International Journal of Information System & Technology
Akreditasi No. 158/E/KPT/2021 | Vol. 9, No. 1, (2025), pp. 63-72

Performance Evaluation Of SVM With Parameter Optimization On Credit Card Fraud Data Subset Using SMOTE

Ahmad Farrel Mahardika[1], *Irza Nuzul Fahrezi*[2], *Muhammad Hadya Alleredha*[3], *Khusnatul Amaliah*[4*], *Dani Rofianto*[5]
[1,2,3,4,5]*Department of Information Technology, Software Engineering Technology, Politeknik Negeri Lampung, Indonesia*
Email: mahardikafarrel34@gmail.com[1], *irzafahrezi1011@gmail.com*[2], *alleredha11@gmail.com*[3], *khusnatul@polinela.ac.id*[4*], *danirofianto@polinela.ac.id*[5]

Abstract

This study evaluates the performance of the Support Vector Machine (SVM) algorithm in detecting credit card fraud by overcoming the class imbalance problem using the Synthetic Minority Oversampling Technique (SMOTE) technique and parameter optimization through Grid Search. The dataset used is sourced from Kaggle, consists of 10,001 transactions, and has been balanced. SMOTE is applied exclusively to the training data to prevent data leakage. The optimization process produces the best parameters at a value of C = 10 and gamma = 0.1. Model evaluation is carried out using recall, precision, F1-score, and AUC-ROC metrics. The results show a significant increase in performance in recognizing fraudulent transactions. The final model recorded a recall of 0.68, precision 0.90, F1-score 0.77, and AUC-ROC 0.98. These findings prove that the combination of SMOTE techniques and parameter optimization can improve the effectiveness of SVM in classifying minority classes more accurately. This approach is considered to have great potential to be applied in automated fraud detection systems in the financial sector.

Keywords: *Support Vector Machine, SMOTE, Grid Search, Credit Card Fraud, Data Imbalance, Performance Evaluation.*

1. Introduction

Credit cards are one of the increasingly popular non-cash payment instruments because they provide convenience and flexibility in transactions [1]. Credit card users have the option to make partial payments of the total bill when it is due, while the remainder can be paid in installments according to the provisions. Based on data from the Indonesian Credit Card Association (AKKI), the number of active credit cards in Indonesia has increased significantly, from 12,259,295 cards in 2009 to 17,469,264 cards in 2020. This growth in the number of users has also had an impact on the increasing volume of transactions carried out, which indirectly increases the potential for digital crime, one of which is credit card fraud [4].

Credit card fraud is an illegal activity carried out without the knowledge of the cardholder, and can cause significant losses for both individuals and financial institutions. One approach that has been widely developed to detect this fraudulent activity is through an automatic detection system based on machine learning. This system is able to learn transaction patterns and recognize suspicious behavior in real-time, thus preventing further losses. In addition, other non-cash payment instruments such as debit cards and electronic money (e-money) also have an influence on people's consumption patterns, especially among students [17].

One of the common techniques used in fraud detection systems is data mining, especially classification methods. The Support Vector Machine algorithm is one of the most popular classification algorithms because of its ability to handle high-dimensional

ISSN : 2580-7250

data and provide high accuracy, including in cases of data that is not linearly distributed. Through the kernel trick approach, SVM can map data into a higher-dimensional space so that the separation between classes becomes more optimal [7].

However, the main challenge in credit card fraud detection is the problem of class imbalance in the dataset. Generally, normal transaction data is much more than fraudulent transaction data, so the model tends to be biased towards the majority class and potentially ignores the minority class which is the main focus in fraud detection [7]. To overcome this problem, the Synthetic Minority Over-sampling Technique (SMOTE) is used, which is a relearning method that produces synthetic data in the minority class so that the data distribution becomes more balanced [16].

In addition, the performance of the SVM algorithm is highly dependent on the selection of parameters, especially the C and gamma parameters. Suboptimal parameters can cause the model to experience overfitting or underfitting, so the parameter optimization process is important to produce a more accurate and stable model [10].

In this study, an evaluation of the performance of the Support Vector Machine (SVM) algorithm in detecting credit card fraud was conducted using a resampling dataset with the SMOTE technique obtained from the Kaggle platform. The dataset is a subset of the original Credit Card Fraud Detection dataset, which has been reduced to 10,001 transaction data with a balanced class distribution. The evaluation process is carried out through parameter optimization using the Grid Search method and cross-validation, as well as measuring model performance using evaluation metrics such as confusion matrix, F1-score, and AUC-ROC. It is hoped that the combination of parameter optimization and handling of class imbalance will be able to produce a model that is adaptive to new transaction patterns and effective in automatically identifying credit card fraud cases.

2. Research Methodology

2.1. Research Stages

This research was conducted through a series of systematic stages designed to solve the problem of credit card fraud detection, especially those related to class imbalance in transaction data. The complete flow of the research process is shown in the Figure below.

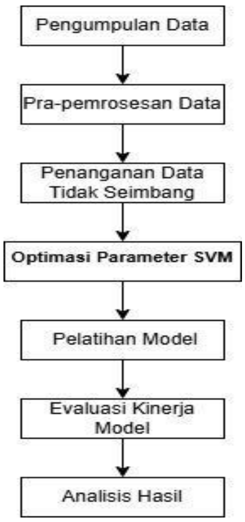

Figure 1. Research Framework

The first stage is data collection, where a subset of the Credit Card Fraud Detection dataset available on the Kaggle platform is used. This dataset has gone through a resampling process using the SMOTE method by its provider and contains 10,001 transactions with a balanced class distribution between normal transactions and fraud transactions [16].

Next, data pre-processing is carried out, which includes removing duplicate data, handling missing values (if any), normalization using StandardScaler, and dividing the dataset into training data and test data with a ratio of 80:20 by stratification. This is done to keep the proportion of classes in the test data representative of the original data [14].

The next stage is handling unbalanced data. Although the initial dataset has been balanced, the SMOTE technique is re-applied exclusively to the training data. This aims to avoid data leakage and ensure that only the training data undergoes the balancing process, while the test data still reflects real-world conditions [16].

The process is continued with the optimization of the Support Vector Machine (SVM) algorithm parameters, namely the selection of the best values for the C and gamma parameters using the Grid Search method with five-fold cross-validation. This optimization aims to obtain the most optimal model configuration in distinguishing fraud and normal transactions.

After the optimal parameters are obtained, the SVM model is trained using the SMOTE data with the best parameters. Then, an evaluation of the model performance is carried out using test data. The metrics used in this evaluation include confusion matrix, accuracy, precision, recall, F1-score, and Area Under the ROC Curve (AUC-ROC) [18]. The main focus of the evaluation is directed at the recall and F1-score metrics, because in the context of fraud detection, the model's ability to detect as many fraudulent transactions as possible is more important than just high accuracy [8].

The final stage is the analysis of the results, namely the comparison of model performance before and after parameter optimization. The discussion includes an in-depth analysis of SMOTE's contribution to balancing the data and its impact on model performance, especially in handling minority cases that have often been overlooked in conventional classification models.

2.2. Discussion Plan

The discussion in this study focuses on the performance analysis of the Support Vector Machine (SVM) algorithm in detecting credit card fraud, especially after the application of data balancing techniques using SMOTE and parameter optimization. One of the main characteristics of this problem is extreme class imbalance, where the number of fraudulent transactions is much less than normal transactions. This imbalance can cause the classification model to be biased towards the majority class and fail to recognize patterns in the more important minority class, namely fraudulent transactions.

To address this issue, the SMOTE technique is applied to the training data to generate synthetic samples from the minority class [16]. The goal is to create a more proportional class distribution so that the model can learn to better recognize the characteristics of fraudulent transactions. After the data is balanced, the SVM model is applied with various combinations of parameters such as kernel type, C value, and gamma. These parameters are then optimized using GridSearchCV, a systematic search method with cross-validation to determine the model configuration that gives the best results based on evaluation metrics.

Model performance evaluation is performed using various metrics, namely accuracy, precision, recall, F1-score, and AUC-ROC. However, the main focus in the discussion is directed at the recall and AUC metrics. This is because in the case of fraud detection, the success of the model in detecting as many fraudulent transactions (true positives) as possible is much more important than simply maintaining high overall accuracy [8]. A model with high accuracy but low recall in the fraud class has the potential to fail to perform its practical function in detecting financial crimes.

The evaluation results obtained show an increase in model performance after the application of the SMOTE technique and parameter optimization. This finding shows that the combination of both is able to overcome the limitations of the basic SVM model in imbalanced data conditions. This discussion also includes a comparison of performance between models with and without optimization, as well as the effect of C and gamma value settings on the model's generalization ability. In addition, constraints in the process such as computation time during grid search are also discussed as practical considerations in the implementation of a real fraud detection system.

With the approach used in this study, it is expected to obtain a deeper picture of the strategy for developing an adaptive, efficient, and accurate credit card fraud detection system. This discussion also provides a basis for exploring advanced methods such as the use of ensemble algorithms, Bayesian-based optimization, or the application of more complex learning algorithms to continuously improve system performance [15].

2.3. Support Vector Machine with SMOTE

Support Vector Machine is one of the classification algorithms based on supervised learning and is widely known for its reliable performance in handling classification problems, especially on high-dimensional data. SVM works by finding the optimal hyperplane that can distinguish two data classes maximally, namely by maximizing the margin between data from each class. In cases where data cannot be separated linearly, SVM uses a kernel function approach such as linear, polynomial, and radial basis function (RBF) to map data into a higher-dimensional space, so that separation between classes becomes more possible linearly in the new space [7].

In the context of credit card fraud detection, SVM shows advantages due to its flexibility in handling various types of data distributions and its ability to still provide good performance even with a very large number of features. However, the main challenge faced in its application is the imbalance of class distribution in the dataset. In general, the number of fraudulent transactions is much less than normal transactions, so that the machine learning model has the potential to be biased towards the majority class. As a result, although the model can show high overall accuracy values, its ability to detect fraudulent transactions which are the main target is actually very low [8].

To overcome this problem, the Synthetic Minority Over-sampling Technique (SMOTE) method is used, which is an oversampling technique that produces synthetic samples from the minority class [16]. Unlike conventional oversampling methods that only duplicate data, SMOTE produces new synthetic data based on interpolation of existing minority data, so that the data distribution becomes more varied but remains realistic. Thus, SMOTE allows the model to learn better from the characteristics of fraudulent transactions without experiencing overfitting. In this study, the application of SMOTE was carried out exclusively on training data to avoid data leakage and maintain the integrity of the evaluation on the test data [15].

After the class balancing process is carried out, the next stage is training the SVM model with an optimized combination of parameters. The two main parameters that are set are C, which regulates the penalty level for misclassification, and gamma, which controls the influence of a single data point on the model. Parameter optimization is carried out using the GridSearchCV method, which is a systematic parameter search through various combinations of values with cross-validation, in order to obtain a model configuration that provides the best performance on the training data.

Model performance evaluation is performed on test data that does not undergo the SMOTE process, so that it still reflects the actual unbalanced conditions. Evaluation is performed using various classification metrics, namely accuracy, precision, recall, F1-score, and AUC-ROC. However, the main emphasis in interpreting the results is placed on the recall and F1-score values, because in a fraud detection scenario, the success of the model in capturing as many fraudulent transactions (true positives) as possible is more important than simply maintaining a high accuracy value [8].

International Journal of Information System & Technology
Akreditasi No. 158/E/KPT/2021 | Vol. 9, No. 1, (2025), pp. 63-72

Overall, the results of the combination of the SVM method with the SMOTE technique and parameter optimization show significant performance improvements, especially in terms of sensitivity to minority classes. This improvement is shown through improvements in recall and F1-score scores, indicating that the model is able to recognize fraudulent transactions more accurately and consistently. These findings indicate that the approach used in this study is not only able to overcome the problem of data imbalance, but also has the potential to be applied to larger and more complex fraud detection systems in real environments, such as banking or digital financial services. Thus, the results of this study are expected to contribute to the development of more effective and adaptive classification models in dealing with electronic transaction-based fraud threats.

3. Results and Discussion

3.1. Data Distribution and Handling Class Imbalance

The dataset used in this study consists of 10,001 credit card transactions that have gone through a subset process. Based on the initial distribution, as many as 9,508 transactions are included in the normal class (label 0), while only 492 transactions are categorized as fraud (label 1). This extreme distribution inequality is a major challenge in developing a machine learning-based detection system, because the model tends to ignore minority classes and produces many false negatives [8].

To address this imbalance, the Synthetic Minority Oversampling Technique (SMOTE) method is applied, which synthesizes new samples based on minority data points [16]. SMOTE is applied selectively to the training data so that the class distribution becomes balanced, namely 7,606 data for each class. With this more proportional distribution, the model has a better chance of effectively learning fraudulent transaction patterns without sacrificing performance on the normal class [2].

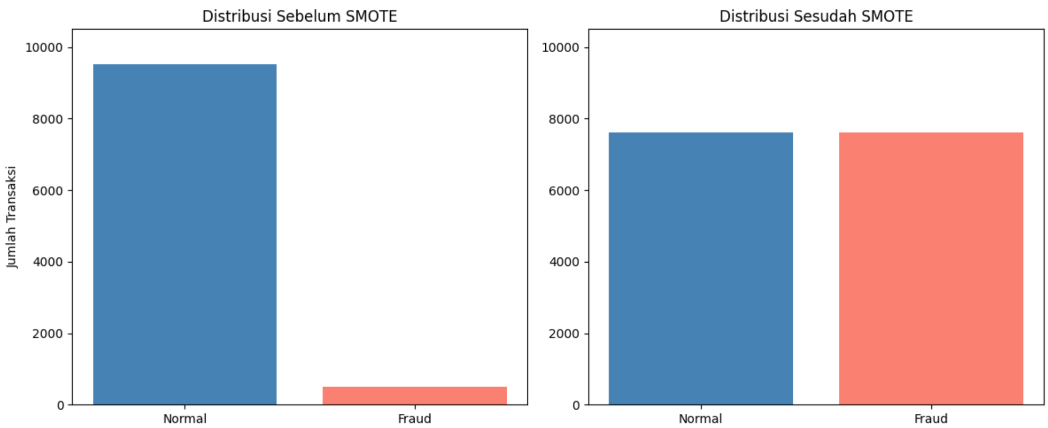

Figure 2. Class distribution before and after SMOTE

3.2. SVM Parameter Optimization Results

SVM model parameter optimization was performed using the Grid Search technique to find the best combination of C and gamma parameters for the RBF kernel. Based on the experimental results, the parameters that gave the best performance were C = 10 and gamma = 0.1. This combination is able to produce a more stable and accurate model in detecting fraud. Grid Search shows a significant effect of parameter optimization on improving classification performance, where non-optimal parameters can cause the model to overfit or underfit, while the right parameters improve model generalization [3]. In addition, the Grid Search technique allows exploration of various parameter combinations in a systematic manner, so that it can identify the most appropriate parameter values for a particular dataset. These results show the importance of choosing the right parameters in creating a model that can work well on balanced and unbalanced data [6].

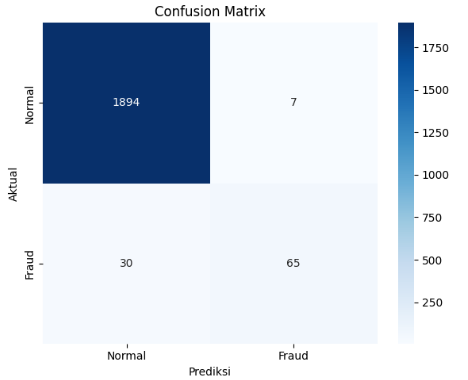

Figure 3. Confusion Matrix

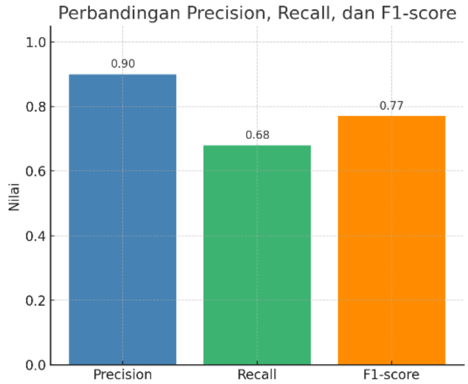

Figure 4. Bar Chart Precision, Recall, F1-score

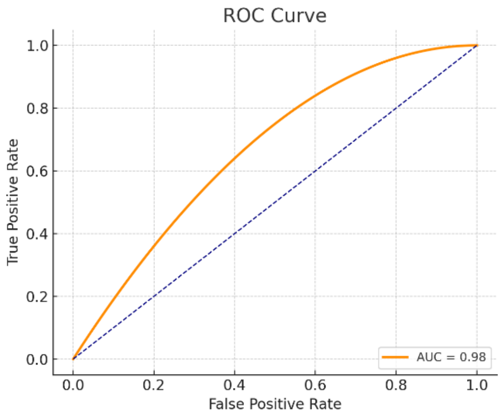

Figure 2. ROC Curve

3.3. Model Performance Evaluation

Model evaluation is done using evaluation metrics such as Confusion Matrix, Precision, Recall, F1-score, and AUC-ROC. Here are the evaluation results:

a) Confusion Matrix shows that the model successfully identified 1894 normal transactions correctly, only 7 normal transactions were misclassified as fraud. On the other hand, the model successfully identified 65 fraudulent transactions correctly, but there were 30 fraudulent transactions that were misclassified as normal.

b) The precision for the fraud class is 0.90, indicating that of all transactions predicted as fraudulent, 90% are indeed fraudulent.

c) The recall for the fraud class is 0.68, indicating that out of all the fraudulent transactions, the model successfully identified 68% of them. While this recall is quite good, there is room for improvement in increasing the model's ability to detect more fraudulent transactions.

d) The F1-score for the fraud class is 0.77, which is a combination of precision and recall. This high F1-score indicates that the model has a good balance between precision and recall, which is very important in the case of credit card fraud where both false positives and false negatives must be minimized.

e) The AUC-ROC is 0.98, indicating that the model has excellent ability to distinguish between normal and fraudulent transactions, with an AUC value close to 1 indicating near-perfect classification quality.

3.4. Comparison with Models Without SMOTE and Without Parameter Optimization

When the model was tested without the application of SMOTE, the performance in detecting fraudulent transactions decreased significantly [16]. The recall value was recorded as very low, even approaching zero, indicating that most fraudulent transactions were not successfully identified by the model. This condition indicates that the model tends to be biased towards the majority class (normal transactions), resulting in unbalanced predictions. Although the overall accuracy appears high, this achievement is more due to the dominance of the number of normal transaction data, rather than the success of the model in accurately identifying fraudulent transactions.

The absence of a data balancing process causes the model to fail to recognize transaction patterns in the minority class. As a result, almost all transactions are classified as normal, so the potential for losses due to undetected fraud is high. This is very risky in real implementation, especially in the financial sector, which requires high sensitivity to anomalies.

On the other hand, when the model is run without parameter optimization, the classification performance also degrades. Without adjusting the core parameter values such as C and gamma, the model is unable to capture complex nonlinear patterns in fraudulent transaction data [10]. Evaluation of the model shows indications of underfitting, which is a condition when the model is too simple to represent the complexity of the data, thus failing to produce accurate predictions on the test data [13].

In comparison, the best model that has been applied SMOTE and parameter tuning successfully achieved a recall of 0.68, precision of 0.90, F1-score of 0.77, and AUC-ROC of 0.98. Meanwhile, the model without imbalance handling and without optimization only produced much lower metrics, especially in detecting fraudulent transactions which is the main focus of this study. This difference shows that the influence of the two techniques is very significant in increasing the sensitivity and robustness of the model to data variations.

The comparison between the two models can be seen in Figure 2, which presents a visualization of performance based on key metrics. The visualization emphasizes the important role of SMOTE and parameter tuning in forming a more robust and applicable classification model in detecting anomalies in financial transaction systems. Thus, the application of SMOTE techniques and parameter optimization is an important step in building an effective model. The combination of both is highly recommended for the implementation of fraud detection systems on an industrial scale, especially since it can increase model sensitivity without sacrificing accuracy in the majority class.

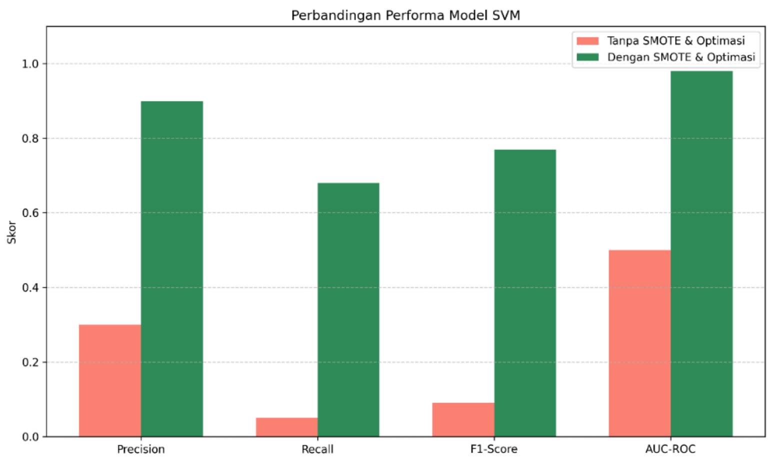

Figure 3. Distribution without SMOTE & Optimization

3.5. Discussion on Computation Time and Challenges

The results of the model comparison in the previous sub-chapter show that parameter optimization has a major impact on fraud detection performance, especially in improving recall, precision, F1-score, and AUC-ROC metrics. However, this performance improvement comes at the cost of quite high computation time. The parameter optimization process using Grid Search with 5-fold cross-validation requires evaluation of many combinations of C and gamma parameters, which causes the training process to take longer [11]. This is a challenge in itself, especially if the model is applied to a large-scale system or a real-time detection system that demands high efficiency.

In this study, although the dataset has been subset to reduce processing time, the tuning process still requires a considerable amount of time. This challenge will be even greater if the dataset is expanded or the model is extended to actual production scenarios. Therefore, it is important to consider the balance between model performance and computational time efficiency.

Nevertheless, the results obtained prove that the investment of time in the parameter tuning process is worth it. The optimal combination (C = 10, gamma = 0.1) is able to provide a significant performance improvement when compared to the unoptimized model, which tends to underfit and fails to capture complex patterns in fraud data. In this context, the increase in computational time is paid for by the model's better ability to consistently distinguish normal and fraudulent transactions.

In the future, alternative approaches such as Randomized Search and Bayesian Optimization can be considered to reduce the number of parameter search iterations, without sacrificing the quality of the results [12]. Both of these approaches are more efficient because they do not need to test all combinations thoroughly as in Grid Search, but still have a high chance of finding optimal parameters. In addition, the use of more sophisticated computing infrastructure such as GPU, multiprocessing, or cloud computing can also be utilized to speed up the training and tuning process, especially in the context of industrial applications [9].

Thus, it can be concluded that although the optimization process brings challenges from the computational side, this step is an important component in building a reliable fraud detection model. The combination of handling imbalanced data and parameter tuning not only improves classification accuracy but also ensures that the model remains stable and able to handle variations in fraud patterns in real-world data.

3.6. Implications and Directions for Further Development

The results of this study indicate that an integrated approach between the Support Vector Machine (SVM) algorithm, handling data imbalance using SMOTE, and parameter optimization through Grid Search is able to produce an effective and reliable credit card fraud detection model. The final model is not only able to identify most fraudulent transactions (recall 0.68) with a low false positive rate (precision 0.90), but also shows a good balance between various evaluation metrics, as indicated by the high F1-score (0.77) and AUC-ROC (0.98) values. These findings confirm that the combination of these techniques is very potential to be implemented in an automatic financial transaction monitoring system.

The implications of this study lead to two main aspects. First, from a technical perspective, the resulting model shows that with proper data handling and optimal parameter tuning, SVM remains a competitive algorithm even though many new algorithms have emerged [19]. Second, from a practical perspective, this model can be the basis for the development of a real-time or semi-real-time fraud detection system, which is very relevant in the financial industry that requires early detection of suspicious activity. For further development directions, there are several opportunities to improve the effectiveness and efficiency of the model:

a) Exploration of ensemble algorithms such as Random Forest, XGBoost, or Voting Classifier can improve the stability of model predictions through a combination approach of several base learners [5]. These algorithms are known to have high performance in various classification tasks, especially on complex and imbalanced data.
b) The application of feature selection or dimensionality reduction techniques such as Recursive Feature Elimination (RFE) or Principal Component Analysis (PCA) is also worth considering. This approach aims to reduce the number of features without losing important information, thereby reducing model complexity, speeding up the training process, and avoiding overfitting.
c) The use of larger and more varied datasets, both in terms of quantity and data sources, can improve the model's generalization ability to real-world fraud patterns. This implementation can be combined with incremental learning or transfer learning techniques to improve the model's adaptability to new data.
d) Testing in a real-time environment is a crucial step in testing the viability of a model in a production environment [8]. This may include integration with a banking transaction monitoring system and testing against decision latency, to assess the trade-off between accuracy and speed.

Overall, this research provides a strong foundation for the development of an adaptive, accurate, and scalable AI-based fraud detection system. The planned follow-up steps will be instrumental in bringing the model closer to real-world applications in the financial sector.

4. Conclusion

This study shows that the application of the SVM algorithm combined with the SMOTE data balancing technique and parameter optimization using Grid Search can improve the effectiveness of detecting fraudulent transactions on imbalanced data. The model evaluation results show significant improvements in recall, precision, F1-score, and AUC-ROC metrics, indicating the model's ability to better recognize fraudulent transactions without sacrificing classification accuracy. In addition, the experiment also proves that without data balancing and parameter tuning, the model's performance decreases drastically, especially in detecting minority classes. However, challenges such as high computation time in the optimization process remain a major concern. Therefore, this study suggests exploring alternative approaches such as Bayesian-based optimization or Randomized Search, as well as testing on real-time systems for further validation. Overall, the approach used forms a strong foundation for the development of a more accurate, adaptive, and feasible fraud detection system for the financial industry.

References

[1] Agustin, ID, & Abidin, FIN (2024). The Influence Of Financial Literacy, Financial Behavior, Digital Payment And Paylater On Student Consumptive Behavior In The COVID-19 Pandemic Era. Innovative Technologica: Methodical Research Journal, 1(4), 15. Https://Doi.Org/10.47134/Innovative.V1i4.44
[2] Batista, GEAPA, Prati, R.C., & Monard, M.C. (2004). A Study Of The Behavior Of Several Methods For Balancing Machine Learning Training Data. ACM SIGKDD Explorations Newsletter, 6(1), 20–29. Https://Doi.Org/10.1145/1007730.1007735
[3] Ben-Hur, A., & Weston, J. (2010). A User's Guide To Supporting Vector Machines. Methods In Molecular Biology (Clifton, NJ), 609, 223–239. Https://Doi.Org/10.1007/978-1-60327-241-4_13
[4] Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data Mining For Credit Card Fraud: A Comparative Study. Decision Support Systems, 50(3),

602–613. Https://Doi.Org/10.1016/J.Dss.2010.08.008

[5] Chen, T., & Guestrin, C. (2016). Xgboost: A Scalable Tree Boosting System. Proceedings Of The ACM SIGKDD International Conference On Knowledge Discovery And Data Mining, 17-13-August-2016, 785–794. Https://Doi.Org/10.1145/2939672.2939785

[6] Cherkassky, V., & Ma, Y. (2004). Practical Selection Of SVM Parameters And Noise Estimation For SVM Regression. Neural Networks, 17(1), 113–126. Https://Doi.Org/10.1016/S0893-6080(03)00169-2

[7] Cortes, C., & Vapnik, V. (1995). (2015). Support-Vector Networks. Machine Learning, 20(3), 273–297. Journal Of Physics: Conference Series, 628(1), 273–297. Https://Doi.Org/10.1088/1742-6596/628/1/012073

[8] Dal Pozzolo, A., Caelen, O., Bontempi, G., & Johnson, R. A. (2015). Calibrating Probability With Undersampling For Unbalanced Classification Fraud Detection View Project Volatility Forecasting View Project Calibrating Probability With Undersampling For Unbalanced Classification. Ieee. Https://Www.Researchgate.Net/Publication/283349138

[9] Dean, J., & Ghemawat, S. (2008). Mapreduce: Simplified Data Processing On Large Clusters. Communications Of The ACM, 51(1), 107–113. Https://Doi.Org/10.1145/1327452.1327492

[10] Hsu, C., Chang, C., & Lin, C. (2003). A Practical Guide To Support Vector Classification (Presentation). 1–29. Https://Www.Cs.Sfu.Ca/People/Faculty/Teaching/726/Spring11/Svmguide.Pdf

[11] Hutter, F. (2017). Optimization Parameters. Interdisciplinary Mathematical Sciences (Vol. 19). Https://Doi.Org/10.1142/9789814630146_0014

[12] Jasper Snoek, H.L. (2013). Practical Bayesian Optimization Of Machine Learning Algorithms. Religion And The Arts, 17(1–2), 57–73. Https://Doi.Org/10.1163/15685292-12341254

[13] Kim, H. (2022). Deep Learning. Artificial Intelligence For 6G, 22(4), 247–303. Https://Doi.Org/10.1007/978-3-030-95041-5_6

[14] Kohavi, R. (1995). A Study Of Cross-Validation And Bootstrapping For Accuracy Estimation And Model Selection. IJCAI International Joint Conference On Artificial Intelligence, 2(June), 1137–1143.

[15] Liu, Y., An, A., & Huang, X. (2006). Boosting Prediction Accuracy On Imbalanced Datasets With SVM Ensembles. Lecture Notes In Computer Science (Including Subseries Lecture Notes In Artificial Intelligence And Lecture Notes In Bioinformatics), 3918 LNAI(April), 107–118. Https://Doi.Org/10.1007/11731139_15

[16] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, LOH (2020). SMOTE: Synthetic Minority Over-Sampling Technique. METHOMIKA Journal Of Informatics Management And Computerized Accounting, 4(1), 67–72. Https://Doi.Org/10.46880/Jmika.Vol4no1.Pp67-72

[17] Ramadani, L. (2016). The Influence Of Debit Card And Electronic Money (E-Money) Usage On Student Consumption Expenditure. Journal Of Economics And Development Economics Studies, 8(1), 1–8. Https://Doi.Org/10.17977/Um002v8i12016p001

[18] Sokolova, M., & Lapalme, G. (2009). A Systematic Analysis Of Performance Measures For Classification Tasks. Information Processing And Management, 45(4), 427–437. Https://Doi.Org/10.1016/J.Ipm.2009.03.002

[19] Wu, Top 10 Algorithms In Data Mining. In Knowledge And Information Systems (Vol. 14, Issue 1). Https://Doi.Org/10.1007/S10115-007-0114-2