International Journal of Autonomous Robotics and Intelligent Systems
Vol. 1, No. 1, June 2025 [81-92]

Machine Learning Techniques in Credit Card Fraud Detection: A Hybrid Supervised and Unsupervised Approach

L. P. Yeoh[1], C. K. Lam[1,2*]

[1]Faculty of Electrical Engineering & Technology, Universiti Malaysia Perlis, 02600 Arau, Perlis, Malaysia.
[2]Centre of Excellence for Intelligent Robotics & Autonomous System (CIRAS), Universiti Malaysia Perlis, 02600 Arau, Perlis, Malaysia.

Received 4 July 2025, Revised 16 July 2025, Accepted 18 July 2025

ABSTRACT

In the dynamic landscape of financial transactions, the escalating threat of fraudulent activities necessitates cutting-edge solutions for real-time detection. This research introduces an innovative approach utilizing the Kaggle credit card dataset, focusing on comparing the effectiveness of hybrid models versus purely supervised learning models. While traditional models rely solely on supervised learning, this study explores the potential performance gains of integrating unsupervised learning (USL) into supervised learning (SL) frameworks. The core investigation centers on whether unsupervised clustering can enhance pattern recognition in unlabeled data and subsequently improve the performance of supervised models. This research not only evaluates the practical benefits of hybrid methodologies in fraud detection, but also advances real-time analytics through Power BI, aiming to provide a more comprehensive and adaptive solution to emerging financial threats. The algorithms yield an accuracy of 99.75% and a remarkably low underkill rate of 0.20%, demonstrating the effectiveness of integrating human oversight with advanced machine learning techniques.

Keywords: Ensemble Model, Fraud Detection, Hybrid Model, Supervised Learning, Unsupervised Learning

1. INTRODUCTION

In the constantly evolving realm of financial world, the pulse of the global economy reverberates through the channels of financial institutions, markets, and transactions. The bedrock of the world economy, the finance sector promotes the growth of economic prosperity. Amid the challenging environment of the financial sector, Financial Technology, or FinTech, has become a disruptive force for change. FinTech, as defined by Investopedia, is the umbrella term for cutting-edge technologies intended to improve the provision and use of financial services. FinTech is primarily utilized to assist businesses, entrepreneurs, and individuals in improving their financial well-being and streamlining their financial operations [1]. According to a remarkable KPMG research, FinTech funding reached a record-breaking $210 billion in 2021 through 5,684 agreements, a huge rise from the $125 billion over 3,674 deals reported in 2020 [2].

Nonetheless, FinTech confronts several significant obstacles as it works to upend and reinvent the financial sector. According to the Association of Certified Fraud Examiners (ACFE), "fraud" is defined as using one's job for personal gain by purposefully misusing or abusing the resources or assets of the employing business [3]. In past few decades, the number of fraud transactions on financial markets has tremendously increased. An estimated 800 billion to 2 trillion dollars are

*lckiang@unimap.edu.my

scammed annually worldwide, representing 2 to 5 percent of the world's gross domestic product [4]. The battle against financial fraud and money laundering is a multifaceted challenge that demands innovative and adaptive solutions. Consequently, combating fraud has emerged as a crucial topic worth investigating [5].

In the current world, artificial intelligence (AI) has become a key component in the identification of frauds. AI can adapt and learn from past data such as machine learning algorithms can identify minute irregularities and patterns that a human might miss. AI systems can examine a wide range of transactions, account patterns, and user activities, quickly identifying questionable conduct and possible fraud. Machine learning algorithms are used by these systems to continuously increase their efficacy and accuracy [6]. This project mainly focuses on harnessing the power of AI, comparing hybrid supervised and unsupervised approach for real-time fraud detection in credit card transactions, using Kaggle credit card datasets [14] to enhance the system's performance and adaptability.

2. RELATED WORKS

Supervised (SL) and Unsupervised (USL) learning are two different approaches in machine learning. In supervised learning, a classifier is trained using a labelled dataset that includes both "fraud" and "nonfraud" entries. This is the most popular method of instruction. The main benefit of SL is that it is simple to apply discriminative pattern classification, and all the class outputs produced by the algorithm are understandable [7][8][9]. Algorithms such as Random Forest (RF), Convolutional Neural Network (CNN) are popular for stands out performance in the field of SL [7]. While in USL, its goal is to extract knowledge from unlabelled data without the assistance of predetermined results. Algorithms using unsupervised learning investigate unlabelled data to uncover hidden patterns and connections [9]. This method works especially well in situations when the underlying structure of the data is not known with certainty, which makes it ideal for the ever-changing and dynamic nature of financial transactions [10].

Through literature reviews, RF and CNN is two of the SL classifier types that have been widely applied to detect fraud in transactions, while K-means is one of the USL clustering methods that has been also widely used for clustering large datasets such as bank transactions. A brief discussion about the classifier and clustering methods are stated in this section.

2.1 Random Forest Classifier

RF is the most used machine learning ensemble method, which builds an ensemble of decision trees by combining random feature selection and bootstrap sampling. This approach, which was first presented by Leo Breiman [11], is very useful in applications like fraud detection because of its durability and excellent predicted accuracy [8][12]. The diversity that RF introduces is what makes it so effective. For example, every decision tree is trained using a different part of the dataset, and every split in the tree construction process takes a different subset of features into account [11]. RF interpretability and adaptation to high-dimensional datasets are key factors in its broad adoption in the fraud detection domain, facilitating the creation of trustworthy and efficient fraud detection systems [7][12]. Figure 1 shows the architecture of RF.

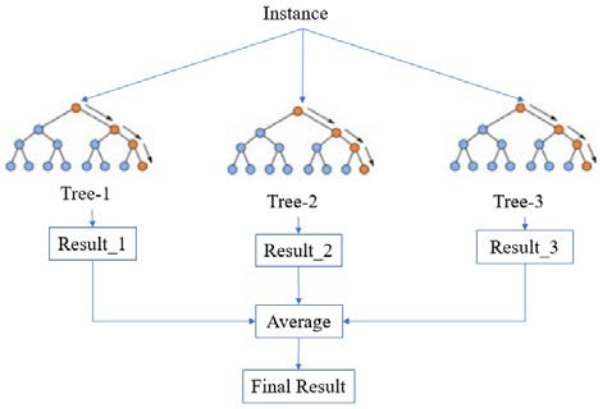

Figure 1. Architecture of RF [6]

2.2 Convolutional Neural Network Classifier

CNN is a popular deep learning model that are frequently used for pattern recognition. Initially hailed for their unmatched abilities in visual pattern recognition, CNNs have gracefully transitioned beyond their original field to become indispensable tools in the complex world of financial fraud detection. Convolutional, pooling, and fully linked layers interact in a carefully planned manner to reveal CNNs' unique design. Every element plays a vital role in the process of extracting hierarchical characteristics from the financial data that is fed into the network, which helps in identifying the subtle but significant patterns that are present in the complex financial transactions [6][7][10][12]. Figure 2 shows the architecture of CNN in fraud detection system.

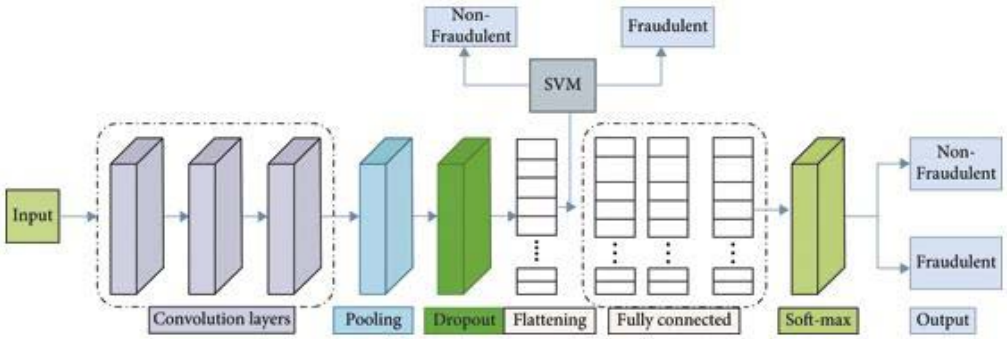

Figure 2. Architecture of CNN [7]

2.3 K-Means Clustering

K-means clustering is a widely utilized technique in USL that categorizes data points into 'k' clusters based on their similarity to one another. The process involves iteratively assigning data points to clusters and adjusting cluster centroids until an optimal grouping is achieved. In the context of fraud detection, K-means clustering can unveil cohesive groups of transactions sharing common features, providing insights into patterns that may signify potential instances of fraud [13]. The adaptability and simplicity of K-means make it a powerful tool for identifying anomalous clusters within financial datasets.

2.4 Current Limitations

In the contemporary landscape of fraud detection [7][9][12][13], there is a predominant emphasis on supervised learning models, which necessitate labelled data for effective training. This reliance presents significant limitations, as acquiring such datasets is resource-intensive and may not capture emerging fraud patterns. Conversely, unsupervised learning methods primarily focus on clustering and lack a comprehensive approach to fraud prevention [6][8][12]. Recognizing this gap, the exploration of a hybrid model is imperative. The envisioned hybrid model leverages unsupervised learning for filtering unlabelled data through clustering techniques to identify potential fraud patterns. This output then transitions into a supervised learning module, adding an additional layer of prevention and enhancing overall efficacy.

Traditional limitations of supervised models, including the inability to achieve 100% accuracy and inherent biases in labelled datasets, necessitate a shift in fraud detection approaches [7][8][10][12]. To address these challenges, a strategic decision was made to implement a dual-model system comprising both machine learning and deep learning approaches. This system integrates the strengths of diverse algorithms, fostering a more robust and adaptable fraud detection mechanism. By combining machine learning and deep learning models, the hybrid system aims to provide a comprehensive understanding of complex fraud patterns, ensuring more effective and accurate detection. This approach enhances the resilience of fraud detection systems beyond the constraints of a singular algorithmic focus.

3. METHODOLOGY

The project's system architecture is delineated into four integral components, each contributing to the overarching goal of building a robust AI-based fraud detection system. The system architecture serves as a conceptual framework to describe the proposed model system's structure and process flow. Figure 3 illustrates the proposed model system's architecture. In the initial phase, unsupervised machine learning is leveraged for clustering unlabelled datasets. The outputs are then proceeded to supervised learning for training and inference. After training supervised learning models, the decision-making system is implemented by ensemble method of both trained models.

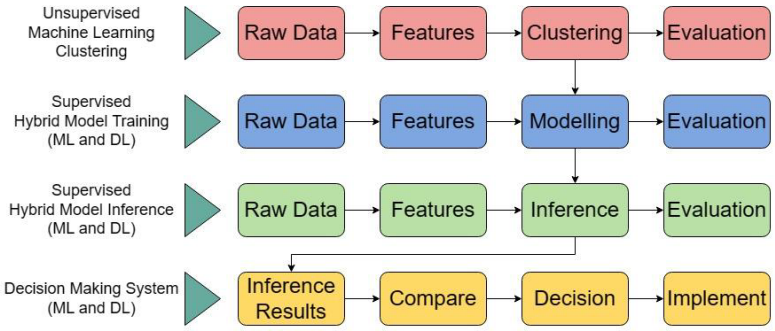

Figure 3. Proposed model system architecture

Kaggle credit card dataset [14] is utilized and unsupervised machine learning turns out to be a crucial component that helps cluster unlabelled data well, which leads to the creation of labelled datasets. This tactical method can be shown in Figure 4, which helps to reduce the workload from daily manual labelling of millions of data. Following rigorous data preprocessing, data cleaning, clustering algorithms such as k-means are designed to clusters similar data points, generating labelled clusters that facilitate grouping the datasets for subsequent analysis.

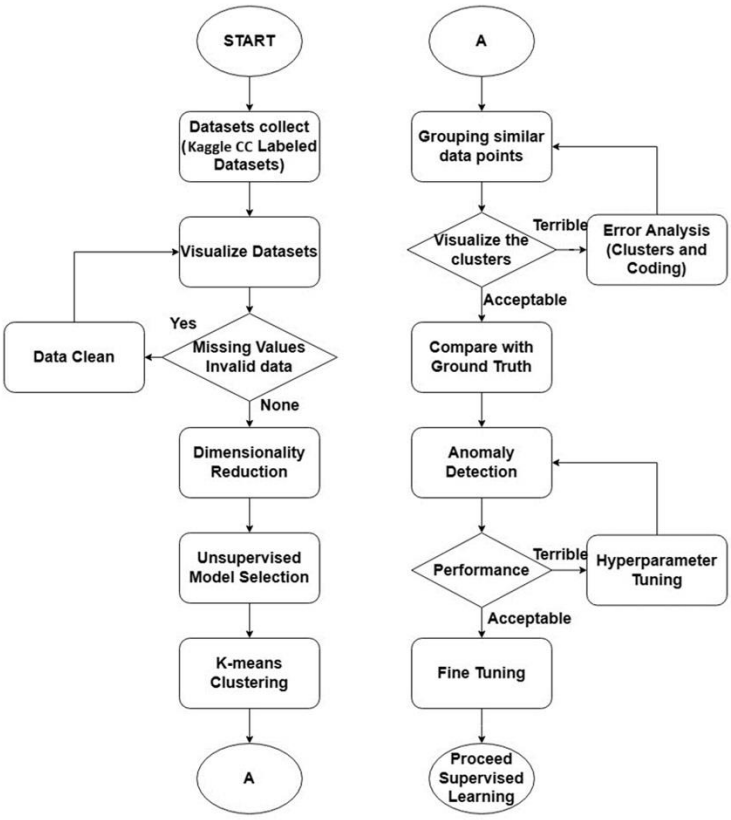

Figure 4. Overview of unsupervised learning setup

Transitioning to the supervised hybrid model training phase, a supervised machine learning model and supervised deep learning model are leveraged and work closely together. The labelled data from unsupervised learning outputs undergoes feature extraction before feeding into a pipeline that effectively integrates RF, a supervised machine learning model, and CNN, a supervised deep learning model. The combination of methodologies aims to capitalize on the strengths and robustness of both models for comprehensive fraud detection. Moving forward, the trained hybrid model is then implemented for inference on training, validation, testing, and new raw datasets to determine the model performance and robustness of the model's ability to classify and identify potential instances of fraud with new unseen data. Overview of supervised learning process setup is shown in Figure 5.

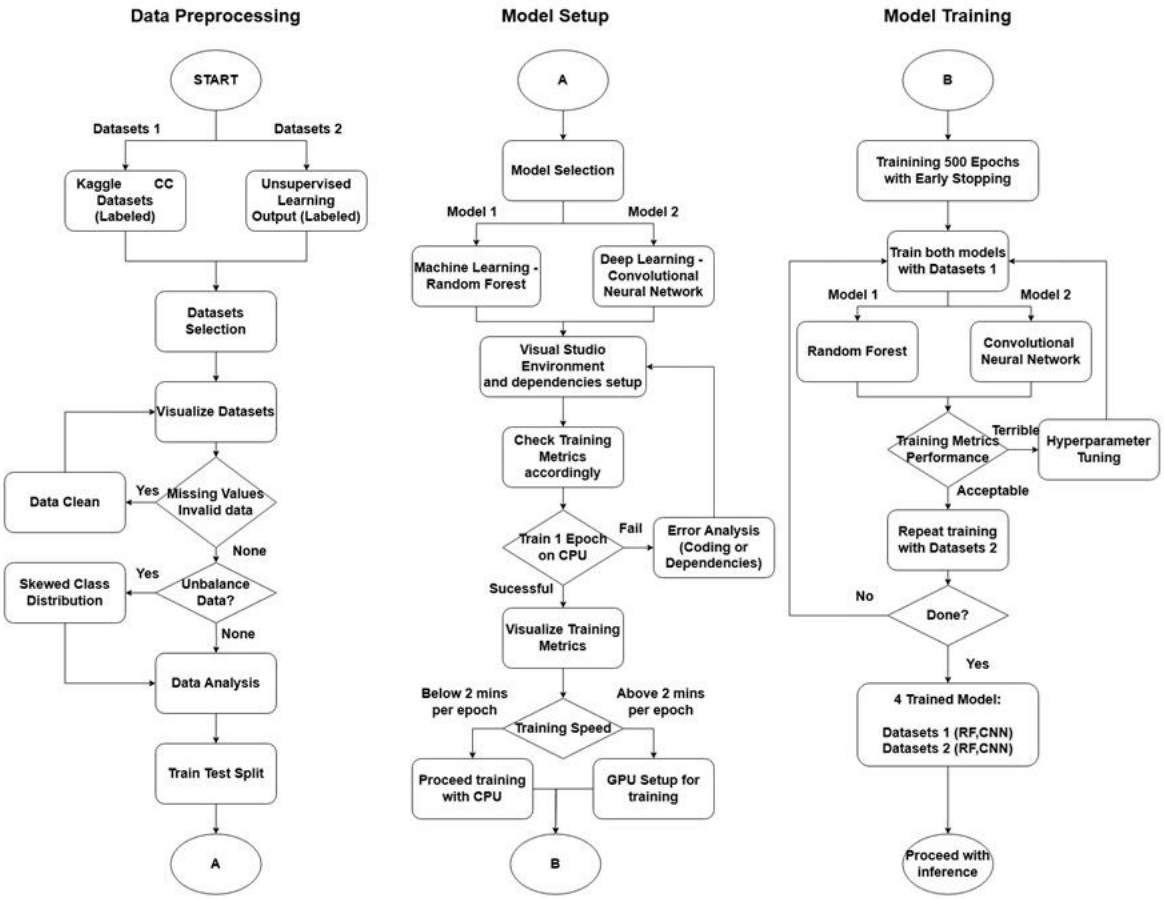

Figure 5. Overview of supervised learning setup

Lastly, to fortify the decision-making process, a sophisticated system is implemented, leveraging both machine learning and deep learning models' inference performance. This dual-model decision-making system is shown in Figure 6, and it ensures an additional layer of inspection with the functionality of; if the predictions from both models align, the system proceeds with the decision [9]. However, if the predictions from both models are different, a human review stage is invoked, tapping into human expertise to address ambiguous or conflicting scenarios. This entire system architecture seamlessly integrates unsupervised learning, supervised learning, and human intervention, creating an adaptive fraud detection system for real-world applications.

In order to measure the performance of proposed models, several evaluation metrics were employed in this research. Primarily, the most important features are False Negative Rate (FNR) and True Positive Rate (TPR), False Negative refers the number of fraud transactions predicted as legal transactions while True Positive refers the number of legal transactions predicted as legal transactions. Besides, accuracy, precision and recall are also being implemented to measure the robustness of the proposed models. Table 1 describe the formula for each evaluation metrics.

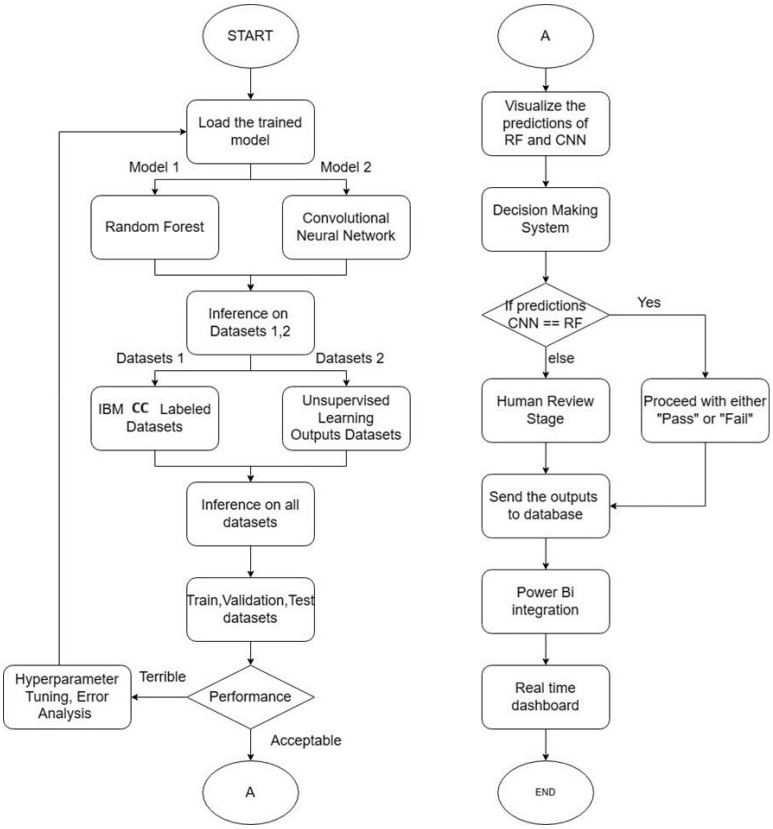

Figure 6. Inference and human review system

Table 1 Formula of evaluation metrics

Evaluation Metrics	Formula/Description
True Positive Rate (TPR)	TP/(TP+FN)
False Negative Rate (FNR)	FN/(TP+FN)
Precision	TP/(TP+FP)
Recall	TP/(TP+FN)
Accuracy	(TP+TN)/(TP+TN+FP+FN)

4. RESULTS AND DISCUSSION

This research aimed to develop a robust fraud detection system by evaluating the performance of various machine learning models using the Kaggle credit card dataset. The study focused on comparing the effectiveness of supervised, and hybrid models, with a particular emphasis on integrating these models within a real-time dashboard for enhanced monitoring and decision-making. The following sections detail the performance and evaluation of these approaches.

4.1 Unsupervised Learning Performance

For USL, a total of 1,000 (500 TP and 500 TN) datasets with label removed was used to evaluate the clustering performance. Despite the Elbow method suggesting a higher number of clusters which is 4, the decision to evaluate k=2 is motivated by the inherent binary classification of the data which is “legal” or “fraud”. Upon evaluation with ground truth data, the clustering performance yields an accuracy of 78.73% on its own datasets, indicating a reasonably accurate partitioning of the data into two clusters. Notably, there are no instances of overkill, implying that

all fraudulent transactions are successfully captured within the clusters. However, the analysis reveals an underkill (FN) of 211 instances, suggesting that some non-fraudulent transactions are misclassified within the fraudulent cluster. Figure 7 shows the clustering results.

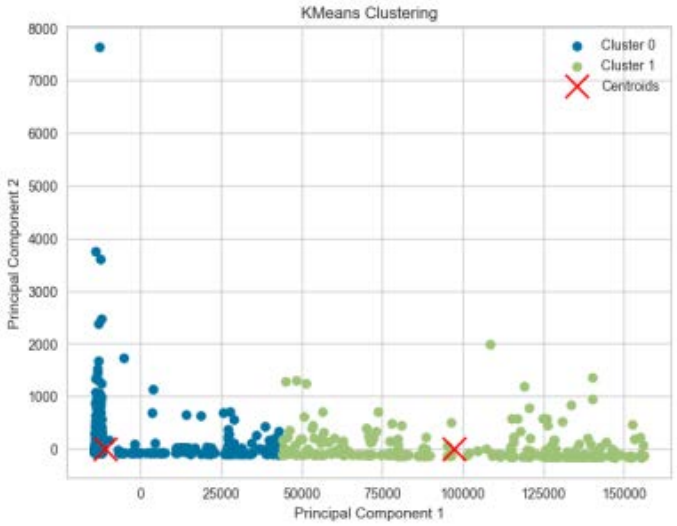

Figure 7. Clustering graph when k = 2

Upon evaluating the performance of clustering with k=4, the results demonstrate a notable improvement in accuracy, reaching 94.35%. This signifies a robust partitioning of the dataset into four distinct clusters, reflecting the enhanced granularity and to identify various pattern of fraud transactions. Importantly, there are no instances of underkill, indicating that all fraudulent transactions are effectively captured within the clusters. Moreover, the analysis reveals a significantly reduced number of underkill instances, totalling only 56. This underscores the improved accuracy achieved through the utilization of four clusters, highlighting the efficacy of this approach in capturing diverse fraud patterns within the dataset, which Cluster 0 is legal transactions, while Cluster 1 to Cluster 3 are different pattern of fraudulent transactions. Figure 8 shows the clustering performance metrics when k=4.

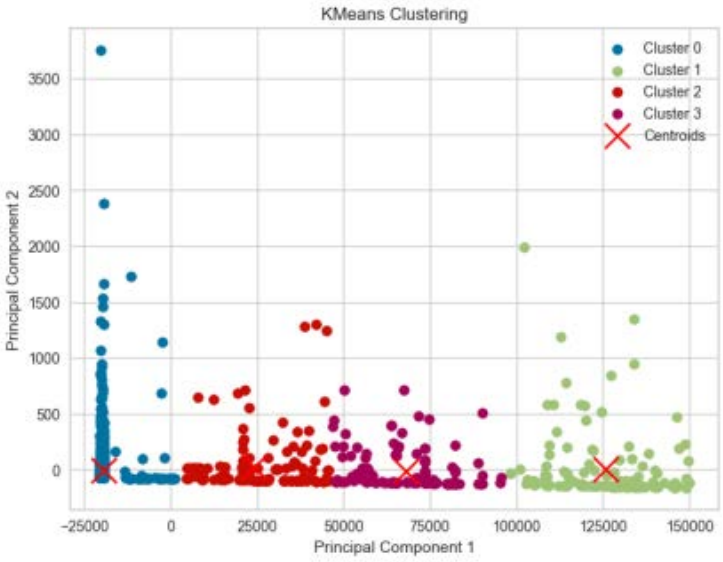

Figure 8. Clustering graph when k = 4

The output of this unsupervised learning is then export for training the supervised learning, but the output of the supervised learning is not as expected, and fall below expectations, hence the results is not being compared and shown. To tackle these issues, PCA is implemented for reduce the dimensionality data of the datasets, while the original datasets is then added to 2500 pass transactions and 500 fraud transactions. The performance has slight increased, hitting accuracy of 97.18% and underkill rate below 11.38% on the original datasets. This output data is being utilized for the following supervised learning training.

4.2 Supervised Learning Performance

The SL alone that trained on original datasets with extracted values of 2500 pass and 500 fraud transactions initially. Then the SMOTE oversampling is applied to address class imbalance, adding additional data to fraud making it both equally distributed. On RF, the performance is surprisingly high with accuracy hit a remarkable 99.96%, FNR below 20.73% on the original raw datasets which consist of 284,807 datasets. On the other hand, CNN undergoes several optimizations such as adding the depth of neural network and fine tuning the batch sizes of the CNN. After hyperparameter tuning of the CNN, it performs very well, with accuracy of 99.06% and FNR below remarkable 2.44%.

Moreover, hybrid approaches combining SL and USL presented mixed results. This is done by training the SL with USL datasets and run inference on the raw datasets which consist of 284,807 datasets. The combination of Random Forest and K-Means showed a significantly lower accuracy of 45.92% with a high underkill rate of 53.25%, indicating inefficiencies in this hybrid approach. On the other hand, the hybrid approach of CNN and K-Means showed inaccurate predictions, which the models are completely biased to the fraudulent transactions, causing it overkill around 90% datasets, with accuracy of only 10.55% with underkill rate of 12.32%.

In contrast, ensemble methods combining RF and CNN without human review achieved high accuracy of 99.74% and a moderate underkill rate of 10.77%. This was achieved by implementing a voting mechanism with two trained supervised models running predictions simultaneously. To further enhance the robustness of the fraud detection models, a human review component was integrated into the ensemble method. This approach involves flagging predictions that fall below a certain threshold for human review. This integration resulted in a further improvement, yielding an accuracy of 99.75% and a remarkably low underkill rate of 0.20%. These results highlight the effectiveness of combining supervised learning models and human oversight, especially when only 3,000 training datasets are used, demonstrating the limitations of USL algorithms and the failure of hybrid approaches. Table 2 shows the performance comparison of different approaches on the raw datasets, consisting of 284,807 records.

Table 2 Summary of inference evaluation

Types	**Hybrid SL + USL**		**SL**			
Metric	**CNN**	**RF**	**CNN**	**RF**	**Ensemble Method**	**Human Review**
TPR	10.42	45.91	99.07	99.99	99.75	99.72
FNR	12.32	53.25	2.44	20.73	10.77	0.20
Precision	0.12	0.50	0.99	0.98	0.99	0.99
Recall	0.10	0.46	0.97	0.89	0.99	0.99
Accuracy	10.55	45.92	99.06	99.96	99.74	99.75

4.3 Real-Time Dashboard

After running the predictive models, the results are loaded into Power BI. In Power BI, a real-time dashboard implemented to allows users or operators to seamlessly load the data, enabling real-time updates. The dashboard effectively displays all key metrics, providing an intuitive overview for the operators.

Firstly, the dashboard highlights the approved transactions by showing both the number of approved transactions and their total value in dollars. It provides insights into fraud or rejected transactions, displaying their count and total value in dollars as well. Another critical component of the dashboard is the human review section. This section is crucial for operators to monitor, as it lists the on-hold transactions that require further decisions by the bank staff. In this section, the total number of on-hold transactions and their cumulative value in dollars are shown. Moreover, a tabular display showcases these on-hold transactions, sorted from the highest to the lowest amount. For better clarity and tracking, the transactions are listed with an "Order ID," which represents a time domain value that has been adjusted due to the original data's application of Principal Component Analysis (PCA), thus starting from 0 to a range instead of real-time values.

Furthermore, the dashboard features a graphical representation of the total transaction amounts over time, allowing operators to visualize transaction trends. A detailed table displays the highest rejected fraud transactions in descending order, including the order ID and amount. This detailed and organized layout ensures that operators can efficiently monitor, review, and act on transaction data, making the dashboard a vital tool for maintaining operational integrity and decision-making processes in real-time. The Power Bi dashboard setup is shown in Figure 9.

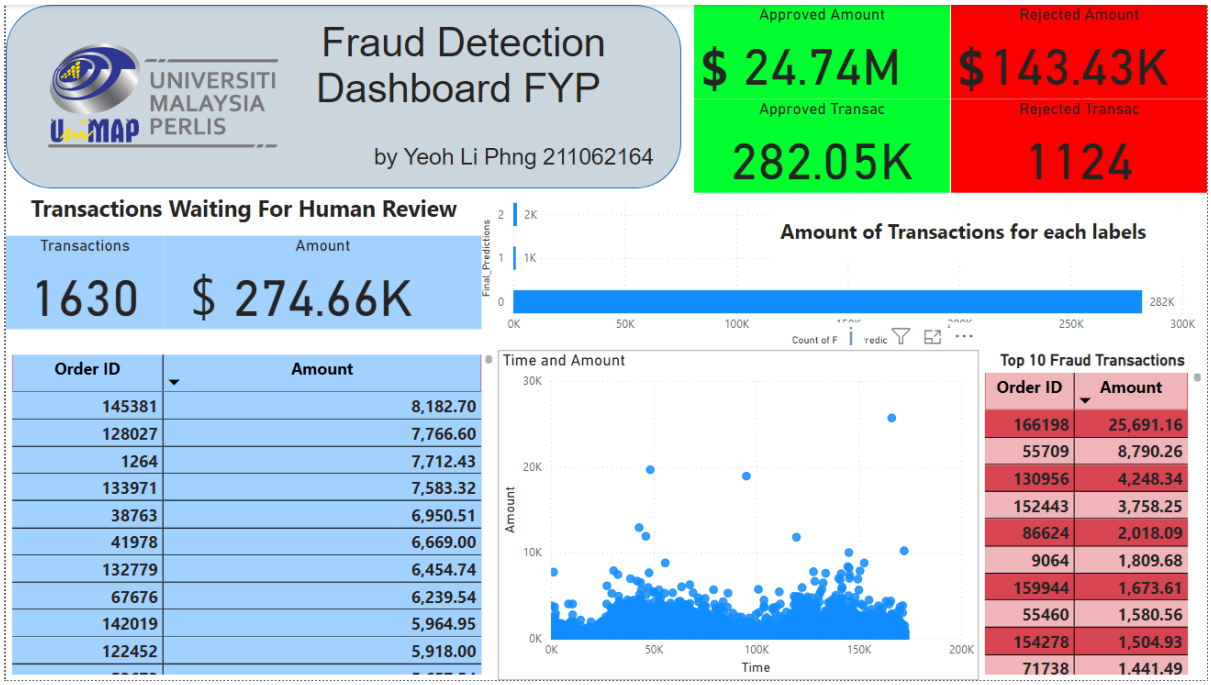

Figure 9. Power BI dashboard

4.4 Discussion and Future Work

The exploration of hybrid models integrating unsupervised learning (USL) with supervised learning (SL) in fraud detection yielded insightful results. While traditional supervised models like Random Forest and Convolutional Neural Network (CNN) demonstrated high accuracy, the integration of unsupervised clustering via K-Means did not enhance performance as anticipated.

Specifically, the hybrid models showed significantly lower accuracy and higher underkill rates, indicating inefficiencies in clustering important features and patterns within high-dimensional data. This suggests that the current methodologies and datasets might not be suitable for hybrid approaches, leading to biased models and suboptimal performance.

One of the critical findings was that the supervised models trained with SMOTE oversampling on imbalanced datasets performed remarkably well, with Random Forest achieving 99.96% accuracy and CNN achieving 99.06% accuracy. The ensemble method combining Random Forest and CNN further improved the robustness, achieving 99.74% accuracy. When human review was integrated into this ensemble approach, the performance slightly increased to 99.75%, and the underkill rate was reduced to 0.20%. This underscores the value of combining machine learning models with human oversight to enhance fraud detection systems.

Future work should focus on several areas to further advance fraud detection capabilities. Firstly, exploring more sophisticated feature selection and dimensionality reduction techniques may help improve the clustering performance of unsupervised learning models. Advanced clustering algorithms, such as DBSCAN or hierarchical clustering, could also be investigated to handle high-dimensional data more effectively. Additionally, expanding the dataset to include more diverse and comprehensive data sources might capture a wider range of fraud patterns, thus improving model training and evaluation.

Moreover, the integration of real-time data streams into the hybrid model can be explored to enhance the timeliness and responsiveness of fraud detection systems. Leveraging advancements in artificial intelligence, such as reinforcement learning and anomaly detection techniques, may also provide new avenues for improving model accuracy and adaptability. Collaborative efforts with financial institutions to refine and validate these models in real-world scenarios would be invaluable, ensuring that the developed solutions are both practical and effective in combating fraud.

5. CONCLUSION

In conclusion, this research successfully developed a robust fraud detection system utilizing supervised learning models and explored the potential of hybrid approaches combining unsupervised and supervised learning. The results demonstrated that while traditional supervised models like Random Forest and CNN achieved high accuracy and low underkill rates, the hybrid models did not perform as well, highlighting the limitations of current unsupervised learning techniques in this context.

The ensemble method, particularly when combined with human review, proved to be the most effective approach, achieving a remarkable accuracy of 99.75% and a very low underkill rate. These findings emphasize the importance of human oversight in enhancing machine learning models for fraud detection.

Future research should focus on improving the clustering performance of unsupervised learning models, exploring advanced algorithms, and expanding the dataset to capture more diverse fraud patterns. Integrating real-time data streams and leveraging new AI advancements will further enhance the robustness and adaptability of fraud detection systems. Collaborative efforts with industry partners will be crucial in refining these models for practical, real-world applications. This research lays a solid foundation for future advancements in the field, aiming to provide comprehensive and adaptive solutions to emerging financial threats.

ACKNOWLEDGEMENTS

Authors are grateful to the Faculty of Electrical Engineering & Technology of UniMAP for providing the resources and support to this work.

REFERENCES

[1] "Financial Technology (Fintech): Its Uses and Impact on Our Lives." [Online]. Available: https://www.investopedia.com/terms/f/fintech [Accessed Nov. 23, 2023].
[2] "The Current State of FinTech | Foley & Lardner LLP." [Online]. Available: https://www.foley.com/insights/publications/2023/09/current-state-fintech/ [Accessed Nov. 23, 2023].
[3] "2002 Report To The Nation; Occupational Fraud And Abuse," 2002.
[4] "United Nations Office On Drugs And Crime Vienna Independent In-depth evaluation of The Global Programme against Money Laundering, Proceeds of Crime and the Financing of Terrorism," 2011. [Online]. Available: www.unodc.org [Accessed Nov. 23, 2023]
[5] S. Kramer, N. Lavrač, P. Flach, 2001. Propositionalization Approaches to Relational Data Mining, Relational Data Mining, pp. 262–291, doi: 10.1007/978-3-662-04599-2_11.
[6] W. Wang, G. Chakraborty, B. Chakraborty, 2021. Predicting the risk of chronic kidney disease (Ckd) using machine learning algorithm, Applied Sciences (Switzerland), vol. 11, no. 1, pp. 1–17, doi: 10.3390/APP11010202.
[7] A. Roy, J. Sun, R. Mahoney, L. Alonzi, S. Adams, P. Beling, 2018. Deep learning detecting fraud in credit card transactions, In *2018 Systems and Information Engineering Design Symposium, SIEDS 2018*, pp. 129–134, doi: 10.1109/SIEDS.2018.8374722.
[8] J. K. Afriyie et al., 2023. A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions," Decision Analytics Journal, vol. 6, doi: 10.1016/j.dajour.2023.100163.
[9] P. Dua, S. Bais, 2014. Supervised learning methods for fraud detection in healthcare insurance, Intelligent Systems Reference Library, vol. 56, pp. 261–285, doi: 10.1007/978-3-642-40017-9_12/TABLES/1.
[10] P. Craja, A. Kim, S. Lessmann, 2020. Deep learning for detecting financial statement fraud, Decis Support Syst, vol. 139, p. 113421, doi: 10.1016/J.DSS.2020.113421.
[11] L. Breiman, 2001. Random forests, Mach Learn, vol. 45, no. 1, pp. 5–32, doi: 10.1023/A:1010933404324/METRICS.
[12] V. N. Dornadula, S. Geetha, 2019. Credit Card Fraud Detection using Machine Learning Algorithms, Procedia Comput Sci, vol. 165, pp. 631–641, doi: 10.1016/J.PROCS.2020.01.057.
[13] R. Xu, D. Wunsch, 2005. Survey of clustering algorithms, IEEE Trans Neural Netw, vol. 16, no. 3, pp. 645–678, doi: 10.1109/TNN.2005.845141.
[14] "Credit Card Fraud Detection." [Online]. Available: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud [Accessed Nov. 27, 2023].