Deep Science Publishing
https://doi.org/10.70593/978-93-49910-19-5_6

DeepScience
Open Access Books

Chapter 6: Designing artificial intelligence -based fraud detection systems capable of handling dynamic threat environments

6.1. Introduction to Fraud Detection

The phenomenal growth of online services with Internet connectivity has led to major changes in the way people conduct business, increasing their frequency and lowering operating costs. However, this reliance on online services has also led to increased incidences of fraud originating from Internet connectivity and by centralizing agencies responsible for revenue collection. This has led to significant amounts of losses on individuals and industries. Fraud occurs because of the anonymity enabled by the Internet, integration of online services, and the movement of funds from one place to another without adequate identification. Emerging technologies such as mobile devices with high computing power and availability of vast amounts of personal information on the Internet make it easier for people to commit fraud. Mobile phone fraud, which mainly deals with fake account creation and abuse of free trial services and promotions, incurs huge losses on mobile carriers. Identity theft by online criminals may result in significant amounts of a person’s identity information being sold online (Phua et al., 2010; Bahnsen et al., 2016; Fiore et al., 2019). Therefore, there is a real need for efficient automated online fraud detection to prevent such misuse and losses.

Fraud detection is the latest challenge to the burgeoning area of processes that detect invalid or anomalous activities occurring in systems or organizations, referred to as anomaly detection. Detecting a fraudulent transaction or activity is a problem of identification of humans who are infeasibly distinct from each other, not a normal activity like money laundering through banking systems. These unusual activities are expected to occur in small amounts of the total activity, and yet some of these unusual activities can lead to huge disasters if they are missed, hence automatic online fraud detection is thus a very challenging problem (West & Bhattacharya, 2016; Roy et al., 2018).

6.1.1. Overview of Fraud Detection Concepts

No organization is completely immune to fraud, errors or mistakes, nor can any guarantee that detection will take place. The success or failure of an organization is often predicated on whether fraudsters are allowed to operate and perpetrate their schemes unnoticed for a period of time. As a result, organizations make attempts to detect fraud as early as possible. However, when fraud is discovered, especially in the case of significant loss, organizations often commit a significant amount of time to find all possible fraudulent transactions. Because of the financial effects of fraud, companies have invested significantly in fraud loss detection capabilities, along with various technologies. Yet, despite these investments, notable examples of fraud have resulted in the loss of billions and billions of dollars.

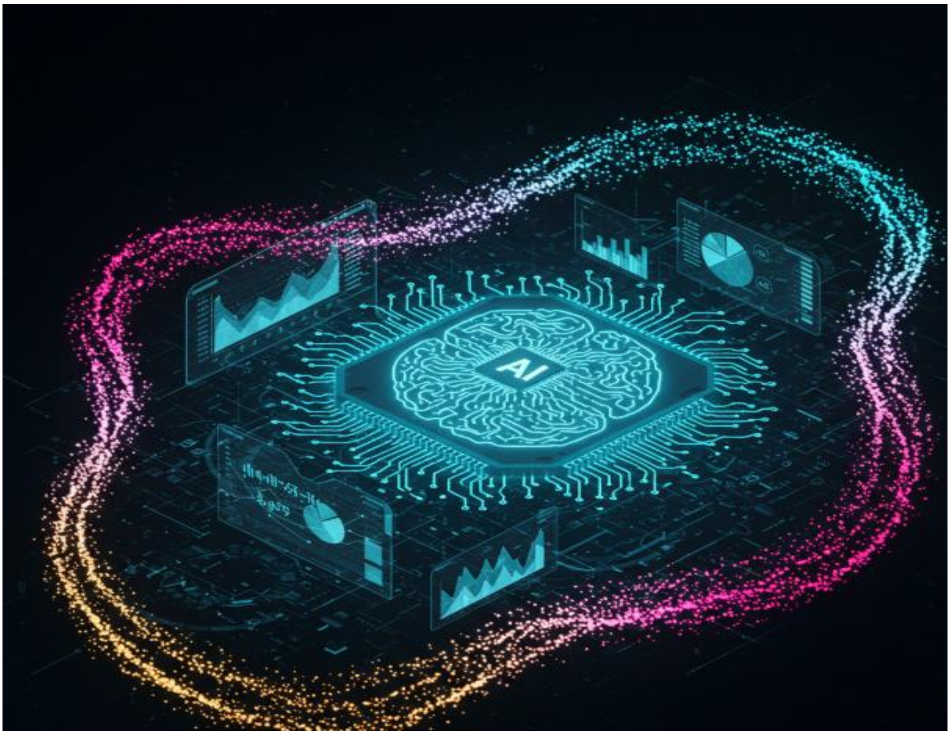

Fig 6 . 1 : AI in Fraud Detection

The purpose of a fraud detection mechanism is to determine the existence of fraudulent characteristics that may or may not be known. In many cases, patterns of fraud are learned first and used subsequently to detect fraud. Statistical classification techniques are consequently commonly used to model fraud. In fact, fraud detection can actually be considered a classification problem; a particular transaction or individual is classified as fraudulent or as non fraudulent. Modelling of normal transactions is preferable since there usually exists many more instances of normal transactions than of fraudulent transactions; thus, models based on normal behavior can prevent classifying all transactions as fraudulent based on the lack of adequate fraud models. In fact, this is a common deficiency of commercial fraud detection systems designed to detect credit card fraud, which typically operate on a transactional basis.

6.2. The Role of AI in Fraud Detection

As a result of the huge volume of transactions taking place on a daily basis, fraud detection systems face high levels of scale, rapid decision-making requirements, increased sophistication of fraudsters using bots to exploit digital vulnerabilities, as well constant change as new methods to currently evade detection previously successful fraud strategies. Properly training a machine learning algorithm is the key to success for fraud detection systems. AI is increasingly being implemented in fraud detection systems by institutions concerned with limiting financial losses. How do we harness the capacity of AI to enable effective monitoring of access and transaction activity in increasingly dynamic decision environments? Artificial Intelligence refers to computerized technologies that are able to simulate the performance of certain tasks similar to that of humans. Machine learning refers to a type of AI that is able to organize large data sets and combine features into categories. Machine learning can thus create a deep learning algorithm that can predict user behavior. Some have argued that humans are still better able to identify and respond to radical novelty or new types of fraud risk, but are somewhat slowed down by not being able to detect subtle dynamic changes in user account activity. These two worlds can increasingly come together, allowing trained humans to quickly review flagged anomalies produced by automated algorithms. In addition, the natural capacity of technology to handle very large amounts of data will allow for the oversight of very large accounts and accounts that have multiple users, which was previously difficult to achieve or nonexistent as oversight was triggered just by conduct violations.

6.2.1. The Impact of AI on Fraud Prevention Strategies

Managing fraudulent activities is a significant area of research for many organizations due to the extent of disruption they cause, both culturally and financially. Fraud has no specific responsibilities and it is difficult to find a single cause of fraudulent activities.

For instance, tax avoidance and tax evasion have no clear motivation, even within an organization tax fraud and embezzlement do not have a common approach therefore, unlike traditional anti-fraud approaches, fraud detection cannot be a one-size-fits-all solution. Government organizations, firms and customers incur both economic losses and mental distress as a result of fraud in various sectors. Corporations incur considerable revenue loss as a result of customers (both internal and external) committing fraud. In recent years, multinationals have faced significant losses as a result of fraud by both their vendors and customers. The growth of fraud demands government organizations identify solutions that protect taxpayers from fraudulent schemes. Criminal organizations amass millions of dollars in illicit profits through serious crimes that inflict incalculable harm on families and communities. Each citizen's safety and security are ensured by government organizations, which serve as a watchdog against housing, Medicaid, and other similar programs. Both businesses and people have suffered from consumers and employees committing fraud against them. Solution providers for related technologies invest heavily across various sectors to meet the increasing challenges.

The prevention and discovery of such actions that are against the organizations is handled by many Artificial Intelligence-based systems. AI-based prevention systems find use in search tools and Investigate Service to download critical evidence from revenue collection, and prevent identity verification fraud. As a business enabler, AI acts as a solution for concerns regarding security compliance, specifically in finding fraudulent activities performed by customers and business staff. AI enhances and streamlines the identification process of fraud detections, which in fact improves upon the lower revenue growth rates of industries such as telecommunications, financial services and public sector services.

6.3. Understanding Dynamic Threat Environments

As artificial intelligence (AI) becomes a core component of many applications, it is natural to consider fraud and cybercrime in association with these systems. In the past few years, AI, especially machine learning (ML) and deep learning (DL), has proliferated into financial fraud detection. Automated machine learning and cloud services have made ML methods easy to use. Cybercriminals are also adopting AI to create more powerful systems. Some academics argue AI is increasing the volume of crime and its efficiency. But, this focus on AI is too narrow. Financial fraud, cybercrime, and money laundering have existed for many years, with changing patterns of attack and evolving methods of detection. In contrast to a traditional multi-investigator crime, these activities are engaged by fluid criminal groups who are oftentimes dispersed around the world, with connections to other illicit operations, including drug trafficking, human trafficking, and terrorism.

The constant evolution of threats, often referred to as a dynamic threat environment (DTE), is a vital consideration for research and development. A common ML method is to learn a model on past data, score new incidents, and mark incidents for audit or for reporting as incidents. At some future time, the alert can be evaluated and either marked false positive or true positive. Fraud such as phishing, skimming, and card-not-present hacking, have been going on for many years but continuously change tactics, targets, and method. Automated methods are already being tested in the field but DTE can be a substantial problem. There are many other areas of AI being evaluated for similar issues including optical character recognition and facial recognition.

6.3.1. Navigating the Landscape of Evolving Threats

The landscape of evolving threats in fraud detection comprises a multitude of techniques and threat vectors. A threat vector defines the route taken by a fraud actor to successfully commit fraud and covers both opportunity and intent aspects of committing fraud. The opportunity aspect represents an abuse that is exploitable, while the intent aspect represents the motivation for exploitation. These threats provide the interested actors with opportunities for intervention to exploit weaknesses associated with one or more components of a fraud system. The opportunities for intervention represent the exercise of free will in choosing to exploit a weak point; thereby, creating unethical profit at the expense of the organization while violating the implicit trust associated with the transaction. The fraud opportunity space is aligned with the opportunity aspect of the definition, unlike the fraud driving space aligned with the intent aspect. Both of these spaces create a model of what is known as a dynamic threat environment because they create boundaries that can be used to delineate the threat landscape. Threats that exist in this space can include fraud-internal, fraud-external, fraud assistance, and fraud overt threats. Fraud-internal threats exploit weaknesses in components associated with detecting or preventing internally committed fraud. Fraud-external threats exploit weaknesses in components associated with detecting or preventing fraud committed by external fraud actors. Fraud assistance threats exploit weaknesses in components associated with detecting or preventing the assistance of fraud committed by an external fraud actor. Fraud overt threats exploit weaknesses in components associated with detecting or preventing the carrying out of a grudge or challenge. Many threats have one or more of these aspects that affect the way that fraud evolves.

6.4. Key Components of AI-Based Fraud Detection Systems

Fraud detection systems play a significant role in the security of many traditional and modern business activities. For instance, in the traditional economy, the use of credit and debit cards has multiplied, exposing both companies and individuals to the risk of fraud by making them dependent on bank online services and on cybercriminals who operate in this area. In addition, the spread of electronic commerce has accelerated the need for effective fraud detection systems to monitor transactions, offer insurance to merchants, and avoid losses caused by false purchases. Today, fraud detection systems are crucial for many products and processes. Their implementation allows the detection of possible fraudulent operations on health care, borrowing and loan, marketing, identity, money laundering, online gambling, insurance, manufacturing, and bankruptcy.

Fraud detection systems have evolved from the usage of traditional statistical techniques to the usage of Artificial Neural Networks, Decision Trees, Random Forests, Support Vector Machines, and Verifiable Computing. The evolution of information and communications technologies has made speculations such as electronic payments possible. However, at the same time, it has made money laundering more accessible, with serious consequences for human health and safety. In addition, the volume and dimensionality of the collected data have increased significantly, resulting in many inherent challenges in handling and processing these data. Fraud is a typical example of class imbalance problems found in supervised learning models, where few fraud instances represent the minority class, while the great majority of transactions are safe. In addition, fraud detection uses customer-specific features that are too sensitive and confidential to be shared. For all the reasons stated above, fraud detection should be dealt with in a personalized way.

6.4.1. Data Collection and Preprocessing

The first phase in building any machine learning system is acquiring the data needed to detect patterns within the task at hand. The most frequently used form of fraud detection data is labeled transaction feature vectors with labels indicating whether the transaction was normal or fraudulent. Sometimes other forms of data have also been used in research, such as miner-coded sequences of actions for which miners identified as fraudulent and others. This research relies on a number of synthetic or real datasets which are openly available, listing even the oldest and also the most recently released ones.

The second, and perhaps most important, phase in building an effective machine learning system is preprocessing the previously collected data so that it is suitable for the task at hand. This is especially true for fraud detection, given the highly imbalanced nature of the underlying data. The most commonly considered methods to balance the feature vector representation of the data with respect to the underlying label distribution consider the use of a random undersampling of the majority class or the use of a random oversampling of the minority class. Different more advanced sampling algorithms, either enhancing these methods or improving better to balance the data, have also been proposed, generally enhancing the detection effectiveness of the model constructed by using them. Other researchers consider the features engineering or the cost-sensitive techniques. Recently, more complex approaches using generative deep learning, clustering, fuzzification-based ensembles, or unsupervised learning have also been analyzed, taking advantage of autoencoders or generative adversarial networks.

6.4.2. Machine Learning Algorithms

There is a huge advancement in proposed machine learning algorithms, including ensemble classifiers like boosting, bagging, random forests, stacking, hybrid methods, semi-supervised and unsupervised methods and in the areas of neural networks, such as deep learning, representation learning, and convolutional networks. These advances are all candidates to help discover false and fraudulent behavior hidden in the large volume of available data. Considerations include the amount of data and time available, and whether the output should be interpretable.

A detailed examination of the large body of work on algorithms generally used to improve predictive performance for fraud detection, and the many variations on these algorithms is way beyond the scope of this survey. Here, we briefly highlight some methods, but are not covering all available methods, not providing reviews of the algorithms, and not reporting the wide variety of domain applications. These include simple classifiers, like logistic regression and Naive Bayes, basic classification trees, the relatively simple K-nearest neighbor, as well as support vector machines. In addition to boosting methods like AdaBoost, and Random Forests, approximate methods like Hoeffding Trees. Variants of neural networks, like multi-layer perceptron, convolutional networks and deep learning as well as variations of self-organizing maps, restricted Boltzmann machines and convolutional deep belief networks, are also included.

Ensemble techniques generally used to generate improved prediction accuracy like Random Forests and boosting-based ensembles have also been chosen for fraud detection. Decision trees are also popular because of their interpretability. Detecting frauds in a set of insurance claims has been studied using other interpretable classifiers like Bayesian networks and Fuzzy Classifiers. Unsupervised anomaly detection methods have been used heavily for fraud detection of credit card transactions, in addition to other applications where labeled training data is scarce or even unavailable.

6.4.3. Real-Time Monitoring

Real-time fraud detection lies at the heart of an AI-based fraud detection system. Most business transactions are temporary phenomena that occur in real time. Each transaction can have several features, which can continuously change under external influences. Furthermore, not every transaction is available for classification, as it may not possess features necessary for the classification task. For example, in predicting whether an insurance claim is fraudulent or not, sometimes the insurer may not have access to details about third-party claim receipts. For analysis or anomaly detection approaches, the model would not know either from which features or component that decision would have come.

Moreover, transactions are often grouped into other hierarchical entities that are also the subject of continuous change. For example, a call detail record can be classified as fraudulent and have alarm trigger points based on aggregation at the CLI level for a given period or aggregation at the country level based on the destination. This shows that activity monitoring of an account can also act as a warning for possible future fraud detection. Such real-time monitoring can be done manually for big external changes that affect life or work on a large scale.

However, for change detection based on machine learning models, as the available data increases, an AI-based approach becomes more useful. Such systems allow the identification of alerts that can trigger action for external change detection based on rules or models. Yet, despite the capability of data-driven approaches, current operational alert monitoring is heavily rule-based. The potentials of adopting machine learning to detect alerts are not sufficiently validated. Some machine learning-based approaches are available for an anomaly or fraud detection alert at scale for credit card fraud.

6.5. Types of Fraud and Their Characteristics

Before discussing the methods and procedures of AI in fraud detection, we present a categorization of types of fraud to be detected. The various fraud types can be classified according to the phases of user activity in an online system in which they occur, as well as according to the relationship of the fraudster to the victim — basically, whether the fraud is perpetrated by insider employees or by outsiders. Three major categories of online fraud are

- A user acts in bad faith and intends to deceive the service provider. The user pretends to be someone else and transacts with the provider's internal systems.

- A user intends to deceive external end users of the service and creates a fake service page to attract victims.

- An internal employee or group of employees collaborates with external users for providing false data to the provider's system.

These categories can further be specialized into various subtypes of fraud depending on each online service and its modus operandi. While it is not practical to survey every possible subtype for every user activity, we present here a short description of some pertinent fraud types and their characteristics. The presented subtypes can also be combined for a more complex fraud case.

Credit Card Fraud

Credit card fraud is an online fraud where someone other than the actual card owner obtains goods or services from a merchant using the card's details for payment. In most cases, credit card fraud is done through online services. Given the explosive growth of e-commerce over the past years, online credit card fraud is a problem that is increasingly alarming merchants, banks, and consumers alike. In fact, fraudulent online card-not-present transactions are growing at a staggering rate. Therefore, for commerce companies, having effective fraud detection systems in place is an essential requirement.

6.5.1. Credit Card Fraud

Credit card fraud has been a problem for consumers, businesses, and financial institutions for over four decades. The incidence of fraudulent credit card transactions has steadily increased since the first criminal acted to take an unauthorized advantage of a cardholder's access to credit. Chargebacks not only create a financial loss for merchants but also damage their reputation with credit card companies. Credit card fraud occurs when the criminal makes unauthorized use of a capable card or its credentials to conduct a fraudulent transaction, usually through a chargeback requested by the real credit card account holder. Carding, which is the fraudulent purchase of goods or services with stolen card details, is one attempt to conduct credit card fraud.

Carding requires the criminal to obtain a database of valid card credentials, preferably with a digital fingerprint for the carding transaction. Both carding and fraudulent transactions via physical credit cards and online redemption of gift certificates are examples of “transaction-initiated” credit card fraud, with a chargeback requested by the credit card account holder. "Account takeover" is the act of changing the credentials for a payment account, such that the true account holder does not have access to the account. When the criminal uses stolen account credentials to access the real victim's payment account in order to transfer funds, that act becomes closely associated with "banking" or "identity" fraud.

Fig 6 . 2 : Enhancing Credit Card Fraud Detection

Criminals adapted quickly to the added layers of protection and complexities presented by card-issuing banks, online payment systems, and the evolution of e-commerce methods. Fraud detection methods are powerful learning systems that are capable of identifying any unusual or transactional activity that deviates from the cardholder's previous patterns. Understanding the various motivations that drive criminal behavior will allow us to create detection models that more appropriately reason over behavioral characteristics.

6.5.2. Identity Theft

Identity theft is the act of using personally identifiable information (PII) of another person to impersonate that person for fraud or deception. In most cases, the perpetrators of identity theft steal the identity without the detection or knowledge of the victim. Because the victim usually does not have knowledge of the theft, this type of fraud can be among the most damaging. By using PII to impersonate the victim, identity thieves often borrow large amounts of money in the victim’s name or make large purchases. These bogus loans or purchases are normally not paid by the thief, leaving the victim responsible for the unpaid debts.

The most frequently stolen forms of PII used in identity theft are Social Security number (SSN), name, and address. Stealing someone's credit card number is credit card fraud, but stealing someone's SSN is not theft of an identity. Identity thieves use the SSN and other PII to fabricate new identities that can be used to apply for loans, credit cards, and other services. Other forms of fraud, including credit card fraud, mortgage fraud, and wire fraud, are used in conjunction with identity theft. Identity fraud also includes the unauthorized act of acquiring services in the victim's name without payment, such as obtaining loans or utility services or creating fake bank accounts to be used in money laundering. College student loan programs, payday loan programs, and mail-order or online third-party vendors are often used to commit identity fraud.

6.5.3. Account Takeover

Account takeover occurs when an attacker uses stolen account credentials—often usernames and passwords stolen in prior data breaches—to access someone else's account with the intent of fraud, theft, damage, or hostile takeover. ATO is different from identity theft in that in identity theft, the attacker impersonates a victim, often by opening a new account, to commit a crime while still in the victim's shoes; in ATO, the attacker impersonates the victim in order to gain access to an existing account. The ATO fraud detection problem typically arises in ongoing transactions; however, these accounts often include nonfinancial accounts linked to financial profiles, creating the opportunity for the perpetrator to commit further fraud even before stealing funds. ATO has become a common, costly, and disruptive form of online fraud, with little warning for users and merchants. The bulk of all fraud arises from ATO of customer accounts. ATO-based fraud against loyalty and gaming accounts has gained special notoriety.

The surge in ATO activity in recent years has been driven by two converging trends: the increased prominence of digital channels, with rising public dependence on mobile apps and digital wallets in particular; and the rise and advance of cybercrime tools and techniques, with fraud-as-a-service offerings and scraped and purchased ATO account credential databases becoming commonplace. Cybercriminals often operate with military precision and the opportunistic air of small-time street thieves; skilled threat actors possess large ATO botnets and are proficient in controlling and directing sophisticated fraud schemes. On the digital side, clients offer an array of accounts, some more attractive; they possess abundant focused ATO attack metrics; the ATO attack surface runs five, nine times larger than non-ATO for preauthorization on-site merchants.

6.6. Challenges in Fraud Detection

Amid the growing popularity of deep learning and other machine learning-based techniques for real-time fraud detection, multiple solutions have started to explore this excitement. However, many of these new-age AI systems have been deployed in situations where these models can work effectively while not actually facing one of the biggest challenges in real-world deployments, the numerous edge cases associated with fraud detection. This is the reason why organizations continue to deploy traditional rules-based systems and domain expert-first analysis in tandem with machine learning systems. Fraudsters are constantly evolving their tactics, requiring continued model refresh efforts to keep detecting fraud effectively. The distinguishing facet of fraud detection is that false positives and false negatives are weighted very differently depending on the domain of application. For instance, a false positive in a bank's automated credit card transaction verification system might mean a delay of a few hours in the customer being contacted to verify the transaction. A false negative in such a case could mean the customer reporting a multi-million-dollar fraudulent transaction after the fact. Both of these results have a large cost, but the company absorbing them in each case is different.

The challenges have been laid out claiming that there are reasons why banks using customer data to create predictive models to automate the fraud detection process cannot. Banks are bound by fiduciary responsibilities and cannot share their customer's sensitive data with third parties. One of the foremost challenges facing global agencies and organizations attempting to share data and build a holistic view of possible terrorist activity and detect fraud is privacy concerns. Concerns have been raised, however, that sharing certain kinds of personal data undermines an individual's rights to data protection, privacy, and non-discrimination. However, automated processes for sharing such concerns could put systems in place to allow stakeholders to share information without violating a given country's data protection laws.

6.6.1. Evolving Fraud Techniques

For an unprecedented time, innovations and technological development have somewhat been observed to be misappropriated with the available resources, such that eventual criminal, fraudulent, oppressive, harmful, and injurious acts are committed against another individual or organization. The various technological scoundrels capitalize on the loopholes of the system in order to breach it for personal gain. Among the several criminal activities that are being seemingly committed under the apparition of technology, manipulation of currency and fund transfer is one. Being the most prone to fraudulent activities, the banking and finance sectors are attacked, sometimes unendingly. However, advances in computer forensic, data mining, and artificial intelligence domain are providing researchers with new tools to develop even more hyper-automated and self-adaptive models. A fundamental reason is that organized crime groups conduct fraud at a global scale, finding victims in several nations and targeting countries where regulation is still inert or where traditional methods of detection are not fully-functional.

With the exhaustive ever-growing ruptures of origination and belief in conducting banking and financial transactions over the Internet, internet banking fraud has generated strong cognizance among researchers in the proxy of network routing and associated network management. Malware is a piece of software that is purposely designed to disrupt, damage, or gain unauthorized access to a computer system. Because the various techniques do evolve, all Internet-enabled devices can be targeted and agent malware installed for either short-term effect or long-term data gathering. Social engineering also utilizes this fraud hack attack scenario in a different way. Rather than ‘doing,’ a fraudster would persuade an individual, in-person or via telephone calls, social networking, or even the Internet.

6.6.2. False Positives and Negatives

False positives and false negatives are key concepts in AI-based fraud detection systems. Due to the nature of fraud, a low threshold is required to minimize the occurrences of false negatives, but at the same time, a high threshold is needed to minimize the occurrences of false positives. The vulnerabilities and losses associated with false positives and false negatives are quite different. False negatives can lead to fraud loss, while false positives can lead to a happier customer turning to another provider. For instance, in predicting that someone will repay a loan, a false negative would involve lending to someone who will not and suffering a loss. However, a false positive would involve denying credit to someone who would have repaid the loan and causing them to seek credit from another lender. For artificial intelligence to be effective in different use cases, it has to strike a balance between minimizing false positives and false negatives.

False positives occur when a fraud detection system uses a detection threshold that is too low and predicts fraud in a legitimate transaction. For fraudsters, this misclassification is of little concern. However, for the victim companies, the impact is significant. Fraudulent transactions directly contribute to company loss and possibly to customer loss when such transactions remain undetected. In addition, if these transactions occur on a regular basis, they will validate the training sets and bias the detector toward the class of fraudsters. Companies will devote their resources to stop the transaction, charge the customers or adjust the accounts, and rebuild their reputation. The misclassification of legitimate transactions as fraudulent will lead the detector to ignore the task of learning the fraudulent transaction behavior. The result will be the increased classified segments for which long-term consequences cannot be anticipated.

On the other hand, detecting real fraudulent actions is a difficult task, and another source of misclassification is false negatives. Generally, in the evolving environment of fraud, and with the fluctuation that financial transactions present, false negatives are thought to be a greater problem. The impact of not detecting fraud is that the company loses resources. Automated fraud detection is needed to limit and minimize losses. If it can be shown that artificial intelligence can detect fraud effectively, many organizations will be willing to invest in this technology.

6.6.3. Data Privacy Concerns

Fraud detection deals with the detection of a sample which is different from the remaining data and at the same time is of great consequence. While addressing different types of frauds, data privacy issues arise in both data acquisition and model building. Anonymizing the data is the first step. In fraud detection, especially insider crimes, it is equally important to safeguard the data owners from adversarial identification. This is more challenging because the fraud sample is an outlier. Any kind of anonymization may lose information that is otherwise available. Moreover, in order to have better accuracy, usually the data owner contributes some features which are sensitive in nature. In different fraud cases, such additional sensitive contributions by the data owners increase privacy concerns.

With the advancement of new technologies, privacy concerns have risen. Biometric data is the most sensitive type of personally identifiable information. Unlike passwords, biometric identifiers cannot be changed or replaced when compromised. Popular authentication systems store biometric records without sufficient protection or security. Encrypting biometric information is a common method to provide protection. But it does not help if the biometric information is compromised before encryption and biometric libraries might also retain non-encrypted records. There have always been moral, ethical, or civil rights issues regarding automated decision-making and privacy violations as common concerns. Therefore, the collection of data must be reported to the public prior to being executed, with a thorough examination of consent-based access.

6.7. Designing an AI-Based Fraud Detection System

The design of an AI-based system for fraud detection requires the consideration of a variety of aspects affecting structure, functioning, operation, and usability of this class of systems. For model-driven systems, the consideration of the principles of MDE can offer an effective approach to foster the development of application models capturing properties and design aspects that can be reused to reduce the time and effort needed for the creation of a specific system. MDE also ensures the enforcement of heuristics and policies that derive from standards about the development of solutions for cyber security in general, and fraud detection in particular.

Specifically, we must consider three questions concerning the design of a fraud detection system integrating AI techniques: which structure must the system have, how it can be integrated with existing systems that manage the data needed by the AI components, and which UI must the system provide to enable analysts and end users to interact properly with the system, ensuring effectiveness in the management of the application domain. We navigate through these questions in the next three subparagraphs.

1. System Architecture

With respect to system architecture, we can assume that an AI-based fraud detection system is composed of a set of components that offer services to other systems or organizational units that interact with the system for exploiting its capabilities. Services can be related to the management of the general workflow concerning fraud incidents and generations of specific outputs. Other services concern the AI-based components that detect or discover fraud patterns, or that support the definition, adaptation, and refinement of the models employed by these components. AI-based components are typically trained and validated by people who understand the fraud domain, AI techniques, and are able to combine AI with the application domain of the component. Their role is particularly critical when knowledge extraction and model generation approaches are used or when classification-based models must be very accurate before being used in production.

6.7.1. System Architecture

AI fraud detection systems comprise several components. These include ingested data from domain-specific streams, an entity resolution module, a purpose-built feature generation module, a predictive power dialect selection module, and a user reporting module. Core to the usability of these systems is an interpretability feature that inputs an explanation request and outputs a natural language explanation. A purpose-built AI fraud detection system is necessary as general-purpose AI architectures lack many of the features that make such systems effective, greatly increasing the potential for false positives in production settings. In a credit card fraud detection application, false positives correspond to unsatisfied customers that did not have fraudulent charges on their cards, while true negatives correspond to satisfied customers that did not experience fraud. While general-purpose AI architectures try to maximize predictive power across classes, the potential mismatch between generating true positives and misclassifying false positives is irreconcilable or highly improbable.

The feature generation stage is critical to the performance of an AI fraud detection system. The current adage on feature generation is that it should be an art not science. The primary goal of AI fraud detection systems is to assist domain-specific users in selecting primary types for reports or summaries. In addition to traditional supervised predictive power, semi-supervised predictive power and unsupervised predictive power should also be used to select features. These systems should also use query-specific predictive power to filter on the predictive power of the input query and output report. In addition, predictive power should also be dialect-specific, as the dialects in the reports or summaries may exhibit predictive power centered on one or a few features. Explanations of the role of features can motivate users to become involved in selecting features for their organizations, increasing trusting usage, and decreasing explanation fatigue.

6.7.2. Integration with Existing Systems

AI-based fraud detection systems may have to be integrated with existing systems for a variety of reasons. First, an organization may already have an extensive set of security policy rules defined that are relevant for fraud detection. The integrative system may use the existing definition first during the detection and the AI-based system can be used for the supplementary role of implementing more sophisticated and flexible reasoning methods, or to reduce false positive decisions made by the original rule definitions. Second, many organizations have developed reactive processes for handling fraud incidents, such as if an alarm is triggered and resources are involved. There may already be involved available resources such as human or machine resources, which are assigned to fraud detection and the integration can enable the system to allocate the appropriate resources efficiently and timely to minimize the loss of fraud. Third, fraud detection may involve multiple different systems across disparate functions involving interaction with many back-end systems such as databases and transaction processing systems. Thus, identifying the interfaces, identifying what functions are to be invoked by the detection system into the external systems, and how to programmatically instantiate these functions is essential for multi-agent collaborative rhetorical and resource allocation during alerts and incident response. Integrating a system requires quality and reliability assurance so that timely performance and correct interfaces into the external systems are accounted for in defining both the conceptual logical model of the integrated system and the necessary engineering capabilities of the assembled working physical system.

6.7.3. User Interface Design

The user interface (UI) of a fraud detection system is crucial for practical deployment and system acceptance. Initially, we would like to point to systems that detect fraud but do not provide warnings to analysts about the fraud instances identified. Clearing small fraud instances in real time or inspection during a reporting period without further analyst involvement is not a UI challenge. Typical systems that fit this description are credit card systems that use transaction patterns and other available characteristics of the user, such as summary transaction statistics, previous actions, and dynamic user behavior modification, to design predictive models that can filter large volumes of transactions. The fraud cases that are identified are usually audited either in real time or during a designated audit period and may not involve any analyst activity. Automated post-validation of detected fraud actions is typical for these systems.

The UI design challenge is addressed when the analyst has to assess the fraud report generated by the system and decide on a critical action. A critical action is one that is likely to change the user status. For instance, in credit card fraud systems, the critical action is a chargeback against the merchant or a report indicating that the card was stolen. Mail fraud systems will focus on an internal employee, while systems that use website visitor logs will concentrate on periodic access attempts. These decisions introduce temporal relations not only to the internal mechanism by which the analyst decides which user/merchant/website behavior is fraudulent but also to the design of the UI. Given the sparse nature of fraud activity, fraud detection performance is not the only measure of predictive model usefulness, and can be even worse in certain situations.

6.8. Evaluating the Effectiveness of Fraud Detection Systems

The evaluation of Fraud Detection Systems (FDSs) has to take into account the particularities of fraud. These highly imbalanced scenarios where a very small proportion of elements belong to the misclassified class and the high cost of misclassifying the minority class dictate the use of special performance measures centered on the performance on the misclassified class. These include Precision and Recall, the F1 Measure, and the Area under the ROC Curve. Also, it is normal to carry out a quantitative performance evaluation in parallel with an error analysis, where one or more validation sets are built for studying the performance of the current FDS as regards the real fraud patterns wherein these patterns are labeled and in which the effect on fraud detection of the fraudulent patterns can be analyzed in detail. It is also common to consider populations affected by different constraints when performing an analysis of the performance of two systems or realizations of a same system in a competitive scenario.

The above considerations lead us to the necessity to build benchmark data sets for studying the performance of FDSs systematically, independently of the type of categorizer used therein. The task is extremely time demanding because it requires the handling of tons of trade data and the study of the provenance of the labeling performed by the human experts, as well as the confirmation that the labeling is correct. Under these circumstances, the availability of publicly labeled benchmark data sets is scarce, and mostly focused on FDSs using traditional categorizing methods. It is also usual that these publicly available data sets have been labeled by a single expert, independently of the label accuracy and the depth of fraud history information that may be mined to label samples correctly.

6.8.1. Performance Metrics

The first metric, accuracy, is a measure of the proportion of correctly classified instances of each class used for evaluating the performance of classifiers. However, due to class imbalances inherent to fraud detection problems, accuracy can be a misleading performance indicator. As such, other commonly utilized performance measures are precision, recall, and F1 score. Precision quantifies the proportion of true positive out of the positive instances predicted by a classifier. Recall indicates the proportion of positive instances correctly identified by the classifier. Importantly, these two metrics are conflicting, as combining a strict fraud prediction rule will increase precision but decrease recall, and vice versa. A compromise between precision and recall can be obtained with the F1 score, which is computed based on the harmonic mean of the two.

The last two metrics, false positive rate (FPR) and false negative rate (FNR), indicate how many negative and positive instances are misclassified, respectively. In fraud detection tasks, FNR is usually minimized as the costs of the fraudster detecting a flaw in the fraud detection system, and then exploiting it to create an attack, can be large. However, the cost incurred on the fraud detection system operator due to false positives can also be considerable, especially in e-commerce applications, as the costs associated with the consequent disruption of transaction processing and/or loss of legitimate sales can often far exceed those incurred from the FNR. As such, the precision and recall measures are more informative, as they can be plotted as a precision-recall curve for different values of classifier score thresholds. Determining the appropriate value of the threshold is a problem in itself and will depend on the costs associated with the misclassification of a transaction to one of the two classes. When the cost of FPR and FNR are equal, the threshold is usually selected to minimize the classification error rate. Various research works selected the score threshold to balance these two metrics based on the cost matrix.

6.8.2. Benchmarking Against Traditional Methods

Benchmarking can be a challenge in a fraud detection context. The first challenge is that the objective should be to detect as many real frauds as possible while minimizing false alarms. Implementing this is problematic, because implementing fraud detection solutions is also hard and requires a special initiative, care, and knowledge. Otherwise, the solution is not likely to be used properly, with the risk of damaging the solution if the wrong operationalizations are in place. Hence, the results when implementing fraud detection solutions usually take a long time to get hold of, thereby increasing the cost of estimating any given method's performance.

For these reasons, a common way to benchmark different fraud detection methods is to use an artificial data set and mimic the act of detecting fraud. Then we have control over the data inside the data set and the way to implement the fraud detection solutions. Note that many fraud detection methods have been created, and are still developed, with inspiration from other application areas. However, this goes the other way as well. To some extent, we can consider fraud detection to be an extension of anomaly detection, and the two fields to be closely related. The idea is that the objective of fraud detection is to find and act upon the detected anomaly, while the implicit assumption in anomaly detection is that the anomaly is indeed an anomaly, that is, not realized in practice.

Besides these challenges and points concerning routing benchmarking, and to the best of our knowledge, there is not much work benchmarking against other areas. Here, we will look a bit deeper into methods that assess anomaly detection methods with modeling common aspects of modern commercial fraud detection systems. It is a start, perhaps a step in what can be called fraud detection modeling.

6.9. Conclusion

The accelerating progress of AI and machine-learning technologies in developing large AI models trained on big data has started to improve decision-making in many real-world applications, such as fraud detection. In this chapter, after reviewing the essential principles of fraud detection, we presented how current developments in AI can enable the design and steering of effective and efficient fraud detection policies and strategies. We would suggest that new concepts, techniques, and tools derived from AI and ML can transform developed countries' Fraud Detection Services from laggards to front-runners of our AI economy. Also, as we know, the world's most advanced AI and ML technologies herald a new AI age of fully automated processes. Trailing the changes in the Digital Fraud business ecosystem driven towards not so much automation, but IVD will require the constant focus of our Banking and Government Services on maintaining the development and investment incentive structure in Digital Economy over the upcoming decades. This change is possible with a step change in available solutions and innovations by partner organizations in the broad ecosystem of trusted players supporting various initiatives and, of course, critical engagement by SMEs, emerging Blockchain companies, and DLT technology innovators.

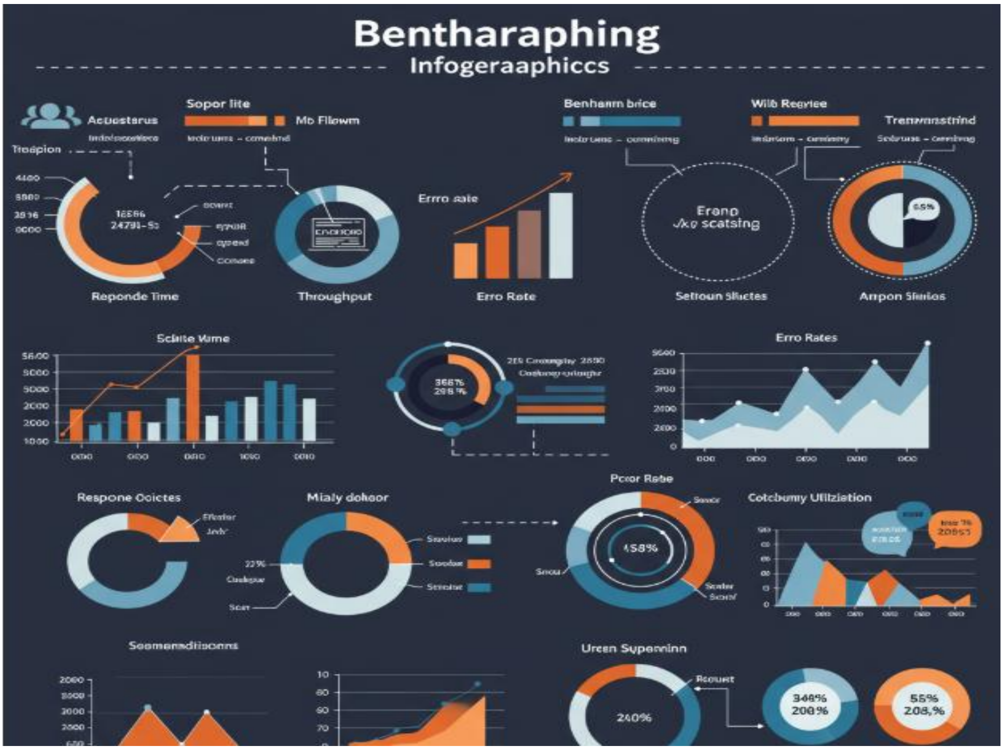

Fig 6 . 3 : Benchmarking Metrics

Research on developing and applying AI technologies in all its forms is critical for efficiently detecting and mitigating fraud detection in today’s complex crimes that simultaneously damage consumers, businesses, and government. Developing and deploying automated enterprise IVD ecosystems based on ML will allow us to achieve the full long-term benefits of our Digital Economies and the trust in them that ultimately allows us to plow the proceeds back into improving our collective way of life and better forecasts and actions upon the Economic Climate we all share.

6.9.1. Final Thoughts and Future Directions in Fraud Detection

Due to the unprecedented advancement of the threats provoked by maleficent actors from various areas and the difficulties in monitoring the ultra-dynamic threat landscape, the sophisticated detection of fraudulent actions is one of the most desired applications of AI techniques, as they are able to face the dilemma of being driven by the absolute need for automation and the impossibility of injecting too many human intelligence assumptions. In this landscape, this chapter has presented some recent and mostly seminal studies, covering the cloud, vehicular, social, mobile, RFID, and defense areas that present AI-based fraud detection efforts in all developed, presenting how scientific work has helped to enhance this specific field. As a future research direction, we highlight the need for the design, experimentation, and comparison evaluation of more distributed models that do not require the elimination of valuable user data, such as multimedia and context. Another possible future direction is related to the extraction of explainable patterns from the models based on the combination of low-accuracy symbolic methods that are capable of extracting patterns from the AI engines. The last highlighted future direction is the investigation into milder and less pessimistic policies that avoid the abusive tracking and level of action that today's systems impose on the legitimate user to ensure their ongoing and additional protection against purely fraudulent actions. This effort collection must help in the reach of the real dream of fraud detection systems in response to aformentioned dichotomy: with increased days' action by only the most suspicious actors, but playing the game of life and the efficacy of their monitoring. Hence, the proposed systems must be efficient not only at the level of fraud detection but also in associating to such a detected account the need for a correct and timely application of penalties that guarantee the maintenance of the accepted justice level and the reduction of the collateral damages of the application.

References

Phua, C., Lee, V., Smith, K., & Gayler, R. (2010). A Comprehensive Survey of Data Mining-Based Fraud Detection Research. arXiv preprint. https://doi.org/10.48550/arXiv.1009.6119

Roy, A., Sun, J., Mahoney, W., Alzahrani, A., & Pal, A. (2018). Machine Learning Techniques for Fraud Detection in Financial Transactions: A Survey. IEEE Access, 7, 63437–63447. https://doi.org/10.1109/ACCESS.2019.2916646

West, J., & Bhattacharya, M. (2016). Intelligent Financial Fraud Detection: A Comprehensive Review. Computers & Security, 57, 47–66. https://doi.org/10.1016/j.cose.2015.09.005

Bahnsen, A. C., Aouada, D., Ottersten, B., & Stojanovic, A. (2016). Feature Engineering Strategies for Credit Card Fraud Detection. Expert Systems with Applications, 51, 134–142. https://doi.org/10.1016/j.eswa.2015.12.030

Fiore, U., De Santis, A., Perla, F., Zanetti, P., & Palmieri, F. (2019). Using Generative Adversarial Networks for Improving Classification Effectiveness in Credit Card Fraud Detection. Information Sciences, 479, 448–455. https://doi.org/10.1016/j.ins.2018.02.060