Journal of Theoretical and Applied Information Technology
15th August 2023. Vol.101. No 15

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING ALGORITHM

[1]MOH. ALIANSYAH, [2]FERALIANA AUDIA UTAMI, [3]RIYANTO JAYADI
[1, 2, 3]Information Systems Management Department, BINUS Graduate Program-Master of Information Systems Management, Bina Nusantara University, Jakarta 11480, Indonesia
E-mail: [1]moh.aliansyah@binus.ac.id, [2]feraliana.utami@binus.ac.id, [3]riyanto.jayadi@binus.edu

ABSTRACT

Credit card fraud is a serious issue in the financial services industry. Billions of dollars are lost each year due to credit card fraud. Credit cards have become one of the fastest growing financial services by banks in recent years. However, with the increasing number of credit card users, banks face an increasing rate of credit card defaults. Credit card fraud is related to the use of prohibited credit card materials for acquisition. The aim of this research is to analyze and compare the usage of three types of machine learning algorithms consisting of Neural Network (NN), K-Nearest Neighbour (KNN), and Random Forest (RF) to analyze and predict credit card fraud security. The credit card transaction dataset was sourced from European cardholders and contained 284,807 transactions. The results of the analysis state that the Neural Network Algorithm has the highest prediction capability (99.80%) and precision rate (100%) compared to the other two machine learning algorithms.

Keywords: *Fraud Detection, Machine Learning, Neural Network, Random Forest, K-Nearest Neighbor*

1. INTRODUCTION

Technological advancements and the pandemic situation have led to a shift towards a cashless lifestyle, where more transactions are processed virtually through credit card payment systems in cryptocurrency [1]. Credit cards are widely used due to their usefulness and direct interaction with the banking system, making them easily accessible at the same time. Credit cards-related crimes are on the rise, and credit card fraud detection uses past data to identify whether a transaction is fraudulent or not [2]. This decision is notoriously tough due to variations in client purchasing behavior, such as during holiday seasons, and in fraudster strategies, particularly those used to adapt to fraud detection tools. In other words, abundant transaction records can be used to train well-performing classifiers to identify fraudulent transactions [3].

Fraudulent activities worldwide present a common global challenge for banks and other financial institutions involved in issuing credit cards [4]. Moreover, the clear exposure of security loopholes in traditional credit card processing systems has dramatically increased credit card fraud, resulting in annual losses of billions of dollars globally. Most fraudsters have designed sophisticated methods to carry out credit card fraud [5].

Fraud is very easy to happen if the targeted person has limited understanding of internet transactions and money transfers. Hackers typically target innocent persons who lack thorough knowledge of and use of online banking applications. Usually, cardholders send their expiration dates and card numbers, and if they forget their bank PIN number, the bank gives them a verification number on their phone or email [6]. Credit card fraud is an easy and risk-free sign, hackers withdraw amounts without the knowledge of the owner and the bank, they transmit the money in a very short period of time and then leave the network. It is difficult to identify the fraudster because they employ very fast instruments and masterminds who have complete knowledge of the owner's credit card and financial transaction moments [7]. Fraudsters are continuously attempting to disguise every fraudulent transaction as legal, making fraud detection a difficult and time-consuming operation. Data analysis enables banks to exchange ambitious data methods in order to deal with actual customer-generated data more effectively. Credit card default prediction is a fundamental estimate. concerns the bank's involvement in calculating credit to better

understand why customers tend to default on payments [5].

The bank wants every small detail of customers to track payment data added to credit history. Credit card fraud detection is a challenging issue that has attracted the attention of machine learning researchers and scientists [8]. Several previous studies analyzed and compared the use of various machine learning algorithms to detect credit card fraud and found different results. For example, Domadula and Geetha found that decision tree, random forest, and logistic regression are capable of producing accurate predictions [9]; Ebiaredoh-Mienye et al. found that batch normalization methods optimized with Adamax Algorithm have high accuracy in detecting credit card fraud [10]; Randhawa et al. proved that machine learning method MCC has the highest accuracy compared to NB, SVM, and DL methods [11]; Nandi et al. stated that the Behavior-Knowledge Space (BKS) method has high effectiveness in detecting credit card fraud [12]; according to Jain et al., Random Forest is a machine learning algorithm that has a higher prediction accuracy rate compared to Decision Tree and XGBOOST [13]; Bhavya et al. proved that Logistic Regression has a better accuracy rate compared to Neural Network and K-Means [14]; Sangeetha et al. proved that Neural Network has a higher accuracy rate compared to SVM [15]; and Asha & Kumar also stated that the accuracy rate of Neural Network in predicting credit card fraud is higher compared to SVM and KNN [16].

Based on the previous research findings and the phenomenon of the increasing trend in credit card fraud, this study will conduct an analysis and comparison of the use of three types of machine learning algorithms, namely K-Nearest Neighbor (K-NN) [17], Neural Networks [18] and Random Forest [7], to be used in detecting credit card fraud. The KNN algorithm was chosen for analysis because it has several advantages, such as being robust to training data with a lot of noise and being effective when the training data is large, which makes it suitable for use in detecting credit card fraud [7]. Random Forest has been proven to perform well in credit card fraud detection in the literature, therefore, it is interesting to study the impact of feature engineering strategies and comparing this machine learning algorithm to other method [19]. Neural Network is a method developed by mimicking the characteristics of the human brain in storing and processing information, so it is considered to be the closest to manual processes in terms of accuracy [6].

Analysis of the use and comparison of these three methods is important to conduct for several reasons. First, there has been no previous research that analyzes and compares these three methods, so the findings of this study can enrich the literature on machine learning algorithms, particularly with the KNN, Random Forest, and Neural Network methods. Second, the study in this research describes the scope of the application of machine learning algorithms in the security field, particularly for detecting credit card fraud. Third, the results of the comparison of the accuracy levels between Neural Network, KNN, and Random Forest can serve as a reference for further research and for practitioners to choose the most appropriate machine learning algorithm to be applied in detecting credit card fraud.

Based on the overall description above, the aim of this research is to analyze and compare the usage of three types of machine learning algorithms consisting of Neural Network (NN), K-Nearest Neighbour (KNN), and Random Forest (RF) to analyze and predict credit card fraud security. Financial transactions containing any kind of criminal activity or fraud can be analyzed to detect such crimes. To build this machine learning model, the dataset from https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud?resource=download was used, which contains 284,807 financial transactions with 492 of them being fraud data. This data contains transactional data records in September 2013 by European cardholders. The principle of credit card companies is to be able to recognize fraudulent credit card transactions so that customers are not charged for items they did not purchase.

2. LITERATURE REVIEW

2.1 Crisp-DM

Data analysis never comes from a momentary process, but from a mature process guided by high standards. The author's statement above is open to debate, but for the standard part of the statement, the author believes that all parties will agree. This is what motivates the author in this article to convey CRISP-DM as one of the standards in data mining [1].

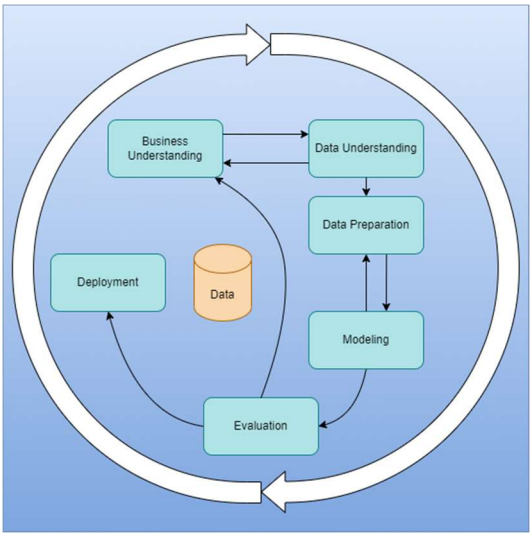

Figure 1: Crisp-DM Framework

Understanding the Business: While the business in this article is now confined to modeling and reporting results, this dataset can be utilized to identify individuals or activities in commercial use cases. Understanding the Data: The data is largely numerical, with some mathematical derivations resulting in later described data, which sets off the classification challenge that must be answered. Data Preparation: As mentioned in the Dataset Description section, the dataset owner has done the most of the heavy work in this dataset. However, data selection and preparation continue to be performed in order to increase cleanliness. Modeling: Because the dataset owner labeled it during the data processing step, unsupervised data mining approaches are not applicable. In the classification situation, neural networks, KNN, and Random Forest are useful since they can deal with more than two output classes. They provide an excellent spectrum of flexible, rigid, good with higher dimensions, and poor with higher dimensions noise [20]. The procedure was performed with standard arguments, as seen. Because the data downsampling to ideal balance pertains to numerous variables, the evaluation is exclusively based on correctness. Deployment: Although there is no deployment to a study case, the model can be implemented correctly and with great accuracy. There is some back and forth over the CRISP-DM cycle, but it stops for this task in the evaluation [18].

2.2 Neural Network

The Neural Network algorithm is a method inspired by the neural network of the human brain, aiming to mimic how the human brain processes and retains information [6]. By expressing a complicated relationship between input and output, the Neural Network technique is used to represent non-linear statistical data.

$$Ij = \sum_{i} wij\, Oi + \theta j$$

Neural Networks consist of several layers, with at least 1 or more Processing Elements (PE) in each layer. Processing Elements are used to simulate the way neurons work in the human brain. Therefore, PEs are often referred to as neurons or nodes, with each PE receiving input from the previous layer [21].

2.3 K-NEAREST NEIGHBOUR (KNN)

The K-Nearest Neighbor (KNN) algorithm is a method for identifying objects based on the training data points that are the nearest to the item. The training data is projected into a multidimensional space, with each dimension representing a data feature [22].

$$\text{Distance} = \sqrt{\sum_{i=1}^{n}\left(X_{training}^{i} - X_{testing}\right)^2}$$

The KNN method has various advantages, including its tolerance to noisy training data and effectiveness when training data is big. However, the disadvantage of KNN is that it requires determining the value of the K parameter (number of nearest neighbors), distance training is unclear about what type of distance should be used and which attributes should be used to get the best results, and the computational cost is quite high because distance calculations are required [7].

2.4 Random Forest

Random Forest is a Decision Tree method development that employs numerous Decision Trees, each of which has been trained using different samples, and each attribute is split on the chosen tree between randomly picked subsets of attributes. Random Forest has several advantages, such as being able to improve accuracy results if there is missing data, being resistant to outliers, and efficient for storing data. Additionally, Random Forest has a feature selection process that can select the best features to improve classification model performance. With feature selection, Random Forest can work effectively on big data with complex parameters [5].

2.5 Previous Studies

Dornadula and Geetha used a new method to identify credit card fraud detection. Various classifiers were used on three advanced evaluation collections and score-modified produced for each classifier type. Self-motivation variations in the structure lead the organization to bias the system. Their findings stated that decision tree, random forest, and logistic regression provided the best results and accuracy [9].

Ebiaredoh-Mienye et al. used machine learning algorithms to classify a large dataset of credit card data. Their findings introduced batch normalization methods to improve the model's accuracy and speed and further prevent overfitting. The model was optimized using the Adamax algorithm [10].

Moradi et al. proposed a dynamic model for credit card fraud risk assessment that outperforms the existing model. Their model has self-motivation tools that evaluate client behavior corruptly on a monthly basis and credit risk that includes fuzzy factors, particularly in financial crises [23].

Randhawa et al. presented credit card fraud detection using machine learning algorithms. Several typical models used were NB, SVM, and DL empirical studies. Their findings proposed the best MCC score was 82% achieved by DL. A hybrid mockup value, noise from 10% to 30%, was added to the model data [11].

Nandi et al. designed a multi-classifier to overcome the challenges of credit card fraud detection. An ensemble model with multiple machine learning classification algorithms was designed, where the Behaviour-Knowledge Space (BKS) was utilized to combine predictions from several classifiers. To ensure the effectiveness of the developed ensemble model, publicly available datasets as well as real financial records were used for performance evaluation. Through statistical tests, their results positively showed the effectiveness of the developed model compared to the majority voting method commonly used to combine predictions from various classifiers in handling noisy data classification and credit card fraud detection issues [12].

Jain et.al. applied a machine learning algorithm to a dataset of credit card fraud, and the performance of three machine learning algorithms was compared for detecting credit card fraud. Their findings stated that the Random Forest machine learning algorithm had the best accuracy compared to the Decision Tree and XGBOOST algorithms [13].

Dhankhad et.al. applied multiple supervised machine learning algorithms to detect credit card fraud transactions using a real-world dataset. Their research also compared the performance of various supervised machine learning algorithms available in the literature against the super classifier implemented in their paper [24].

Roy et.al. analyzed a comprehensive guide for sensitivity analysis of model parameters related to fraud detection performance. Their research also discussed a framework for topology parameter tuning of Deep Learning for credit card fraud detection to enable financial institutions to reduce losses by preventing fraudulent activities [25].

Melo-Acosta et.al. analyzed a Spark platform and Hadoop file system supporting our proposed solution, sequentially to enable scalability of the proposed solution. The proposed approach achieved an absolute improvement of around 24% in geometric mean compared to the standard random forest learning strategy [26].

Zhuhai analyzed two types of Random Forests to train normal and abnormal transaction behavior features. They compared two different random forests based on their basic classifiers and analyzed their performance in credit card fraud detection. The data used in this experiment was obtained from an e-commerce company in China [27].

The research conducted by Bhavya et.al. aimed to compare the accuracy and precision levels of Logistic Regression, K-Means, and Neural Networks models in detecting credit card fraud. The analysis results showed that the model with the highest accuracy rate was Logistic Regression, with an accuracy rate of 99.88%; followed by Neural Network with an accuracy rate of 99.61%; and lastly K-Means, which only had an accuracy rate of 54.27%. Based on these results, it can be concluded that the Logistic Regression model is better than the Neural Network and K-Means models because Logistic Regression has a higher accuracy rate [14].

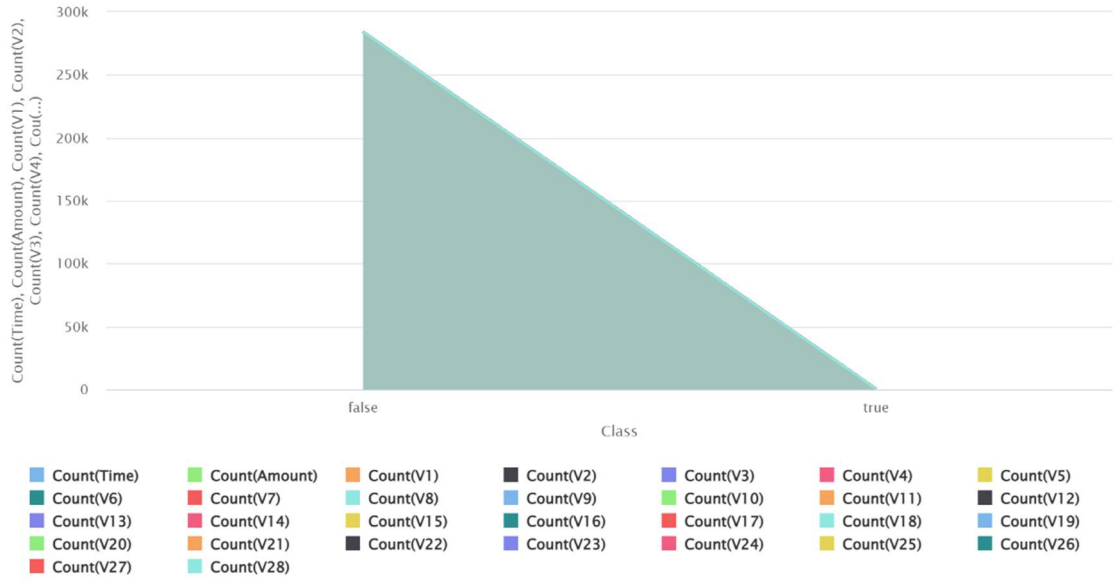

Figure 2: Observation Data Sample

Sangeetha et. al. compared the use of Artificial Neural Network (ANN) and Support Vector Machine (SVM) Algorithm classification techniques for credit card fraud detection. The analysis results showed that the ANN classification technique had an accuracy rate above 95%, while the SVM accuracy rate was only 93%. Based on these results, it can be concluded that the Neural Network classification technique is better because it has a higher accuracy rate in predicting credit card fraud [15].

Asha & Kumar analyze the use of several machine learning algorithms consisting of Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Artificial Neural Network to detect credit card fraud. The research results stated that the Neural Network is the most appropriate machine learning algorithm to detect credit card fraud. This is based on the highest accuracy rate of the Neural Network compared to others, which is 99.92%. The accuracy rate of SVM is 93.49%, while KNN has an accuracy rate of 99.82% [16].

3. METHOD

3.1 Data Understanding

The initial data used in this study was transaction data performed by European credit card holders in September 2013. This dataset presents transactions that occurred over two days, where we have 492 fraud cases out of 284,807 transactions. The results of the descriptive analysis of the data can be seen in Figure 2.

3.2 Data Preparation

The data preparation stage includes data selection, transformation, and cleaning activities. Preprocessing the data with the following steps to ensure that the data is suitable for further modeling

Table 1: Data Attribute Description

Attributes	Data Types	Description
Class	Integers	Response variable and requires a value of 1 if fraud occurs and 0 otherwise
Time	Integers	Contains the elapsed time between each transaction and the first transaction in the data set
Amount	Real	Number of transactions, this feature can be used for cost-sensitive learning that depends on examples
V1-V28	Real	Principal component obtained with PCA

and evaluation processes. Data selection is the stage of selecting the data needed to meet the modeling objectives. The data used in this study uses the data in figure 1. In the data cleansing stage, it is known that there are no missing values in each attribute of Class, Time, Amount, and V1-V28.

balanced, I performed down-sampling by artificially duplicating each class to be the same. So, I do not need to lose data and do not require special support, although it takes a lot of time to balance the data.

3.3 Modeling

The neural network (NN), k-nearest neighbor (K-NN), and Random Forest (RF)

Name	Type	Missing	Statistics			Filter (31 / 31 attributes): Search for Attributes
Time	Integer	0	Min 0	Max 172792	Average 94813.860	
Class	Integer	0	Min 0	Max 1	Average 0.002	
V1	Real	0	Min -56.408	Max 2.455	Average 0.000	
V2	Real	0	Min -72.716	Max 22.058	Average 0.000	
V3	Real	0	Min -48.326	Max 9.383	Average -0.000	
V4	Real	0	Min -5.683	Max 16.875	Average 0.000	
V5	Real	0	Min -113.743	Max 34.802	Average 0.000	

Figure 3: Missing Value

Table 2: Data Set

V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20	V21	V22	V23	V24	V25	V26	V27	V28	Time	Amount	Class
-1.35981	-0.07278	2.536347	1.378155	-0.33832	0.462388	0.239599	0.098698	0.363787	0.090794	-0.5516	-0.6178	-0.99139	-0.31117	1.468177	-0.4704	0.207971	0.025791	0.403993	0.251412	-0.01831	0.277838	-0.11047	0.066928	0.128539	-0.18911	0.133558	-0.02105	0	149.62	0
1.191857	0.266151	0.16648	0.448154	0.060018	-0.08236	-0.0788	0.085102	-0.25543	-0.16697	1.612727	1.065235	0.489095	-0.14377	0.635558	0.463917	-0.1148	-0.18336	-0.14578	-0.06908	-0.22576	-0.63867	0.101288	-0.33985	0.16717	0.125895	-0.00898	0.014724	0	2.69	0
-1.35835	-1.34016	1.773209	0.37978	-0.5032	1.800499	0.791461	0.247676	-1.51465	0.207643	0.624501	0.066084	0.717293	-0.16595	2.345865	-2.89008	1.109969	-0.12136	-2.26186	0.52498	0.247998	0.771679	0.909412	-0.68928	-0.32764	-0.1391	-0.05535	-0.05975	1	378.66	0

Data selection is determining the attributes for the data to be used in the final dataset to be processed using Rapid Miner. The data used are as follows, which can be seen in Table 2.

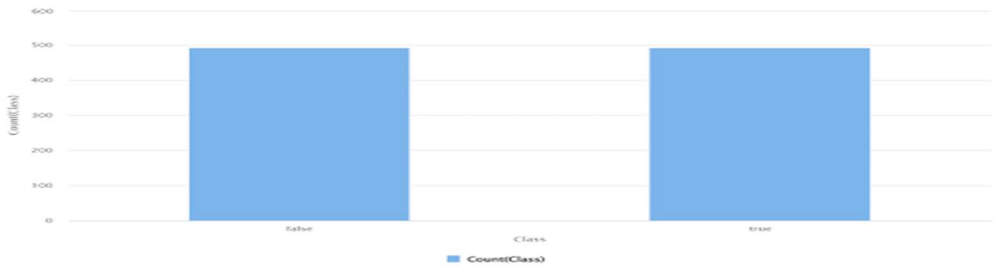

Figure 4: Data Balance based on Fraud and non-Fraud Category

There are 492 (or 0.172%) fraudulent transactions in the data. This means that the data is highly imbalanced with respect to the target variable class. After realizing that my dataset is not

algorithms are examples of machine learning/deep learning techniques. Artificial Intelligence (AI) includes Neural Networks (NN), which is a learning model affected by biological neuron activities [26].

A neural network is made up of nodes that operate on input data and provide output to other nodes. Each node's output is described as activation or node value. Weights are assigned to nodes to assist the network in adapting to changes. These weights describe the magnitude of an input's influence on the output.

The K-Nearest Neighbor (K-NN) algorithm is a method for classifying objects based on existing training data that is the nearest to the object [28]. As a result, in order to make predictions with K-NN, we must first devise a metric for measuring the distance between the query point and the cases from the example samples. The Euclidean system is one of the most prevalent methods for measuring this distance.

The Random Forest (RF) machine learning technique is quite beneficial. It is mostly utilized in classification, regression analysis, and other domains. During training, the RF algorithm generates a large number of decision trees [21]. RF is a supervised learning technique that uses testing data to train the model. It generates a random forest for a set of problems and then uses this random forest to find solutions.

Figure 5 illustrates the process of finding the best model using Machine Learning/Deep Learning. The process consists of several parts. The first part is the data source, which in this experiment is the credit card company dataset. Next is the data preparation, which involves handling missing values and imbalanced data. Numerical to Binominal is used to convert numeric data in transaction data to binomial data "true" and "false". Multiply is used to make copies of objects in RapidMiner. This operator takes an object from the input port and sends its copy to the output port.

Table 3: IT Balance Scorecard

PRESPEKTIF	KPI	TARGET
CORPORATE CONTRIBUTION	REALTIME TRANSACTIONS	100%
USER ORIENTATION	UAT (USER ACCEPTANCE TESTING)	100%
OPERATIONAL EXCELLENCE	AVAILABILITY OF SERVICE APPLICATIONS	100%
	SERVICE APPLICATION RELIABILITY	100%
	SERVICE APPLICATION CAPABILITIES	100%
FUTURE ORIENTATION	BACK UP	100%
	DRC (DISASTER RECOVERY CENTER)	100%
	DRP (DISASTER RECOVERY PLAN)	100%
	BCP (BUSINESS CONTINUITY PLAN)	100%

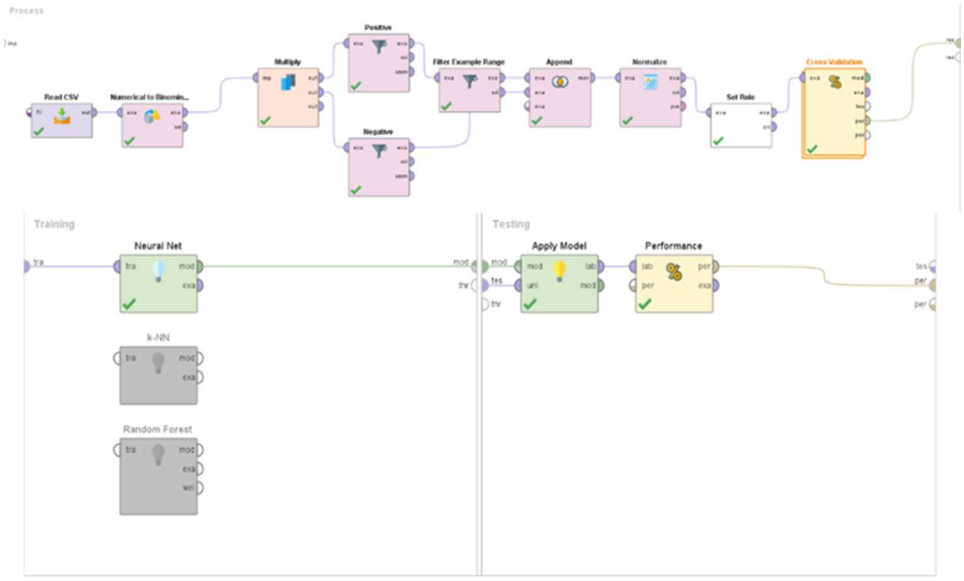

Figure 5: Experimental Design

Each connected port creates an independent copy. So, changing one copy does not affect the other copies.

The Filter Example returns those that match the condition specified by the "class," for example, setting positive "true" and negative "false." The Filter Example Range takes a set of examples and can reduce the number of examples in an example set but does not affect the attributes. Append is used to create an example set that is combined from 2 or more example sets that are compatible by adding all the combinations of example sets. Normalize is used to scale values to fit within a specific time range. Set Role explains how other operators handle this attribute. The default role is regular, while other roles are classified as special. Various types of roles are explained below in the parameter section.

Cross-validation is mainly used to estimate how accurate a model (learned by a particular learning operator) will perform in practice. There are two subprocesses: Training and Testing. The Training subprocess is used to train the model, which is then applied in the Testing subprocess. The model's performance is measured during the Testing phase. Data Split Training and Testing split the data into two parts: Training data is used to train the algorithm, while Testing data is used to determine the trained algorithm's performance. The algorithm used to find the best model is the Neural Network (input layer used with a value of 2), KNN, and Random Forest Classification models.

The best accuracy selection will be analyzed from the experiments performed, and the model with the best accuracy will be used to run the testing data. The model with the best accuracy is run on the testing data, and the results obtained will be analyzed for implementation purposes.

To measure the performance of Information Technology (IT) in this credit card company, a strategy for improving business efficiency based on the Balanced Scorecard-Grembergen four perspectives is implemented. The perspectives are Corporate Contribution, User Orientation, Operational Excellence, and Future Orientation. Corporate Contribution focuses on the strategic contribution of the company and the performance of IT services that are executed. This is taken from the perspective of the company management as the unit that manages all business activities. User Orientation focuses on how users perceive the company's credit card application from both the perspective of the users and the internal system maintenance, who aim to ensure user satisfaction.

Operational Excellence: This perspective emphasizes how the IT performance measurement process in a credit card company focuses on the operational system in credit card company applications. The goal is to pay attention to the consequences that may occur with the current operational system. Future Orientation: This perspective evaluates the IT system in terms of future implementation and personnel who use it, with the aim of exploring the company's performance in understanding current trends and opportunities to develop existing services as business assets.

Although the credit card company can carry out its business activities in its business unit, its IT division still needs to plan and establish indicators, initiatives, or activities that can align and support the company's overall strategies and determine these indicators and how significant the role of IT is in supporting the company's vision, mission, and strategy. By considering the balanced scorecard consisting of financial measures showing the results of actions taken as shown in the three other operational measure perspectives such as customer satisfaction, internal processes, and organizational capabilities in making improvements.

4. Results

4.1 Evaluation

The results of the rapid miner for the confusion matrix based on the three machine learning methods can be seen in Figures 6, 7, and 8.

accuracy: 99.80% +/- 0.43% (micro average: 99.80%)

	true false	true true	class precision
pred. false	492	2	99.60%
pred. true	0	490	100.00%
class recall	100.00%	99.59%	

Figure 6: Confusion Matrix Neural Network

accuracy: 95.23% +/- 1.92% (micro average: 95.22%)

	true false	true true	class precision
pred. false	484	39	92.54%
pred. true	8	453	98.26%
class recall	98.37%	92.07%	

Figure 7: Confusion Matrix KNN

accuracy: 99.80% +/- 0.43% (micro average: 99.80%)

	true false	true true	class precision
pred. false	491	1	99.80%
pred. true	1	491	99.80%
class recall	99.80%	99.80%	

Figure 8: Confusion Matrix Random Forest

Based on these figures, it can be seen that the highest accuracy (99.80%) was achieved by Neural Network, while the highest precision (100%) was also achieved by Neural Network using the same dataset combination. All three methods were treated with the same dataset, and we can see all of their results. Meanwhile, the highest ROC-AUC (100%) was achieved by Neural Network.

Neural Network provided us with good accuracy with efficient performance. The proposed method worked with 99.80% accuracy using the Neural Network algorithm, and this work was done with several different steps.

Table 4 shows a comparison between models. The machine learning and deep learning accuracy values obtained were best with 99.80% for Neural Network and Random Forest compared to the KNN model. The fraud data available is imbalanced data. To balance the dataset and obtain optimal values, upsampling can be used. The accuracy value for Neural Network yielded the best results, namely 99.80%, compared to other models. This accuracy value is better than the 0.172% sum.

Table 4: Comparison of Results among Models

MODEL	ACCURACY SCORE	ROC-AUC	PRECISION SCORE	RECALL SCORE
NEURAL NETWORK	99,80%	100%	100%	99,63%
KNN	95,23%	98,10%	98,36%	92,09%
RANDOM FOREST	99,80%	100%	99,80%	99,82%

Journal of Theoretical and Applied Information Technology
15th August 2023. Vol.101. No 15

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

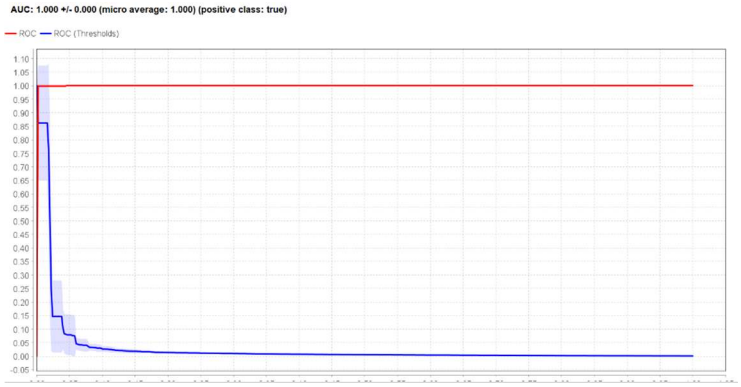

Figure 9: ROC/AUC Neural Network

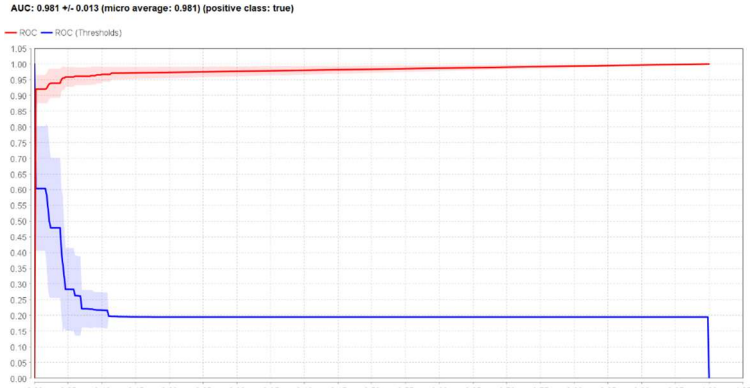

Figure 10: ROC/AUC KNN

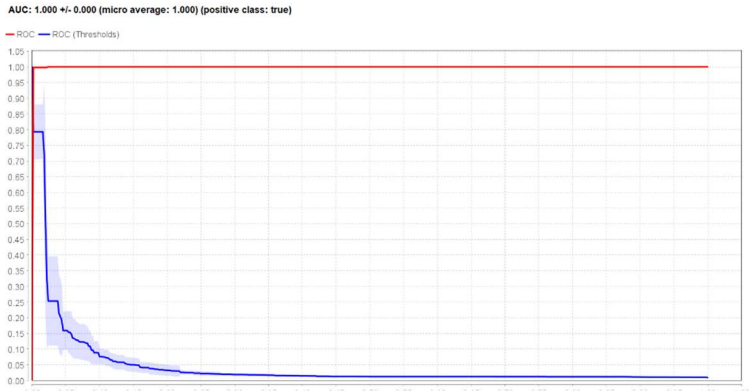

Figure 11: ROC/AUC Random Forest

In fraud detection, not only accuracy values but also precision scores should be considered as this value indicates how accurate the model is in detecting a fraud transaction compared to all actual fraud values. The higher the precision score, the more secure the model will be if implemented in fraud prevention. A normal transaction being classified as fraud is far more dangerous than a fraudulent transaction being classified as normal. These normal transactions will greatly harm the user, and financial losses will be suffered by the customer. Therefore, the appropriate model for fraud detection is Neural Network (NN). These findings are in line with the previous research by

Sangeetha et.al. and Asha & Kumar, who also analyzed the use of Neural Network algorithms and several other classification methods to detect credit card fraud. The research results of Sangeetha et.al. stated that Neural Network has the highest accuracy rate, which is up to 95% compared to Support Vector Machine, which has an accuracy rate of 93%. Similarly, the findings of Asha & Kumar stated that the accuracy rate of the Neural Network algorithm is the highest, which is 99.92%, while SVM has an accuracy rate of 99.49%, and KNN 99.82%. However, these research findings are not in line with previous research by Bhavya et.al., who found that the Neural Network classification method has an accuracy of 99.61%, which is below the accuracy of the Logistic Regression method with an accuracy rate of 99.88%.

Based on the explanation, it can be stated that the use of Neural Network method to detect credit card fraud is empirically proven to have a better accuracy rate compared to other methods such as Random Forest, KNN, and SVM. However, the accuracy rate of Neural Network may still need to be compared again with the Logistic Regression method which, according to previous research, has a slightly better accuracy rate.

4.2 Deployment

At this stage, I created a model and analyzed the business process to build a company based on the available data. This descriptive analysis is useful for creating predictive projects to contribute to the information provided based on the company's situation. My modeling results in a credit card company can be seen in Figure 11. By increasing the utilization of ITS for the company's sustainability, it will increase.

We refer to Effectiveness (the extent to which information technology infrastructure can provide benefits from the organization's perspective) and Efficiency (the extent to which the investment cost for information technology infrastructure can provide maximum throughput capacity in line with the demand from the business itself). Effectiveness can also be the alignment between the vision, mission, and long-term objectives towards the strategic suitability and governance of information technology so that a harmonious situation can be formed. The dynamism of vision, mission, and objectives is not impossible to occur, but it still remains a part that needs to be considered in the formulation of strategies that certainly affect IT planning.

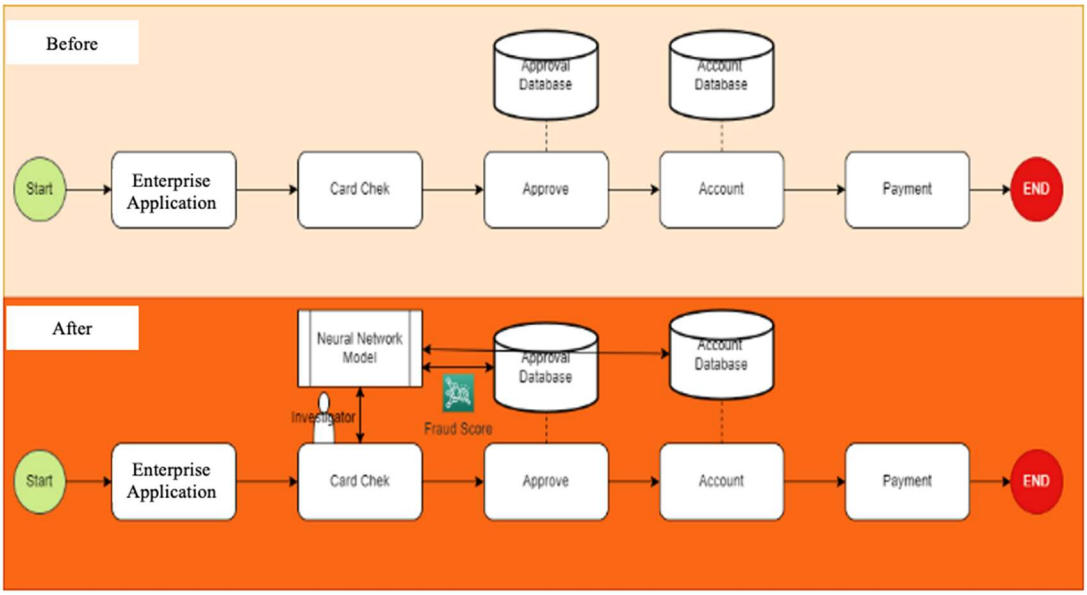

Figure 12: As-IS and To-Be Business Process

Efficiency is related to the maximization of the already invested cost so that maximum throughput can be obtained and the business can feel the best value from the invested investment. In some implementation cases, risk factors are also considered, so the tendency is not only towards the extent of the benefits that can be felt from an IT

investment but also towards long-term risks that may occur.

The Business Value of IT focuses on how much benefit can be felt from the implementation of information technology in an organization, measured from the perspective of effectiveness and efficiency. All business value measures relate to the relationship between information technology costs and its contribution to improving organizational performance, which can be measured in three dimensions: financial performance, business performance, and strategic performance.

Financial Performance is measured by the financial performance indicators of an organization. Profitability is obtained from productivity and efficiency that reduces operational costs; Revenue is measured by the organization's revenue presentation, divided by the indicator of the correct level of information technology expenditure; Productivity is the relationship between the output produced and the input used to produce that output from a system.

Business Performance can be measured by non-financial performance indicators as a complement and combination with financial performance measures. Non-financial performance measures are widely used by businesses for internal control and to link information technology costs with improved business performance, such as competitive advantages, sales of new products, development time for new products, customer satisfaction, etc.

The measurement of business value of information technology by linking information technology costs and the strategic performance of the organization, which can only be measured by the total cost and how the organization can realize Critical Success Factors (CSFs), the most critical activities that contribute to the greatest business success, and ultimately if critical activities can be carried out well, competitive performance can be guaranteed for the organization.

The next point as part of BtripleE highlights the effectiveness of information technology, which is slightly different. This effectiveness focuses more on how other business units see information technology within the scope of its processes but still on the ability of information technology to provide all the functions needed with the availability of existing costs. Existing functional departments, such as marketing, procurement, sales, etc., see the suitability of needs between resource availability so that work can be carried out better in three limitations: speed, capacity, and reduced human interaction. The cost incurred for this can continue to be adjusted until an increase in speed, capacity, and less human interaction is felt, which is not too significant in increasing per unit of currency. This can mean that the level of investment has reached the optimal point and no further increase in information technology investment is needed.

The third point in BtrippleE is IT Supply Effectiveness and IT Supply Efficiency. These two keywords emphasize different aspects. Every organization has its own goals reflected in its vision, ways to achieve them reflected in its mission, and objectives that are a set of targets aligned with the long-term mission of the company. Although there are two keywords in this point, they both highlight the same thing, which is IT Supply. IT Supply can be interpreted as the expenses incurred in the operational use of information technology. Effectiveness looks at how well the level of suitability of the IT strategy (detailed in the IT Expenditures) can be linked to the long-term goals of the company. Meanwhile, efficiency refers to the minimum cost incurred in pursuing these

Table 5: B3EEE Framework of the Companies

BTripleE Factor	KPI	Target
Business Management	Real Time	100%
	Cost per unit	<30%
	Manufacturing Cycle Effectiveness	100%
IT Management	System Availability	100%
	System Reliability	100%
	Capacity System	100%
	User Satisfaction	10 (1-10)
	User Acceptance Testing	100%
IT Supply Management	Backup	100%
	DRC (Disaster Recovery Center)	100%
	DRP (Disaster Recovery Plan)	100%
	BCP (Business Continuity Plan)	100%

targets so that they can be easily monitored by comparing the expenses and business improvement related to information technology.

To connect a business goal with the IT/IS investment to be implemented, a thorough planning by the company is necessary in order to achieve the goal.

Based on the IT BSC and BTripleE Framework perspectives, a credit lending company measures KPIs in detail so that the implementation of IT can be maximized or optimized sustainably. This triggers the success factors of fulfilling the presence of organization strategies that are constantly updated in both short and long term.

5. CONCLUSION

In this research, various machine learning techniques and methods were used to analyze and predict credit card fraud security. An anti-fraud approach can be adopted to prevent banks from significant damages and minimize threats. The aim of this study was taken differently from typical classification problems as we had cost of misclassification variables. Three machine learning algorithms, namely Neural Network, KNN, and Random Forest, were compared in terms of accuracy using a dataset for credit card fraud detection.

Based on the results and experimental findings, it was proven that the Neural Network Algorithm has highest accuracy and precision in predicting credit card fraud with an accuracy of 99.80% and accuracy of 100%. While Random Forest has 99,80% accuracy score and 99,80% precision score; and KNN has 95,23% accuracy score and 98,36% precision score.

The findings of this research indeed prove the high accuracy and precision of the Neural Network machine learning algorithm compared to KNN and Random Forest. However, the analysis was not conducted to compare Neural Network and Logistic Regression, which according to previous research, has a higher level of accuracy. Based on this, further research can try to analyze the use of four machine learning algorithms at once, namely Neural Network, Random Forest, KNN, and Logistic Regression, and compare their accuracy and precision in detecting credit card fraud.

REFERENCES:

[1] J. Vater, P. Schamberger, A. Knoll, and D. Winkle, "Fault Classification and Correction based on Convolutional Neural Networks exemplified by laser welding of hairpin windings," in *2019 9th International Electric Drives Production Conference (EDPC)*, IEEE, Dec. 2019, pp. 1–8. doi: 10.1109/EDPC48408.2019.9012044.

[2] F. Carcillo, Y.-A. Le Borgne, O. Caelen, Y. Kessaci, F. Oblé, and G. Bontempi, "Combining unsupervised and supervised learning in credit card fraud detection," *Inf Sci (N Y)*, vol. 557, pp. 317–331, May 2021, doi: 10.1016/j.ins.2019.05.042.

[3] Z. Li, G. Liu, and C. Jiang, "Deep Representation Learning With Full Center Loss for Credit Card Fraud Detection," *IEEE Trans Comput Soc Syst*, vol. 7, no. 2, pp. 569–579, Apr. 2020, doi: 10.1109/TCSS.2020.2970805.

[4] T. A. Olowookere and O. S. Adewale, "A framework for detecting credit card fraud with cost-sensitive meta-learning ensemble approach," *Sci Afr*, vol. 8, p. e00464, Jul. 2020, doi: 10.1016/j.sciaf.2020.e00464.

[5] S. Bagga, A. Goyal, N. Gupta, and A. Goyal, "Credit Card Fraud Detection using Pipeling and Ensemble Learning," *Procedia Comput Sci*, vol. 173, pp. 104–112, 2020, doi: 10.1016/j.procs.2020.06.014.

[6] F. Zamachsari and N. Puspitasari, "Penerapan Deep Learning dalam Deteksi Penipuan Transaksi Keuangan Secara Elektronik," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 2, pp. 203–212, Apr. 2021, doi: 10.29207/resti.v5i2.2952.

[7] S. J. Saleh, S. Q. Ali, and A. M. Zeki, "Random Forest vs. SVM vs. KNN in classifying Smartphone and Smartwatch sensor data using CRISP-DM," in *2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)*, IEEE, Oct. 2020, pp. 1–4. doi: 10.1109/ICDABI51230.2020.9325607.

[8] M. Ali, R. Prasad, Y. Xiang, and R. C. Deo, "Near real-time significant wave height forecasting with hybridized multiple linear regression algorithms," *Renewable and Sustainable Energy Reviews*, vol. 132, p. 110003, Oct. 2020, doi: 10.1016/j.rser.2020.110003.

[9] V. N. Dornadula and S. Geetha, "Credit Card Fraud Detection using Machine Learning Algorithms," *Procedia Comput*

Sci, vol. 165, pp. 631–641, 2019, doi: 10.1016/j.procs.2020.01.057.

[10] S. A. Ebiaredoh-Mienye, E. Esenogho, and T. G. Swart, "Artificial neural network technique for improving prediction of credit card default: A stacked sparse autoencoder approach," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 5, p. 4392, Oct. 2021, doi: 10.11591/ijece.v11i5.pp4392-4402.

[11] K. Randhawa, C. K. Loo, M. Seera, C. P. Lim, and A. K. Nandi, "Credit Card Fraud Detection Using AdaBoost and Majority Voting," *IEEE Access*, vol. 6, pp. 14277–14284, 2018, doi: 10.1109/ACCESS.2018.2806420.

[12] A. K. Nandi, K. K. Randhawa, H. S. Chua, M. Seera, and C. P. Lim, "Credit card fraud detection using a hierarchical behavior-knowledge space model," *PLoS One*, vol. 17, no. 1, p. e0260579, Jan. 2022, doi: 10.1371/journal.pone.0260579.

[13] V. Jain, M. Agrawal, and A. Kumar, "Performance Analysis of Machine Learning Algorithms in Credit Cards Fraud Detection," in *2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, IEEE, Jun. 2020, pp. 86–88. doi: 10.1109/ICRITO48877.2020.9197762.

[14] L. Bhavya, S. Reddy, A. Mohan, and S. Karishma, "Credit card fraud detection using classification, unsupervised, neural network models," *International journal of engineering research & technology*, vol. 9, no. 4, pp. 806–810, 2020.

[15] S. Sangeetha, T. Selvi, M. Sirija, and R. Reena, "Credit card detection using artificial neural network (ANN) algorithm," *International research journal of engineering and technology*, vol. 9, no. 7, pp. 2565–2570, Jul. 2022.

[16] R. B. Asha and K. R. S. Kumar, "Credit card fraud detection using artificial neural network," *Global Transitions Proceedings*, vol. 2, no. 1, pp. 35–41, Jun. 2021, doi: 10.1016/j.gltp.2021.01.006.

[17] M. Sadali, Y. K. Putra, and Mahpuz, "Evaluation Of Lecturer Education And Teaching Performance Through E-Monevin Using K-Nearest Neighbor (K-NN) Algorithm," *J Phys Conf Ser*, vol. 1539, no. 1, p. 012017, May 2020, doi: 10.1088/1742-6596/1539/1/012017.

[18] T. Küfner, F. Döpper, D. Müller, and A. G. Trenz, "Predictive Maintenance: Using Recurrent Neural Networks for Wear Prognosis in Current Signatures of Production Plants," *International Journal of Mechanical Engineering and Robotics Research*, pp. 583–591, 2021, doi: 10.18178/ijmerr.10.11.583-591.

[19] Y. Lucas *et al.*, "Towards automated feature engineering for credit card fraud detection using multi-perspective HMMs," *Future Generation Computer Systems*, vol. 102, pp. 393–402, Jan. 2020, doi: 10.1016/j.future.2019.08.029.

[20] N. Rtayli and N. Enneya, "Enhanced credit card fraud detection based on SVM-recursive feature elimination and hyper-parameters optimization," *Journal of Information Security and Applications*, vol. 55, p. 102596, Dec. 2020, doi: 10.1016/j.jisa.2020.102596.

[21] V. Aithal and , Roshan David Jathanna, "Credit Risk Assessment using Machine Learning Techniques," *International Journal of Innovative Technology and Exploring Engineering*, vol. 9, no. 1, pp. 3482–3486, Nov. 2019, doi: 10.35940/ijitee.A4936.119119.

[22] W. Yustanti, "Algoritma K-Nearest Neighbour untuk memprediksi harga jual tanah," *Jurnal matematika, statistika, & komputasi*, vol. 9, no. 1, pp. 58–68, Jul. 2012.

[23] S. Moradi, A. Rashidi, and M. Golmohammadian, "The Effectiveness of Positive Thinking Skills on Academic Procrastination of High School Female Students Kermanshah City," *Interdisciplinary Journal of Virtual Learning in Medical Sciences*, vol. 8, no. 1, Mar. 2017, doi: 10.5812/ijvlms.11784.

[24] S. Dhankhad, E. Mohammed, and B. Far, "Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: A Comparative Study," in *2018 IEEE International Conference on Information Reuse and Integration (IRI)*, IEEE, Jul. 2018, pp. 122–125. doi: 10.1109/IRI.2018.00025.

[25] A. Roy, J. Sun, R. Mahoney, L. Alonzi, S. Adams, and P. Beling, "Deep learning detecting fraud in credit card transactions," in *2018 Systems and Information Engineering Design Symposium (SIEDS)*,

IEEE, Apr. 2018, pp. 129–134. doi: 10.1109/SIEDS.2018.8374722.

[26] G. E. Melo-Acosta, F. Duitama-Munoz, and J. D. Arias-Londono, "Fraud detection in big data using supervised and semi-supervised learning techniques," in *2017 IEEE Colombian Conference on Communications and Computing (COLCOM)*, IEEE, Aug. 2017, pp. 1–6. doi: 10.1109/ColComCon.2017.8088206.

[27] M. IEEE Systems and Institute of Electrical and Electronics Engineers, "ICNSC 2018," Zhuhai, China, Mar. 2018.

[28] A. Bayhaqy, S. Sfenrianto, K. Nainggolan, and E. R. Kaburuan, "Sentiment Analysis about E-Commerce from Tweets Using Decision Tree, K-Nearest Neighbor, and Naïve Bayes," in *2018 International Conference on Orange Technologies (ICOT)*, IEEE, Oct. 2018, pp. 1–6. doi: 10.1109/ICOT.2018.8705796.