IJIREEICE

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

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

Vol. 8, Issue 12, December 2020

DOI 10.17148/IJIREEICE.2020.81213

Big Data-Driven Fraud Detection in Digital Banking Platforms

Dileep Valiki

Independent Researcher, India

Abstract: Simulating the market demand and supply interaction on individual banks and applying econometric techniques on the simulated bank data to statistically establish a significant operational relationship between the banking industry and the real sector of the economy. Stress-testing models have also been developed for some aggregate banking sector indicators for work. Cost and benefits of fraud detection solutions were evaluated to substantiate the business justification and return on investment for building, maintaining and supporting fraud detection solutions. The NIST Big Data Interoperability Framework and the DAMA-DMBOK framework were utilized to define the data ecosystem along six perspectives—data sources, data governance, data quality, data storage, data privacy and data provenance. The analysis included a taxonomy of financial fraud, and the definitions and characteristics of fraud detection, risk assessment and fraud monitoring. Some of the trends emerging in the digital banking ecosystem were also discussed. Banks need to achieve a balance in the prevention and detection of fraud, which is seen as never ending and deserves constant attention. The playbook defines how the operation teams respond to fraud. Such resources should ideally reside in the same location and have clear visibility for decision-making.

Given the extensive amount of available data, fraud detection solutions for banks require solid scientific foundations. Data-driven approaches that leverage machine learning and data-mining techniques are therefore being explored as tools for improving the prediction and detection of fraud events. Supervisory authorities require banks to incorporate fraud detection systems into their digital banking ecosystems. Such Data Driven Decision Systems must be capable of automatically detecting suspicious activity within acceptable limits of cost, accuracy, risk, coverage and performance level without too many false positives and as few wrong flags as possible.

Keywords: Banking; Big Data; Cloud Computing; Data Mining; Digital Forensics; Decision Theory; Risk Management; Real-time Analytics; Telecommunication; Fraud Detection.eva.

I. INTRODUCTION

The banking and financial services sector is experiencing rapid digital transformation, demonstrated by the widespread adoption of online and mobile banking platforms. Much like e-commerce, which has witnessed skyrocketing growth in recent years, the digital banking ecosystem is not impervious to malicious activities. The prevalence of fraud—enabled by the convenience and anonymity of digital banking platforms—poses a substantial risk to customers, banks, and the overall economy. It ultimately erodes customer trust and confidence in the banking system. Fraud detection is a key business capability of digital banking platforms—as necessary as payment and transaction authorization. Burgeoning digital interactions (transactions, accounts, devices, applications, and services) and advances in Big Data technologies present a fertile ground for adopting a data-driven model for transaction fraud detection and prevention in digital banking. The convergence of Big Data sets and technologies with data-driven decision-making methodologies, including Artificial Intelligence (AI) and machine learning, enables sophisticated fraud detection and prevention solutions. These methods leverage data from a wide range of internal and external sources, including an organization's operations, customers, employees, partners, suppliers, market, and ecosystem. Data-driven models are able to identify novel and known fraud patterns, employing playbooks and rule-based detection systems. Sensitivity to fraud detection performance is ensured by integrating business intelligence and analytics, financial metrics, and customer experience insights into a decisioning framework.

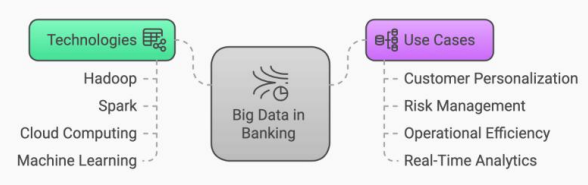


Fig 1: Big Data in Banking

IJIREEICE

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

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

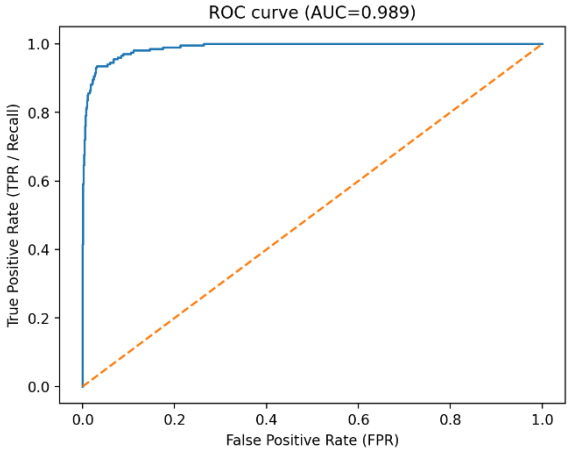
Vol. 8, Issue 12, December 2020

DOI 10.17148/IJIREEICE.2020.81213

1.1. Background and Significance

In the aftermath of the financial crisis of 2008, banks and other financial institutions began investing heavily in information technology, hoping to derive new business models and avenues for profitability. These investments, aided by the advent of big data and cloud computing technologies, led to the proliferation of digital banking model. Today, a bank's ATM and Internet banking channels are not just alternate channels, but are the primary channels for most customers. The convenience and lower cost of these products were key factors in its success. With rapid evolution of technology in these enabling channels, the nature of interaction with the banking systems has changed. Customers now rely more on the bank's mobile banking and internet banking application or a third-party payment gateway to conduct banking transactions. Banks continue to encourage customers to transact through these digital channels because it reduces operation and transaction costs and allows for more personalized services. However, the growing importance of digital channels has also attracted fraudsters to exploit weaknesses in the banking system and technology to steal customer credentials, create unauthorized transactions, and gain from the banking system fraudulently.

Consequently, fraud detection in digital banking channels has become a critical area of research for a safe and secure digital banking ecosystem. This research is especially relevant given the continued use of digital channels during the COVID-19 pandemic. To ensure a secure digital banking environment, there is a need to adopt and operationalize Big Data-based fraud platforms that incorporate a fraud detection and alert generation ecosystem for early warning and detection of digital banking fraud. An ecosystem-specific fraud framework for digital banking, covering aspects such as data sources, data quality, model building, operationalization, impact on customers and the economy, and the challenges and limitations of a Big Data-based fraud detection system, is explored.



1.2. Research design

Fraud detection in digital banking systems constitutes an interdisciplinary domain connecting banking and data science. A holistic perspective brings together research in fraud detection, banking, and Big Data technologies. A formal system-of-systems framework capable of integrating aspects of these domains facilitates data-driven fraud detection in digital banking environments. Scanning the literature on these topics unveils the nature and scope of the problem, the data ecosystem and governance, methods and algorithms, evaluation criteria and benchmarking, governance and risk management, and the impacts of fraud-detection models on fraudsters and genuine users.

Fraud detection and management in banking systems encounter inherent difficulties. Fraudsters' intent is to exploit emerging technologies, systems, and their vulnerabilities to perpetrate malicious acts. They continually adapt their modus operandi to avoid detection, while the chance of wrongful allegation increases for innocent users. Technology companies do not possess the banking expertise to identify fraud. Banks are experienced in detecting fraud but lack the scale, expertise, and technology to leverage data for adequately predictive models. The solutions built in isolation by the banking ecosystem are largely ineffective. Thus, a unique system of modern Big Data technology, complete with data-capture, storage, analysis, and actuation capabilities, is essential for data-driven banking fraud detection.

II. THEORETICAL FOUNDATIONS OF FRAUD DETECTION IN DIGITAL BANKING

Fraud encompasses intentionally deceptive practices that target enterprises and individuals, often exploiting the asymmetrical access to information and capabilities. Potential fraudsters are often enabled by an organizational structure

IJIREEICE
ISSN (Online) 2321-2004
ISSN (Print) 2321-5526
International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering
Vol. 8, Issue 12, December 2020
DOI 10.17148/IJIREEICE.2020.81213

that allows for breaches of trust. The space of fraud in financial services is considerably wider than pure fraud against users, illustrated by the following taxonomy from the Australian Fraud and Cyber Crime in the Financial Sector. Here fraud is categorized into eight types with relevant sub-types: (1) Identity theft and impersonation fraud (Application fraud, Account takeover, In-person impersonation fraud, Triage impersonation fraud); (2) Authentication manipulation fraud (Account credential theft, Phishing fraud, SIM swapping fraud); (3) Account takeover fraud; (4) Cyber fraud beyond the financial institution; (5) Employee fraud (Dishonest insiders, Abuse of position for gain, Facilitation of fraud); (6) Fraud against the organizations' customers; (7) Fraud against the organizations' funders; (8) Insider fraud.

The challenge of detecting and preventing such fraud in real time is growing with the increasing reliance on technology to evolve and conduct transactions. Authorities also monitor and assess how disruptive technologies, industry changes, and consumer behaviors contribute to enabling, preventing, and detecting fraud. Fraud is becoming a major concern across digital ecosystems, leading to rising concerns for customers and adverse cost–benefit ratios for organizations, customers, and ecosystems reconfiguring to incorporate formal BDA command centers and teams, regardless of industry The Big Data Analytics ecosystem reflects the existing, available, and future repository of digital (structured and unstructured) transactional data for organized investigation, detection, and prevention of these illicit practices with technology supported capability for processing and analyzing at an enterprise level.

2.1. Definitions and Taxonomy of Financial Fraud

Fraud constitutes a wide range of deceptive acts. It is difficult to capture all instances of fraud with a comprehensive definition due to the diversity of motives, methods of execution, and areas of influence. Despite this, most definitions center on four essential concepts: deceiver, deceived, method of deception, and resulting damage. VBA defines fraud as a deliberate portrayal of falsehoods intended to benefit oneself and harm others. The American Law Institute, citing moral culpability, emphasizes deceit as the basis of punishable fraud. Koller provides a broader categorization of fraud in banking, distinguishing between external fraud (e.g. identity theft) and internal fraud (e.g. embezzlement).

A formal classification of fraud forms the basis for developing a fraud information system, where financial crime is classified into money laundering, fraud, corruption, and production fraud, and banking fraud is categorized into external and internal types. The digital banking business processes most affected by fraud are identified, with an academic examination of the implications of banking product development selecting fraud prevention as the most essential design variable. Cybernetic theory forms the basis for defining financial fraud as an information transmission and control imbalance between the deceiver and person deceived, which is exemplified by ten key forms of cyber fraud.

Financial fraud is also presented in relation to other banking and business functions, which permits a strategic analysis of fraud management under the four categories of banking crime scheme, crime typology, fraud detection system (FDS) configuration, and bank functional process. The well-known fraud triangle specifies three conditions that shape conditions suitable for fraud: motive (such as financial problems), opportunity (due to organizational weaknesses or overrides), and rationalization (such as belief that “everyone else is doing it”).

Equation 1: Precision, Recall, F1: derived step-by-step

Start from conditional probability:

$$\text{Precision} = P(y = 1 \mid \hat{y} = 1)$$

Convert probability to counts:

- The event $\hat{y} = 1$corresponds to **all predicted fraud** $= TP + FP$.
- Within those, true fraud are **TP**.

So:

$$\boxed{\text{Precision} = \frac{TP}{TP + FP}}$$

Recall answers: “Among true fraud transactions, how many did I catch?”

Start from conditional probability:

$$\text{Recall} = P(\hat{y} = 1 \mid y = 1)$$

IJIREEICE
ISSN (Online) 2321-2004
ISSN (Print) 2321-5526
International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering
Vol. 8, Issue 12, December 2020
DOI 10.17148/IJIREEICE.2020.81213

Convert to counts:

- $y = 1$corresponds to **all actual fraud** $= TP + FN$.
- Caught fraud are **TP**.

So:

$$\boxed{\text{Recall} = \frac{TP}{TP + FN}}$$

The F-score family is:

$$F_\beta = (1 + \beta^2)\frac{PR}{\beta^2 P + R}$$

where P =precision, R =recall.
For $\beta = 1$:

$$F_1 = 2\frac{PR}{P + R}$$

Now substitute $P = \frac{TP}{TP+FP}$and $R = \frac{TP}{TP+FN}$:

$$F_1 = 2 \cdot \frac{\frac{TP}{TP + FP} \cdot \frac{TP}{TP + FN}}{\frac{TP}{TP + FP} + \frac{TP}{TP + FN}}$$

Factor TPin denominator:

$$F_1 = 2 \cdot \frac{\frac{TP^2}{(TP + FP)(TP + FN)}}{TP\left(\frac{1}{TP + FP} + \frac{1}{TP + FN}\right)}$$

Cancel one TP:

$$F_1 = 2 \cdot \frac{\frac{TP}{(TP + FP)(TP + FN)}}{\frac{(TP + FN) + (TP + FP)}{(TP + FP)(TP + FN)}}$$

Cancel $(TP + FP)(TP + FN)$:

$$\boxed{F_1 = \frac{2TP}{2TP + FP + FN}}$$

2.2. Big Data Paradigms and Infrastructure for Banking

Big data paradigms provide the technological foundations for fraud detection in digital banking platforms. The advent of the cloud, IoT, Big Data Analytics, Social Media, and Artificial Intelligence underpin a new digital paradigm. Digital banking platforms operate in distinct but intertwined banking ecosystems: the bank–customer ecosystem; the bank–bank ecosystem; and the bank–law enforcement and regulatory agency ecosystem. Detecting or preventing fraud requires orchestrating the decision process across the three ecosystems, supported by a digital data ecosystem for fraud detection.

Digital banking fraud detection can be conceptualized as a risk management process based on Big Data. A digital risk management architecture provides the blueprint for representing the preliminary game plan. Digital fraud detection constitutes only one of the many processes and data-empowered capabilities needed to transform traditional resources, risk management, and banking methodologies into a digital-first environment. Banks and banking regulators are gradually shifting their operational mindset toward thinking in terms of a data-as-a-service strategy, which allows them to focus on data ecosystems for a wide spectrum of services, products, and high-impact use cases. A data ecosystem for fraud detection integrates multiple types and sources of data from a wide array of partners, harnessing the datasets through Big Data paradigms, technologies, and analytical platforms.

IJIREEICE
ISSN (Online) 2321-2004
ISSN (Print) 2321-5526
International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering
Vol. 8, Issue 12, December 2020
DOI 10.17148/IJIREEICE.2020.81213

III. DATA ECOSYSTEM AND GOVERNANCE FOR FRAUD DETECTION

A comprehensive fraud detection framework in digital banking relies on data from multiple internal and external sources. Data sources such as transactions, device information, network patterns, and location information provide a context for assessing fraud risk. Data from social networks helps to obtain user information and rank that information, while data from the internet of things indicates a user's everyday behaviors. However, fraud detection systems frequently focus on internal data, leading to fraudsters exploiting the evident patterns.

Detailed information about the user fraud ecosystem is essential for decisioning systems to adopt an attack-playbook methodology, wherein the risks and modes of operation for different fraud scenarios are established. Proximity playbooks identify fraud by exploiting the consumer's relationship with another party involved in the transaction. The accuracy and reliability of data for decision-making is vital. Continuous monitoring of data sources and strict governance policies ensures that only data with the due diligence sign-off is utilized for implementation. Similarly, provenance tracking aids regulatory and compliance requirements, ensuring that customer and third-party data privacy laws are not violated.

3.1. Data Sources and Collection Methods

All types of big data and fraud detection require data sources and collection methods to be defined and planned. For big data, a variety of external sources including data feeds, clouds, and web scraping mechanisms can augment internal sources. Data sources must therefore be identified (including cybersecurity data), data collection methods and mechanisms determined, and be integrated into a data ecosystem. A multitude of data sources currently exist within banks: e.g. transaction logs; usage logs; customer profile databases; performance databases; credit Bureau databases; customer declarations; account opening data; KYC/AML data; behavioral biometrics; cookies (including geo-location); device fingerprints; and social networks. A smart fraud solution can take all these internal data sources (and possibly still many others) into account. In addition, such a solution should consider taking many external sources into account and combining them to enrich the decisioning framework and improve fraud detection.

The analysis of digital financial services has demonstrated that a multitude of big data sources can be captured for the fraud detection system. External data sources that can be collected from the web and the dark web can also be applied to the fraud detection framework. Indeed, the definition of data sources and collection mechanisms can be formalized for any type of big data. Consequently, the identification of data sources is only a small part of the quest for data quality. Data-pipeline quality also includes the reliable sourcing of data (data provenance), the treatment of sensitive information such as personal or financial data in line with data protection regulations (data privacy), and the design and management of the data environment in line with data governance best practices.

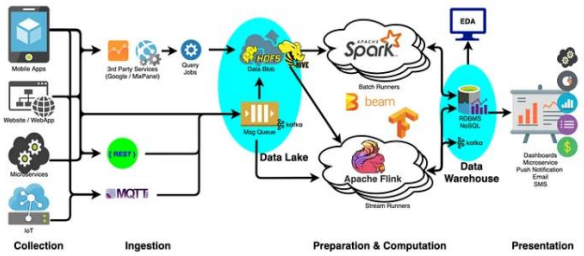


Fig 2: Big Data Architecture for Fraud Detection and Prevention in Banking

3.2. Data Quality, Provenance, and Privacy

Fraud detection in digital banks requires sufficient data of high quality, both to train models effectively and to make real-time predictions. Data testing is essential to ensure quality. Unfortunately, while testing can detect discrepancies in data ranges and patterns, it provides no assurance of detection for data quality in real time or production. For both training and deployment, data must be exhaustively validated and any known bias minimized. Provenance represents the lineage of information, detailing how it was created. In fraud detection, provenance establishes trust. Finally, preservation of privacy is an absolute requisite during data collection, storage, and access.

Data quality is discussed first in terms of common characteristics and criticality. These issues are followed by an exploration of provenance, or information lineage, and privacy, both of which shape perception to the end user. Quality is more about data testing and validation, while provenance and privacy are addressed as enablers supporting trust in decisioning frameworks. A poor-quality dataset results in model training that does not yield a good fit and thus a weak model. In deployment, untested models can still produce results; however, these results will not be reliable. The only

IJIREEICE
ISSN (Online) 2321-2004
ISSN (Print) 2321-5526
International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering
Vol. 8, Issue 12, December 2020
DOI 10.17148/IJIREEICE.2020.81213

type of quantitative assessment data quality undergoes prior to modelling is descriptive, which provides the user with high-level information on the data set or merely reveals unexpected or implausible data.

IV. METHODOLOGIES AND ALGORITHMS FOR FRAUD DETECTION

Fraud detection systems call for sophisticated detection methodologies. The choice of methodology depends on the characteristics of the underlying data and the amount of task-specific labelled data available for training. Fraud detection faces the challenge of an increased class imbalance between legitimate and fraudulent accounts, with few instances of fraud being known. The detection of rare events in the dataset requires appropriate methodologies.

Supervised and semi-supervised fraud detection methods use the labelled fraud instances to extract parameters used for detection, while unsupervised and anomaly-based detection methodologies avoid the challenge of class imbalance by considering Gaussian distribution for legit data description. Supervised and semi-supervised approaches are particularly effective when sufficient labelled fraudulent transactions are available. Such methodologies start by transforming the raw transaction data into a new feature space boosted through domain expertise. Feature engineering involves the use of custom transformations and aggregations, e.g., frequency-based, importance-weighted, or time-since-last granularity, resulting in a bank-specific latent space optimally adapted for fraud detection. Feature enrichment provides the detective framework with a comprehensive set of customer behavior parameters. It serves both direct modelling and data-driven selection of auxiliary data components. A combination of functions can be employed for data preparation using cutting-edge Big Data tools (e.g., Spark).

Equation 2: FPR and ROC curve, then AUC
FPR answers: "Among legit transactions, how many did I wrongly flag?"

$$\text{FPR} = P(\hat{y} = 1 \mid y = 0) = \frac{FP}{FP + TN}$$

If your model outputs a **score** $s(x)$and you classify fraud when $s(x) \geq t$(threshold t), then each $t$gives a pair:

$$(\text{FPR}(t), \text{TPR}(t))$$

where $\text{TPR}(t) = \text{Recall}(t)$.
Plotting $\text{TPR}(t)$vs $\text{FPR}(t)$over all thresholds gives the ROC curve.
Geometrically:

$$\boxed{\text{AUC} = \int_0^1 \text{TPR}(\text{FPR})\, d(\text{FPR})}$$

Numerically (typical implementation), use trapezoids across points $(x_i, y_i) = (FPR_i, TPR_i)$:

$$\boxed{\text{AUC} \approx \sum_{i=1}^{m-1} (x_{i+1} - x_i) \cdot \frac{y_{i+1} + y_i}{2}}$$

4.1. Supervised and Semi-Supervised Approaches

SEMA has become a platform for sales and marketing fraud amorphosizing where advertisements are technologies are installed for the product base commercialization where datacenters of comprehensive traffics are built accomplished with sustainable transactional operations. KEDRAP is the only core solution product which is site specific and connected to the web dynamically with KEDRA as the master server masking the sectoral vision together contributing more towards the classifier concept.

The task Assist offering 'product three marketing' is drastically different from the conventional marketing system of consumers seeking famous and branded products. The NEW GENERATION firms/companies are putting their hard work for years to replenish their product all over the world at less margin. This make the task Assist act as a medium to source the brands from the manufacturers instead of searching the products of awareness. The genuine customers as per the active KEDRA scanned traffics scoping same product are pushed to the product through their KEDRA installed accounts along their scope activity.

IJIREEICE

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

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

Vol. 8, Issue 12, December 2020

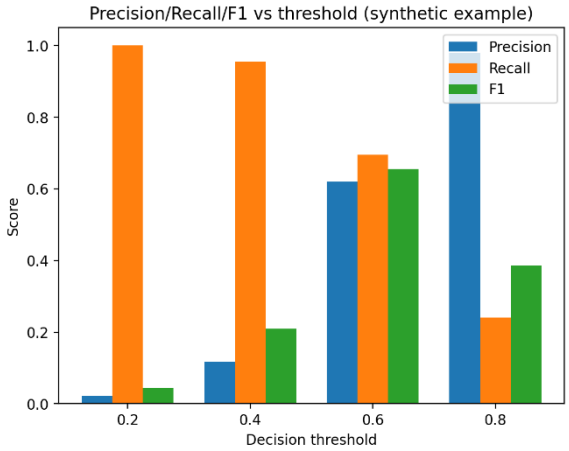
DOI 10.17148/IJIREEICE.2020.81213

Formal and informal request of bulk purchase on the task Assisting site is done by the customers and the price will be negotiated by the product owner to the concerned genuine consumer. In e-commerce no dependence on files and test accounts but automated. For SEMASM using internet marketing no credit cards and mobile no. are integrated and no need of online bank transfer. Legitimate members seeking "product" do not need to branch/submit anything for admission. SYSTEM members selling items duly registered for online marketing have no dependency to wait for buyers.

4.2. Unsupervised and Anomaly-Based Methods Unsupervised learning techniques exploit unlabelled data and reveal underlying patterns for later detection purposes. The lack of label also excludes supervised evaluation during training. Silhouette score, Dunn index, Davies-Bouldin and Calinski-Harabasz metrics enable clustering quality assessment, while various methods to identify and classify outliers can be applied. A big advantage inculcated by unsupervised approaches is noise resilience, since they are trained on true positive as well as true negative examples. Data distribution and quantity can be leveraged for patterns emergence via self-training or co-training. While self-training assigns pseudo-labels and the classifier self-learns, co-training uses distinct classifiers and annotates conflicting predictions, e.g. through a graph-based approach.

Anomaly detection aims to distinguish rare items. Unsupervised clustering tools (k-NN, k-Means) or rapid autoencoders build simple models for a 'normal' class. A similar paradigm adopts a supervised approach, using real change events (negative class) to learn a classifier. A density-distribution-based approach uses a differentiable distribution model as a partitioning tool and is ameliorated by KD-Trees. Anomaly detection using deep generative models jointly minimizes the data reconstruction loss and a data-oblivious Kullback-Leibler divergence to train, while procedure based on a Wasserstein distance associates labels with the nearest high-density regions via a two-level risk function.

Transaction networks also disclose useful relations. The attribute-pattern-relationship triplet associated to each network component can classify nodes or edges by trait or pressure type. Communities serve to generate new nodes and edges, impute attributes and detect malicious nodes, or to dispatch volume and time-weighted anomalous edges.



V. EVALUATION, VALIDATION, AND BENCHMARKING

Fraud detection, like other ML tasks, can be assessed in terms of the four basic properties of predictive models: calibration, discrimination, outcome accuracy, and prediction accuracy (Disease, 2008). However, these general properties should be considered in light of the specific business, regulatory, stakeholder, and operational requirements of the detection and prevention process within the broader context of the governing life cycle playbooks for fraud.

Dangers of Over-Sensitive Detection
The development and monitoring of any detection model should be preceded by a stakeholder workshop that agrees on the acceptable metrics for performance evaluation. For supervised detection, stakeholder agreement should then be followed by investigation, detection–decisioning playbook creation, and benchmarking against data obtained prior to the use of any playbooks and used for expense-booking and regulatory reporting purposes.

Discriminative metrics are most useful when assessing any playbook at a level above the component detection model. High accuracy rates should be treated with caution, since it is the cost of false positives that usually drives the creation of a playbook rather than avoidance of false negatives. Conversely, a very low detection rate or success ratio is almost

IJIREEICE
ISSN (Online) 2321-2004
ISSN (Print) 2321-5526
International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering
Vol. 8, Issue 12, December 2020
DOI 10.17148/IJIREEICE.2020.81213

certainly a cause for concern and should require additional scrutiny, even when any loss can be handled by a fraud-booking model. In an ethical environment, detection should be considered for improvement when the cost of false negative frauds exceeds the cost of false positive false alarms.

Isolated detection components in such playbooks should be assessed using precision, recall, and F-score (Cong, 2015). Such examination also enables insight into which data fields in playbooks warrant augmentation via change detection to increase the proportion of fraud cases detected. These three metrics should be focused upon highly unstable detection fields, where prediction accuracy is low and variations in pattern and model threshold level unreliable. These data fields are often excellent candidates for inclusion in detailed anomaly detection playbooks.

5.1. Metrics for Fraud Detection Performance

The performance of fraud detection systems must be evaluated with care due to the critical economic, reputational, and regulatory consequences of misclassifications. In applications like credit card, insurance, or online transaction applications, predicting negatives is much easier than predicting positives. The quantity of negative cases on a typical labeled dataset is usually many orders of magnitude larger than the number of positives. Even with large Datasets, Fraud models tend to see only a fraction of these negative cases; therefore, high true-negative prediction rates are usually not that insightful. This is especially problematic when estimating performance using cross-validation. For many individuals, it can be thought of as shooting fish in a barrel, but for the target group, especially during the timeframe of those highlighted events, those fishes can be a prized catch, bringing fraudulent return. False positives also carry some economic cost, but for these individuals, the targeting is inversely related to the probability that they are indeed a backdoor suspect.

The detection of fraudulent online activities can tackle four key areas: detection of blackhat SEO, click-fraud detection, spam detection, and bot detection. For detecting fraudulent online activities, specialized datasets spanned over the four key areas are collected and Multi-class Cross Domain Embedded Costume–based-paradigm is employed. The context information of the data risk domains is utilized to realize direct supervision to address the challenge of multi-class imbalance. The outcomes are evaluated in the light of area under the ROC curve (AUC), precision, recall, F1-score, and Cohen's kappa and utilized for analyzing detection across domains.

5.2. Datasets, Standards, and Benchmarking Platforms

The evaluation and validation of fraud detection algorithms remain as challenging for practitioners as they are for researchers. Data are often proprietary; hence, researchers do not have access to the many tagged examples that would permit central authorities to define necessary performance benchmarks. Decision-makers, for their part, may be engaged in a criminal financial activities themselves. Efforts to create standard datasets that would enable better validation and systematic comparisons remain few.

GAIA, a GAIA Data Curation & Enrichment Framework styled clustering-centric data store, is mentioned in the data source discussion; relevant naturalistic dataset is also referenced there. SNAP contains datasets for general-propose anomaly detection. DeblurFAS is relevant for face anti-spoofing, supporting learning and testing both on real and simulated suspended data. CVC-FaceLive 3D, part of CVC-Face, can be used to train detection schemes in a supervised way. Speaking more generally, the AMIRAD-D dataset is specifically designed to allow the testing of the face anti-spoofing problem in both the seen and unseen settings. MNIST, while not directly application-centric, serves for evaluation/verification purposes.

Fig 3: An approach to benchmark fraud detection

IJIREEICE

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

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

Vol. 8, Issue 12, December 2020

DOI 10.17148/IJIREEICE.2020.81213

VI. GOVERNANCE, RISK MANAGEMENT, AND OPERATIONALIZATION

A successful strategy for fraud detection requires not only the quality of algorithmic detection of fraud events but also the associated governance and operational processes. These processes may be defined by, but are not limited to, the deployment of fraud playbooks, model monitoring and decisioning frameworks, and the infrastructure for fraud detection performance management.

Fraud detection is not a standalone activity. The fraud detection results influence business decisions within the enterprise, both directly and indirectly. Such decisions are either automated or elevated to decision-makers within the enterprise. As such, each major type of fraud detection activity can be associated with a fraud playbook, a repository that contains the various dimensions of the fraud detection cycle specific to that type of fraud detection, linking detection to action. The fraud playbook contains sections on detection, model development, model monitoring, and decisioning frameworks.

6.1. Fraud Playbooks and Decisioning Frameworks

Fraud detection models inform business stakeholders of potential fraud when transactions with sufficient confidence satisfy a detected fraud risk. These outputs are supported by technical fraud playbooks—decisioning frameworks that define risk levels, probabilities, characteristics, conditions, actions, and rationales—structured for business interaction. The playbooks define the nature and extent of fraud in business terms, enabling sensors to deliver risk alerts and detect limitations. Based on playbook specifications, a filtering model identifies cost-effective alerts based on business actions. Alerts passing the filter are routed for action assignment and operational response management.

Operational response plays are closely aligned with pre-integration business action requirements. Each play defines roles and responsibilities for enacting the response to underlying business needs. The plays represent the course of action and are designed to address alerts with other activated response pathways. New plays can be defined, based on operational management requirements, and form part of operational test/dev frameworks, enabling proactive identification and resolution of grouping issues within alerts.

Equation 3: Cohen's kappa: derived step-by-step

$$p_o = \frac{TP + TN}{N}$$

Let:

- True positive rate in data: $p_{y=1} = \frac{TP+FN}{N}$
- True negative rate in data: $p_{y=0} = \frac{TN+FP}{N}$
- Predicted positive rate: $p_{\hat{y}=1} = \frac{TP+FP}{N}$
- Predicted negative rate: $p_{\hat{y}=0} = \frac{TN+FN}{N}$

Chance agreement:

$$p_e = p_{y=1}p_{\hat{y}=1} + p_{y=0}p_{\hat{y}=0}$$

$$\boxed{\kappa = \frac{p_o - p_e}{1 - p_e}}$$

6.2. Model Deployment, Monitoring, and Explainability Model deployment, monitoring, and explainability are crucial aspects in the operationalization of big data-driven fraud detection systems in digital banking platforms. The development of a fraud detection model is usually just one piece in a complex ecosystem of people, processes, and technology designed to minimize the occurrence of financial fraud while limiting the impact on legitimate customers. To realize business value, success means running and embedding the model into operations in a way that fraud detection decisions are made quickly, accurately, and consistently for every transaction. However, even the best fraud detection models suffer from the problem of spurious correlations that can remain undetected for several months until a fraud playbook (decision framework) is created for a specific model output. A decision framework determines when to make a high-cost decision based on the output score of the fraud detection model.

Once deployed, models must be monitored to uncover any underlying frauds that may not have been detected. Common reasons for the operational failure of production models include model drift in the statistical properties of the input data

IJIREEICE
ISSN (Online) 2321-2004
ISSN (Print) 2321-5526
International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering
Vol. 8, Issue 12, December 2020
DOI 10.17148/IJIREEICE.2020.81213

(distribution shift), model degradation (loss of predictive accuracy), concept drift (change in the relationship between the input data and the predicted output), environment changes or other new factors present in the real-world data, and unknown unknowns (changes in the environment that have not yet been included in the data). Auditability, interpretability, and explainability are burgeoning terms frequently used by business stakeholders and regulatory bodies for evaluating machine-learning models (particularly black-box models such as neural networks). Auditability refers to the design of models such that their operation can be subjected to review and control. Monitoring focuses on the observation of specific metrics that indicate the risk level of using the model. Explainability is concerned about the extent to which a human could understand why the model took a specific decision.

VII. IMPACTS ON DIGITAL BANKING ECOSYSTEMS

Digital banking fraud detection directly affects customers, banks, insurance firms, and the entire ecosystem. Losses from undetected fraud incidents reduce customer funds, while investigating suspected fraud events raises operational costs. Research indicates that a single suppression can save operating costs exceeding the transaction amount. Consequently, customers' willingness to pay depends more on the performance of the fraud detection system than the risk of a false positive. Strong performance in these playbooks not only supports business goals but also enhances customer trust.

The growing number of systemic online fraud cases has resulted in substantial losses for banks' insurers. Although banks generally do not incur direct losses from fraud, they still factor insurance costs into customer loans. Consequently, higher reported fraud losses exert upward pressure on bank financing costs, prompting banks to take action and maintain insurance claims within reasonable levels. Big data algorithms and playbooks that reduce fraud detection-related losses will also lessen the financial burden on insurers, ultimately benefitting customers and the whole economy.

7.1. Customer Experience and Trust

The capability of fraud detection systems to achieve a high detection rate with minimized False Positive Rate (FPR) is critically important for a good customer experience. Despite their considerable impact on customer satisfaction and business reputation, fraud detection systems are not able to demonstrate the same level of consideration for customer experience that other platforms like ChatGPT do. It has been shown that fraud detection systems exert significant influence on customer churn and that among customers who leave the platform, a higher proportion is dissatisfied with the ML-based fraud detection compared to the business-as-usual policy. Not only automated systems suffer from this effect, but also human analysts often create troubles for legitimate customers. Specifically, empirically-based decision-making frameworks for detected fraud attempts from third parties allow for better customer experience than rules with high adherence. The lack of a satisfactory experience reflects on customers' perceived trustworthiness of the bank system. Lessons from the AutoFraud project exemplarily demonstrate that trustworthy fraud detection systems contribute positively to customer perceived trust toward the digital banking ecosystem. Thus, trustworthiness should be considered when devising fraud detection, prevention, and risk management playbooks.

As financial fraud, in all its forms, is today considered one of the major impediments to the actual development of digital banking ecosystems, extensive reviews summarize the possible impacts on banking operations. On the one hand, the losses incurred by these criminal acts represent a significant decrement in the expected revenues; on the other hand, the effort must be supported in terms of personnel, tools, and image capital. Although the efforts to avoid being defrauded must be recognized as an operational cost, they remain hidden to the customers. When fraud events occur, especially when fraud is attempted but not successful, customers typically become dissatisfied with the institution, even if they are not legitimate victims.

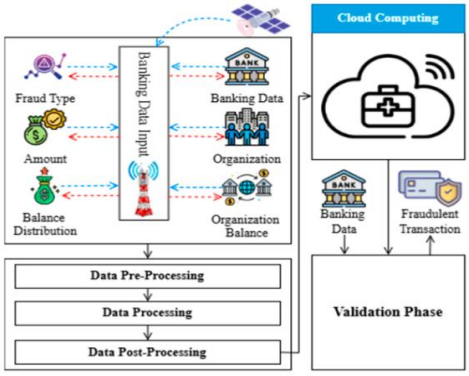


Fig 4: Customer Experience and Trust

IJIREEICE

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

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

Vol. 8, Issue 12, December 2020

DOI 10.17148/IJIREEICE.2020.81213

7.2. Economic Implications and Cost-Benefit Considerations

Economic considerations are a primary driver for fraud detection investments. Fraud detection must be risk-based and solution costs must be weighed against benefits. Businesses incur a proportion of losses from fraud, such as merchants absorbing card-not-present fraud losses.

For fraud incidents resulting in losses, the total cost, rather than just the loss amount, should be considered. Lost revenue represents the sum of future profit associated with the lost customer. Fraud detection should include the associated monitoring cost. Investments that surpass a fraud limit represent a poor business decision, resulting in lost revenue exceeding the detect-and-mitigate costs.

Conversely, ignoring fraud and risk management bears its own costs. Customers expect e-business to provide low fraud loss rates, which they link to the level of security provision. High vulnerability and unwillingness to pay for trustworthiness will discourage customer use. Therefore, developing suitable detection models is important for business.

VIII. CONCLUSION

The extensive use of the Internet, smartphones, tablets, and sensors in various areas of society has resulted in the generation of a large volume of data. The e-commerce business is constantly growing and merchants are increasingly accepting payments through the Internet. Nevertheless, the opportunities offered to the honest users of these online services are also tempting for malicious agents. Fraud, which is an illegal activity carried out with malicious intent, always occurs in society and manifests itself through various means, which can be financial, corporate, credit card, identity, etc. Banks and financial institutions are constantly in front of collecting, analyzing, and responding to the data generated. Fraud detection constitutes one of the more challenging elements. The quality of a bank's fraud-detection model can be associated with its survival and can reflect the institution's financial strength.

Since the 1970s, fraud detection mechanisms have been developed and used in banks with varying performances and applications. The adoption of Big Data in banks and payment providers has allowed the continuous and real-time honing of these fraud-detection models. However, although banks store and manipulate terabytes of data every day, many still do not have an ideal fraud-detection model that is operationally applied for all areas of a digital banking ecosystem. In recent studies, Big Data-based fraud detection has been systematically reviewed with a focus on external payment fraud, or partner fraud, in digital banking ecosystems.

8.1. Emerging Trends

Emerging trends affecting the digital banking ecosystem are reflected in five key areas: improved customer experience, evolving decisioning frameworks, risk-based governance, product evolution, and cost-benefit challenges and solutions.

First, continuous improvement in user experience remains a top priority given the competitive nature of the industry. In this regard, improved fraud detection and prevention mechanisms are especially important for enhancing customer satisfaction and trust. Decisioning frameworks evolve to incorporate advanced capabilities in real-time data integration, risk and fraud management, machine learning, and cognitive analytics, allowing banks to successfully balance customer experience and risk, deploying the best possible combination of trust versus friction for any set of interactions.

Second, with fraud ever-present in the digital arena, risk and fraud management are now regarded as fundamental elements of the digital banking experience. Fraud playbooks have emerged as repositories of decisioning rules, heuristics, and patterns pertaining to the operational handling of fraud risk across product lines and customer journeys. Built upon these structures, optimally configured integrated fraud risk frameworks enable a bank to maintain a robust digital experience in real-time while safeguarding stakeholder interests.

Third, as banks continually update and enhance their digital offers, a flexible fraud detection and prevention solution provides significant support. Such an infrastructure connects to various digital data sources, exposes online and batch integration capabilities, and evaluates rules, models, and playbooks as simple, configurable, modular services. Finally, preparing for the next major economic downturn is a focus of every bank. Globally, banks continue to add operational expenditure, but revenue growth is difficult. As a result, a strong focus on cost-benefit analysis and impact assessment has emerged, especially for new investments, with technology capital and operational spend carefully reviewed.

IJIREEICE

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

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

Vol. 8, Issue 12, December 2020

DOI 10.17148/IJIREEICE.2020.81213

REFERENCES

[1]. Ahmed, M., Mahmood, A. N., & Hu, J. (2016). A survey of network anomaly detection techniques. *Journal of Network and Computer Applications, 60*, 19–31.
[2]. Kummari, D. N. (2020). Machine Learning Applications inRegulatory Compliance Monitoring forIndustrial Operations. Global Research Development (GRD) ISSN: 2455-5703, 5(12), 75-95.
[3]. Baesens, B., Van Vlasselaer, V., & Verbeke, W. (2015). *Fraud analytics using descriptive, predictive, and social network techniques: A guide to data science for fraud detection*. Wiley.
[4]. Segireddy, A. R. (2020). Cloud Migration Strategies for High-Volume Financial Messaging Systems.
[5]. Bahnsen, A. C., Aouada, D., & Ottersten, B. (2016). Example-dependent cost-sensitive logistic regression for credit scoring. *Expert Systems with Applications, 42*(9), 4246–4256.
[6]. Vadisetty, R., Polamarasetti, A., Guntupalli, R., Rongali, S. K., Raghunath, V., Jyothi, V. K., & Kudithipudi, K. (2020). Generative AI for Cloud Infrastructure Automation. International Journal of Artificial Intelligence, Data Science, and Machine Learning, 1(3), 15-20.
[7]. Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. *Statistical Science, 17*(3), 235–255.
[8]. Keerthi Amistapuram , "Energy-Efficient System Design for High-Volume Insurance Applications in Cloud-Native Environments," International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering (IJIREEICE), DOI 10.17148/IJIREEICE.2020.81209
[9]. Bose, I., & Mahapatra, R. K. (2001). Business data mining—A machine learning perspective. *Information & Management, 39*(3), 211–225.
[10]. Integrating Big Data and AI in Cloud-Based Healthcare Systems for Enhanced Patient Care and Disease Management. (2020). Global Research Development(GRD) ISSN: 2455-5703, 5(12), 19-42. https://doi.org/10.70179/g32nmm07
[11]. Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J. (2000). LOF: Identifying density-based local outliers. *ACM SIGMOD Record, 29*(2), 93–104.
[12]. Annapareddy, V. N. (2020). Integrating Solar Infrastructure with Cloud Computing for Scalable Energy Solutions. Global Research Development (GRD) ISSN: 2455-5703, 5(12), 152-170.
[13]. Carcillo, F., Dal Pozzolo, A., Le Borgne, Y.-A., Caelen, O., Mazzer, Y., & Bontempi, G. (2019). Combining unsupervised and supervised learning in credit card fraud detection. *Information Sciences, 557*, 317–331. Chakilam, C., Koppolu, H. K. R., Chava, K. C., & Suura, S. R. (2020). Integrating Big Data and AI in Cloud-Based Healthcare Systems for Enhanced Patient Care and Disease Management. Global Research Development (GRD) ISSN: 2455-5703, 5(12), 19-42.
[14]. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys, 41*(3), 1–58.
[15]. Inala, R. Designing Scalable Technology Architectures for Customer Data in Group Insurance and Investment Platforms.
[16]. Dal Pozzolo, A., Caelen, O., Johnson, R. A., & Bontempi, G. (2015). Calibrating probability with undersampling for unbalanced classification. *2015 IEEE Symposium Series on Computational Intelligence*, 159–166.
[17]. Meda, R. (2020). Designing Self-Learning Agentic Systems for Dynamic Retail Supply Networks. Online Journal of Materials Science, 1(1), 1-20.
[18]. [Dal Pozzolo, A., Caelen, O., Le Borgne, Y.-A., Waterschoot, S., & Bontempi, G. (2014). Learned lessons in credit card fraud detection from a practitioner perspective. *Expert Systems with Applications, 41*(10), 4915–4928.
[19]. Meda, R. (2020). Data Engineering Architectures for Real-Time Quality Monitoring in Paint Production Lines. International Journal Of Engineering And Computer Sci
[20]. Duman, E., & Ozcelik, M. H. (2011). Detecting credit card fraud by genetic algorithm and scatter search. *Expert Systems with Applications, 38*(10), 13057–13063.
[21]. Rongali, S. K. (2020). Predictive Modeling and Machine Learning Frameworks for Early Disease Detection in Healthcare Data Systems. Current Research in Public Health, 1(1), 1-15.
[22]. 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.
[23]. Nuka, S. T. (2020). Predictive Modeling in Healthcare: Early Diagnosis and Patient Risk Profiling Using Machine Learning. Global Research Development (GRD) ISSN: 2455-5703, 5(12), 96-115.
[24]. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems, 27*, 2672–2680.
[25]. Dwaraka Nath Kummari, Srinivasa Rao Challa, "Big Data and Machine Learning in Fraud Detection for Public Sector Financial Systems," International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), DOI: 10.17148/IJARCCE.2020.91221

IJIREEICE

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

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

Vol. 8, Issue 12, December 2020

DOI 10.17148/IJIREEICE.2020.81213

[26]. Hand, D. J. (2009). Measuring classifier performance: A coherent alternative to the area under the ROC curve. *Machine Learning*, *77*(1), 103–123.
[27]. Inala, R. (2020). Building Foundational Data Products for Financial Services: A MDM-Based Approach to Customer, and Product Data Integration. Universal Journal of Finance and Economics, 1(1), 1-18.
[28]. Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*.
[29]. Gottimukkala, V. R. R. (2020). Energy-Efficient Design Patterns for Large-Scale Banking Applications Deployed on AWS Cloud. power, 9(12).
[30]. Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P.-E., He-Guelton, L., & Caelen, O. (2018). Sequence classification for credit-card fraud detection. *Expert Systems with Applications*, *100*, 234–245.
[31]. Burugulla, J. K. R. (2020). The Role of Cloud Computing in Scaling Secure Payment Infrastructures for Digital Finance. Global Research Development (GRD) ISSN: 2455-5703, 5(12).
[32]. Kirkos, E., Spathis, C., & Manolopoulos, Y. (2007). Data mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications*, *32*(4), 995–1003.
[33]. Varri, D. B. S. (2020). Automated Vulnerability Detection and Remediation Framework for Enterprise Databases. Available at SSRN 5774865.
[34]. Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. *International Conference on Learning Representations*.
[35]. Inala, R. (2020). Big Data-Driven Optimization of Retirement Solutions: Integrating Data Governance and AI for Secure Policy Management. Global Research Development (GRD) ISSN: 2455-5703, 5(12).
[36]. Kou, Y., Lu, C.-T., Sirwongwattana, S., & Huang, Y.-P. (2004). Survey of fraud detection techniques. *IEEE International Conference on Networking, Sensing and Control*, 749–754.
[37]. Pamisetty, V. (2020). Optimizing Unclaimed Property Management through Cloud-Enabled AI and Integrated IT Infrastructures. Universal Journal of Finance and Economics, 1(1), 1–20. Retrieved from https://www.scipublications.com/journal/index.php/ujfe/article/view/1338
[38]. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.
[39]. Meda, R. End-to-End Data Engineering for Demand Forecasting in Retail Manufacturing Ecosystems.
[40]. Li, J., Huang, K.-Y., Jin, J., & Shi, J. (2008). A survey of statistical methods for financial fraud detection. *In* C. Yi (Ed.), *Statistical fraud detection* (pp. 1–45). Springer.
[41]. Pamisetty, A. (2019). Big Data Engineering for Real-Time Inventory Optimization in Wholesale Distribution Networks. Available at SSRN 5267328.
[42]. Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. *IEEE International Conference on Computer Vision*, 2980–2988.
[43]. Gadi, A. L. The Role of Digital Twins in Automotive R&D for Rapid Prototyping and System Integration.
[44]. Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation Forest. *2008 IEEE International Conference on Data Mining*, 413–422.
[45]. Adusupalli, B., Singireddy, S., & Pandiri, L. Implementing Scalable Identity and Access Management Frameworks in Digital Insurance Platforms.
[46]. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*, 4765–4774.
[47]. Balaji Adusupalli, Sneha Singireddy, Lahari Pandiri, "Implementing Scalable Identity and Access Management Frameworks in Digital Insurance Platforms," International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), DOI: 10.17148/IJARCCE.2020.91224
[48]. McKinney, W. (2010). Data structures for statistical computing in Python. *Proceedings of the 9th Python in Science Conference*, 51–56.
[49]. Koppolu, H. K. R. Beyond the Bedside: Examining the Influence of Family-Integrated Care Practices on Patient Outcomes and Satisfaction in Diverse Clinical Settings.
[50]. Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review. *Decision Support Systems*, *50*(3), 559–569.
[51]. Preethish Nandan, B. (2020). Advanced Testing Frameworks for Next - Generation Semiconductor Devices Using Machine Learning. International Journal of Science and Research (IJSR), 1911–1920. https://doi.org/10.21275/sr20125160704
[52]. OECD. (2020). *Cybersecurity policy making at a turning point: Analysing the increasing role of states*. OECD Publishing.
[53]. Recharla, M. (2020). Targeted Gene Therapy for Spinal Muscular Atrophy: Advances in Delivery Mechanisms and Clinical Outcomes. International Journal of Science and Research (IJSR), 1921–1934. https://doi.org/10.21275/sr20126161624

IJIREEICE

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

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

Vol. 8, Issue 12, December 2020

DOI 10.17148/IJIREEICE.2020.81213

[54]. Phua, C., Lee, V., Smith, K., & Gayler, R. (2010). A comprehensive survey of data mining-based fraud detection research. *arXiv preprint arXiv:1009.6119*.

[55]. Pallav Kumar Kaulwar, "Designing Secure Data Pipelines for Regulatory Compliance in Cross-Border Tax Consulting," International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering (IJIREEICE), DOI 10.17148/IJIREEICE.2020.81208

[56]. Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M. P., Shyu, M.-L., Chen, S.-C., & Iyengar, S. S. (2018). A survey on deep learning: Algorithms, techniques, and applications. *ACM Computing Surveys, 51*(5), 1–36.

[57]. Balaji Adusupalli, Lahari Pandiri, Sneha Singireddy, "DevOps Enablement in Legacy Insurance Infrastructure for Agile Policy and Claims Deployment," International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering (IJIREEICE), DOI 10.17148/IJIREEICE.2019.71209

[58]. Quionero-Candela, J., Sugiyama, M., Schwaighofer, A., & Lawrence, N. D. (Eds.). (2009). *Dataset shift in machine learning*. MIT Press.

[59]. Machine Learning Applications inRegulatory Compliance Monitoring forIndustrial Operations. (2020). Global Research Development(GRD) ISSN: 2455-5703, 5(12), 75-95. https://doi.org/10.70179/tqqm2y82

[60]. Ravisankar, P., Ravi, V., Rao, G. R., & Bose, I. (2011). Detection of financial statement fraud and feature selection using data mining techniques. *Decision Support Systems, 50*(2), 491–500.

[61]. Nandan, B. P., Sheelam, G. K., & Engineer Sr, I. D. Data-Driven Design and Validation Techniques in Advanced Chip Engineering.

[62]. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144.