Enhancing Enterprise Financial Fraud Detection using Machine Learning

Mustafa Mohamed Ismail
Department of Computer Science, Al Majmaah University, Saudi Arabia
441104754@s.mu.edu.sa

Mohd Anul Haq
Department of Computer Science, Al Majmaah University, Saudi Arabia
m.anul@mu.edu.sa (corresponding author)

Received: 8 April 2024 | Revised: 24 April and 8 May 2024| Accepted: 12 May 2024

 DOI: https://doi.org/10.48084/etasr.7437

ABSTRACT

The present research aims to improve the detection of financial frauds in enterprises through the utilization of Artificial Intelligence (AI) methods. The proposed framework employs machine learning algorithms and data analytics to accurately identify patterns, anomalies, and signs of fraudulent activity. Exploratory data analysis approaches were employed to identify instances of missing values and imbalanced data. The random forest was based on its ability to consistently capture intricate patterns and efficiently tackle the multicollinearity problem. The isolation forest approach yielded an accuracy of 99.7%, while the local outlier factor method achieved an accuracy of 99.8%. Similarly, the random forest algorithm demonstrated an accuracy of 99.9%..

Keywords-financial fraud; anomaly detection; internal fraud; fraudulent behavior; formatting

I. INTRODUCTION

As the landscape of traditional financial activities undergoes a transformative change, technology is playing a pivotal role in facilitating a shift toward computer-centric financial services. This shift marks a significant transition for financial institutions, which are moving away from a human-centered model to one that is increasingly reliant on computerized systems. The advent of technology has enabled financial institutions to streamline their operations, improve efficiency, and reduce costs while simultaneously providing customers with a more seamless and personalized experience. The adoption of technology in financial services has become imperative for institutions that seek to remain competitive in the modern digital age [1].

Enterprise fraud is a widespread danger in various industries, significantly affecting firms' financial health, reputation, and operational strength. This complex type of fraud involves a variety of deceitful actions conducted within or directed at a firm, encompassing not only employees, but also contractors, customers, and external individuals. Enterprise fraud has serious consequences, including substantial financial losses, compromised data security, regulatory violations, and a decline in customer trust. Companies can lose up to 5% of their annual income due to fraudulent actions. Traditional detection approaches for occupational fraud, insider threats, and financial crime are frequently insufficient. These strategies concentrate on specific aspects of organizational data and specialize in particular types of fraud, leading to delayed identification of manipulative behaviors that could endanger the organization's survival. To effectively combat enterprise fraud, a comprehensive approach is needed that combines advanced technology, strong internal controls, ongoing monitoring, and proactive risk management..

Advanced analytics and Machine Learning (ML) play a crucial role in combating fraud. These technologies allow analyzing many data sources, such as financial transactions, employee records, customer behavior, and communication networks. They reveal hidden links and detect suspicious actions immediately. It is becoming increasingly important to prevent fraud schemes from developing, which requires a proactive defense. Even the most fundamental paradigms informing financial regulation now need to be rethought [2]. One of the biggest challenges currently facing the financial industry is the integration of Artificial Intelligence (AI) into existing systems and processes of financial institutions [3].

II. RELATED WORK

The integration of AI with financial fraud detection has been the subject of considerable research and development. In recent years, regulatory authorities have increased their focus on detecting and preventing financial crimes, see Table I. This is particularly important for banks and financial institutions, as they have multiple business processes that are vulnerable to fraud [4] and procedures that are a good source for criminal actors [5]. The related work on this issue is extensive, demonstrating a wide range of techniques, procedures, and technology targeted at improving the identification of

fraudulent activity at the company level. Scholars and practitioners have investigated numerous elements of financial data analysis, including how AI might be used to detect aberrant patterns and questionable transactions. The importance of preprocessing and feature engineering in extracting useful insights from financial information is emphasized in the literature. Numerous studies have shown that using powerful ML techniques to discriminate between genuine and fraudulent transactions might help to strengthen business security procedures.

Authors in [6] explored the transformative role of AI and ML in the finance industry, focusing on their applications in intelligent advising, lending, monitoring, and customer service. [6]. Authors in [7] studied the rising occurrence of financial fraud and its significant cost to institutions and consumers. Fraudsters' expertise in exploiting flaws in existing preventative mechanisms are utilized across multiple fields, including credit cards, insurance, money laundering, stock and commodities fraud, and insider trading. The study highlights the limits of fraud prevention systems and emphasizes the rising importance of powerful fraud detection systems. The authors methodically evaluate anomaly detection strategies, stressing the progression from mostly supervised learning algorithms to the addition of semi-supervised and unsupervised learning models to overcome supervised learning difficulties. The survey provides an in-depth analysis of the most efficient methods for identifying anomalies in financial fraud, with a particular emphasis on recent advances in semi-supervised and unsupervised learning, contributing valuable insights to the ongoing debate on improving financial fraud detection. Authors in [5] used AI and ML to detect money laundering financial transactions.

To identify suspicious activity, supervised and unsupervised learning methodologies, classification models, and anomaly detection tecniques are emploeyd. In [8], the accuracy, precision, recall, and F1 score of several classifiers, including Artificial Neural Networks (ANNs), Random Forest (RF), and Logistic Regression (LR) are examined. The study emphasizes the potential for combining ML approaches with feature engineering to increase financial crime detection rates, and it offers further studies, such as investigating consumer attributes and creating intelligent feature engineering. The goal is to design a model using annual financial statements to detect the likelihood of significant financial irregularities in enterprises. The research compares the financial records of 54 corporations renowned for accounting problems in the twentieth century against those of 58 similar "honest" enterprises. LR, linear discriminant analysis, deep ANNs, Naive Bayes models, RF, and gradient-boosted Decision Trees (DTs) were among the ML and AI technologies used. The results suggested that the gradient-boosted DTs and the RF were the most successful algorithms. Authors in [9] employed SVM, DT, and CHAID to identify financial crimes by building an effective detection model. According to the findings, the C5.0-SVM model yielded the greatest outcomes. The C5.0-SVM model had the highest accuracy rate of 83.15%, followed by SVM-SVM whose accuracy was 81.91%, C5.0-CHAID with an accuracy of 80.93%, and SVM-CHAID with 77.16% accuracy.

Authors in [10] explore the impact of AI on financial services, highlighting its potential in risk assessment, stock trading, and credit lending. It highlights its benefits, including improved loan services, simplified stock trading, and enhanced fraud detection. Authors in [11] examined the issue of detecting financial frauds. Traditional solutions often use rule-based algorithms or manual feature extraction, but these fail to account for the extensive Multiview network interactions that exist among financial service consumers. SemiGNN, is introduced, a semi-supervised attentive graph neural network that detects fraud using both labeled and unlabeled data. To grasp the relationships between distinct neighbors and perspectives, a novel hierarchical attention mechanism is used, which makes the model's judgments interpretable. The algorithm exhibited encouraging results when tested on Alipay, a prominent Chinese payment site. SemiGNN outperformed state-of-the-art algorithms for user default prediction, with an AUC of 0.807 and a KS of 0.464. The findings underscore the significance of certain applications, nicknames, and addresses in generating these predictions.

The continuation of Financial Statement Fraud (FSF) poses significant concerns to global capital market stability [12]. Many predictive and investigative strategies have been investigated to combat FSF, but their practical implementation remains difficult because of the complexities of real-world settings. Research employing 18 financial datasets depicting a fraud triangle used both supervised and unstructured approaches to identify FSF in China's stock market data. This study fully examined the following supervised approaches: multi-layer feed-forward ANN, probabilistic ANN, SVM, multinomial log-linear model, and discriminant analysis. MFFNN was particularly good in detecting fake financial statement data, according to empirical studies. The results demonstrated that MFFNN was very effective at detecting false financial statement data. This strategy integrated supervised and unsupervised methods, providing a practical way for building prediction models that are adaptive across different financial statement datasets. Authors in [13] delved into the creation of a fraud detection model spanning a decade (2004-2014) with data from 160 Taiwanese firms in the quest for a sustainable financial market.

In the current paper, we make several significant contributions to the field of financial fraud detection. Firstly, we conduct a comprehensive comparative analysis of three prominent algorithms—RF, Isolation Forest (IF), and Local Outlier Factor (LOF), providing valuable insights into their respective strengths and weaknesses. Secondly, we emphasize the importance of considering not only accuracy, but also other performance metrics, such as computational efficiency and interpretability when choosing a fraud detection method suitable for enterprise applications. Thirdly, our study highlights the potential of ensemble approaches and advanced ML techniques for enhancing fraud detection capabilities. Lastly, the implications of our findings for future research and practical applications are discussed, paving the way for more robust and adaptive fraud detection systems in the financial sector.

TABLE I. SUMMARY OF THE LITERATURE REVIEW

Ref.	Method	Application	Dataset	Evaluation metric	Limitation
[6]	AI, ML	Financial System Transformation	Not specified	Not specified	Broad perspective lacks detailed analysis, specific case studies
[7]	Anomaly Detection: SL, SSL, USL	Credit cards, insurance, money laundering, stock and commodities fraud	General financial sectors	Anomaly detection	Need for unsupervised learning due to limitations of supervised learning
[5]	ANNs, RF, LR	Money Laundering Detection	Banking and non-banking stats, regulatory requirements, professional insights	Accuracy, Precision, etc.	Not mentioned
[8]	LR, DNN, Naive Bayes, RF, Gradient Boosted Trees	Identifying unfair corporate culture in companies	Financial records of 112 companies	RF and Gradient Boosted Trees were top performers	Not mentioned
[9]	SVM, DTs, CHAID	Financial Reporting Fraud Detection	Data from Taiwan Economic Journal from 2007 to 2016	Accuracy, C5.0-SVM: 83.15%, SVM-SVM: 81.91%	Emphasis on using financial and non-financial information
[10]	AI	Impact on Financial Services	Secondary sources and surveys with experts	Not specified	Not mentioned
[11]	Semi-supervised Attentive Graph ANN	Financial Fraud Detection	Alipay user data	AUC: 0.807, KS: 0.464, precision varied, Top 1% Precision	Traditional methods often overlook extensive multiview network interactions
[12]	MFFNN, supervised and unsupervised methods	Financial Statement Fraud Detection	18 financial datasets from China's stock market	Not specified	Complexity of real-world settings affects practical implementation
[13]	ANNs, SVM, CART, CHAID, C5.0, QUEST	Financial Statement Fraud Detection	Data from 160 Taiwanese firms (2004-2014)	Accuracy: Over 90% (for the combination of ANN and CART)	Not specified
[14]	LSTM, RNNs	Financial Statement Fraud Detection	Data from 153 TWSE/TPEx listed firms from 2001-2019	Accuracy LSTM: 94.88%, RNN: 87.18%	Not mentioned
Present study	RF, IF, LOF	Financial Fraud Detection	PaySim contains 11 attributes and 6362620 records	Accuracy: Random Forest (99.95%) Isolation Forest (99.82%) Local Outlier Factor (99.75%)	More adaptability may be required for upcoming fraud patterns

III. METHODOLOGY

Detecting frauds in real-time is a significant challenge due to the constantly changing common and fraudulent behaviors. Manual fraud detection techniques are often inaccurate and require significant time and resources to identify patterns. In the current study, ML models were used for fraud detection. This method is designed to provide a more complete understanding of a research problem by leveraging the strengths of all the considered methods. Figure 1 illustrates the implementation process for the methodology.

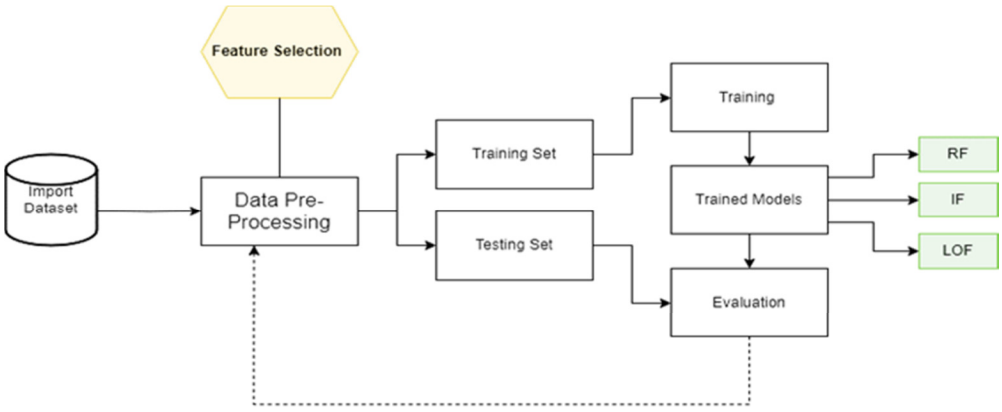

Fig. 1. Flowchart of the proposed methodology.

A. Model Building

For real-time fraud detection, three ML algorithms were developed and implemented

1) Random Forest

RF is an ensemble classifier that includes multiple DTs, each of which maintains the same distribution across the ensemble by utilizing a random vector. This mitigates the individual overfitting tendencies of each tree [15]. The process of DT creation can be improved by merging predictions from multiple trees. This technique is particularly useful for datasets with a high number of dimensions, as it eliminates the need for complex methods such as dimensionality reduction or feature selection. Additionally, it allows for faster training and simplifies parallel model usage.

2) Isolation Forest

IF is unique because it focuses on isolating unusual or rare data points, which are often indicators of fraudulent activities. By creating random splits in the data, it can quickly identify these anomalies and separate them from the majority of legitimate transactions. This approach proved effective in our study, showing strong performance in identifying fraudulent transactions amidst a sea of normal ones. Its ability to pinpoint outliers makes it a valuable tool in fraud detection systems.

IF and RF are both tree-based ML algorithms, but they serve different purposes. IF is designed for anomaly detection, focusing on isolating outliers by randomly partitioning the data. This makes it effective for spotting rare events like fraudulent transactions. On the other hand, RF is a supervised learning algorithm used for classification and regression tasks..

3) Local Outlier Factor

LOF is an unsupervised ML algorithm used for detecting outliers or anomalies in data. Unlike supervised methods that require labeled data, LOF works by comparing the density of points in the vicinity of each data point to determine its degree of "outlierness." Points with significantly lower density compared to their neighbors are considered outliers. In the context of financial fraud detection, LOF can be particularly useful for identifying subtle variations and anomalies in transaction data that may indicate fraudulent activities. The LOF algorithm can be divided into four parts:

- k-Distance and k-Neighbors: The hyperparameter k determines the number of neighbors to consider, determining the distance between observations. A small value increases noise sensitivity, while a large value may not recognize local anomalies.
- Reachability Distance: Reachability Distance (a, b) is the maximum distance between two points, using Euclid, Minkowski, Manhattan, and other distance measures.
- Local Reachability Density: Reachability distances are calculated for all k closest neighbors of a point, and the values are summed and divided by k. The inverse of this value is used to calculate the desired local accessibility density.
- LOF Calculation: The local reachability densities a are compared to the local reachability densities of its nearest k neighbors, and the density of each neighbor is summed and divided by a's density.

IV. DATA DESCRIPTION

Obtaining real financial data for any company can be difficult due to the private nature of the financial transactions. This makes it difficult for researchers to obtain publicly available datasets. The dataset employed for training and testing the models is Lopez-Rojas's PaySim dataset on Kaggle [16]. It contains 11 attributes and 6362620 records. Table II shows a sample of the dataset with some of its attributes.

TABLE II. DATASET SAMPLE

Step	Type	Amount	nameOrig	oldbalanceOrg	newbalanceOrig	nameDest	OldbalanceDest
1	PAYMENT	9839.64	C1231006815	170136	160296.4	M1979787155	0
1	PAYMENT	1864.28	C1666544295	21249	19384.72	M2044282225	0
1	PAYMENT	181	C1305486145	181	0	C553264065	0
1	PAYMENT	181	C840083671	181	0	C38997010	21182
1	PAYMENT	11668.14	C2048537720	41554	29885.86	M1230701703	0
1	PAYMENT	7817.71	C90045638	53860	46042.29	M573487274	0
1	PAYMENT	7107.77	C154988899	183195	176087.2	M408069119	0
1	PAYMENT	7861.64	C1912850431	176087.2	168225.6	M633326333	0
1	PAYMENT	4024.36	C1265012928	2671	0	M1176932104	0

- **Step**: maps a unit of time in the real world. In this case 1 step is 1 hour of time, with 744 total steps (30 days simulation).
- **Type**: CASH-IN, CASH-OUT, DEBIT, PAYMENT and TRANSFER.
- **Amount**: amount of the transaction in local currency.
- **NameOrig**: customer who started the transaction.
- **oldbalanceOrg**: initial balance before the transaction
- **newbalanceOrig**: new balance after the transaction.
- **nameDest**: customer who is the recipient of the transaction.
- **oldbalanceDest:** initial balance recipient before the transaction. Note that there is not information for customers that start with M (Merchants).
- **newbalanceDest**: new balance recipient after the transaction. Again, there is no information for customers that start with M (Merchants).

- **isFraud**: These are the transactions made by the fraudulent agents inside the simulation. In this specific data set the fraudulent behavior of the agents aims to profit by taking control of our customers' accounts and try to empty the funds by transferring to another account and then cashing out of the system.
- **isFlaggedFraud**: any attempt to transfer more than 200.000 in a single transaction is flagged as illegal.

A. Data Preprocessing and EDA

Combining the regular preprocessing techniques with the Exploratory Data Analysis (EDA) represents a valuable step in the current financial fraud detection analysis. Data preprocessing is essential for preparing data for ML models, including fraud detection. It involves cleaning the data by handling missing values and removing outliers to ensure accuracy. Additionally, features like scaling and normalization help standardizing data, making sure all parts contribute equally to the model. Encoding categorical data and feature engineering transform non-numerical information into a usable format for the algorithms. Overall, proper data preprocessing improves the model's performance by highlighting patterns and making it easier to distinguish between normal and fraudulent transactions.

1) Data Exploration

EDA is essential in preparing financial transaction datasets for fraud detection analysis. It provides insights into data characteristics, detects anomalies, guides feature selection and engineering, assesses data quality, and informs modeling decisions. By exploring the dataset, the specific time frame when the fraud occurred and determining the maximum amount stolen by the fraudster can be analyzed. Figure 2 shows the count of the types of transaction in the dataset (i.e. payment, transfer, cash out, debit, and cash in). Figure 3 shows the most common fraud types used by fraudsters. The correlation among several features is shown in Figure 4, which offers valuable insights into the interconnections among features.

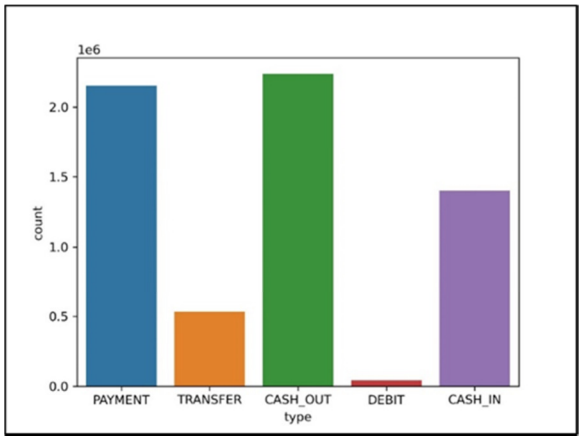

Fig. 2. Transaction types.

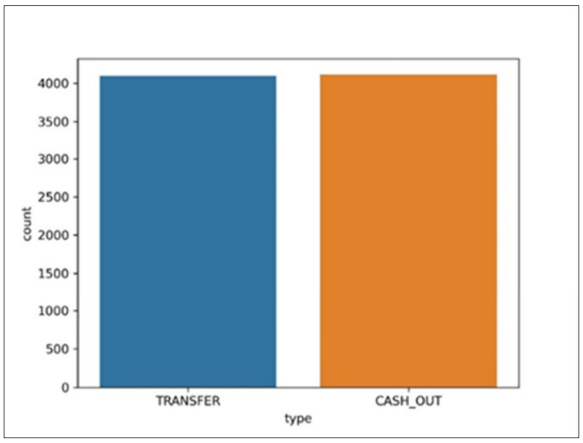

Fig. 3. Most common types used by fraudsters.

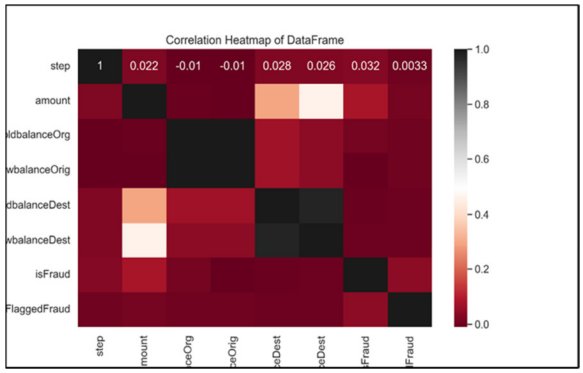

Fig. 4. Correlation between features.

2) Data Cleaning

An important aspect of the data preprocessing phase is to clean the data and eliminate any anomalies that might impede the efficacy of subsequent analysis. It includes handling missing values, and removing duplicates, or outliers to ensure the integrity and reliability of the dataset. Such measures allow obtaining optimum accuracy in the modeling phase, enhancing the robustness of the fraud detection mechanisms.

- Handling Missing Values: Impute missing values using techniques such as mean, median, or mode imputation, or consider more advanced methods like interpolation.
- Duplicate Data Removal: Duplicate records are identified and removed to ensure data integrity.
- Outlier Detection and Treatment: Statistical methods or ML algorithms are employed to identify outliers and decide whether to remove, transform, or treat them separately.

3) Data Balancing

There were challenges with the dataset that could influence the results if not resolved. The utmost apparent problem was the very unbalanced classes as of the 6362620 transactions, 6362620 (99.8709%) were legitimate and only 8213 (0.1291%) were fraudulent. These imbalanced class proportions could lead to

inaccurate predictions due to bias towards the abundant legitimate transactions. Therefore, it is essential to employ techniques such as oversampling, under-sampling, or synthetic data generation to balance the classes. The present investigation used the SMOTE technique to address the class imbalance issue. SMOTE works by generating synthetic samples in the minority class, helping to balance the dataset and improve the model's ability to identify fraudulent activities [15].

4) Feature Extraction

The transaction features are the input of the classification models. Each transaction type is associated with several features. Transaction features may include the frequency of transactions and the value of each transaction. Such features may also include the destination account, time, origin's and destination's geographical location, amount, accumulated fund flow, and accumulated transaction amount, entity (person or company doing the transfer, average person), and type of transaction (cash, money transfer). All are recognized as critical for fraud detection and classification. Once the semantic features are extracted from the raw transaction data, they are fed into the classifier.

The main attributes of a transaction were considered. These features will make a vector for each transaction consisting of the type and the amount of the transaction, origin old balance, origin new balance, destination old balance, and destination new balance. For the current research, feature engineering included three steps:

- Feature Selection: selecting the most relevant features that have the potential to predict fraudulent activities.
- Encoding Categorical Variables: Converting categorical variables into numerical format by using various techniques such as one-hot encoding or label encoding.
- Data Transformation: such as normalization to scale numerical features to a similar range to prevent certain features from dominating others during model training.

In the current research, after the feature selection phase, only the most important features were included, namely, type of transaction, amount of transaction, origin old balance, origin new balance, destination old balance, and destination new balance.

5) Data Splitting

The dataset was split into three subsets: 70% for training, 15% for validation, and 15% for testing. The training set was used to train the model, the validation set to fine-tune the hyperparameters and monitor performance during training, and the test set for the final evaluation of the model's generalization capability. This division ensures the model is assessed on unseen data, reducing the risk of overfitting where the model memorizes the training data without generalizing well to new data.

V. RESULTS AND DISCUSSION

The current study examined three methods to detect financial fraud: RF, IF, and LOF. RF stood out with an impressive accuracy of 99.95%, demonstrating its excellence in distinguishing between genuine and fraudulent transactions. Following closely, the IF achieved an accuracy of 99.82%, making it effective in identifying unusual and potentially fraudulent activities. On the other hand, LOF reached an accuracy of 99.75%, indicating its capability to detect subtle anomalies in financial transactions. When selecting a fraud detection approach for businesses, it is crucial to consider factors beyond just accuracy. This includes evaluating the method's computational efficiency, ease of interpretation, and overall performance metrics. Table III offers further insights and implications regarding these methods. Collaborative efforts between researchers, industry experts, and regulators could also establish standardized frameworks for more effective and transparent fraud detection in the future.

The insights from this study can serve as a valuable reference for enhancing fraud detection systems in industries such as healthcare, e-commerce, and cybersecurity [24-29]. However, adapting the optimal model and parameters to each unique application area's characteristics and challenges may be necessary. Thus, while our focus is on financial fraud detection, the methodologies discussed offer broad guidelines for improving fraud detection across various sectors.

TABLE III. ELABORATION AND IMPLICATION OF MODELS

Model	Elaboration and Implications	Accuracy Score
RF	RF achieved an accuracy score exceeding 99.95%, showcasing its robustness in capturing complex financial patterns.	99.95%
IF	IF demonstrated an accuracy score of around 99.82%, excelling in isolating anomalies within financial datasets.	99.82%
LOF	LOF attained a strong accuracy score of approximately 99.75%, specializing in detecting local irregularities.	99.75%

VI. CONCLUSION

In pursuing strengthening of financial fraud detection within enterprise systems, the present investigation used three ML models, namely RF, IF, and LOF. Each model exhibited remarkable accuracy in distinguishing between fraudulent and non-fraudulent transactions, showcasing the potential for practical deployment in the complex landscape of financial security. The RF model, known for its ensemble learning

capabilities, demonstrated an exceptional accuracy score of approximately 99.95%. This underscores its robustness in handling intricate relationships within financial data, positioning it as a highly effective tool for identifying fraudulent activities. Similarly, the IF model, designed to isolate anomalies efficiently, achieved a commendable accuracy score of around 99.82%. Its ability to pinpoint outliers within the dataset makes it an asset for detecting irregularities indicative of fraudulent behavior. The LOF model, leveraging density-based techniques, achieved an accuracy score of approximately 99.75%.

The present investigation has demonstrated the effectiveness of RF, IF, and LOF models in detecting financial fraud, the future direction lies in continuous research, refinement, and adaptation of these models to counter emerging fraudulent tactics and evolving financial technology. As financial technology evolves, ongoing research and refinement of these models will be imperative to stay ahead of sophisticated fraudulent activities and safeguard the integrity of enterprise financial systems using advanced models including but not limited to [20-23]. The findings lay the groundwork for informed decision-making in selecting models that align with the unique requirements and challenges posed by the dynamic landscape with advanced ML/AI models in various areas [30-34].

REFERENCES

[1] R. Lacasse *et al*., "A digital tsunami: FinTech and crowdfunding," in *International Scientific Conference on Digital Intelligence*, 2016, pp. 1–5.

[2] D. W. Arner, J. Barberis, and R. P. Buckey, "FinTech, RegTech, and the Reconceptualization of Financial Regulation," *Northwestern Journal of International Law & Business*, vol. 37, no. 3, pp. 371–414, 2017.

[3] J. Truby, R. Brown, and A. Dahdal, "Banking on AI: mandating a proactive approach to AI regulation in the financial sector," *Law and Financial Markets Review*, vol. 14, no. 2, pp. 110–120, Apr. 2020, https://doi.org/10.1080/17521440.2020.1760454.

[4] A. Beheshti, B. Benatallah, and H. R. Motahari-Nezhad, "ProcessAtlas: A scalable and extensible platform for business process analytics," *Software: Practice and Experience*, vol. 48, no. 4, pp. 842–866, 2018, https://doi.org/10.1002/spe.2558.

[5] Z. Rouhollahi, "Towards Artificial Intelligence Enabled Financial Crime Detection." arXiv, May 23, 2021, https://doi.org/10.48550/arXiv.2105.10866.

[6] M. Xie, "Development of Artificial Intelligence and Effects on Financial System," *Journal of Physics: Conference Series*, vol. 1187, no. 3, Dec. 2019, Art. no. 032084, https://doi.org/10.1088/1742-6596/1187/3/032084.

[7] W. Hilal, S. A. Gadsden, and J. Yawney, "Financial Fraud: A Review of Anomaly Detection Techniques and Recent Advances," *Expert Systems with Applications*, vol. 193, May 2022, Art. no. 116429, https://doi.org/10.1016/j.eswa.2021.116429.

[8] J. Wyrobek, "Application of machine learning models and artificial intelligence to analyze annual financial statements to identify companies with unfair corporate culture," *Procedia Computer Science*, vol. 176, pp. 3037–3046, Jan. 2020, https://doi.org/10.1016/j.procs.2020.09.335.

[9] D.-J. Chi, C.-C. Chu, and D. Chen, "Applying Support Vector Machine, C5.0, and CHAID to the Detection of Financial Statements Frauds," in *15th International Conference on Intelligent Computing Methodologies*, Nanchang, China, Aug. 2019, pp. 327–336, https://doi.org/10.1007/978-3-030-26766-7_30.

[10] R. Pothumsetty, "Implementation of Artificial Intelligence and Machine learning in Financial services," *International Research Journal of Engineering and Technology*, vol. 7, no. 3, pp. 3186–3193, 2020.

[11] P. H. dos Santos Teixeira and R. L. Milidiu, "Data stream anomaly detection through principal subspace tracking," in *ACM Symposium on Applied Computing*, Sierre, Switzerland, Mar. 2010, pp. 1609–1616, https://doi.org/10.1145/1774088.1774434.

[12] D. Wang *et al.*, "A Semi-Supervised Graph Attentive Network for Financial Fraud Detection," in *International Conference on Data Mining*, Beijing, China, Nov. 2019, pp. 598–607, https://doi.org/10.1109/ICDM.2019.00070.

[13] M. Omidi, Q. Min, V. Moradinaftchali, and M. Piri, "The Efficacy of Predictive Methods in Financial Statement Fraud," *Discrete Dynamics in Nature and Society*, vol. 2019, May 2019, Art. no. e4989140, https://doi.org/10.1155/2019/4989140.

[14] C. Jan, "An Effective Financial Statements Fraud Detection Model for the Sustainable Development of Financial Markets: Evidence from Taiwan," *Sustainability*, vol. 10, no. 2, Feb. 2018, Art. no. 513, https://doi.org/10.3390/su10020513.

[15] M. A. Haq, "Smotednn: A novel model for air pollution forecasting and aqi classification," *Computers, Materials and Continua*, vol. 71, no. 1, pp. 1403–1425, 2022, https://doi.org/10.32604/cmc.2022.021968.

[16] E. Lopez-ROjas, "Synthetic Financial Datasets For Fraud Detection." kaggle, Accessed: May 15, 2024. [Online]. Available: https://www.kaggle.com/datasets/ealaxi/paysim1.

[17] W. Koehrsen, "Random Forest Simple Explanation," Medium. Accessed: Mar. 20, 2024. [Online]. Available: https://williamkoehrsen.medium.com/random-forest-simple-explanation-377895a60d2d

[18] Y. Regaya, F. Fadli, and A. Amira, "Point-Denoise: Unsupervised outlier detection for 3D point clouds enhancement," *Multimedia Tools and Applications*, vol. 80, no. 18, pp. 28161–28177, Jul. 2021, https://doi.org/10.1007/s11042-021-10924-x.

[19] I. Popchev, R. Ketipov, and V. Angelova, "Risk Averseness and Emotional Stability in e-Commerce," *Cybernetics and Information Technologies*, vol. 21, no. 3, pp. 73–84, Jun. 2021, https://doi.org/10.2478/cait-2021-0030.

[20] B. Alshawi, "Utilizing GANs for Credit Card Fraud Detection: A Comparison of Supervised Learning Algorithms," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12264–12270, Dec. 2023, https://doi.org/10.48084/etasr.6434.

[21] S. S. Taher, S. Y. Ameen, and J. A. Ahmed, "Advanced Fraud Detection in Blockchain Transactions: An Ensemble Learning and Explainable AI Approach," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 12822–12830, Feb. 2024, https://doi.org/10.48084/etasr.6641.

[22] M. Anwer, S. M. Khan, M. U. Farooq, and Waseemullah, "Attack Detection in IoT using Machine Learning," *Engineering, Technology & Applied Science Research*, vol. 11, no. 3, pp. 7273–7278, Jun. 2021, https://doi.org/10.48084/etasr.4202.

[23] E. Yilmaz and O. Can, "Unveiling Shadows: Harnessing Artificial Intelligence for Insider Threat Detection," *Engineering, Technology & Applied Science Research*, vol. 14, no. 2, pp. 13341–13346, Apr. 2024, https://doi.org/10.48084/etasr.6911.

[24] M. A. Haq, "DBoTPM: A Deep Neural Network-Based Botnet Prediction Model," *Electronics*, vol. 12, no. 5, Jan. 2023, Art. no. 1159, https://doi.org/10.3390/electronics12051159.

[25] J. Gyani, M. A. Haq, and A. Ahmed, "Analyzing the Impact of Lockdown on COVID-19 Pandemic in Saudi Arabia," *International Journal of Computer Science & Network Security*, vol. 22, no. 4, pp. 39–46, 2022, https://doi.org/10.22937/IJCSNS.2022.22.4.6.

[26] M. A. Haq and A. Ahmed, "On Interesting Correlation between Meteorological Parameters and COVID-19 Pandemic in Saudi Arabia," *International Journal of Computer Science & Network Security*, vol. 22, no. 4, pp. 159–168, 2022, https://doi.org/10.22937/IJCSNS.2022.22.4.20.

[27] M. A. Haq, M. A. R. Khan, and M. Alshehri, "Insider Threat Detection Based on NLP Word Embedding and Machine Learning," *Intelligent Automation and Soft Computing*, vol. 33, no. 1, pp. 619–635, 2022, https://doi.org/10.32604/iasc.2022.021430.

[28] S. Kumar, M. A. Haq, C. Jason, N. R. Moparthi, N. Mittal, and S. Alzamil, "Multilayer Neural Network Based Speech Emotion Recognition for Smart Assistance," *Computers, Materials and Continua*, vol. 74, pp. 1523–1540, Aug. 2022, https://doi.org/10.32604/cmc.2023.028631.

[29] A. Alabdulwahab, M. A. Haq, and M. Alshehri, "Cyberbullying Detection using Machine Learning and Deep Learning," *International Journal of Advanced Computer Science and Applications*, vol. 14, pp. 424–432, Oct. 2023, https://doi.org/10.14569/IJACSA.2023.0141045.

[30] J. Gyani, A. Ahmed, and M. A. Haq, "MCDM and Various Prioritization Methods in AHP for CSS: A Comprehensive Review," *IEEE Access*, vol. 10, pp. 33492–33511, 2022, https://doi.org/10.1109/ACCESS.2022.3161742.

[31] G. Revathy, S. A. Alghamdi, S. M. Alahmari, S. R. Yonbawi, A. Kumar, and M. Anul Haq, "Sentiment analysis using machine learning: Progress in the machine intelligence for data science," *Sustainable Energy Technologies and Assessments*, vol. 53, Oct. 2022, Art. no. 102557, https://doi.org/10.1016/j.seta.2022.102557.

[32] M. Suresh, A. S. Shaik, B. Premalatha, V. A. Narayana, and G. Ghinea, "Intelligent & Smart Navigation System for Visually Impaired Friends," in *12th International Conference onAdvanced Computing*, Hyderabad, India, Dec. 2022, pp. 374–383, https://doi.org/10.1007/978-3-031-35641-4_30.

[33] S. Merugu, K. Jain, A. Mittal, and B. Raman, "Sub-scene Target Detection and Recognition Using Deep Learning Convolution Neural Networks," in *1st International Conference on Data Science, Machine Learning and Applications*, 2020, pp. 1082–1101, https://doi.org/10.1007/978-981-15-1420-3_119.

[34] A. Bathula, S. kr. Gupta, S. Merugu, and S. S. Skandha, "Academic Projects on Certification Management Using Blockchain- A Review," in *International Conference on Recent Trends in Microelectronics, Automation, Computing and Communications Systems*, Hyderabad, India, Dec. 2022, pp. 1–6, https://doi.org/10.1109/ICMACC54824.2022.10093679.

[35] A. Bathula, S. Muhuri, S. kr. Gupta, and S. Merugu, "Secure certificate sharing based on Blockchain framework for online education," *Multimedia Tools and Applications*, vol. 82, no. 11, pp. 16479–16500, May 2023, https://doi.org/10.1007/s11042-022-14126-x.