IDENTITY THEFT DETECTION IN CREDIT CARD APPLICATION FORMS USING MULTILAYER ALGORITHMS BASED ON DATA MINING TECHNIQUES

Mr. Amol Jagdish Shakadwipi[1]*, Dr. Dinesh Chandra Jain[2], Dr. S. Nagini[3]

Article History: **Received: 02/07/2022** **Revised: 10/08/2022** **Accepted: 11/09/2022**

Abstract:
Owing to the volatile nature of the market, the slowdown in economic growth, and the rapid upsurge of digital e-commerce, the issue of fraud has become increasingly prevalent. The continuous evolution of electronic commerce technology has led to a significant increase in credit card usage, establishing it as the favored method of payment for both online and offline transactions. Consequently, the rise in credit card fraud has become apparent, affecting individuals seeking smart cards and loans, who now have the choice of applying for credit cards through online channels or traditional paper forms. Unfortunately, these application processes have exposed instances of fraudulent activities, notably identity theft, which poses a significant concern for both credit cardholders and financial institutions. Dishonest actors are unlawfully obtaining customer's identities and gaining unauthorized access to credit cards, thereby posing substantial risks for both customers and financial entities.However, the existing strategies that rely on business rules and scorecards for fraud detection, without utilizing data mining, have shown limitations. In response to these shortcomings, this study introduces an innovative method for real-time fraud detection during the application phase. This method involves implementing a novel multi-layer fraud tracking system based on data mining algorithms. This system comprises two distinct algorithms, namely communal tracing and spike tracing, working together to improve the accuracy, speed, and efficiency of fraud detection procedures. By validating applications in real-time upon submission, this system acts as a robust deterrent against the approval of fraudulent credit card applications before issuance.

Keywords: Multi-layer fraud tracking system, Communal tracing, Spike tracing, Market volatility

[1]*Research Scholar, Department of Computer Science and Engineering, Oriental University, Indore, amolshakadwipi@gmail.com
Orcid Id : 0009-0009-1477-2005
[2]Professor, Department of Computer Science and Engineering, Oriental University, Indore,
Email: dineshjain25210@gmail.com, Orcid Id :0000-0001-8990-8127
[3]Professor, Department of Computer Science and Engineering, VNR Vignana Jyothi Institute of Engineering & Technology, Hydrabad,
[3]Department of Computer Science and Engineering, Oriental University, Indore, Email: nagini_s@vnrvjiet.in
Orcid Id: 0000-0002-5699-3801

***Corresponding Author:** Mr. Amol Jagdish Shakadwipi
*Research Scholar, Department of Computer Science and Engineering, Oriental University, Indore, amolshakadwipi@gmail.com
Orcid Id : 0009-0009-1477-2005

DOI: 10.53555/ecb/2022.11.10.46

Introduction :

In an age dominated by the swift progression of information technology and communication mediums, there has been a global surge in fraudulent activities. This has made it increasingly crucial for organizations to intensify their efforts in tracing fraud, given its significant impact on operational costs. Among these fraudulent practices, identity theft shines as a particularly concerning issue—a type of fraud where an individual assumes someone else's identity, using their personal information to gain unauthorized access to resources or obtain credit and benefits in the victim's name. The consequences can be severe if the victim is wrongly held accountable for the perpetrator's actions, potentially leading to harsh penalties. This deceitful act occurs when an individual illicitly uses someone else's identifiable information such as their name, social security number, or credit card details for fraudulent or illegal activities.

Identity fraud, particularly in credit applications, has risen due to the abundance of personal data on the internet and the vulnerabilities in traditional mail systems. Culprits can easily hide their true identities by exploiting both online and physical credit application processes, taking advantage of inadequate security measures or disposal methods. The impact of these fraudulent activities extends to national economies. To combat these deceitful practices and minimize financial losses, many businesses have adopted advanced analytical methods. Among these approaches, data mining has emerged as a powerful tool in uncovering fraudulent activities. By carefully analyzing extensive datasets, data mining reveals crucial patterns and insights, enabling the creation of predictive models. This complex procedure involves meticulous curation and exploration of data to reveal hidden trends and patterns. While not a new concept and already employed in credit scoring and fraud prevention by financial institutions, data mining encompasses numerous techniques for tracing fraud, including counterfeit detection in online credit card applications. This study aims to emphasize the effectiveness of data mining techniques in credit application fraud tracing systems, integrating spike tracing and communal tracing methodologies to enhance overall effectiveness.

Literature Survey :

In today's interconnected world, the internet has seamlessly woven itself into our daily lives, enabling a wide array of activities and collaborations through web-based platforms and services. This evolution has propelled the widespread adoption of web services, powering applications ranging from online banking, social networks, and cloud computing to e-commerce.

The escalating global prevalence of fraud, characterized by intentional deception for personal gain or inflicting harm on others, has amplified the urgency for effective tracing methodologies due to resultant increased operational costs. Identity theft, a prevalent form of fraud, involves perpetrators illicitly using stolen personal information to assume the identities of innocent individuals. This pilfered data might be authentically acquired from the victim (identity theft) or artificially crafted by adversaries to deceive systems (synthetic identity fraud). This issue extends to situations where unauthorized individuals exploit victims' financial details, such as withdrawing funds or applying for loans under false identities. Instances of identity theft often entail the use of stolen data or fraud committed by individuals known to the victim who have access to their financial documents.

In the realm of digital transactions, identity theft poses a serious concern, sparking the quest for innovative techniques to counter fraudulent activities. Credit card application forms have emerged as prime targets for identity theft, necessitating robust detection mechanisms to safeguard users and financial institutions. This survey of literature explores the research landscape concerning the detection of identity theft in credit card applications, particularly focusing on the application of data mining techniques, notably the use of multilayer algorithms to enhance accuracy and effectiveness.

The surge in digitalization and online transactions has introduced unprecedented convenience and efficiency but has concurrently created opportunities for malicious actors seeking to exploit vulnerabilities. The primary objective of this survey is to comprehensively explore various studies, methodologies, and advancements in detecting identity theft within credit card applications. Leveraging data mining, researchers and practitioners aim to unveil concealed patterns and irregularities in application forms, thereby enabling timely fraud prevention.

Scholars have recognized the severity of identity theft and its ramifications for both individuals and financial institutions. Traditional methods of fraud detection, such as rule-based systems, have exhibited limitations in adapting to evolving attack vectors. Consequently, the shift towards data mining techniques has gained prominence due to their ability to process vast volumes of data and extract meaningful insights.

The exploration commences with a retrospective analysis of seminal works addressing fraud detection and identity theft, particularly within the context of credit card application processes. Early endeavors aimed to establish foundational classification algorithms to differentiate between legitimate and fraudulent applications. As the field progressed, researchers ventured into multilayer algorithms, demonstrating superior performance by amalgamating multiple detection methods, each targeting distinct facets of identity theft. This layered approach contributes to a comprehensive and efficient detection mechanism.

Additionally, studies have underscored the significance of feature selection and data preprocessing to enhance the accuracy of detection models. Techniques such as ensemble learning and deep learning have surfaced, further refining detection systems by incorporating diverse information sources and capturing intricate patterns.

The integration of multilayer algorithms within the data mining framework has showcased promising outcomes in minimizing false positives and false negatives, thus bolstering the overall reliability of fraud detection systems. Extensive experiments conducted on real-world credit card application datasets validate the efficacy of these multilayer approaches in identifying suspicious activities indicative of identity theft.

While existing literature highlights significant progress, ample room remains for future research. The dynamic nature of fraud necessitates continuous adaptation and innovation in detection techniques. Moreover, the incorporation of explainable AI and ethical considerations surrounding user privacy and data security stand as areas warranting further exploration.

In conclusion, this literature survey underscores the escalating significance of data mining techniques, particularly multilayer algorithms, in detecting identity theft within credit card application forms. The collated studies collectively contribute to a deeper comprehension of the challenges, advancements, and potential avenues for further research in the critical domain of cybersecurity and fraud prevention.

Regarding the enumerated steps:

ii) Scrutinizing each application's link concerning the whitelist enables the identification of communal relationships and subsequently reduces their Link scores.
iii) Aggregating scores from previous applications with the current application's score, using link information and prior application scores, helps determine the present application's score.
iv) Modifying the value of a randomly selected parameter aims to strike a balance between efficiency and effectiveness, leading to the creation of a new whitelist based on prevailing Mini-discrete stream linkages.

Working and Architectural daigram :

The proposed system introduces an innovative approach comprising two distinct layers, known as Communal Tracing (CT) and Spike Tracing (ST), designed to bolster the security of credit card transactions within application processes. This research stands out for strategically integrating these dual data-mining layers to enhance the detection of fraudulent activities across multiple dimensions.

A. Communal Tracing:

The Communal Tracing layer implements a whitelist-driven methodology that utilizes a predefined set of attributes to differentiate authentic social connections and diminish suspicion scores, thereby fortifying resistance against tampering attempts involving synthetic social relationships [1]. The CT algorithm facilitates a comparison of all linkages with the whitelist, enabling the identification of communal associations and subsequently reducing their linkage scores. However, a potential limitation of CT lies in its attribute threshold, which necessitates at least three matching values within the dataset for detection, potentially missing instances where crucial values are duplicated by malicious entities. CT also considers attribute weights and conducts comparisons with previous applications within a moving window to compute the score of the current application. At the conclusion of each Mini-discrete data stream, a randomly selected parameter's value is adjusted systematically to maintain a balance between efficiency and effectiveness, resulting in a refreshed whitelist grounded in the current linkages.

The iterative steps involved in the CT algorithm are as follows:
i) Assess each application value against a record of preceding application values to unveil linkages.
ii) Examine each application's link concerning the whitelist, enabling the identification of communal relationships and subsequent reduction of their Link scores.
iii) Combine scores from previous applications with the current application's score, leveraging link information and prior application scores to ascertain the present application's score.
iv) Adjust the value of a randomly selected parameter to achieve equilibrium between efficiency and effectiveness, thereby generating a new whitelist based on the prevailing Mini-discrete stream linkages.

Communal Tracing Algorithm mathematical modeling :

Input:

- Dataset of attribute values for credit card applications
- Whitelist of predefined attributes for genuine social connections
- Attribute threshold value

Output:

- Link scores for identifying communal associations

1. Initialize attribute_match dictionary to store attribute match values.
2. For each attribute X_i in the current application:

a. Compare X_i against corresponding attributes in preceding applications.
b. Determine if there is a match with at least three matching values:

- If matched, attribute_match[X_i] = 1
- If not, attribute_match[X_i] = 0

3. Calculate scores based on communal associations:

- For each application, sum the attribute_match values to generate link scores.
- Reduce suspicion scores for communal associations identified in the whitelist.

4. Adjust attribute weights considering attribute importance and their impact on communal linkages.
5. At the end of each Mini-discrete data stream:

- Modify a random parameter's value to maintain balance between efficiency and effectiveness.
- Refresh the whitelist based on updated linkages and attribute values.

Through the integration of the Communal Tracing layer with the subsequent Spike Tracing layer, this research aims to establish a more comprehensive and adaptable mechanism for detecting fraudulent activities within credit card application processes.

In contrast to the Communal Tracing (CT) approach, Spike Tracing operates on a different principle, focusing on identifying spikes to elevate the suspicion score while safeguarding attributes from probing attempts. This strategy aims to reduce the risk of malicious actors accessing attributes crucial for computing the Spike Tracing (ST) score. Employing an attribute-oriented framework, ST adopts a selective nature, targeting attributes that strike a balance between sparsity and density. It computes the ST suspicion score and periodically eliminates redundant attributes to enhance efficiency.

B. Spike Tracing:

Spike Tracing Algorithm:

Input:

- Dataset of attribute values for credit card applications
- Threshold criteria for spike detection
- Attribute weights for computation

Output:

- Spike suspicion scores for identifying irregularities

1. Define function spikeDetected(value) to determine if a spike is present:

a. Calculate the difference between the current value and the average of preceding values.
b. If the difference exceeds the threshold, mark as a spike.

2. Initialize spike_scores array to store spike suspicion scores for each application.
3. For each application in the dataset:
4. Sequentially match each attribute value against preceding application values.

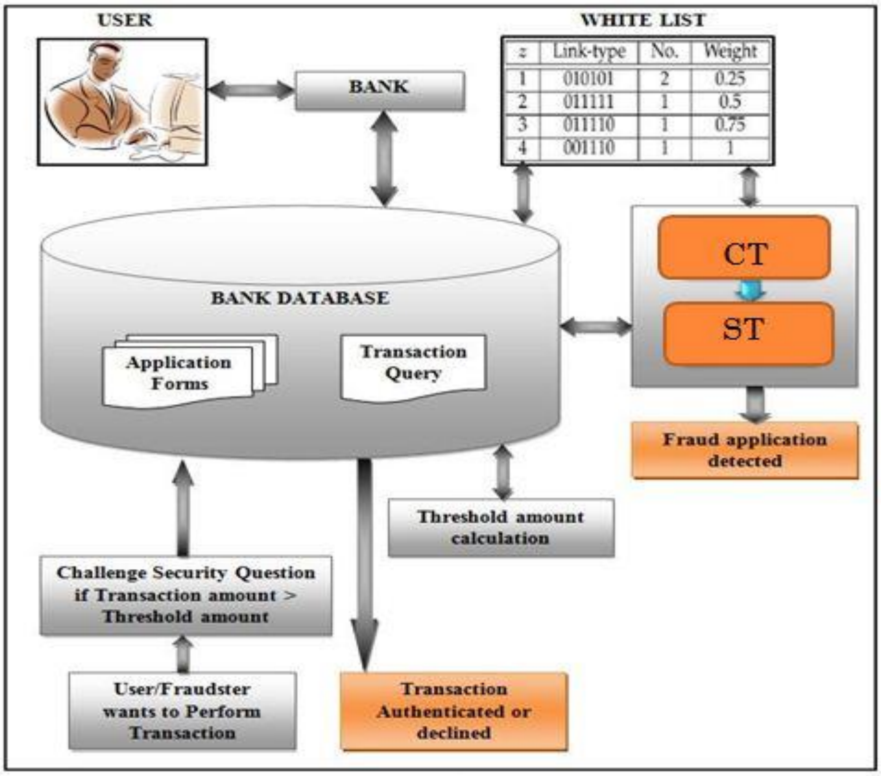
a. Employ spikeDetected() function to detect spikes in attribute values.
b. Calculate the spike suspicion score for the application based on the number of detected spikes.
c. Compute the application's score considering attribute weights.

4. Identify pivotal attributes influencing the ST suspicion score:

a. Analyze the impact of attributes on spike detection and score computation.

b. Refine attribute weights based on the significance of pivotal attributes.

The integration of insights garnered from Spike Tracing with the preceding Communal Tracing methodology aims to present an inclusive and adaptable framework for identifying fraudulent activities within credit card application protocols. This comprehensive approach intends to fortify fraud detection mechanisms and enhance the security of credit card transactions.


USER
BANK
WHITE LIST
BANK DATABASE
Application Forms
Transaction Query
CT
ST
Fraud application detected
Threshold amount calculation
Challenge Security Question if Transaction amount > Threshold amount
User/Fraudster wants to Perform Transaction
Transaction Authenticated or declined

z	Link-type	No.	Weight
1	010101	2	0.25
2	011111	1	0.5
3	011110	1	0.75
4	001110	1	1

Figure1.1. : Architectural diagram for credit card application form fraud detection system

The system comprises several key components:
Intuitive Graphical Interface: This facilitates users in entering credit card application details.
Data Acquisition and Storage: The system collects and stores user-provided data.
Linkage Comparison: It executes a comparison between incoming applications and existing ones in the bank's database, represented by a binary string indicating matched or unmatched attributes.
Initial White List Generation: This includes validated applications, linkage types, associated application counts, and corresponding weights.

Communal Tracing Steps:

a) The system compares a new application with the white list records.
b) The system proficiently identifies similarities among applications within the Communal Tracing (CT) layer.
c) CT allocates decreased suspicion if the new application matches four or more fields in the white list.
d) If no match occurs, the new application is integrated into the white list and updated.
e) The Suspicious Transaction (ST) layer receives the suspicion score assigned to the new application.

Spike Tracing Progression:

a) Within the ST layer, a meticulous evaluation of matched fields is conducted, giving priority to unique identification attributes.
b) Presence of unique IDs elevates suspicion scores, labeling the application as suspicious and leading to rejection.
c) Absence of matching unique IDs adds the application to the white list, prompting an update.

Additionally, the system calculates the transaction amount threshold based on the user's historical transactions.

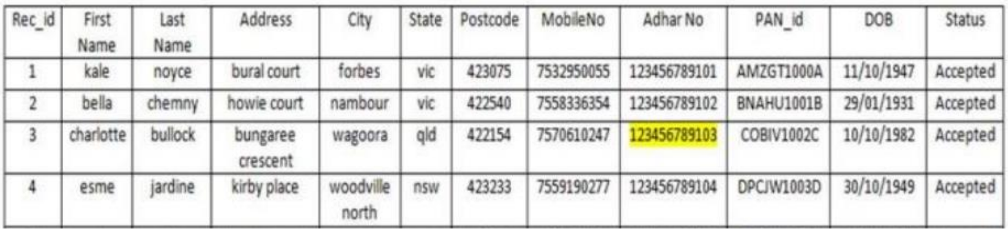
Results:

In an era of rapid e-commerce expansion, combatting fraud remains crucial. The present system employs proactive data mining techniques to thwart credit card fraud, shifting from reactive to preemptive measures. It dynamically selects mining techniques to refine accuracy, speed, and cost-effectiveness.

Validation of credit card applications involves challenge security questions to strengthen security for online transactions. Operationalizing the system utilizes a synthetic dataset of 50,000 credit applications, undergoing real-time scrutiny against user-initiated applications, contributing to the white list. The system provides CT and ST suspicious scores, amalgamated for unified evaluation.

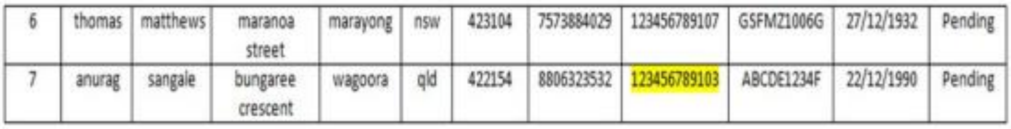
Dynamic updates in CT and ST layers reinforce the system's resilience and adaptability against potential fraud attacks. The real-time credit application fraud tracing mechanism is founded on data mining layers, establishing a methodical search for patterns during transactions.

Core principles include resilience, adaptivity, and data quality, aiming to fortify credit application security using cutting-edge data mining strategies.



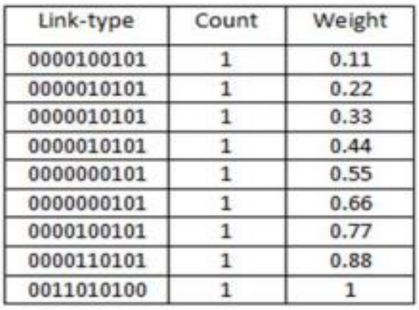
Rec_id	First Name	Last Name	Address	City	State	Postcode	MobileNo	Adhar No	PAN_id	DOB	Status
1	kale	noyce	bural court	forbes	vic	423075	7532950055	123456789101	AMZGT1000A	11/10/1947	Accepted
2	bella	chemny	howie court	nambour	vic	422540	7558336354	123456789102	BNAHU1001B	29/01/1931	Accepted
3	charlotte	bullock	bungaree crescent	wagoora	qld	422154	7570610247	123456789103	COBIV1002C	10/10/1982	Accepted
4	esme	jardine	kirby place	woodville north	nsw	423233	7559190277	123456789104	DPCJW1003D	30/10/1949	Accepted

Figure 1.2 Dataset of accepted list



6	thomas	matthews	maranoa street	marayong	nsw	423104	7573884029	123456789107	GSFMZ1006G	27/12/1932	Pending
7	anurag	sangale	bungaree crescent	wagoora	qld	422154	8806323532	123456789103	ABCDE1234F	22/12/1990	Pending

Figure 1.3 Input Dataset



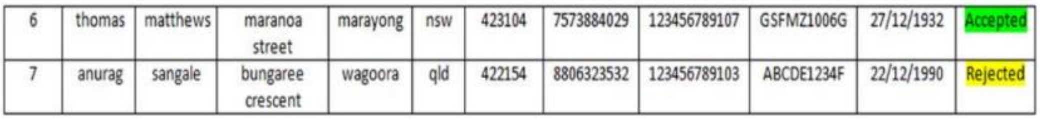
Link-type	Count	Weight
0000100101	1	0.11
0000010101	1	0.22
0000010101	1	0.33
0000010101	1	0.44
0000000101	1	0.55
0000000101	1	0.66
0000100101	1	0.77
0000110101	1	0.88
0011010100	1	1

Figure 1.4 White List

The illustration presented in Figure 1.4 showcases an example of the white-list generated from the credit applications outlined in Figures 1.2 and 1.3. The Spike detection algorithm's outcome generates the ST suspicious score. Both CT and ST scores combine to form a unified score. Importantly, the ST mechanism contributes to enhancing CT attribute weights.

The dataset comprises approved credit card applications, specifically a subset of five entries labeled as "accepted." These records consist of 12 attributes, notably including the Adhar card number and PAN ID, both linked to the "accepted" status.

Records numbered 6 and 7 represent the input dataset to the system, currently in a pending status. Figure 1.4 visualizes the White-List constructed by the CT algorithm during the processing of this Input dataset.



6	thomas	matthews	maranoa street	marayong	nsw	423104	7573884029	123456789107	GSFMZ1006G	27/12/1932	Accepted
7	anurag	sangale	bungaree crescent	wagoora	qld	422154	8806323532	123456789103	ABCDE1234F	22/12/1990	Rejected

Figure 1.5 Processed Applications

Upon employing the CT and ST algorithms, the system conducts computations to determine the link type and attribute weight. Subsequently, after the successful execution of these procedures, the record's status transitions from pending to either accepted or rejected.

Conclusion:

This project's primary aim is to develop a multi-tiered fraud detection framework tailored to identify fraudulent activities within credit applications. It leverages advanced data mining techniques, specifically the Communal Tracing (CT) and Spike Tracing (ST) layers. These integrated layers collaborate to ensure secure real-time transactions by detecting instances of fraudulent actions and authenticating social connections. Regular updates to both CT and ST layers aim to preempt potential attacks by fraudsters. By establishing a real-time credit application fraud tracing system anchored in data mining, this research seeks to execute systematic pattern searches in real-time, enhancing transactional security for credit applications.

A key focus of this study is on executing principled real-time pattern searches to actively safeguard credit applications during transactions. Guided by principles such as resilience (employing multiple protective layers), adaptivity (addressing evolving fraudulent and legal behavior), and data quality (real-time rectification of inaccuracies), this system emphasizes comprehensive, adaptable, and high-quality fraud detection methodologies.

References :

1. E.M.S.W. Balagolla et al., "B-Box.com: Blockchain-based secure credit card fraud prevention," in Proceedings of the 2021 IEEE 16th International Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS), pp. 63-68, 2021.
2. W. Wang et al., "A blockchain-based approach for identity crime detection in Byzantine environments," in Proceedings of the 2019 IEEE International Conference on Blockchain (Blockchain), pp. 570-576, 2019.
3. A. Asgaonkar and B. Krishnamachari, "A blockchain-based approach for securing credit card transactions," in Proceedings of the 2019 IEEE International Conference on Blockchain (Blockchain), pp. 682-689, 2019.
4. K. Vidhya and P. D. Kumar, "Fraud detection in credit card transactions using spike tracing algorithm," in Proceedings of the 2013 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), pp. 1-5, 2013.
5. A. Herenj and S. Mishra, "Communal detection and spike detection algorithms for secure credit card transaction mechanisms," in Proceedings of the 2013 IEEE International Conference on Data Mining and Advanced Computing (SAPIENCE), pp. 339-344, 2013.
6. S. K. K et al., "Comparative analysis of classification algorithms for fraud detection," in Proceedings of the 2012 IEEE International Conference on Data Science and Data Intensive Systems (DSDIS), pp. 327-332, 2012.
7. J. O. Awoyemi et al., "Anomaly-based detection of credit card fraud: A machine learning approach," International Journal of Data Analysis Techniques and Strategies, vol. 9, no. 3, pp. 297-314, 2017.
8. S. Wang, K. Zhao, and J. Zhou, "A comprehensive review of data mining-based fraud detection research," IETE Technical Review, vol. 27, no. 5, pp. 367-385, 2010.
9. G. Apparao, P. S. Kumar, and P. R. Prasad, "Fraudulent financial statement detection in the advanced big data environment," Procedia Computer Science, vol. 93, pp. 60-67, 2016.
10. D. Yue, J. Wu, and J. Huang, "Fraudulent statement of financial position detection: A comprehensive framework," Decision Support Systems, vol. 43, no. 1, pp. 272-282, 2007.
11. E. Kirkos, C. Spathis, and Y. Manolopoulos, "Data mining techniques for the detection of fraudulent financial statements," Expert Systems with Applications, vol. 32, no. 4, pp. 995-1003, 2007.