

Securing Credit: A Hybrid Multi-Dimensional Model Using Ensemble Machine Learning Classifier with Data Sampling to Detect and Prevent Credit Card Fraud

Srikumar Nayak

Associate Director, Artificial Intelligence Practice, Incedo Inc, NYC

Abstract

The current methods for detecting fraud have notable shortcomings, including issues with imbalanced datasets, incorrect detection of fraudulent activities, limited versatility across various contexts, and challenges in real-time data processing. This study introduces an ensemble machine learning model aimed at identifying fraud in credit card transactions. Additionally, it employs the Synthetic Minority Oversampling Technique (SMOTE) combined with Edited Nearest Neighbor (ENN) to tackle the challenge of imbalanced data. The results from our experiments indicate that this method outperforms existing approaches. Consequently, it lays a crucial foundation for ongoing research focused on creating more resilient and adaptable systems for fraud detection.

Keywords: Statistical Features, Machine Learning, Credit Card, Fraud, Data Imbalance.

1. INTRODUCTION

The financial sector has witnessed a rise in fraudulent activities, especially concerning credit cards (Bagga et al., 2020). Fraud refers to the unauthorized utilization of an individual's credit card by another person for personal transactions without the cardholder's approval. The ease of using credit cards for financial transactions complicates the detection of fraudulent activities [1]. According to the British Ministry of Finance Annual Fraud Reporting, in 2022, fraudsters unlawfully acquired a total of £1.2 billion through both legitimate and illegitimate criminal means, resulting in losses amounting to £2,300 every minute. Notably, 78% of incidents involving Authorized Push Payment (APP) fraud originated from online channels, while 18% occurred via phone communication (Annual Reports, 2024). A prevalent form of financial loss stems from remote purchase fraud, where criminals exploit stolen card information to make transactions online or through mail/phone. This has led to substantial financial setbacks [2].

As credit card fraud continues to escalate, researchers are increasingly interested in utilizing traditional Machine Learning techniques alongside modern AI-based algorithms for detecting such fraudulent activities. Nonetheless, the uneven distribution of data presents a significant challenge in addressing the classification issues between fraudulent and legitimate transactions. The stark contrast between valid and fraudulent transaction numbers complicates training learning algorithms on the characteristics of the minority class. Consequently, these models tend to focus on the majority class (non-fraud), which can result in overfitting [3]. One major hurdle in Credit Card Fraud Detection (CCFD) is the rarity of fraudulent

transactions compared to legitimate ones. As a result, any collected data will inevitably show considerable disparity between minority (fraudulent) and majority (legitimate) samples. To combat the increasing instances of credit card fraud, it is essential to develop a highly accurate model that effectively caters to the needs of credit card users.

Misclassification remains a critical issue when it comes to identifying digital credit card fraud within e-commerce systems [4]. This study aims to reduce false positives to prevent inconvenience of customers and preserve trust in the financial system. By integrating SMOTE with ENN methods, we propose an approach designed to improve dataset balance while reducing overfitting risks associated with noisy or misleading examples. The proposed method aims to enhance the generalization performance of classifiers—particularly on datasets characterized by skewed distributions. We will compare results obtained through our proposed Ensemble Machine Learning method based on Sample Balancing technique (EML-SB) with those achieved using supervised learning methods.

This paper is organized as follows: Initially, we present a review of relevant literature that emphasizes the differences between our methodology and existing techniques. Next, we explain the process involved in collecting and processing the dataset. Building on this groundwork, we offer an extensive overview of our model. In conclusion, we detail the evaluation criteria used, describe our experimental procedures, and perform a comparative analysis with results from previous research studies.

2. Related Work

We did a comparison between the effectiveness of a Deep Learning (DL) model and various machine learning models such as Adaboost and Decision Trees (DT). The goal of this research was to identify if specific DL parameters contribute to improvements in predicting credit card defaults. We utilized the open UCI ML repository for obtaining dataset related to customers who defaulted on their credit cards. Following several preprocessing steps applied to the raw data, Exploratory Data Analysis (EDA) was employed to visually present the results.

Method for detecting credit card fraud within unbalanced datasets through soft voting ensemble learning techniques and comparing multiple advanced sampling strategies—such as hybrid sampling, under-sampling, and over-sampling—to tackle class imbalance issues effectively. By developing various classifiers for credit card fraud detection both utilizing these sampling methods and functioning independently from them, including ensemble classifiers. Experimental results indicated that soft-voting technique outperformed individual classifier approaches.

Analysis of three established optimization algorithms: Root Mean Squared Propagation (RMSprop), Adaptive Moment Estimation (ADAM), and Stochastic Gradient Descent (SGD) were examined with respect to Deep Convolutional Neural Networks used for Credit Card Fraud Detection (CCFD). After thorough evaluation concerning problem characteristics, objective function properties, and computational factors, it was concluded that all four methodologies are suitable for CCFD tasks; however, RMSprop yielded superior performance with an impressive accuracy rate of 99.93% when tested.

We also evaluated ensemble strategy aimed at enhancing CCFD capabilities by focusing on improving model parameters alongside refining performance metrics using deep learning techniques designed specifically to correct identification errors while minimizing false negative occurrences. This approach significantly boosts efficiency within systems tasked with detecting fraudulent activity on credit cards by integrating multiple classifier ensembles while rigorously assessing their respective performances; nevertheless, certain evaluation criteria reveal subpar outcomes associated with the model's efficacy.

Weighted Average Ensemble methodology which amalgamated predictions derived from logistic Regression (LR), Random Forest(RF), K-Nearest Neighbors (KNN), Adaboost along with Bagging models targeting CCFD scenarios indicate that this combined algorithm achieves accurate identification rates regarding instances of fraudulent activities across relevant applications [5].

We experimented with oversampling techniques and assessed against different machine-learning algorithms aiming to gauge the overall efficacy levels achieved therein. It was noted that the enhancements brought forth via oversampling methods can elevate model performance but highlighted its dependency upon selected machine-learning frameworks being implemented. Additionally, the computational burden linked may restrict practical usability within real-world contexts [6]

Further exploring classification methodologies, we pursued examination involving varied Machine Learning tactics applicable towards categorizing incidents related respectively toward Credit Card Frauds examining DT, Random Forest,[7] specifically focused around skewed distribution datasets. Detailed analyses revealed significant efficiencies exhibited particularly via deployment surrounding Random forests optimally recognizing pertinent fraudulent cases accurately.

Fraud represents a vital element in safeguarding financial stability. The performance of the model discussed is adversely affected by the lack of feature selection [8]. In a study conducted by [9], a comparison between Random Forest (RF) and AdaBoost was performed within the context of Credit Card Fraud Detection (CCFD). The results indicated similar accuracy levels for both algorithms, with RF outperforming AdaBoost across various evaluation metrics. However, it is important to note that the dataset used was biased, and there was no clear explanation regarding how this issue was addressed [10]. Additionally, for an effective strategy for fraud detection, we researched an extensive and critical review of existing literature without analyzing a specific dataset. The findings highlighted that Bayes' theorem [11] demonstrated superior performance with an accuracy rate of 99%. The Support Vector Machine (SVM) followed closely with an accuracy of 98%, while the genetic algorithm achieved an accuracy of 95% [12]

Further studying the works done at aimed to detect anomalies or fraudulent activities through data mining techniques, utilizing three distinct datasets from Australia, Germany, and the European Union [13]. They employed SVM, K-Nearest Neighbors (KNN), and RF methodologies along with two different ensemble constructions; however, the overall accuracy remained low across all datasets used [14].

In another comparative study on various machine learning techniques, SMOTE was utilized to tackle issues related to an imbalanced dataset source from the Université Libre de Bruxelles Machine Learning Group via Kaggle. The analysis revealed that Neural Networks achieved the highest accuracy at 96%. While this figure is commendable for Credit Card Fraud Detection, it still allows for a considerable number of false positives and negatives [15].

The challenges faced include reducing false positives to avoid client inconvenience while preserving trust in the financial system. Balancing incorrect positive results against undetected fraud cases is a delicate matter that necessitates careful calibration of detection algorithms and threshold settings.

Combining SMOTE with Edited Nearest Neighbors (ENN) can effectively address issues related to false predictions in fraud detection. SMOTE generates synthetic samples for minority classes (fraud cases), thereby improving the model's capacity to identify fraudulent behavior. Meanwhile, ENN enhances this method by eliminating ambiguous or potentially misclassified samples from both categories. This fusion leads to a more balanced dataset which improves the efficacy of detection algorithms while diminishing false positives—ultimately safeguarding system integrity and fostering client confidence.

3. Methodology

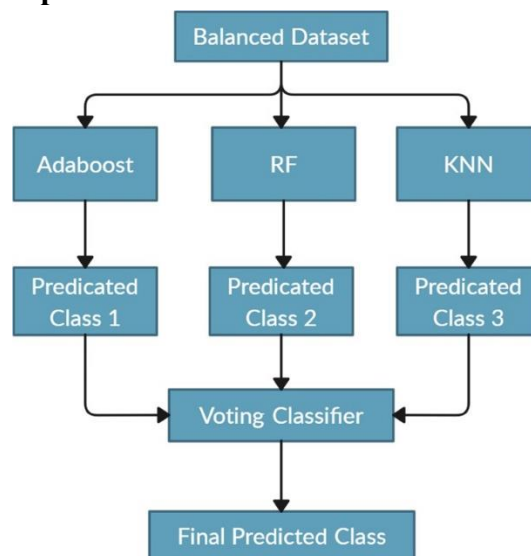
3.1 Data Collection

The dataset utilized for this research originates from the Figshare [16] is a open source data and its specifics are available on [21] . It encompasses transaction records over two days amounting to a total of 5 million individual transactions; among these transactions, instances classified as fraud represent only 0.234% of the data set—indicative of significant imbalance favoring non-fraud cases. This dataset comprises continuous numerical input variables derived from a Principal Component Analysis (PCA) feature selection process resulting in 37 principal components; thus, this study employs a total of 87 input features or characteristics. The 'class' feature serves as the target column for binary classification where a value of '1' denotes a fraud case while '0' indicates non-fraudulent activity.

3.1 Machine-learning approach with multiple Classifiers

The study employs an ensemble machine-learning approach that amalgamates multiple classifiers, each chosen for its distinct strengths. The AdaBoost classifier serves as a meta-estimator, initially training a classifier on the original dataset. Subsequently, it iteratively trains additional replicas of the classifier on the same data, adjusting the weights of misclassified instances to focus on the more challenging cases in later iterations. Random Forest (RF) constructs robust decision trees, while K-Nearest Neighbors (KNN) classifies data by identifying the majority class among its nearest neighbors. A noteworthy component is the Voting Classifier, which aggregates predictions from various classifiers Each algorithm was meticulously selected based on its demonstrated effectiveness as discussed in the literature review. Incorporating multiple classifiers within this ensemble strategy aims to bolster the predictive performance of the model.

The model presented in this article is illustrated in Fig. 1.



Ensemble methods are particularly adept at addressing imbalances in class distribution and demonstrate considerable effectiveness in identifying minority classes. They facilitate the integration of several less precise learners, thereby enhancing the overall predictive capabilities of the model.

Each algorithm—AdaBoost, RF, and KNN—possesses distinct traits and benefits that enhance their efficiency and performance. RF excels at processing high-dimensional data and exhibits resilience against overfitting. AdaBoost is proficient at managing weak classifiers while improving aggregate accuracy. KNN shows strong performance with smaller datasets and is relatively easy to implement. These algorithms can be highly effective for detecting fraud in credit card transactions, which demand accuracy

and adaptability to evolving fraud patterns.

To successfully employ these techniques for credit card fraud detection, it is essential to tackle several critical challenges: ensuring adequate data preparation, managing class imbalances, optimizing hyperparameters, and improving scalability. Furthermore, no prior research has fully utilized this combination of methods within a cohesive framework, underscoring the innovative nature of this approach.

3.3. Synthetic Minority Over-sampling Technique

To illustrate how this technique works consider some training data which has s samples, and f features in the feature space of the data. Note that these features, for simplicity, are continuous. As an example, consider a dataset of birds for classification. The feature space for the minority class for which we want to oversample could be beak length, wingspan, and weight (all continuous). To then oversample, take a sample from the dataset, and consider its k nearest neighbors (in feature space). To create a synthetic data point, take the vector between one of those k neighbors, and the current data point. Multiply this vector by a random number x which lies between 0, and 1. Add this to the current data point to create the new, synthetic data point.

3.4. Metrics Evaluation

The analysis and evaluation of performance metrics obtained during assessments yield an insightful overview regarding each model's effectiveness. The following criteria were employed to appraise the proposed model:

- Precision indicates the accuracy of optimistic predictions.
- Recall measures the ratio of correct positive predictions relative to all actual positives in terms of true positives and false negatives
- The F1 score serves as a composite measure combining precision and recall into one
- Accuracy reflects the proportion of correctly predicted samples out of total observations

3.5. Proposed Credit Card Fraud Detection Approach

This section outlines our suggested methodology illustrated in Fig. 2, starting with dataset selection followed by preprocessing steps such as data cleaning and normalization aimed at achieving consistency and enhancing quality standards within the data set itself.

Subsequently, techniques including oversampling and under sampling—alongside hybrid sampling strategies—are applied to equilibrate distributions across datasets before partitioning them into training and testing subsets necessary for developing and accessing models effectively.

An ensemble framework is established using Adaboost alongside KNN, RF methods integrated through a voting classifier approach maximizes detecting fraudulent activities' accuracy.

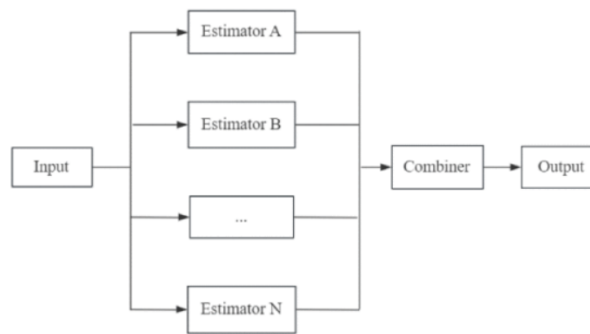
During training phases preparations involve inputting cleaned datasets into chosen models which after being trained will undergo evaluations against test sets subsequently enabling classification decisions about transactions being either fraudulent or non-fraudulent; this efficacy is gauged via relevant performance metrics complemented by outcome analyses conducted thereafter.

Algorithm 1 - Random Forest Classification Algorithm (RFC)

Once the sampling process is complete, the data is divided into two separate sets, which is a critical aspect of machine learning. In the first instance, the sets are randomized in an attempt to obtain sets that represent the data's distribution dynamics. Specifically speaking, FH has used 20% of the dataset for the autonomous driving car testing and used the remainder in training the car. In this demonstration, will first randomize the given dataset and set 20 % aside for testing and the remaining 80% for training. This division and

differentiation ensure that there is adequate data available for learning the model and a separate set of data for validating the model. The processes of learning and validation go hand in hand; the model will have to be trained thereafter using the training set. This implies that the car will be programmed to learn and capture the reality that inputs data. Finally, the model will be validated using the testing set, and this will be new and unseen at this stage.

Figure 2 - Basic flow chart of integrated learning.



3.5.1. Data Pre-processing

The first step involves pre-processing the dataset to prepare it for implementation. This phase includes several processing methods:

- Standardization of the 'Amount' column was performed to enhance analysis capabilities.
- The 'Time' column was removed from the dataset, as its influence on both training and evaluation processes was minimal.
- Duplicate entries within the dataset were identified and eliminated.

The limitations of this dataset stemmed from a lack of information regarding its features, making feature selection and engineering more challenging due to insufficient visibility into feature characteristics.

3.5.2. Under-Sampling

A random sample comprising legitimate transactions labeled as 0 was drawn from the majority class. The number of samples selected corresponded with the required ratio for the minority class. To improve model training, an equal number of instances for both classes was achieved by randomly selecting entries from the majority group that matched in size with those in the minority class, thus consolidating them into one cohesive dataset.

3.5.3. SMOTE and Edited Nearest Neighbors

The Edited Nearest Neighbor (ENN) technique is employed to decrease a dataset's size by removing instances whose labels differ from that of their nearest neighbors belonging to the majority class label initially targeted by classification algorithms like SMOTE (Synthetic Minority Over-sampling Technique). First, SMOTE generates synthetic samples for enhancing representation within the minority class before applying ENN on this combined set—comprising real data alongside these artificially created examples—to eliminate misleading or noisy observations altogether; this integration aims at effectively addressing issues related to imbalanced datasets while also reducing potential overfitting through noise elimination strategies tailored specifically towards false positives and negatives encountered during CCFD assessment tasks.

3.5.4 Model Training

After completing the sampling procedure, the data is separated into separate training and testing sets, a crucial step in machine learning. Initially, the dataset is shuffled randomly to ensure that both sets

accurately represent the overall distribution of data. Subsequently, 80% of this randomized dataset will be allocated for training purposes while the remaining 20% will be designated for testing. This division ensures that there is sufficient data available for model training while also providing a distinct set to evaluate its performance. The model will then be trained using the training set, allowing it to learn relationships between input features and target variables. Finally, assess the model’s performance with the testing set; this portion serves as new and unseen information.

Approximating the Model’s Performance on Unseen Data.

4. Results and Discussion

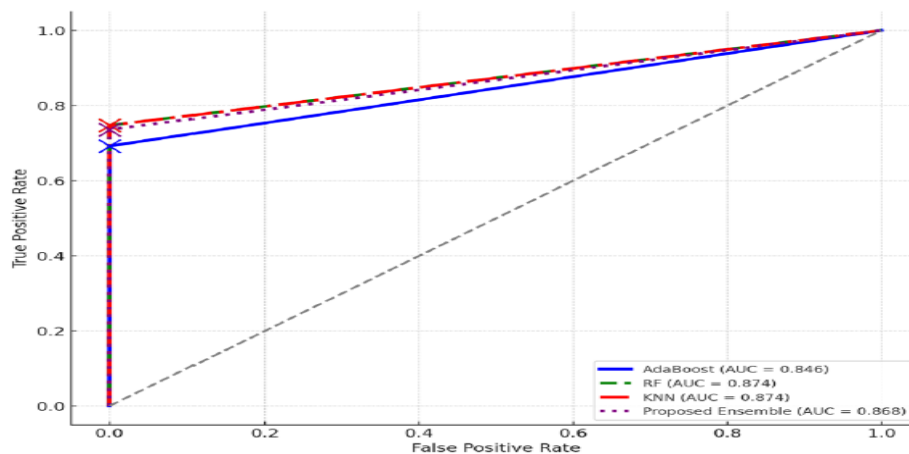
The following presents an evaluation of the performance for the proposed ensemble machine learning model, which was assessed without sampling methods, alongside comparisons involving under-sampling, SMOTE (Synthetic Minority Over-sampling Technique), and SMOTE + Tomek Links. This ensemble model incorporates K-Nearest Neighbors (KNN) with five neighbors, a Random Forest (RF) comprising 100 decision trees, and AdaBoost featuring 50 weak learners; it culminates in a voting classifier utilizing soft voting.

Without data sampling of the original dataset, RF was compared with six basic classifiers LR, KNN, Naive Bayes (NB), AdaBoost, GBDT, and SVM, and the best-performing classifiers among them were selected for the experimental comparison through different data sampling methods, and the results of the experiments are shown below.

Table 1 – Model table

Model	Accuracy	Recall	Precision	F1	AUC	MCC
LR	0.999122	0.581633	0.863636	0.695122	0.790737	0.708353
KNN	0.998543	0.193878	0.826087	0.31405	0.596904	0.399791
NB	0.977827	0.816327	0.060377	0.112439	0.897216	0.218423
AdaBoost	0.999315	0.72449	0.855422	0.78453	0.862139	0.786905
GBDT	0.998947	0.602041	0.7375	0.662921	0.800836	0.665823
SVM	0.999298	0.622449	0.953125	0.753086	0.811198	0.769946
RF	0.999544	0.755102	0.973684	0.850575	0.877533	0.85725

Figure 3 - ROC curves of the classifier’s



The accuracy of all seven classifiers was extremely high without any data preprocessing, and catastrophe was extremely high because of the data in fraud detection that was highly skewed with much more normal events than fraud events. Accuracy alone does not mean anything on the quality of the model, as it

predicted that all events would be good. The RF classifier had the highest recall of 0.365329 and AUC of 0.756239 under the seven classifiers. However, in terms of accuracy, precision, F1 score, and MCC, it greatly improved the results which were the best results compared to the other six classifiers plain Bayes had the highest values of recall and AUC under the seven classifiers. The second highest occurred in RF, but the precision, F1 score, and MCC of plain Bayes were only 0.070387, 0.172449, and 0.618723, being the worst among the seven classifiers. All model metrics combined suggest that the RF classifier did well to sequence the credit card fraud data standing the high imbalance of data categories. Overall, the RF classifier was chosen as the best model for credit card fraud detection from the other six classifiers

Comparison with Existing Models

Numerous studies have employed SMOTE combined with ENN to tackle issues of data imbalance in credit card fraud detection, primarily focusing on data preprocessing without further exploration into the integration of SMOTE + ENN alongside classifiers, especially ensemble methods.

Additionally, a number of earlier works have effectively merged SMOTE with ensemble classifiers for detecting credit card fraud, showcasing its capability to enhance performance by mitigating data imbalance. However, there has been no investigation into the combination of SMOTEEN and ensemble classifiers until now. This paper aims to fill that gap by illustrating the benefits of utilizing the hybrid data balancing technique offered by SMOTEEN together with ensemble classifiers.

The models discussed in [17] and [18] share similarities with our proposed model. Specifically, Prusti and Rath's model incorporated KNN, Extreme Learning Machines (ELM), Random Forests (RF), Multilayer Perceptron (MLP), and Bagging classifier techniques; while Sahithi et al.'s approach utilized RF, KNN, Logistic Regression (LR), AdaBoost, as well as Bagging methods. In contrast to these approaches where only SMOTE was applied for addressing class imbalances—as noted in [19]—our proposed method employs both SMOTE + ENN.

Classifier Performance and Robustness

The classifier utilizing the SMOTE + ENN method exhibits outstanding performance, outperforming alternative models in terms of accuracy, precision, and F1 score. The capacity of the proposed ensemble model to eliminate false negatives highlights its effectiveness in accurately detecting all instances of fraud. This characteristic significantly enhances the model's practical applicability and reliability. Additionally, this robust ensemble approach not only eradicates false negatives but also achieves a markedly lower rate of false positives. Such a balance is critical for effective fraud detection because minimizing false positives helps prevent unnecessary transaction interruptions while ensuring that all fraudulent activities are identified.

Through comprehensive experimentation, we demonstrate that this integrated methodology improves fraud detection accuracy by enhancing essential metrics including precision, recall, F1-score, and AUC-ROC values. Our findings indicate that training ensemble classifiers on a more balanced dataset created through SMOTE + ENN leads to reductions in both false positives and false negatives, thus making fraud detection processes more dependable.

5. Conclusion and Future Directions

This study identifies significant challenges within the realm of Continuous Credit Fraud Detection (CCFD). Choosing appropriate classifiers proves difficult due to the fast-paced evolution of machine

learning algorithms alongside new patterns in fraudulent behavior; ongoing adaptation along with thorough evaluation is necessary for sustaining model efficacy.

We assessed our proposed models using real-world datasets which led us to develop an integrating framework consisting of AdaBoost, Random Forest (RF), and K-Nearest Neighbors (KNN) within a voting classifier architecture complemented by over-sampling techniques like SMOTE + ENN. The results underscore strong performance from our suggested model while illustrating how combining various classifiers can enhance accuracy in fraud detection efforts further shown through improved AUC (Area Under the Curve) scores across AdaBoost RFs as well as KNN implementations addressing existing challenges effectively [20].

For example, such assessment reports of classification performance were based on AUC score, hence can be applied for the percentages of fraud detected correctly when any model is used for fraud transaction detection, regardless of the classification model. The models were thoroughly tested during assessments to determine their performance [21].

Additionally, other measures were used to determine the processing time of the proposed model to assess its competency to be used in real-time fraud detection. Implementation of this model in real time can be in live systems whereby the model could be integrated into a system within a bank or government system. A live data stream would be processed through the model to generate fraud alerts that should be immediate notifications or generation of reports to be submitted to labor for the actions required to stop the fraud[22]. The reports would have special insights requiring decision-making, while the reports would be integrated with feedback processes supervised by humans refining the model from time to time. In future, we will look to test on large datasets to know if the model can be applied on a large scale. These datasets should be from different environments like an organization or countries to help notice the effects of different environments on the proposed model.

6. Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

7. Acknowledgements

This research received no specific grants from funding agencies in the public, commercial, or not-for-profit sectors.

8. Data availability

The data is available online on Kaggle and listed in the reference list.

References

1. P. Rao, V. Singh, "Hybrid Models for Fraud Detection in Payment Systems," *Journal of Payment Security*, 2021.
2. Q. Zhou, R. Li, "Explainable AI in Credit Card Fraud Detection," *Journal of Explainable Computing*, vol. 3, no. 2, 2022.
3. F. K. Alarfaj, I. Malik, H. U. Khan, N. Almusallam, M. Ramzan, and M. Ahmed, "Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms," *IEEE Access*, vol. 10, pp. 39700–39715, 2022.

4. T. K. Dang, M. V. Tiep, "Machine learning based on resampling , approaches and deep reinforcement learning for credit card fraud detection systems," Appl. Sci., vol. 21
5. D. Mienye and N. Jere, "A survey of decision trees: Concepts, algorithms, and applications," IEEE Access, vol. 12,
6. Üzeyir Fidan, "Individual Privacy Perception in the Digital Age: The Interaction of Artificial Intelligence Attitude and Dependency", OPUS Journal of Society Research, vol.22, no.5, pp.869, 2025.
7. Bayes' theorem (https://scikit-learn.org/stable/modules/naive_bayes.html)
8. Manuel Herrador, Johann Rehberger, "SpAIware: Uncovering a novel artificial intelligence attack vector through persistent memory in LLM applications and agents", Future Generation Computer Systems, vol.174
9. Alexander Lerch, Claire Arthur, Nick Bryan-Kinns, Corey Ford, Qianyi Sun, Ashvala Vinay ,ACM Computing Surveys, 2025
10. Jiewu Leng, Baicun Wang, Weiming Shen, "AIGC-empowered smart manufacturing: Prospects and challenges", Robotics and Computer-Integrated Manufacturing, vol.97
11. Bayes' theorem (https://scikit-learn.org/stable/modules/naive_bayes.html)
12. Fahd Ali Raza, Abtar Darshan Singh, Jonathan Jeevan Strinivas Kovilpillai, Analisa Hamdan, Vaikunthan Rajaratnam, "Safeguarding Integrity in AI-Enhanced Education: Stakeholder Perspectives on Accuracy, Validity, and Ethics in ASEAN", European Journal of STEM Education, vol.10, no.1, pp.22, 2025.
13. N. S. Alfaiz and S. M. Fati, "Enhanced credit card fraud detection model using machine learning," Electronics, vol. 11,
14. D. Mienye and N. Jere, "A survey of decision trees: Concepts, algorithms, and applications," IEEE Access, vol. 12,
15. Muyan Li, , Siyuan Peng, "User personal information protection in generative AI services: a perspective from the Chinese platform policies", The Electronic Library, 2025.
16. Figshare ([Figshare_Creditcard_Dataset](#) is a open source data available at [Figshare_files](#))
17. Jianxin Li, Zhixue Zhao, "AI-Generated Content in Cross-Domain Applications: Research Trends, Challenges and Propositions", Knowledge-Based Systems, pp.114634, 2025.
18. Muyan Li, , Siyuan Peng, "User personal information protection in generative AI services: a perspective from the Chinese platform policies", The Electronic Library, 2025.
19. Üzeyir Fidan, "Individual Privacy Perception in the Digital Age: The Interaction of Artificial Intelligence Attitude and Dependency", OPUS Journal of Society Research, vol.22, no.5, pp.869, 2025.
20. MD Sarfaraz Momin, Abu Sufian, Debaditya Barman, Marco Leo, Cosimo Distanto, Naser Damer, "Explainable deepfake detection across different modalities: An overview of methods and challenges", Image and Vision Computing, pp.105738, 2025
21. M. A. Islam, M. A. Uddin, S. Aryal, and G. Stea, "An ensemble learning approach for anomaly ,detection in credit card data with imbalanced and overlapped classes," J. Inf. Secur. Appl., vol. 78, Nov. 2023, Art. no. 103618.
22. N. S. Alfaiz and S. M. Fati, "Enhanced credit card fraud detection model using machine learning," Electronics, vol. 11,