processors, and new biometric innovations will radically change the approach toward fraud and risk mitigation.

With the development and commercial introduction of IoT-enabled devices, the collection of data around customer-specific habits, preferred routines, health conditions, and ways of conducting business will increase significantly. Coupled with powerful analytic tools, these devices will allow companies to create digital profiles that are specific to a customer and their routine. As a result, AI can be proactively trained to detect outliers in order to prevent the occurrence of fraudulent transaction patterns or

minimize the risk by using specific customer authentication protocols for specific transactions post-event. Because all this data is highly sensitive, privacy regulations need to be established and incorporated into the related analytical tools. Organizations need to conduct trainings and collaborate with trusted third parties in order to be security compliant in the landscape of rapidly growing data availability.

2.11.2. The Role of Blockchain

Blockchain enables what are called "smart contracts," self-executing code that lives on the Blockchain and performs specific tasks automatically, with only minimal outside interaction. Smart contracts are used to facilitate transactions of cryptocurrency as well as in many other ways, and they can help increase the efficiency of fraud prevention and detection, by eliminating human interpretation from many transactional operations. The fraud detection and prevention that must be performed to prevent fraudulent transactions from occurring is reduced, since the Blockchain is immutable and cannot be altered once information is added. All transactions are verified by the entire network before they are recorded on the Blockchain, so all verified transactions are immutable. This increases public confidence in the system. Blockchain technology will also provide the ability to both verify and access a person's past identity – the document being presented for verification and acceptance, and its prior presence on the Blockchain, creating a digital identity resume. Thus, Blockchain has the potential to revolutionize identity verification by providing a means to validate that you are who you say you are by using Blockchain's digital signature capabilities. This would eliminate the problems with repetitive "know your customer" requests, where firms spend huge sums verifying the identities of their customers, only to have them request onboarding from a different firm a few days later, beginning the process all over again. The value of digital identity verification will be explored in the next section.

2.12. Conclusion

While research in real-time fraud detection using machine learning techniques has been carried out for specific applications such as financial transaction verification, healthcare fraud detection, telecommunication fraud detection, payment fraud detection, insurance fraud detection, and ledger operation verification, this work emphasized the generalization of all of the recognized approaches and methodologies into a clearer and more coherent whole, subsequently addressing the major factors and dimensions influencing application in all domains. Based on this work, a taxonomy to all such real-time fraud detection systems is built, separately outlining the business needs and considerations surrounding real-time deployment of machine learning algorithms. Based

on the insight gained throughout the whole of this work, an architecture for enterprise fraud detection operations is additionally proposed. The suggested architecture, made universally agnostic to organizational flow and protocol modifications, minimizes friction during the system training-to-implementation cycle while minimizing model feedback loop time. Fraud detection is a challenging and interesting problem that can benefit immensely from deep technology advances. It is not only a challenging data science problem in itself, but also a practice where technology advances can be leveraged to realize automation and efficiency improvements including reduced detection as well as feedback loop times. We highlighted the areas where technology advances are required. The survey of detection problems discussed in this report also suggests there are many areas where detection accuracy can be improved upon, whether through innovation or via nudging. Such challenges are not easy to transition from proof of concept to production systems due to the extreme challenges related to volume data, decision making as well as model feedback loop delays. It is our hope that insight from this work helps practitioners to achieve tangible progress in reducing fraud costs by leveraging the power of AI and researchers to continue pushing the frontiers of AI research to help accelerate transitions from models in research to model operationalization in production fraud detection systems.

References:

Y. Yazici, "Approaches to Fraud Detection on Credit Card Transactions Using Artificial Intelligence Methods," arXiv, 2020.

M. A. Hossain et al., "AI-Enhanced Fraud Detection in Real-Time Payment Systems: Leveraging Machine Learning and Anomaly Detection to Secure Digital Transactions," Australian Journal of Machine Learning Research & Applications, 2025.

K. Yamini et al., "An Intelligent Method for Credit Card Fraud Detection using Improved CNN and Extreme Learning Machine," in 2023 8th International Conference on Communication and Electronics Systems (ICCES), IEEE, 2023.

O. Wang, "Explainable AI for Credit Card Fraud Detection: A Review," IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 5, pp. 2301-2316, May 2023.

H. L. Liu and H. Li, "Deep Learning for Credit Card Fraud Detection: A Review," IEEE Transactions on Emerging Topics in Computing, vol. 10, no. 1, pp. 237-247, 2022.([arxiv.org][1], [sydneyacademics.com][2], [academia.edu][3])