Detecting Suspicious Foreign Currency Transactions Using SOM and K-Means++ Algorithm

Michelle Elfalin, Sandy Kosasi

STMIK Pontianak, Pontianak, Indonesia

Correspondence : *e-mail*: piscesmesh14@gmail.com

Abstrak

Dalam konteks perekonomian global yang terus berkembang dan kompleks, aktivitas jual beli uang kertas asing melibatkan volume transaksi yang besar dan beragam, sehingga meningkatkan risiko terjadinya transaksi keuangan mencurigakan. Berbagai penelitian sebelumnya telah memanfaatkan metode klasifikasi dan clustering, namun sebagian masih terbatas pada satu algoritma sehingga kurang optimal dalam mendeteksi pola anomali secara akurat. Penelitian ini bertujuan untuk mengembangkan sistem deteksi transaksi mencurigakan menggunakan kombinasi algoritma Self-Organizing Map (SOM) dan K-Means++ agar dapat mengidentifikasi pola transaksi abnormal secara cepat dan akurat. Data transaksi periode 2021–2022 digunakan sebagai data latih, kemudian dianalisis melalui tahapan praproses, normalisasi, pelatihan model, dan evaluasi menggunakan confusion matrix untuk menghitung akurasi, presisi, recall, spesifisitas, dan F1-score. Hasil pengujian menunjukkan bahwa model mampu mendeteksi transaksi mencurigakan dengan tingkat akurasi 95,45% pada data latih dan 77,75% pada data uji, dengan recall tinggi yang menandakan sensitivitas deteksi terhadap transaksi fraud. Temuan ini menegaskan bahwa metode SOM dan K-Means++ efektif sebagai alat bantu identifikasi transaksi mencurigakan, sekaligus menyediakan landasan untuk pengembangan sistem deteksi otomatis yang lebih adaptif pada sektor keuangan.

Kata kunci: *Self-Organizing Map, K-Means++, deteksi transaksi, clustering, uang kertas asing.*

Abstract

In the context of an increasingly complex global economy, foreign currency trading involves large and diverse transaction volumes, thereby raising the risk of suspicious financial activities. Previous studies have applied classification and clustering methods, yet many remain limited to a single algorithm, resulting in suboptimal anomaly detection accuracy. This study aims to develop a suspicious transaction detection system by combining Self-Organizing Map (SOM) and K-Means++ to enhance the precision of abnormal pattern identification. Transaction data from the 2021–2022 period was used as training data through preprocessing, normalization, model training, and evaluation using a confusion matrix to calculate accuracy, precision, recall, specificity, and F1-score. The experimental results demonstrate that the proposed model successfully detected suspicious transactions with an accuracy of 95.45% on the training data and 77.75% on the test data, with a high recall indicating strong sensitivity toward fraudulent transactions. These findings confirm that the integration of SOM and K-Means++ is effective as a tool for detecting suspicious transactions, while also contributing to the development of more adaptive automated detection systems in the financial sector.

Keywords*: Self-Organizing Map, K-Means++, transaction detection, clustering, foreign banknotes.*

1. Introduction

Foreign exchange trading (forex trading) is a crucial financial activity in the increasingly complex global economy [1]. This activity involves the trading of foreign banknotes in international financial markets with the aim of gaining profit from exchange rate differences. Such transactions are massive, involving millions of activities daily, conducted by banks, financial institutions, licensed non-bank foreign currency exchange businesses (KUPVA BB), multinational corporations, individual speculators, and

investors [2]. However, behind this high transaction volume lies the risk of suspicious financial transactions (SFT), characterized by unclear economic purposes, repeated extensive cash usage, and unusual patterns.

PT Kenisha Anugerah Perdana, a licensed KUPVA BB in Pontianak authorized by Bank Indonesia (License No. 23/13/KEP.GBI/PTK/2021), faces challenges in detecting and reporting SFT. According to the Integrated Customer Service Information System (SIPESAT) from 2020 to 2024, a total of 7,127 customers with thousands of potentially suspicious transactions were recorded (Figure 1a). Manual reporting is prone to input errors, reporting delays, and inefficiencies in the audit process. The complexity increases with potential risks such as money laundering, terrorist financing, fraud, and other illegal activities. [3], as reflected in the Terrorist Financing Suspect Information System (SIPENDAR) watchlist, which recorded 7,526 individuals and 492 corporations during 2020–2024 (Figure 1b).

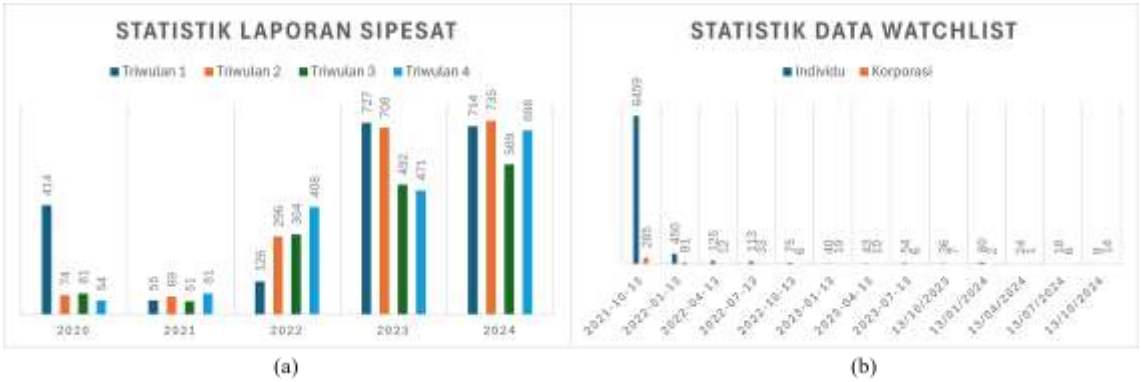

Figure 1 Statistical Data, (a) SIPESAT, (b) SIPENDAR
Source: (PT Kenisha Anugerah Perdana, 2024), (SIPENDAR, 2024)

The challenge of detecting SFT lies not only in transaction volume but also in the complexity of patterns that resemble normal activities, making them difficult to identify with conventional methods [4]. Recent studies have proposed hybrid approaches, such as Gaussian Mixture Models, SOM, and adaptive K-Means, which are more adaptive in detecting financial anomalies [5]. Another challenge is imbalanced datasets, which affect detection accuracy, as well as the need for data visualization that is easily interpretable by management and regulators [6]. These issues demand the application of machine learning technologies for effective, fast, and accurate data management and analysis, thereby enabling early warnings for high-risk transactions [7].

Previous studies on fraud and suspicious transaction detection employed methods such as Logistic Regression, Naive Bayes, and Random Forest, yet struggled to detect fraudulent transactions similar to legitimate ones [8]. Other studies used a Self-Organizing Map (SOM) to project data into a grid, facilitating outlier identification in credit card fraud [9]. Hybrid approaches such as SVESOM, which combines Support Vector Machine (SVM) and Emergent SOM (ESOM), have proven effective in handling imbalanced datasets compared to conventional methods [10]. This research introduces an innovative combination of SOM and K-Means++ in a web-based detection system developed using Python Streamlit and MySQL, capable of automated detection, visualization of suspicious transaction patterns, and reporting in compliance with Bank Indonesia regulations.

2. Research Method

This study adopts an experimental research method with a case study approach on foreign banknote trading transactions at PT Kenisha Anugerah Perdana. The method applies SOM and K-Means++ to transaction data, comparing results before and after clustering. The objective is to identify standard transaction patterns and detect anomalies systematically. The study is descriptive and exploratory, focusing on evaluating the accuracy of suspicious transaction detection.

The research procedure, shown in Figure 2, illustrates the workflow of the algorithms. The first stage is initial mapping with SOM [11], which includes random weight initialization (Equation 1), Euclidean distance calculation to determine the Best Matching Unit (BMU) (Equation 2), and weight updating of the BMU and its neighbors using Gaussian-based neighborhood functions (Equations 3–4). The second stage is clustering optimization with K-Means++ [12], which initializes centroids using probability distribution based on squared distances (Equations 5–6), assigns clusters to each data point (Equation 7), and updates centroids as the mean of cluster members until convergence (Equation 8).

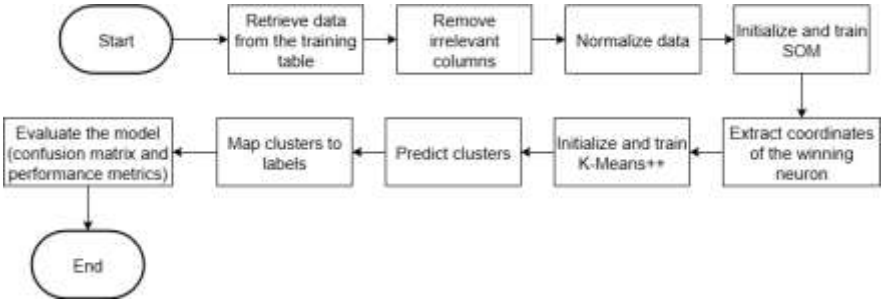

Figure 2 Algorithm Workflow

$$W_j = Random\ \forall j \tag{1}$$

$$D(X, W_j) = \sqrt{\sum_{i=1}^{n}(X_i - W_{j,i})^2},\ j^* = \arg\min_j D(X, W_j) \tag{2}$$

$$W_j(t+1) = W_j(t) + \alpha(t) \cdot h_{j,j^*} \cdot (X - W_j(t)) \tag{3}$$

$$h_{j,j} = exp\left(-\frac{d_{j,j}^2}{2\sigma^2}\right) \tag{4}$$

$$C_1 = X_{i_1}\ dengan\ i_1 \in \{1,2,\dots,N\} \tag{5}$$

$$D(X_i) = \min_{k\in\{1,\dots,K\}} \|X_i - C_k\|^2,\ P(X_i) = \frac{D(X_i)}{\sum_{i=1}^{N} D(X_i)} \tag{6}$$

$$Assign(X_i) = \arg\min_k \|X_i - C_k\|^2 \tag{7}$$

$$C_k = \frac{1}{|C_k|}\sum_{X_i \in C_k} X_i \tag{8}$$

Data collection was conducted in two forms: primary data, obtained from PT Kenisha Anugerah Perdana consisting of customer identity, transaction time, currency type, exchange rate, nominal value, and transaction total; and secondary data, derived from literature, journals, books, and documents relevant to suspicious transaction detection and the application of SOM and K-Means++.

System testing employed Scenario-Based Test-Cases [13], simulating fundamental user interactions with variations in transaction amounts, exchange rates, and time. Evaluation was conducted using confusion matrices to compute accuracy, precision, recall, specificity, and F1-score [14]:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{9}$$

$$Precision = \frac{TP}{TP+FP} \tag{10}$$

$$Recall = \frac{TP}{TP+FN} \tag{11}$$

$$Specificity = \frac{TN}{TN+FP} \tag{12}$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{13}$$

3. Result and Discussion

The development of a suspicious financial transaction (SFT) detection system began with a needs analysis at PT Kenisha Anugerah Perdana, which included mapping transaction flows and identifying obstacles in manual detection. The system was designed to be integrated with customer and transaction data management through Create, Read, Update, and Delete (CRUD) features. The main features include login with access rights, an interactive dashboard, data tables for transactions and customers, a suspicious transaction detection module based on SOM and K-Means++, statistical visualization, and monthly reports. The system is also equipped with user management, user guidance, developer information, and license display to support operational processes and legal compliance.

From an architectural perspective, the system was built using a Three-Tier Architecture approach based on the Streamlit framework. The frontend layer functions as a responsive and user-friendly interface. The backend is responsible for managing application logic, integrating detection algorithms, and connecting the system with the database. The data tier layer uses MariaDB as a centralized and structured data storage server (Figure 3). This approach provides advantages in terms of modularity, security, and maintainability, allowing the system to be further developed without disrupting its overall functionality.

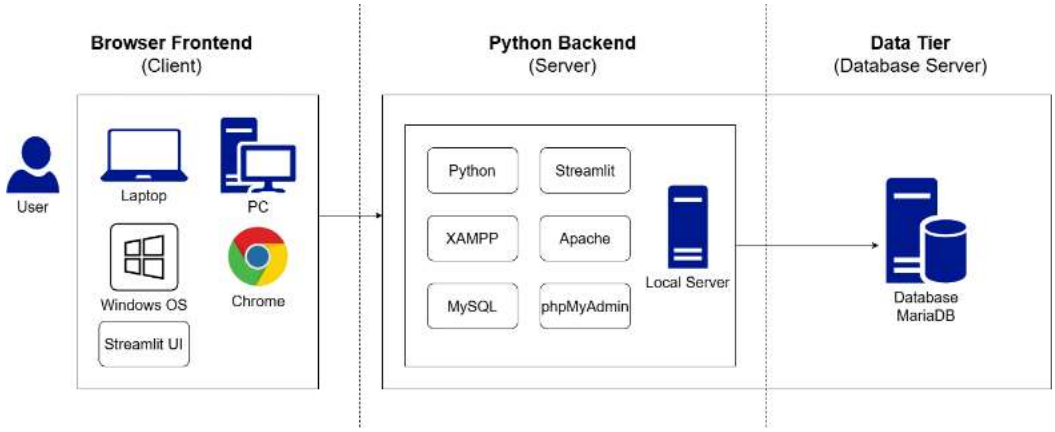

Figure 3 Software Architecture

The detection method was implemented through a combination of SOM and K-Means++. SOM was used for the feature extraction stage, namely, mapping high-dimensional transaction data into a two-dimensional representation that is easier to analyze. The process began with retrieving training data from the database, cleaning irrelevant columns, and normalizing values using MinMaxScaler. The SOM model was built with the MiniSom library using a 10×10 grid parameter, sigma 1.0, learning rate 0.1, and 3600 iterations. Training was carried out randomly (train_random) with Best Matching Unit (BMU) selection based on Euclidean distance. The BMU coordinates became the input for the clustering stage.

The next stage was clustering using the K-Means++ algorithm. The model was initialized with two clusters (fraudulent and non-fraudulent), n_init = 20, and a fixed random_state to ensure consistent results. After clustering, label mapping was performed using the majority method to align clustering results with the fraud detection context. This approach allows the system to perform learning without explicit labels during training, while also leveraging SOM's ability to map non-linear patterns and K-Means++'s strength in generating stable clusters.

System testing was conducted using a Scenario-Based Test Case approach to verify the functionality of each module, and model performance evaluation was carried out using a confusion matrix along with evaluation metrics. The test scenarios covered six aspects: model training using training data, evaluation of detection results, visualization of fraud percentages, display of confusion matrix, classification metrics assessment, and data summaries based on currency type (Table 1). All scenarios were successfully executed with results consistent with expectations, indicating that the system is stable and ready for operational implementation.

Table 1 Scenario Testing of SFT Detection System

No	Scenario Name	Action	Expected Output	Actual Output	Test Result
1	Model training from training data	Click "Start Training"	SOM+K-Means++ model trained, evaluation displayed (table & graph)	Model successfully trained, evaluation displayed	Valid
2	Evaluation of detection results	Use labeled training data	fraud_pred column matches the actual label	Column displayed, majority matches label	Valid
3	Visualization of fraud percentage		Pie chart fraud vs non-fraud	Pie chart fraud 28.6%, non-fraud 71.4%	Valid
4	Confusion Matrix		Matrix displayed (TP, TN, FP, FN correct)	TN=461, TP=169, FP=20, FN=10	Valid
5	Evaluation of classification metrics		Accuracy, Precision, Recall, Specificity, and F1 were displayed	Matches classification results	Valid
6	Summary by currency type		Summary of fraud/non-fraud by currency type and transaction type	Complete summary displayed (AUD, MYR, USD, etc.)	Valid

Model evaluation on the training data was carried out with varying numbers of training iterations to determine the optimal performance point. The best performance was achieved at the 3600th iteration with True Positive (TP) of 169, True Negative (TN) of 461, False Positive (FP) of 20, and False Negative

(FN) of 10 (Table 2). When tested on test data consisting of 3,564 transactions, the results differed. The model successfully classified 1,611 fraudulent transactions and 1,160 non-fraudulent transactions correctly, but also produced 775 FP (everyday transactions classified as fraudulent) and 18 FN (fraudulent transactions not detected) (Table 3).

Table 2 Confusion Matrix of Training Data

	Actual Positive	Actual Negative
Predicted Positive	169	20
Predicted Negative	10	461

Table 3 Confusion Matrix of Test Data

	Actual Positive	Actual Negative
Predicted Positive	1611	775
Predicted Negative	18	1160

Based on evaluation metrics, the model achieved 95.45% accuracy, 89.42% precision, 94.41% recall, 95.84% specificity, and 91.85% F1-score on the training data. These results demonstrate a good balance between precision and recall, as well as strong capability in identifying everyday transactions. However, on the test data, accuracy decreased to 77.75%, precision to 67.52%, recall increased to 98.90%, specificity declined to 59.95%, and F1-score to 80.25%. The very high recall value indicates that the model was effective in capturing the majority of suspicious transactions. However, the high false positive rate highlights an issue with precision caused by overfitting and imbalanced data distribution (Table 4).

Table 4 Model Evaluation Results on Training Data and Test Data

Dataset	Accuracy	Precision	Recall	Specificity	F1-Score
Training Data	95,45%	89,42%	94,41%	95,84%	91,85%
Test Data	77,75%	67,52%	98,90%	59,95%	80,25%

The strength of the SOM method lies in its ability to detect patterns without relying on labels during training (unsupervised learning), while its integration with K-Means++ helps stabilize centroid initialization. This finding is consistent with studies comparing the effectiveness of SOM and K-Means in detecting Value Added Tax (VAT)-based fraud, where SOM was proven to be superior in identifying non-linear anomaly patterns [15]. Its limitations include high sensitivity to training parameters and potential overfitting if the training data is too specific. This result is consistent with previous studies, which reported that SOM is effective for anomaly detection, but its performance may decline when applied to test data with different distributions [16]. Recent research has even proposed a variant called Growing SOM (GSOM), which improves efficiency in large-scale data processing [17]. Other studies also emphasize that SOM variants can help address imbalanced data issues, although additional techniques are still required to improve precision [18]. In addition, prior research confirmed SOM's usefulness in anomaly pattern visualization to support detection analysis [9]. Another study also demonstrated the application of SOM in industrial fault detection, highlighting its superiority in terms of clustering interpretability [19].

Nevertheless, most previous studies have focused on single-method applications. Therefore, this research presents novelty in the integration of the SOM and K-Means++ algorithms to detect suspicious financial transactions in foreign banknote trading. By combining SOM's strength in extracting non-linear features with K-Means++'s ability to generate more stable centroid initialization, the developed detection system can provide more representative clustering results. Furthermore, the system has been realized in the form of a web-based application using Python Streamlit integrated with a MySQL database, making it directly applicable for KUPVA BB providers to support transaction reporting in compliance with Bank Indonesia regulations.

4. Conclusion

This study developed a suspicious transaction detection system for foreign banknote trading at PT Kenisha Anugerah Perdana by integrating SOM for feature extraction and K-Means++ for clustering optimization. Experimental results on training data demonstrated high performance with 95.45% accuracy, 89.42% precision, 94.41% recall, 95.84% specificity, and 91.85% F1-score. On testing data, accuracy dropped to 77.75%, with recall remaining high (98.90%) but lower precision (67.52%) and specificity (59.95%) due to false positives. The system successfully automated suspicious transaction detection,

visualized transaction patterns, and generated reports in compliance with Bank Indonesia regulations. Future work should consider data balancing techniques, adaptive thresholds, and ensemble methods to improve precision without compromising sensitivity. Similar approaches, such as SMOTE-KMeans and ensemble learning, have demonstrated strong results, achieving an AUC of 0.96 in large-scale fraud detection [20].

References

[1] Y. Safitri and D. Z. Putri, 'Analisis Determinan Cadangan Devisa di Indonesia', *Jurnal Kajian Ekonomi dan Pembangunan*, vol. 3, no. 4, pp. 97–108, 2021, [Online]. Available: http://ejournal.unp.ac.id/students/index.php/epb/index

[2] Bank Indonesia, *Peraturan Bank Indonesia Nomor 18/20/PBI/2016 Tahun 2016 tentang Kegiatan Usaha Penukaran Valuta Asing Bukan Bank*. Indonesia: LN 2016/NO 194; PERATURAN.GO.ID : 30 HLM, 2016. Accessed: Sep. 20, 2024. [Online]. Available: https://peraturan.bpk.go.id/Details/135678/peraturan-bi-no-1820pbi2016-tahun-2016

[3] Y. Aditia, *Financial Integrity Rating on Money Laundering and Terrorist Financing*. Jakarta: Pusat Pelaporan dan Analisis Transaksi Keuangan (PPATK), 2021.

[4] S. M. Darwish, A. I. Salama, and A. A. Elzoghabi, 'Intelligent Approach to Detecting Online Fraudulent Trading with Solution for Imbalanced Data in Fintech Forensics', *Sci Rep*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-01223-8.

[5] A. Vaid, C. Reddy, and S. Prabhagaran, 'A Hybrid Framework for Dynamic Clustering and Anomaly Detection in SAP ERP Systems', *International Journal of Computer Science and Mobile Computing*, vol. 13, no. 12, pp. 23–34, Dec. 2024, doi: 10.47760/ijcsmc . 2024.v13i12.003.

[6] Y. Chen, C. Zhao, Y. Xu, C. Nie, and Y. Zhang, 'Year-over-Year Developments in Financial Fraud Detection via Deep Learning: A Systematic Literature Review', *Journal of Financial Data Science*, vol. 7, no. 2, pp. 45–59, Jul. 2025, [Online]. Available: http://arxiv.org/abs/2502.00201

[7] W. N. Dilla and R. L. Raschke, 'Data visualization for fraud detection: Practice implications and a call for future research', *International Journal of Accounting Information Systems*, vol. 16, pp. 1–22, Mar. 2015, doi: 10.1016/j.accinf.2015.01.001.

[8] S. Bagga, A. Goyal, N. Gupta, and A. Goyal, 'Credit Card Fraud Detection using Pipeling and Ensemble Learning', in *Procedia Computer Science*, Elsevier B.V., 2020, pp. 104–112. doi: 10.1016/j.procs.2020.06.014.

[9] D. Olszewski, 'Fraud Detection Using Self-Organizing Map Visualizing the User Profiles', *Knowl Based Syst*, vol. 70, pp. 324–334, Nov. 2014, doi: 10.1016/j.knosys.2014.07.008.

[10] Y. Y. Nguwi and S. Y. Cho, 'An Unsupervised Self-Organizing Learning with Support Vector Ranking for Imbalanced Datasets', *Expert Syst Appl*, vol. 37, no. 12, pp. 8303–8312, 2010, doi: 10.1016/j.eswa.2010.05.054.

[11] I. Nunes *et al.*, *Artificial Neural Networks: A Practical Course*. Switzerland: Springer International, 2017. doi: 10.1007/978-3-319-43162-8.

[12] D. Arthur and S. Vassilvitskii, 'K-Means++: The Advantages of Careful Seeding', in *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, New Orleans, Louisiana: Society for Industrial and Applied Mathematics (SIAM), Jan. 2007, pp. 1027–1035.

[13] N. B. N. Nyakundi, S. M. Reynolds, and H. Reza, 'Scenario-Based Approach to Systematically Derive Test Cases for Systems', in *2023 IEEE International Conference on Electro Information Technology (eIT)*, IEEE, May 2023, pp. 51–58. doi: 10.1109/eIT57321.2023.10187246.

[14] O. Rainio, J. Teuho, and R. Klén, 'Evaluation Metrics and Statistical Tests for Machine Learning', *Sci Rep*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-56706-x.

[15] F. Bankole and Z. Vara, 'Artificial Intelligence System for Value Added Tax Collection via Self Organizing Map (SOM)', *J Forensic Sci & Criminal Inves*, vol. 18, no. 2, 2024, doi: 10.19080/JFSCI.2024.18.555981.

[16] A. Nowak-Brzezinska and C. Horyn, 'Self-Organizing Map Algorithm as a Tool for Outlier Detection', in *Procedia Computer Science*, Elsevier B.V., 2022, pp. 6162–6171. doi: 10.1016/j.procs.2022.09.276.

[17] S. Ruiz-Moreno, A. Núñez-Reyes, A. García-Cantalapiedra, and F. Pavón, 'Prototype Generation Method using a Growing Self-Organizing Map Applied to the Banking Sector', *Neural Comput Appl*, vol. 35, no. 24, pp. 17579–17597, Aug. 2023, doi: 10.1007/s00521-023-08630-w.

[18] G. Douzas, R. Rauch, and F. Bacao, 'G-SOMO: An Oversampling Approach Based on Self-Organized Maps and Geometric SMOTE', *Expert Syst Appl*, vol. 183, Nov. 2021, doi: 10.1016/j.eswa.2021.115230.

[19] L. Concetti, G. Mazzuto, F. E. Ciarapica, and M. Bevilacqua, 'An Unsupervised Anomaly Detection Based on Self-Organizing Map for the Oil and Gas Sector', *Applied Sciences (Switzerland)*, vol. 13, no. 6, Mar. 2023, doi: 10.3390/app13063725.

[20] Y. Wang, 'A Data Balancing and Ensemble Learning Approach for Credit Card Fraud Detection', in *2025 4th International Symposium on Computer Applications and Information Technology (ISCAIT)*, New York: IEEE, Mar. 2025, pp. 386–390. doi: https://doi.org/10.48550/arXiv.2503.21160.