Pakistan Journal of Life and Social Sciences
www.pjlss.edu.pk

https://doi.org/10.57239/PJLSS-2024-22.1.00316

RESEARCH ARTICLE

Integrated Machine Learning Algorithms Applied to the Detection of Financial Accounting Fraud, Cases of Peruvian Companies

Jorge M. Chávez-Díaz[1*], Mark Barrueta-Pinto[2], Manuel Antonio Murillo Luna[3], Alejandro Valencia Arias[4], José L. Arquero[5]

[1,2,3]Universidad Peruana de Ciencias Aplicadas
[4]Institución Universitaria ITM, Colombia
[5]Universidad de Sevilla, España

ARTICLE INFO

Received: May 22, 2024

Accepted: Jul 10, 2024

Keywords

Local Outlier Factor

Isolation Forest

Forensic accounting

Fraud detection

Support vector machine

***Corresponding Author:**

pccajoch@upc.edu.pe

ABSTRACT

This paper presents the result of the integrated use of three of the main algorithms proposed by the literature for fraud detection: the local outlier factor (LOF), the isolation forest (ISF) and the support vector machine (SVM). The study integrates them and applies them to financial information of Peruvian companies of different sizes (small, medium and large companies). The inductive generalization method is followed, whose level of knowledge is to describe the application of the algorithms described on certain databases. The methodological keys are the data collection, the application of the algorithms and for the evaluation a specialized software and the F1-Score model were used. The results showed that the application of Machine Learning algorithms detect clues for fraud detection, which should be part of a comprehensive procedure library according to the research planning. In addition, the results showed that the algorithms detect more than 50% of anomalies and that materiality thresholds should be considered to refine conclusions. The results promote the improvement of fraud detection and provide the first studies on the subject in Peruvian companies of different sizes through the application of machine learning algorithms. Future work could evaluate other algorithms on different financial accounting information.

INTRODUCTION

Financial fraud is one of the most significant risks facing companies globally. Whether it is inflated expenses, embezzlement, CEO conflicts of interest, false payroll payments, bribes or even tax evasion. The proposed research shows how the use of machine learning helps substantially in the early detection of these malpractices.

This research work incorporates the use of machine learning algorithms applied to various financial accounting information to detect anomalies leading to possible cases of fraud. The proposed algorithms are taken from the conclusions of the latest scientific articlespublished (obtained from the Scopus database) and constitute basic knowledge that the forensicaccountant (chartered accountant) should consider for their work in today's competitive, technological, and digital industry such as Industry 4.0.

It is of great importance that the use of the proposed algorithms be incorporated into thetraining of forensic accountants to apply them to the multiple fields in which they finddatabases to be examined. The conclusions of this work will allow the implementation of machine learning routines in the day-to-day work of forensic accountants, whether at the labor,financial, tax or even in the expertise of the various fiscal and judicial processes.

Some works have been identified on the application of ML algorithms in financial transactions with F1-Score higher than 95% (Dávila-Morán et al., 2023) and LOF and ISF applications that show positive results in the detection of fraud onfinancial information (Vijayakumar et al., 2020). Research of this type will continue to be produced given the business needs to achieve sustainability and even more important in contexts of high perception of corruption (Chávez-Díaz et al., 2023).

As a result of the theoretical contributions, the following three research questions are presented:

(Q1) What are the results obtained in the application of combined Machine Learning (ML) algorithms for fraud detection?

(Q2) How to implement the findings obtained by application of combined ML algorithms withother forensic fraud detection procedures?

(Q3) What is the degree of reliability of the results obtained in the application of combined ML algorithms for fraud detection?

To answer research questions Q1 and Q2, machine learning algorithms will be run on the proposed databases, so that the performance will be evaluated according to the size of the database. Regarding research question Q3, simulations will be performed and the results obtained will be measured with the F1-Score.

THEORETICAL FRAMEWORK

Fraud

According to the Association of Certified Fraud Examiners (ACFE) (2020) fraud is conceptualized in one of three ways: as misappropriation or misappropriation of an entity's resources; the second, corporate, which relates to fraudulent reporting; and the third, identified with corruption. And fraud detection is defined as the applied knowledge-intensive activities that are executed to prevent financial resources from being obtained through non-legal pretexts (Pitchayatheeranart & Phornlaphatrachakorn, 2023).

White-collar crimes refer to fraudulent actions carried out by people of high reputation, with strong personality, with a high orientation to success and achievement of goals, with some traits of psychopathy. These individuals, generally male, could be attractive and with a lot of charisma, such that they gain people's trust. Individuals that generate protection networks, being cultured, and educated, with professional studies and often with good dress and personalappearance (Clavería Navarrete & Carrasco Gallego, 2023).

Forensic Accounting

Forensic accounting is a profession that goes beyond standardized audit approaches to detect and prevent fraud (Afriyie et al., 2023) and is established in first-world countries such as Australia and varies according to different contexts (Al Shbeil et al., 2023). Although the field of forensic accounting is still relatively small compared to other accounting specializations, but there is a great deal of interest in it worldwide (Ozili, 2023).

In times of globalized economy, forensic accounting is an effective tool to detect and prevent fraud (Capraş & Achim, 2023), its use involves applying reliable principles and methods to obtain sufficient facts for prosecution in a court of law (Afriyie et al., 2023).

Forensic Accountant

Forensic accountants play a crucial role in investigating fraud, providing litigation support, performing business valuations, and detecting cybercrime (Al Shbeil et al., 2023). They need to possess skills such as communication, presentation, analytical thinking and objectivity, independence, attention to detail, courtroom procedures, IT skills, among others (Al-Daoud et al., 2023).

Various tools and techniques are used to uncover financial irregularities and include proactive audit, investigation and fraud detection methods (Afriyie et al., 2023). Traditional forensic accounting models focus on behavioral characteristics of executives or usenumerical approaches based on financial data. However, more recent models combine big data analytics with psychological intuitions (Honigsberg, 2020).

Customer interviews to gather information in financial fraud cases are widely used techniques (Fico & Walsh, 2023). The skills and attributes of a forensic accountant, aswell as the techniques used in forensic accounting, make them an effective tool for detecting and preventing fraud (Capraş & Achim, 2023).

In this regard, data analysis techniques are used to detect financial fraud by applying reliable principles and methods to obtain sufficient facts or data for prosecution (Afriyie et al., 2023). They need extensive knowledge of accounting, law, auditing, internal auditing, business management, psychology, criminal science, and computer technologies, among others (Kılıç, 2020). The use of big data techniques, data analytics and algorithms has improved the effectiveness of fraud inspections (Kılıç, 2020).

Techniques such as data mining and text mining are effective tools for detecting and preventing fraud (Capraş & Achim, 2023). However, data analytics as an effective tool in the fight against fraud is underutilized (Aboud & Robinson, 2022). The intention to adopt forensic accounting practices is influenced by coercive, mimetic, and normative pressures (Habis Alrawashedh, 2023).

Machine Learning

As a fundamental part of Artificial Intelligence, machine learning is a computer programthat can automatically learn patterns and trends from historical data without being explicitly programmed by humans (Huang & Wang, 2023).

Machine learning algorithms fall into two categories: supervised and unsupervised learning. Supervised learning uses the program to predict the outcomes of future instances based on historical data. The data must include predefined outcomes based on which the machine learning algorithms can learn and find the correct class. Unsupervised learning algorithms learn normal patterns based on historical data without pre-defined classes (Huang & Wang, 2023).

Machine learning plays an important role in improving the effectiveness of data analysisfor forensic accountants in detecting financial fraud, (Ayad et al., 2023; Bao et al., 2022; Hu & Sun, 2022; Odia & Akpata, 2020). Machine learning algorithms improve the quality of accounting information by analyzing financial statements and detecting frauds (Ayad et al., 2023).

The use of machine learning in fraud detection is particularly relevant in the context of accounting fraud (Bao et al., 2022). Data science techniques, including machine learning, provide forensic accountants with the ability to extract, analyze, and visualize large volumes of data, allowing them to stay ahead of fraudsters (Odia & Akpata, 2020).

Machine learning is gradually being applied in tasks such as source document review, business transaction analysis, and risk assessment in the accounting and assurance profession, (Cho et al., 2020). The use of machine learning in financial fraud detection has shown promising results, with

techniques such as support vector machines, artificial neural networks,and decision trees being commonly used, (Ramírez-Alpízar et al., 2020). These techniques have demonstrated effectiveness in detecting real cases of fraud, with accuracy values rangingfrom 70 % to 99, 9 %, (Ramírez-Alpízar et al., 2020).

Machine learning algorithms in forensic accounting. A brief review

Recent research has used machine learning algorithms in data analytics to detect financial fraud. However, different approaches can be found in the forensic accounting literature:

Table 1: Machine learning algorithms in forensic accounting

Algorithm used	Applications	Author
Parameter-Fitted Automatic Set Learning (FFD-PTEML)	Fraud detection in the financial industry	(Atassi et al., 2024)
XGboost, LGBM, CatBoost, Random Forest (RF) and Decision Tree (DT).	On customer data privacy in the financial sector.	(Dasari & Kaluri, 2024)
Quantum Machine Learning Model (QML).	Comparative study with models such as variational quantum classifier (VQC), estimator quantum neural network (QNN) and sampler quantum neural network (QNN Sampler).	(Innan et al., 2024)
Genetic Algorithm (GA)	Financial fraud in credit card transactions	(Mantena et al., 2024)
Logistic regression, Naive Bayes, KNN, Perceptron techniques and others.	Combating financial fraud in banking systems. Exploratory analysis.	(Moreira et al., 2022)
Support Vector Machines (SVM), artificial neural networks and decision trees.	Fraud detection in financial statements. Systematic mapping of 67 studies	(Ramírez-Alpízar et al., 2020)
Random Forest, XGBoost, Naive Bayes, K-Nearest Neighbors and Support Vector Machine	Ongoing fight against financial fraud and provides valuable guidance for future investigative efforts.	(Sinap, 2024)
Egret Swarm Optimization Algorithm (ESOA) integrated with machine learning method	Development of a fraud detection framework.	(Yi et al., 2023)

Local Outlier Factor

Early work indicated that, in many cases, it is more useful to assign each object a degreeof outlier. This degree is called the local outlier factor (LOF) of an object. It is local in the sensethat the degree depends on how isolated the object is with respect to the surrounding neighborhood. Observations with a high LOF value are anomalous (Stripling et al., 2018).

Isolation Forest (IForest, ISF)

Compared to general anomaly identification methods, the IForest algorithm does not require the sample data to fit the characteristics of the Gaussian distribution. Moreover, compared to other methods such as one-class SVM, IForest can capture more complexstructural information. Therefore, in system integration, the IForest algorithm becomes our firstchoice for capturing implicit anomalous transactions. The algorithm first uses decision trees tobuild forests. Then, a feature is randomly selected and a random partition value between the minimum and maximum value of the selected feature is selected to create an isolated partition.This algorithm demonstrates higher accuracy and reliability in practical applications, which helps to identify and prevent potential asset laundering risks more effectively (Yang et al., 2023).

Support Vector Machine (SVM)

A literature review found that supervised analytical models such as SVM support vector machines can help mitigate threats in the marketplace and prevent millions of dollar losses (Ortíz et al., 2022). SVM is one of the most widely used techniques for financial statement fraud detection, with studies showing high effectiveness and accuracy rates ranging from 70 % to 99.9 % (Ramírez-Alpízar et al., 2020).

SVM has also been used in combination with other models, such as the M-S kernel and the Z-S kernel, to accurately predict financial statement fraud (Phong et al., 2022). In addition, SVM has been used to detect financial reporting fraud, which has yielded high accuracy rates (Chi et al., 2019). Several authors argue that SVM is a valuable tool for identifying patterns of financial fraud in forensic accounting (Chi et al., 2019; Ortíz et al., 2022; Phong et al., 2022; Ramírez-Alpízar et al., 2020).

Integration or combination of algorithms

Isolation Forest is only sensitive to global outliers and is weak in dealing with local outliers. Although LOF performs well in detecting local outliers, it has high time complexity. Therefore, there are studies showing that the joint application of Local Outlier Factor and Isolation Forest can significantly improve the outlier detection rate and greatly reduce the timecomplexity (Cheng et al., 2019).

The results considered anomalies are:

a) Upper limit exceptions, referring to the highestvalues above a given threshold.

b) Lower limit exceptions, where the lowest values of the groupare identified.

c) Unicorn alert, identifies a single value with the highest anomaly and

d) Others,where values whose distance to other values is significant are included.

METHOD

The inductive generalization method is followed, whose level of knowledge is to describe the application of machine learning algorithms on a selected set of relevant databases, then classify the characteristics, identify irregularities and, therefore, project the directions of implementation of their results in the audit and forensic accounting processes.

The key aspects of the methodology are the collection and preparation of data, which were obtained from companies of different sizes as measured by revenue level. They were then subjected to the Local Outlier Factor, Isolation Forest and Support Vector Machine algorithms separately in a preparatory manner, and then applied together. Finally, for performance evaluation purposes, specialized software that follows the F1-Score model was used to measure fraud detection performance.

Databases used

Four databases were used for the work. The first one referred to the information of the purchase records of an SME company by monthly period in PLE format (Electronic Books Program, defined by the Peruvian Tax Administration - SUNAT), periods 2019 to 2023. The total number of observations was 51 with a total of 2,139 records evaluated. The tests performed were based on Local Outlier Factor (LOF), Isolation Forest (ISF) and Support Vector machine (SVM) analysis on theTP_CDP category (type of payment voucher) and the TOTAL numeric field (total amount of the document recorded).

A second group of data refers to savings bank account information of a non-profit organization, periods 2019 to 2020. The total number of observations was 24 with a total of 61,420 records evaluated. The tests performed were based on Local Outlier Factor (LOF), Isolation Forest (ISF) and Support Vector machine (SVM) analyses on the Period category and the numeric fields Charges and Credits.

The third group of data worked on are the class 6 expenses of the Daily Book of a large industrial company, in PLE format (Electronic Book Program, defined by the Peruvian Tax Administration - SUNAT), period 2022. The total number of observations was 12 with a totalof 209,190 records evaluated. The tests performed were based on Local Outlier Factor (LOF), Isolation Forest (ISF) and Support Vector machine (SVM) analysis on the "Period" category and the numeric field Debit.

The fourth group of data worked on is the class 6 expenses from the journal of a large industrial company, in PLE format, period 2020. The total number of observations was 12 witha total of 46,512 records evaluated. The tests performed were based on Local Outlier Factor (LOF), Isolation Forest (ISF) and Support Vector machine (SVM) analyses on the "Two-digit accounting account" category andthe numeric field Debit. This database will be used to carry out the simulation proposed in the third research question. For which a total of 24 manipulated records will be added.

Table 2: Basic accounting information for analysis

No.	Entity	Information	Period	Numeric field	Tables	Total, records
01	SME Company	Purchase register	2019-2023	Total	51	2,139
02	Non-profit organization	Bank statements - savings	2019-2020	Charge and Credit	24	61,420
03	Industrial company 1	Journal - Class 6	2022	Debit	12	209,190
04	Industrial company 2	Journal - Class 6	2020	Debit	12	46,512
TOTAL					99	319,261

Caseware's CAAT IDEA analysis tool and its IDEALabs - Anomalies add-on was used.An example of the interactions with the software to point out the categories and the numericalfield corresponding to the first group of data are presented in Figure 1, and to set the parametersof the Machine Learning algorithms are shown in Figure 2.

The general statistics were generated in IDEA and the results of the evaluations by tablewere appended in a general table according to the source of information to be tabulated and finally exported to MS Excel for presentation purposes.

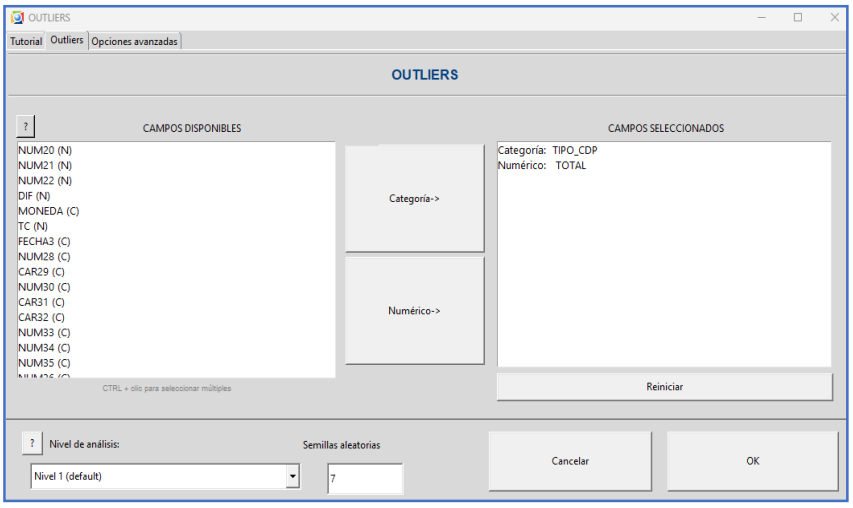

Figure 1: Set Categories and Numbers

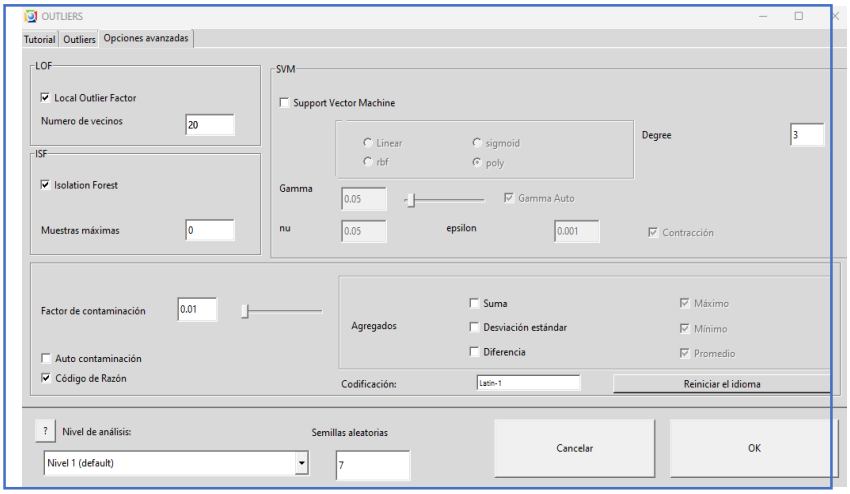

Figure 2: Set algorithm and parameters.

F1-Score

A metric recognized as a standard for measuring fraud detection performance will be used (Dávila-Morán et al., 2023). Based on the confusion matrix shown in Figure 3, formulas are developed to evaluate accuracy, precision, sensitivity, and the F1-Score as a robust conjunction of sensitivity and precision.

Modelo A		Prediction	
		No fraud	Fraud
Reality	No Fraud	True Negative (TN)	False Positive (FP)
	Fraud	False Negative (FN)	True Positive (TP)

accuracy	(TN+TP) / (TN+TP+FN+FP)
precision	TP / (TP+FP)
recall	TP / (TP+FN)
F1-Score	2*recall*precision/ (recall+precision)

Figure 3: Confusion matrix and F1-Score formula.

RESULTS AND DISCUSSION

First research question - Performance of machine learning algorithms

Entity 1. It was obtained that up to 98% of the monthly records are categorized as normaloperations; upper limit exceptions up to 33%; unicorn alert up to 12.5%; lower limit exceptionsup to 9% and others up to 12.5%, as presented in Table 3. The monthly records that make up the purchases range from 4 to 93. The number of monthly records exceeds 20 transactions; therecords classified as normal range from 92% to 98%. The number of records lower than 20 increases the number of anomalies because there are fewer reference points.

Table 3. Results in small companies

Entity 1	Normal	Outliers: 1	Outliers: 2	Total, general	Range of records per month (times)	Range % max. max. - min.
No comments	51			51	93 - 4	98.46% - 66.67%
Upper limit exception		35	4	39	2(5) - 1(34)	33.33% -- 1.05%
Unicorn Alert		34	4	38	1(38)	12.5% - 1.05%
Lower limit exception		7		7	1(7)	9.09% - 1.52%
Other		4		4	1(4)	12.5% - 5.56%
Total, general	51	80	8	139		

Entity 2. It was found that up to 98% of the monthly records are normal operations; the upper limit exceptions are up to 1%. In this case, it is observed that the more records evaluated, the more anomalies detected stabilize around 1% for each type of anomaly, as shown in Table4

Table 4. Result in median entity

Entity 2	Normal	Outliers: 1	Total, general	Records per month max - min	Range perc. % max-min
Another		24	24	54 - 11	1.34% - 1.01%
Upper limit exception		24	24	45 - 11	1.03% - 0.73%
No comments	24		24	4843 - 1054	98.03% - 97.91%
Total, general	24	48	72		

Entity 3. It was obtained that up to 98% of the monthly records are categorized as normaloperations; the upper limit exceptions up to 1% and others up to 1.5%, as presented in Table 5.

Table 5: Results in large company 1

Entity 3	Normal	Outliers: 1	Total, general	Records per month max - min	Range perc. % max-min.
No comments	12		12	18593 - 15404	98.01% - 97.99%
Upper limit exception		12	12	168 - 85	0.93% - 0.51%
Another		12	12	257 - 196	1.48% - 1.08%
Total, general	12	24	36		

Entity 4. It was obtained that up to 98.54% of the monthly records are categorized as normal operations; upper limit exceptions up to 1.06%; lower limit exceptions up to 0.67%; unicorn alert up to 0.38% and others up to 0.96%, as presented in Table 6.

Table 6: Results in large company 2

Entity 4	Normal	Outliers: 1	Outliers: 2	Total, general	Records per month max. - min	Range perc. % max. – mín.
Upper limit Exception		12	1	13	49 - 1	1.06% - 0.03%
No comments	12			12	5140 - 1483	98.54% - 7.96%
Another		12		12	49 - 2	0.96% - 0.13%
Lower limit exception		11		11	19 - 1	0.67% - 0.03%
Unicorn Alert		5	1	6	15 - 1	0.38% - 0.02%
Total, general	12	2	40	54		

The following characteristics can be inferred from the results obtained:

1. Lower limit exceptions. Results were obtained only in one case for evaluations in the Journal - Class 6 (Table 5). The ML algorithms identified transactions with zero amountin the debit field. In the other database of the Journal - Class 6 (Table 4) no results wereobtained since the category used was Period and not the two-digit Account.

Thus, it identifies transactions if analysis categories are specified. From the four evaluations performed, it can be concluded that the contribution would not be significantit follows materiality-oriented procedures.

2. Upper bound exceptions. ML algorithms establish records of larger amounts from a threshold, which is usually around 1% of the records in a table. It is efficient for analyzinginformation from 20 records onwards. From the evaluations performed, this classificationis important because it ensures significant amounts subject to further specific analysis.

3. Unicorn Alert. It only identified records with this type of classification in the case of entity 4 (Table 5). In this case, the two-digit account was considered as a category and the records identified showed a numerical value of zero (account 69).

From the four evaluations carried out, the contribution is not significant.

4. Others. It is observed for all four databases that records with unique unusual values are obtained. It does not distinguish material numerical amounts, i.e. it identifies values of any amount. The larger the database, the more efficient its results will be. This classification could be very useful if a materiality threshold is added in the audit planningor execution of the expertise, to refine the result.

According to the results obtained, the first research question is solved. The Machine Learning algorithms (LOF, ISF, SVM) provide significant results since they detect clues (materialand outliers) that allow a deeper analysis in the process of detecting financial accounting fraud. Thus, it corroborates the findings of (Cheng et al., 2019; Vijayakumar et al., 2020).

Second research question.

The findings obtained can be implementedin conjunction with other fraud detection procedures such as the following:

- Benford's Law of first digit. For the evaluation of fragmentation of expenses, especially in journal entries - Class 6 (Antunes et al., 2023).
- Application of duplicates. For evaluations of invoices recorded in the Purchase Register, Asset IDs or Codes in the Asset Ledger, invoices in the Sales Register, minor provisions inthe Journal, etc.
- Application of omissions. For fields with correlative coding characteristics such as journal entry numbers in the Journal or payment voucher numbers in the Sales Register or Warehouse Receipt Parts for Inventories.
- Evaluation of round amounts and zero decimal places. Applied to accruals in the Journal or purchase invoices in the Purchase Register.

Third research question.

Two simulations were carried out. In the firstone, records were entered in the database of entity 4, consisting of expenses for the same account and of repetitive high amounts.

The result showed that the manipulated records were not detected by the algorithms. This case

provides clues regarding the nature of the LOF, ISF and SVM algorithms, which detect anomalies,or outliers. With this approach, then, the added records do not constitute outliers and, therefore, the model provides confidence in not detecting false positives. The F1-Score did not show any results, as detailed in Figure 4.

Simulation 1		Prediction	
		No fraud	Fraud
Reality	No fraud	45,656	861
	Fraud	49	-

accuracy	0.98046
precision	-
recall	-
F1-Score	N/D

Figure 4: Result Simulation 1

In a second simulation, 24 records with anomaly characteristics were added to the samedatabase for entity 4. In this case, the model did detect more than 50% of the manipulated records (13 of 24). The F1-Score obtained is 0.029, still low (Figure 5). In this regard, it is explained that the resulting detections correspond to records with non-significant amounts, i.e.,without materiality.

Simulation 2a		Prediction	
		No fraud	Fraud
Reality	No fraud	45,714	849
	Fraud	11	13

accuracy	0.98154
precision	0.01508
recall	0.54167
F1-Score	0.02935

Figure 5: Simulation result 2a

In a variant of simulation 2, for a material analysis, the materiality threshold was established to obtain meaningful results. In this case, it was S/5,000. The result is more in linewith the practice of the forensic or expert witness. The F1-Score improved considerably (0.17808). However, a classification ability of 17.8% is far from other studies that obtain an ability of up to 95% (Dávila-Morán et al., 2023). The model may be improved with the joint application of other algorithms in ML.

In this way, the usage model can become more meaningful as usage parameters are established based on the forensic expert's experience. It should be noted that the recalled detections could be evaluated in a separate sub-process using Benford tests, to close the risk of fragmentation of significant amounts. The model reports a recall of 0.54167, i.e., it continues to detect more than 50% of the manipulated records, as detailed in Figure 6, which in monetary terms means that the entity is reducing the risk of fraud by that percentage.

Simulation 2b		Prediction	
		No fraud	Fraud
Reality	No fraud	46,454	109
	Fraud	11	13

accuracy	0.99742
precision	0.10656
recall	0.54167
F1-Score	0.17808

Figure 6: Simulation result 2b

CONCLUSIONS

This paper presents its results on a multi-algorithmic approach in Machine Learning (LOF, ISF, SVM), which proved to provide a deep analysis of the information that makes it an effective tool in obtaining clues in the detection of financial and accounting fraud. Its efficiencyis evidenced as the volume of information is greater and its application is executed on specificcategories of information, such as, for example, monthly periods, two-digit accounting account,complete accounting account, cost center, bank account, etc.

For a successful implementation of the algorithms, it is necessary to use it in conjunctionwith other fraud detection procedures, and to add some parameters that the experience of the auditor or expert advises.

Initial research looked ahead to further research into a multi-algorithmic approach that would allow for the classification and evaluation of large datasets (Wheeler & Aitken, 2000).Therefore, further research is recommended to analyze other algorithms such as Bayesian Networks or others that engineering and data science recommend in fraud detection applied to the field of business, management, and accounting.

FUNDING: Research Direction of the Universidad Peruana de Ciencias Aplicadas. Code: A-046-2024.

CONFLICT OF INTEREST: The authors declare no conflict of interest.

REFERENCES

Aboud, A., & Robinson, B. (2022). Fraudulent financial reporting and data analytics: an explanatory study from Ireland. *Accounting Research Journal*, *35*(1), 21–36. https://doi.org/10.1108/ARJ-04-2020-0079

Afriyie, S. O., Akomeah, M. O., Amoakohene, G., Ampimah, B. C., Ocloo, C. E., & Kyei, M. O. (2023a). Forensic Accounting: A Novel Paradigm and Relevant Knowledge in Fraud Detection and Prevention. *International Journal of Public Administration*, *46*(9), 615–624. https://doi.org/10.1080/01900692.2021.2009855

Afriyie, S. O., Akomeah, M. O., Amoakohene, G., Ampimah, B. C., Ocloo, C. E., & Kyei, M. O. (2023b). Forensic Accounting: A Novel Paradigm and Relevant Knowledge in Fraud Detection and Prevention. *International Journal of Public Administration*, *46*(9), 615–624. https://doi.org/10.1080/01900692.2021.2009855

Al Shbeil, S., Alshurafat, H., Taha, N., & Al Shbail, M. O. (2023). What Do We Know About Forensic Accounting? A Literature Review. In *Lecture Notes in Networks and Systems* (Vol. 557, pp. 629–636). https://doi.org/10.1007/978-3-031-17746-0_49

Al-Daoud, K., Abuorabi, Y., Darwazeh, R., Nawaiseh, M. Y., Saifan, N. M., & Al-Hawary, S. I. S. (2023). Electronic Financial Crimes: The Required Skills, Education and Qualifications for Forensic Accountants to Predict and Prevent. *Information Sciences Letters*, *12*(3), 1237–1248. https://doi.org/10.18576/isl/120315

Antunes, A. M., Teixeira, D., & Sousa, F. (2023). Benford's Law: the fraud detection's left hand. *2023 18th Iberian Conference on Information Systems and Technologies (CISTI)*, 1–6. https://doi.org/10.23919/CISTI58278.2023.10211738

Association of Certified Fraud Examiners (ACFE). (2020). *Report to the Nations, 2020*. https://legacy.acfe.com/report-to-the-nations/2020/

Atassi, R., Zikriyoev, A., Turayev, N., & Botırovna, S. G. (2024). Boosting Financial Fraud Detection Using Parameter Tuned Ensemble Machine Learning Model. *Journal of Cybersecurity and Information Management*, *13*(2), 66–74. https://doi.org/10.54216/JCIM.130205

Ayad, M., El Mezouari, S., & Kharmoum, N. (2023). Impact of Machine Learning on the Improvement of Accounting Information Quality. In *Lecture Notes in Networks and Systems: Vol. 637 LNNS*. https://doi.org/10.1007/978-3-031-26384-2_43

Bao, Y., Hilary, G., & Ke, B. (2022). Artificial Intelligence and Fraud Detection. In *Springer Series in Supply Chain Management* (Vol. 11). https://doi.org/10.1007/978-3-030-75729-8_8

Capraş, I. L., & Achim, M. V. (2023a). An Overview of Forensic Accounting and Its Effectiveness in the

Detection and Prevention of Fraud. In *Contributions to Finance and Accounting: Vol. Part F1313*. https://doi.org/10.1007/978-3-031-34082-6_13

Capraș, I. L., & Achim, M. V. (2023b). An Overview of Forensic Accounting and Its Effectiveness in the Detection and Prevention of Fraud. In *Contributions to Finance and Accounting: Vol. Part F1313*. https://doi.org/10.1007/978-3-031-34082-6_13

Chávez-Díaz, J. M., Bonilla Migo, A., Monterroso Unuysuncco, N. I., & Romero-Carazas, R. (2023). Gestión para la recaudación de impuestos municipales: diagnóstico y propuesta. *Revista Venezolana de Gerencia*, *28*(103), 1052–1067. https://doi.org/10.52080/rvgluz.28.103.9

Cheng, Z., Zou, C., & Dong, J. (2019). Outlier detection using isolation forest and local outlier. *Proceedings of the 2019 Research in Adaptive and Convergent Systems, RACS 2019*, 161–168. https://doi.org/10.1145/3338840.3355641

Chi, D.-J., Chu, C.-C., & Chen, D. (2019). Applying Support Vector Machine, C5.0, and CHAID to the Detection of Financial Statements Frauds. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 11645 LNAI*. https://doi.org/10.1007/978-3-030-26766-7_30

Cho, S., Vasarhelyi, M. A., Sun, T., & Zhang, C. (2020). Learning from machine learning in accounting and assurance. *Journal of Emerging Technologies in Accounting*, *17*(1), 1–10. https://doi.org/10.2308/jeta-10718

Clavería Navarrete, A., & Carrasco Gallego, A. (2023). Forensic accounting tools for fraud deterrence: a qualitative approach. *Journal of Financial Crime*, *30*(3), 840–854. https://doi.org/10.1108/JFC-03-2022-0068

Dasari, S., & Kaluri, R. (2024). An Effective Classification of DDoS Attacks in a Distributed Network by Adopting Hierarchical Machine Learning and Hyperparameters Optimization Techniques. *IEEE Access*, *12*, 10834–10845. https://doi.org/10.1109/ACCESS.2024.3352281

Dávila-Morán, R. C., Castillo-Sáenz, R. A., Vargas-Murillo, A. R., Dávila, L. V., García-Huamantumba, E., García-Huamantumba, C. F., Cajas, R. F. P., & Paredes, C. E. G. (2023). Application of Machine Learning Models in Fraud Detection in Financial Transactions. *Data and Metadata*, *2*. https://doi.org/10.56294/dm2023109

Fico, J., & Walsh, D. W. (2023). Investigative Interviews Conducted by Forensic Accounting and Auditing Professionals. In *The Past, Present, and Future of Accountancy Education and Professions* (pp. 66–81). https://doi.org/10.4018/978-1-6684-5483-1.ch004

Habis Alrawashedh, N. (2023). Factors affecting organizational intention to adopt forensic accounting practices: A case of Jordan. *Problems and Perspectives in Management*, *21*(3), 343–350. https://doi.org/10.21511/ppm. 21(3).2023.27

Honigsberg, C. (2020). Forensic accounting. In *Annual Review of Law and Social Science* (Vol. 16). https://doi.org/10.1146/annurev-lawsocsci-020320-022159

Hu, H., & Sun, T. (2022). The Applications of Machine Learning in Accounting and Auditing Research. In *Encyclopedia of Finance, Third Edition*. https://doi.org/10.1007/978-3-030-91231-4_91

Huang, F., & Wang, Y. (2023). Introducing machine learning in auditing courses. *Journal of Emerging Technologies in Accounting*, *20*(1), 195–211. https://doi.org/10.2308/JETA-2022-017

Innan, N., Khan, M. A.-Z., & Bennai, M. (2024). Financial fraud detection: A comparative study of quantum machine learning models. *International Journal of Quantum Information*, *22*(2). https://doi.org/10.1142/S0219749923500442

Kılıç, B. İ. (2020). The effects of Big Data on Forensic Accounting Practices and Education. In *Contemporary Studies in Economic and Financial Analysis* (Vol. 102). https://doi.org/10.1108/S1569-375920200000102005

Mantena, S. S., Rao, T. V., Babu, B. R. C., Prasad, T., & Naresh, D. (2024). PREVENTION OF CREDIT CARD FRAUD TRANSACTION USING GA FEATURE SELECTION WITH MACHINE LEARNING. *African Journal of Biological Sciences (South Africa)*, *6*(7), 232–239. https://doi.org/10.33472/AF5BS.6.7.2024.232-239

Moreira, M. Â. L., Junior, C. de S. R., Silva, D. F. de L., de Castro Junior, M. A. P., Costa, I. P. de A., Gomes, C. F. S., & dos Santos, M. (2022). Exploratory analysis and implementation of machine learning techniques for predictive assessment of fraud in banking systems. *Procedia Computer Science*, *214*(C), 117–124. https://doi.org/10.1016/j.procs.2022.11.156

Odia, J. O., & Akpata, O. T. (2020). Role of data science and data analytics in forensic accounting and fraud detection. In *Handbook of Research on Engineering, Business, and Healthcare Applications of Data Science and Analytics*. https://doi.org/10.4018/978-1-7998-3053-5.ch011

Ortíz, M. M., Marín, L. M. G., Villegas, H. H. J., & Escobar, C. C. P. (2022). Analytical models to identify financial crime patterns: a systematic literature review | Modelos analíticos para identificar patrones de delitos financieros: Una revisión sistemática de la literatura. *RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao*, *2022*(E49), 586–598.

Ozili, P. K. (2023). Forensic accounting research around the world. *Journal of Financial Reporting and Accounting*. https://doi.org/10.1108/JFRA-02-2023-0106

Phong, N. A., Tam, P. H., & Thanh, N. P. (2022). Fraud Identification of Financial Statements by Machine Learning Technology: Case of Listed Companies in Vietnam. In *Studies in Systems, Decision and Control* (Vol. 427). https://doi.org/10.1007/978-3-030-98689-6_28

Pitchayatheeranart, L., & Phornlaphatrachakorn, K. (2023). Forensic Accounting and Corporate Productivity in Thailand: Roles of Fraud Detection, Risk Reduction and Digital Capability. In *MANAGEMENT AND ACCOUNTING REVIEW* (Vol. 22).

Ramírez-Alpízar, A., Jenkins, M., Martínez, A., & Quesada-López, C. (2020). Use of data mining and machine learning techniques for fraud detection in financial statements: A systematic mapping study | Uso de técnicas de minería de datos y aprendizaje automático para la detección de fraudes en estados financieros: Un mapeo sistemá. *RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao*, *2020*(E28), 97–109. https://www.proquest.com/openview/b64fe75fd94f1f38e005e83624dbb0bc/1?pq-origsite=gscholar&cbl=1006393

Sinap, V. (2024). Comparative analysis of machine learning techniques for credit card fraud detection: Dealing with imbalanced datasets. *Turkish Journal of Engineering*, *8*(2), 196–208. https://doi.org/10.31127/tuje.1386127

Stripling, E., Baesens, B., Chizi, B., & vanden Broucke, S. (2018). Isolation-based conditional anomaly detection on mixed-attribute data to uncover workers' compensation fraud. *Decision Support Systems*, *111*, 13–26. https://doi.org/10.1016/j.dss.2018.04.001

Vijayakumar, V., Divya, N. S., Sarojini, P., & Sonika, K. (2020). Isolation Forest and Local Outlier Factor for Credit Card Fraud Detection System. *International Journal of Engineering and Advanced Technology*, *9*(4), 261–265. https://doi.org/10.35940/ijeat.D6815.049420

Wheeler, R., & Aitken, S. (2000). Multiple algorithms for fraud detection. *Knowledge-Based Systems*, *13*(2), 93–99. https://doi.org/10.1016/S0950-7051 (00)00050-2

Yang, G., Liu, X., & Li, B. (2023). Anti-money laundering supervision by intelligent algorithm. *Computers and Security*, *132*. https://doi.org/10.1016/j.cose.2023.103344

Yi, Z., Cao, X., Pu, X., Wu, Y., Chen, Z., Khan, A. T., Francis, A., & Li, S. (2023). Fraud detection in capital markets: A novel machine learning approach. *Expert Systems with Applications*, *231*, 120760. https://doi.org/10.1016/j.eswa.2023.120760