Hasina
JURNAL AKUNTANSI DAN BISNIS SYARIAH
Edisi 4: Vol. 3 Nomor 1 (Januari 2026)
e-ISSN: 3110-5971

Anomaly Detection in Financial Data: Comparative Analysis of Statistical Methods and Machine Learning in Simulation Studies

Pardomuan Robinson Sihombing[1]

[1] BPS Statistics Indonesia, robinson@bps.go.id

Abstract:

The increasing prevalence of financial fraud in the digital era necessitates the development of reliable and efficient detection methods. This study aims to conduct a systematic comparative analysis of the effectiveness and consistency of seven anomaly detection techniques applied to simulated financial transactional data. The examined methods encompass a broad spectrum of approaches, including Benford's Law, robust statistical techniques based on Median Absolute Deviation (MAD), supervised machine learning models such as Logistic Regression and Random Forest, and unsupervised machine learning methods including K-Means Clustering and Isolation Forest. This study employs a simulation-based approach using R software to generate a dataset consisting of 20,000 financial transactions, with 5% intentionally manipulated to represent fraudulent activities through structured fraud scenarios. The findings reveal that nearly all methods successfully identified anomalies with high detection performance. Supervised models, particularly Binary Logistic Regression, demonstrated near-perfect performance, while unsupervised methods such as K-Means Clustering and MAD-based Robust Distance achieved recall rates exceeding 95%. Despite the overall strong performance, variations were observed in the trade-off between recall and false positive rates, highlighting the importance of selecting detection methods aligned with specific business objectives and operational needs. This study concludes that a layered detection approach combining statistical screening techniques with advanced machine learning models provides the most comprehensive strategy for financial fraud mitigation.

Keywords:

Benford; Cluster; Fraud Detection; Isolation ForestLogistic

1. Introduction

Digital transformation has revolutionized the financial sector, offering unprecedented efficiency and accessibility, but it has also opened new avenues for illegal activities such as fraud (Huang, 2017) (Deepa & Murugesakumar, 2023). Losses due to fraud not only impact the financial stability of companies but also erode public trust. The scale of this problem is significant; the Association of Certified Fraud Examiners (ACFE, 2022) consistently finds in its reports that organizations typically lose 5% of their annual revenue due to various forms of fraud. As transaction volumes and speeds increase exponentially, manual auditing methods and static rule-based detection systems that were once relied upon are no longer adequate. These

systems are often too rigid to adapt to dynamic and increasingly sophisticated fraud patterns, allowing many anomalies to go undetected (Abdallah et al., 2016) (Ramadhan & Adhim, 2021).

This situation has prompted a paradigm shift towards using data analysis and *machine learning* as the front line in modern fraud defense systems. This approach has given rise to various techniques, ranging from classical statistical methods to complex *machine learning* algorithms. Previous studies have extensively explored and proven the effectiveness of applying various algorithms for this task. For example, a comprehensive study by West & Bhattacharya (2016) and a survey by Carneiro et al. (2017) demonstrate the effectiveness of various *machine learning* models*,* including Random Forest, in detecting fraud in financial reports and transactions. Similarly, in the context of the Indonesian financial landscape, research by Abdurrochman et al. (2024) successfully applied a *hybrid* method that combines *sampling* techniques with *machine* learning to effectively detect anomalies in imbalanced data. These studies are strong evidence that analytical approaches can significantly improve detection accuracy compared to traditional methods.

However, most existing studies tend to focus on evaluating one or two types of models in a particular class, for example, comparing several *supervised learning* models with each other (Zareapoor & Shamsolmoali, 2015). A gap in the literature comprehensively compares the consistency and relative performance of a broader spectrum of methods—including classical statistical methods (Benford's Law), robust statistics, *supervised learning,* and modern *unsupervised* learning—in a controlled experimental framework. Understanding the extent to which the results of these approaches complement or contradict each other is still very limited. Therefore, this study aims to fill this gap by systematically implementing, evaluating, and comparing seven representative anomaly detection methods. Using simulated data, this study will provide deep insights into the strengths, weaknesses, and potential synergies among various methods, which may ultimately guide the development of more holistic and robust fraud detection strategies.

2. Research Method

Digital transformation has revolutionized the financial sector, offering unprecedented efficiency and accessibility, but it has also opened new avenues for illegal activities such as fraud (Huang, 2017) (Deepa & Murugesakumar, 2023). Losses due to fraud not only impact the financial stability of companies but also erode public trust. The scale of this problem is significant; the Association of Certified Fraud Examiners (ACFE, 2022) consistently finds in its reports that organizations typically lose 5% of their annual revenue due to various forms of fraud. As transaction volumes and speeds increase exponentially, manual auditing methods and static rule-based detection systems that were once relied upon are no longer adequate. These systems are often too rigid to adapt to dynamic and increasingly sophisticated fraud patterns, allowing many anomalies to go undetected (Abdallah et al., 2016) (Ramadhan & Adhim, 2021).

This situation has prompted a paradigm shift towards using data analysis and *machine learning* as the front line in modern fraud defense systems. This approach has given rise to various techniques, ranging from classical statistical methods to complex *machine learning* algorithms. Previous studies have extensively explored and proven the effectiveness of applying various algorithms for this task. For example, a comprehensive study by West & Bhattacharya (2016) and a survey by Carneiro et al. (2017) demonstrate the effectiveness of various *machine learning* models*,* including Random Forest, in detecting fraud in financial reports and transactions. Similarly, in the context of the Indonesian financial landscape, research by Abdurrochman et al. (2024) successfully applied a *hybrid* method that combines *sampling* techniques with *machine*

learning to effectively detect anomalies in imbalanced data. These studies are strong evidence that analytical approaches can significantly improve detection accuracy compared to traditional methods.

However, most existing studies tend to focus on evaluating one or two types of models in a particular class, for example, comparing several *supervised learning* models with each other (Zareapoor & Shamsolmoali, 2015). A gap in the literature comprehensively compares the consistency and relative performance of a broader spectrum of methods—including classical statistical methods (Benford's Law), robust statistics, *supervised learning,* and modern *unsupervised* learning—in a controlled experimental framework. Understanding the extent to which the results of these approaches complement or contradict each other is still very limited. Therefore, this study aims to fill this gap by systematically implementing, evaluating, and comparing seven representative anomaly detection methods. Using simulated data, this study will provide deep insights into the strengths, weaknesses, and potential synergies among various methods, which may ultimately guide the development of more holistic and robust fraud detection strategies.

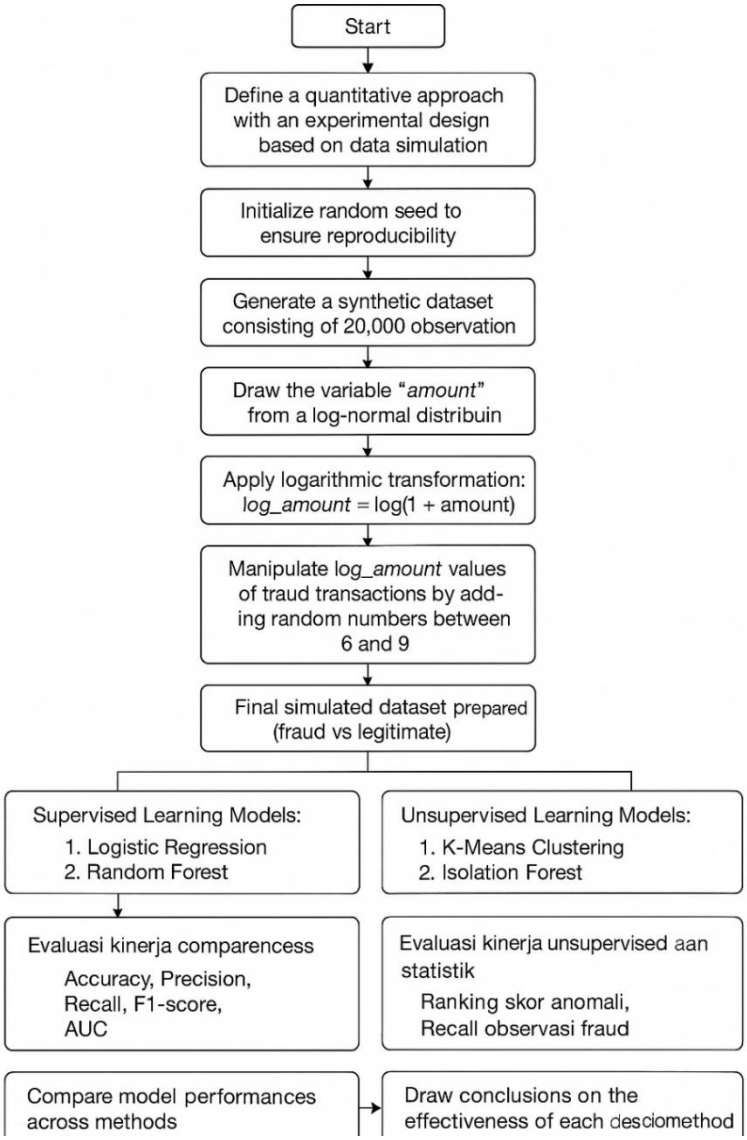

Figure 1. Flowchart of The Result Methodology

3. Result and Discussion

Descriptive analysis of the simulation data, as presented in Table 1, shows that the manipulation process successfully created a dramatic difference between legitimate and fraudulent transactions. The average value of fraudulent transactions (around 3.2 million) far exceeded the average value of legitimate transactions (around 1,236), which became the basis for detection models to perform discrimination.

Category	n	mean	sd	Min	Max
Valid	19000	1236.6	3423.5	1.3	129,527.9
Fraud	1000	3,231,631.0	9,553,631.9	2767.1	174926274.3

Table 1 . Descriptive Statistics Based on Transaction Category

Applying Benford's Law to Table 2 shows a significant deviation from the theoretical distribution, with a Chi-Square test result of $p < 0.001$. This result validates the role of Benford's Law as a sensitive initial screening tool.

Digit	FE (%)	FO (%)	diff (%)
1	30.1	30	-0.1
2	17.6	18	0.4
3	12.5	12.8	0.3
4	9.7	9.4	-0.3
5	7.9	7.8	-0.1
6	6.7	6.9	0.2
7	5.8	5.6	-0.2
8	5.1	5.2	0.1
9	4.6	4.2	-0.4
Chi-square ($p < 0.001$)			

Table 2 . Benford's Law Analysis Results

In the supervised *machine learning* domain, both models showed very high performance. Table 3 shows that Logistic Regression achieved 99.6% accuracy and 95.7% *recall*, with a perfect AUC of 1.000. Meanwhile, the Random Forest Model also performed excellently with 99.4% accuracy and 95.0% *recall*. This result indicates that both models can capture the relationship between predictors and fraud classes well, thanks to the obvious simulation scenario.

Metrics	Binary Logistic	Random Forest
Accuracy	0.996	0.994
Precision	0.963	0.922
Recall	0.957	0.950
F1-Score	0.960	0.936
AUC	1.000	0.988

Table 3. Performance Matrix of Binary Logistic Regression and Random Forest

The performance of the unsupervised method is also robust. Table 4 shows that the Robust Distance (MAD) method successfully identified 94.8% of total fraud by reviewing only 5% of data with the highest anomaly scores, outperforming Isolation Forest (82.7%). This result shows that robust statistical metrics such as MAD can detect anomalies defined by value deviation.

Method	X Fraud.Detected Top 5 Anomaly.Score.
Isolation Forest	82.7
Robust Distance (MAD)	94.8

Table 4. Anomaly Ranking Method Performance

The K-Means Clustering results in Table 5 show an almost perfect separation capability, with a fraud capture rate of 99.2%. However, this success comes at the "cost" of 238 legitimate transactions being misclassified as anomalies (*false positives*).

Cluster	Anomaly	Normal	Total
Valid	238	18,762	19,000
Fraud	992	8	1000
Total	1230	18,770	20,000
Accuracy	98.77	Recall	99.2

Table 5. Distribution of K-Means Clustering Results

The final comparative analysis in Table 7, which focuses on *recall* in the test data, presents interesting findings that require in-depth discussion. K-Means Clustering ranks highest with near-perfect *recall* (99.7%). This superior performance can be explained by the nature of the partitioning algorithm, which, in a highly separated data scenario such as this simulation, can effectively create a single *cluster* specifically for the fraud data group, whose distribution has been "shifted" far from the normal data. However, this *recall* advantage must be viewed in the context of Table 6, which shows that K-Means also produces 238 *false positives*. This result illustrates the classic trade-off in fraud detection: K-Means is very aggressive in capturing anomalies, but at the risk of falsely accusing legitimate transactions.

Method	% Fraud Detected (Recall on Test Data)
K-Means Clustering	99.7
Robust Distance (MAD)	96
Logistic Regression	95.70
Random Forest	95.00
Isolation Forest	83.30

Table 6. Comparison of Fraud Detection Rates (Recall) Among Methods

On the other hand, statistical methods such as Robust Distance (MAD) and *supervised* models (Logistic Regression, Random Forest) show very competitive *recall* performance (95-96%) with higher precision rates (as shown in Tables 3). The result aligns with the findings of Bolton & Hand (2002), who stated that *supervised* models often provide a better balance between detecting fraud and minimizing false alarms. The relatively lower performance of Isolation Forest (83.3%) in this metric can also be explained theoretically. This algorithm isolates *individual* anomalous *data points* (Liu et al., 2008). In our simulation, the fraud data formed a dense *cluster* despite being distant. This condition meant that some points within the fraud *cluster* were not "isolated" from each other, so their anomaly scores were slightly lower than if they were truly isolated *outliers*. These findings underscore the main argument in the anomaly detection literature: there is no single "best" method, and the choice of method depends heavily on the specific goal—whether to maximize fraud capture at all costs or achieve balanced and reliable detection.

4. Conclusion

This study successfully demonstrates that in data scenarios with clear separation between normal and anomaly classes, various supervised and unsupervised detection methods can achieve very high performance. The main finding shows that although *unsupervised* methods such as K-Means can achieve the highest recall, this often comes at the cost of lower precision. In contrast, *supervised* models offer more balanced performance. The success of all methods depends on proper data preprocessing, in this case, logarithmic transformation, which proved crucial for normalizing the data distribution. The practical implication of this study is that the choice of fraud detection method should be tailored to the business's risk tolerance for *false positives* versus *false negatives,* and a layered approach combining multiple techniques remains the most robust strategy.

For future research, several directions for development are suggested. First, this comparative framework will be tested on complex, high-dimensional real-world financial datasets, where fraud signals may be much more subtle and overlap with normal behavior. Second, the analysis can be expanded by incorporating other transactional features (e.g., location, time, frequency) to move from univariate to multivariate detection. Third, explore applying *deep learning* techniques such as *Autoencoders* or *Long Short-Term Memory* (LSTM) to capture more sophisticated sequential or non-linear fraud patterns that traditional algorithms cannot model.

5. References

Abdallah, A., Maarof, M. A., & Zainal, A. (2016). Fraud detection system: A survey. *Journal of Network and Computer Applications*, *68*, 90–113. https://doi.org/10.1016/j.jnca.2016.04.007

Abdurrochman, M., Wibowo, A., Rosal, D., & Setiadi, I. M. (2024). Optimized Machine Learning Model for Credit Card Fraud Detection Using Smote-Tomek and Feature Engineering. *Journal of Applied Informatics and Computing (JAIC)*, *8*(2), 580. http://jurnal.polibatam.ac.id/index.php/JAIC

ACFE. (2022). Occupational Fraud 2022: A Report To The Nations. In *Association of Certified Fraud Examiners*.

Ahmed, M., Mahmood, A. N., & Islam, M. R. (2016). A survey of anomaly detection techniques in financial domain. *Future Generation Computer Systems*, *55*, 278–288. https://doi.org/10.1016/j.future.2015.01.001

Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. *Statistical Science*, *17*(3), 235–255. https://doi.org/10.1214/ss/1042727940

Breiman, L., Friedman, J. ., Olshen, R. ., & Stone, C. . (1984a). *Classification and Regression Trees, The Wadsworth Statistics and Probability Series*. Wadsworth International Group.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984b). *Classification and Regression Trees*. Wadsworth International Group.

Carneiro, N., Figueira, G., & Costa, M. (2017). A data mining based system for credit-card fraud detection in e-tail. *Decision Support Systems*, *95*, 91–101. https://doi.org/10.1016/j.dss.2017.01.002

Deepa, P., & Murugesakumar, B. (2023). Role of Big Data Analysis in Predicting Financial Market. *Tuijin Jishu/Journal of Propulsion Technology*, *44*(6), 1001–4055. https://ssrn.com/abstract=3827106

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression* (3rd ed.). John Wiley & Sons.

Huang, S. C. (2017). A big data analysis system for financial trading. *Global Business and Finance Review*, *22*(3), 32–44. https://doi.org/10.17549/gbfr.2017.22.3.32

Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation forest. *Proceedings - IEEE International Conference on Data Mining, ICDM*, *January 2009*, 413–422. https://doi.org/10.1109/ICDM.2008.17

Nigrini, M. J., & Wells, J. T. (2012). *Benford's Law: Applications for forensic accounting, auditing, and fraud detection*. John Wiley & Sons.

Ramadhan, M. S., & Adhim, C. (2021). Fraud detection in the procurement of goods and services. *Journal of Contemporary Accounting*, *3*(3), 113–129. https://doi.org/10.20885/jca.vol3.iss3.art1

West, J., & Bhattacharya, M. (2016). Intelligent financial fraud detection: A comprehensive review. *Computers and Security*, *57*, 47–66. https://doi.org/10.1016/j.cose.2015.09.005

Zareapoor, M., & Shamsolmoali, P. (2015). Application of credit card fraud detection: Based on bagging ensemble classifier. *Procedia Computer Science*, *48*(C), 679–685. https://doi.org/10.1016/j.procs.2015.04.201