International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
Volume 12 Issue 12 – December 2024 - Pages 1-10
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395

IDENTIFYING THE IDEAL MACHINE LEARNING MODEL FOR CREDIT CARD FRAUD DETECTION

Sai Siddhish Chandra Sekaran[1]

Centennial High School, 4300 Centennial Ln., Ellicott City, MD 21043, USA.
saisiddhish@gmail.com

Abstract: Close to 8 billion dollars are stolen every year in the United States due to credit card fraud. This type of fraud is a huge issue in today's ever-increasing technological world, so ensuring consumer protection and providing better cybersecurity have become relevant, modern issues that need to be urgently addressed. Even with banks' detection systems, billions of dollars are still being stolen, which proves the potential and scale of the issue. This paper is going to employ the experimentation of several AI models to identify the model best at credit card fraud detection. The implication of this research is to better prevent and detect credit card fraud to allow for better cybersecurity and assurance of money. The results show that there is a lot of room for improvement when it comes to credit card fraud detection and prevention. Results: The Support Vector Machine achieved 99.85% accuracy and is the best at credit card fraud detection from the models tested.

Keywords: Credit Card Fraud; Artificial Intelligence; Machine Learning; Linear Regression; Logistic Regression; Random Forest; Support Vector Machine; Cybersecurity

1. Introduction

Many people have heard of many different types of fraud, from literal theft to online fraud with hacking and advanced criminals. Going a bit deeper into that spectrum, credit card fraud can be found, which is the fraudulent activities that are aided by or done through the use of a credit card [1]. Though credit card fraud seems like a simple issue, it can be very complex and can never be completely solved. However, measures can be taken to get closer to a near-perfect system for credit card fraud detection and prevention [2].

Today, banks use security measures that are inadequate compared to the skills of malicious hackers and cybercriminals attempting to hack their systems [3]. Banks typically "only use the most basic of statistical analyses to develop the fraud rules." This leads to a set of simple if-statements and conditions. If-statements are a fundamental piece of Computer Science that runs code only when a given condition is true. In this case, banks only flag a transaction as fraud when one of a few simple conditions

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
Volume 12 Issue 12 – December 2024 - Pages 1-10
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395

is met [4]. Some of these conditions may include the transaction amount being over an arbitrarily determined threshold or multiple large transactions happening successively or over a short period.

To determine the best method for detecting credit card fraud, this study focuses on analyzing several machines learning models, including logistic regression, decision trees, random forests, and neural networks. We compare various performance indicators such as accuracy, precision, recall, and computing efficiency in order to identify the model that provides the optimum trade-off between operational viability and fraud protection. The paper is organized as follows, the literature review section elaborates about previous works and approaches using various machine learning and deep learning techniques to effectively identify credit card fraud. The methods and materials section describes the dataset used, pre-processing techniques, models used for evaluation followed by the results section emphasizing the findings and its comparison, Future works section addressing our limitations and areas of improvement in future and finally the conclusion section summarizing all the key findings and the overall metrics.

2. Literature Review

There exist many solutions and models for detecting credit card fraud. A few of these solutions will be shown in this section. In [5] the authors used an ensemble hidden Markov model (EHMM) which was the main credit card fraud predictor. PCA and MRE (Mean, Relative Amplitude, and Entropy) were used to extract features. The results showed that PCA predicts credit card fraud but increases EHMM computational time. MRE boasts reduced complexity while still making accurate predictions. PCA-EHMM and MRE-EHMM were tested using recall/sensitivity, specificity, precision, and F1-score. Researchers in [6] utilized various supervised machine learning methodologies to identify credit card fraud, including Decision Trees (DT), Random Forests (RF), Artificial Neural Networks (ANN), Naïve Bayes (NB), and Logistic Regression (LR). The findings demonstrated that the Decision Tree technique attained the highest accuracy among the assessed models, proficiently identifying multiple types of credit card fraud, such as identity theft, skimming, counterfeit cards, mail interception fraud, and lost or stolen cards. The techniques employed by Rohith et al. [7] involve the integration of the application with financial institutions and law enforcement organizations, facilitating effective collaboration in cases of suspected fraud. The application utilizes sophisticated features including customized fraud detection strategies, machine learning algorithms, and real-time data processing to improve user awareness and readiness against fraudulent transactions. These technologies yield real-time fraud notifications that encompass transaction information, probable fraudulent behaviors, and recommended actions, thus enabling users to proactively oversee their accounts and react promptly to suspicious activities. The authors in [8] utilized a machine learning model for credit card fraud detection, employing three correlation types: Pearson, Spearman, and Kendall, for feature selection to improve the fraud detection process. The testing findings on datasets exhibited exceptional accuracy rates, attaining category accuracy of 99.95% and 99.58%, so exceeding other methodologies. Furthermore, it was determined that Kendall correlation was the most efficacious among the three correlation types for attribute selection across all sanctioned datasets. The authors in work [9] employed machine learning algorithms, notably convolutional neural network (CNN) models, to assess heat maps derived from credit card transaction data, surpassing conventional methods such as random forest and logistic regression. The findings suggest

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
Volume 12 Issue 12 – December 2024 - Pages 1-10
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395

that the utilization of CNN models is an innovative method for assessing credit card fraud detection, potentially enhancing the efficacy of recognizing fraudulent actions.

3. Methods and Materials

In this section, we will have a brief discussion of the methods and materials used in our research including the dataset used, preprocessing techniques, and models used. The flow of our methodology is depicted in Figure 1.

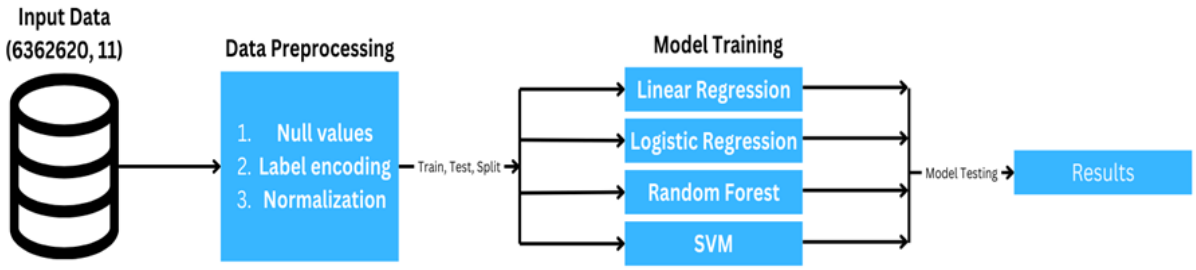

Figure 1. Methodology diagram

3.1. Data Collection

The dataset was used for predicting fraudulent transactions and developing an actionable plan to detect credit card fraud. The data was available in CSV format and had 6362620 rows and 11 columns [10]. There were three columns with integers, five columns with floats, and three columns with strings. The target column was “isFraud.” The dataset collected had a considerable amount of missing values, strings, and extreme values that had to be preprocessed.

3.2. Data Pre-Processing

To make this dataset usable and effective, it had to be preprocessed. We did three things, which were replacing null values using the mean method, converting strings to integer values using label encoding [11], and some values were in the extremes and were normalized using the minmax normalization technique [12].

3.3. Models Used

We trained and developed four different models for predicting on our credit card transactions dataset.

3.3.1. Linear Regression

In linear regression, a linear line is used to best model the relationship between the independent variables (X) and the dependent variable (Y) [13]. The equation for linear regression is shown below:

$$\hat{Y} = w_X + b_Y \qquad (1)$$

- $\hat{Y}$: Predicted output (dependent variable)

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
Volume 12 Issue 12 – December 2024 - Pages 1-10
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395

- X: Input feature (independent variable)
- w: Weight (slope of the line)
- b: Bias (intercept of the line)

The goal is to find the best values for the weights and the bias that minimizes the error between the predicted values of the line ($\hat{Y}$) and the actual values (Y) [14].

3.3.2. Logistic Regression

We used Logistic regression for binary classification. Logistic regression outputs a probability (0 or 1) that the given input belongs to a particular class [15]. The equation of a Logistic regression model is as shown below:

$$z = w_1X_1 + w_2X_2 + \cdots + w_nX_n + b \quad (2)$$

Logistic regression also uses the sigmoid function (σ) to take the output of the model equation above and put it in a range of 0 to 1 [16].

$$\sigma(z) = \frac{1}{1 + e^{-z}} \qquad (3)$$

- $w_1, w_2, \ldots, w_n$: Weights (coefficients) for each input feature
- $X_1, X_2, \ldots, X_n$: Input features (independent variables)
- b: Bias (intercept)
- $P(Y = 1 \mid X)$: The probability that the output is 1, given the input X

3.3.3. Random Forest

Random Forest is an ensemble algorithm that we used for our regression task of credit card fraud prediction. It's based on creating a group (forest) of decision trees—which makes predictions by splitting the dataset into conditions based on feature values and creating a tree-like structure—with each tree trained on a random subset of the data [17]. Random Forests use a technique called Bagging to reduce overfitting by randomly sampling and splitting the dataset into several subsets, with each decision tree being trained on its subset. On top of Bagging, the Random Subspace Method is also used [18], which selects a random subset of features for the subsets from Bagging to make the decision trees less alike and reduce overfitting. The final output is the most common output amongst the predictions from the trees. Since Random Forests take from multiple models, it ensures a flexible algorithm that can handle complex datasets with high accuracy and robustness [19]. The final output of a Random Forest model can be put into the equation below:

$$\hat{Y} = Mode(Y_1, Y_2, \ldots, Y_T) \qquad (4)$$

- $Y_1, Y_2, \ldots, Y_T$: Predictions from the T decision trees.
- $\hat{Y}$: Final predicted class (the mode of the predicted classes from all trees).

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
Volume 12 Issue 12 – December 2024 - Pages 1-10
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395

3.3.4. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised machine learning algorithm that we used for our classification task. The core of SVM is to find the optimal hyperplane (boundary) that gives the maximum margin between the data points of different classes [20]. In 2D, the hyperplane is a line, in 3D it is a plane, and in 4D and above is called a hyperplane. The data points closest to the hyperplane, which are the most difficult to classify, are called support vectors. These points are the only ones the SVM looks at and relies on to directly influence the position and orientation of the hyperplane [21].

$$w_1X_1 + w_2X_2 + \cdots + w_nX_+{}_n + b = 0 \quad (5)$$

- $w_1, w_2, \ldots, w_n$: Weights (coefficients) for each feature
- $X_1, X_2, \ldots, X_n$: Input features (independent variables)
- b: Bias term

4. Results and Discussion

In this study, we evaluated the performance of four different machine learning models—Linear Regression, Logistic Regression, Random Forest, and Support Vector Machine (SVM)—on the test dataset. The evaluation metrics used include accuracy, F1 score, precision, and recall, with confusion matrices provided to further analyze each model's performance. These metrics provide insights into how well each model performs, particularly on our classification problem. The table of results is provided below.

Table 1. Results comparison

Model Name	Testing Accuracy	F1 Score	Precision	Recall
Lin Reg	0.9892	0.9891	0.9978	0.9807
Log Reg	0.9978	0.9978	0.9981	0.9974
RF	0.9935	0.9935	0.9980	0.9891
SVM	0.9985	0.9985	0.9986	0.9983

The Linear Regression model achieved a testing accuracy of 0.9892, making it one of the lower-performing models in this study. Its F1 score of 0.9891 reflects a balance between precision and recall; however, the model still has a weakness in recall.

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
Volume 12 Issue 12 – December 2024 - Pages 1-10
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395

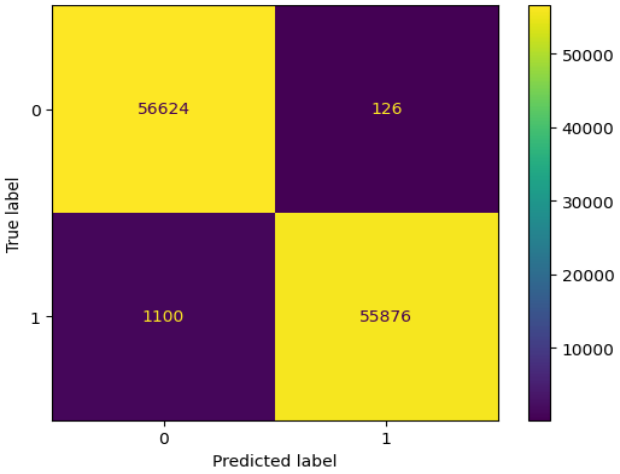

Figure 2. Confusion matrix of Linear Regression

Figure 2 illustrates 1,100 false negatives, the L:inear Regression model missed a lot of true positive cases, resulting in a recall of 0.9807. This means that out of all the actual positive cases, approximately 2% were not captured by the model, which could lead to significant consequences as every missed fraudulent transaction could be a loss amounting to a significant sum of money. Linear Regression, with 56,624 true positives and 126 false positives, demonstrates that the model performs exceptionally well in correctly identifying positive cases when it does predict them. The relatively low number of false positives (only 126) gives the model a high precision of 0.9978, meaning nearly all the positive predictions were correct. However, the false negative rate of 1,100 is a notable shortcoming, as it represents instances where the model failed to flag a positive case. Given these observations, Linear Regression might not be the best option in our scenario, where minimizing false negatives is crucial. Logistic Regression performed significantly better than Linear Regression, achieving a testing accuracy of 0.9978 and an F1 score of 0.9978. This indicates that the model has a well-balanced performance in terms of precision and recall. The precision of 0.9981 is particularly strong, as only 104 false positives were predicted, making this model highly reliable when predicting positive instances.

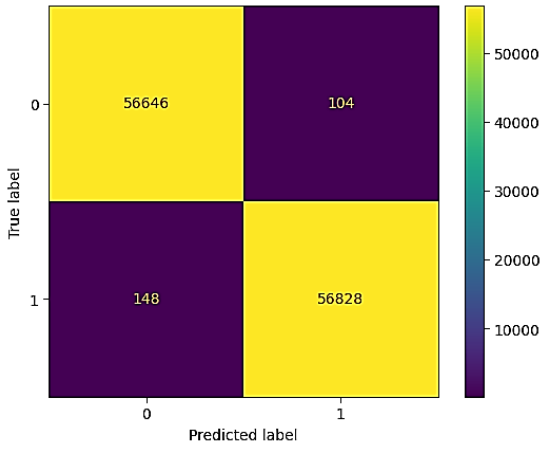

Figure 3. Confusion matrix of Logistic Regression

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
Volume 12 Issue 12 – December 2024 - Pages 1-10
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395

Figure 3 shows 56,646 true positives and 148 false negatives. The relatively low false negative rate results in a recall of 0.9974, signifying that the model correctly identified over 99% of all actual positive cases. This makes Logistic Regression a valid contender as it has minimized both false negatives and false positives. Compared to Linear Regression, the reduction in false negatives from 1,100 to 148 shows a dramatic improvement, demonstrating that Logistic Regression is more sensitive to true positive cases, making it more effective in our fraudulent transaction detection where missing positive cases could have critical consequences. Furthermore, the confusion matrix highlights how well Logistic Regression performs at identifying true negatives, with 56,828 true negatives, meaning it correctly identifies most negative cases while minimizing false positives. The Random Forest model achieved a testing accuracy of 0.9935 and an F1 score of 0.9935, placing it between Logistic Regression and Linear Regression in terms of overall performance.

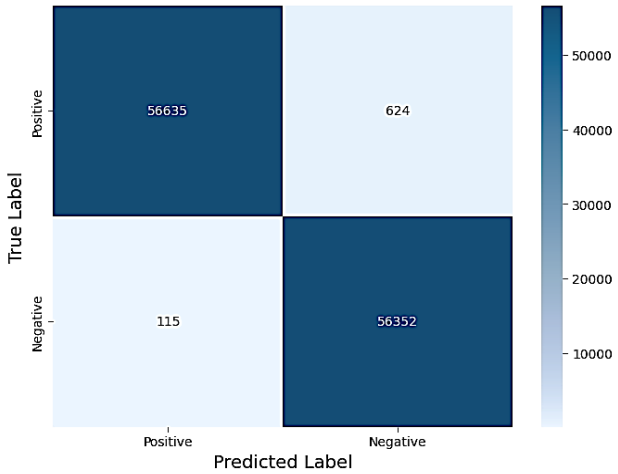

Figure 4. Confusion matrix of Random Forest

Considering figure 4, one of its key strengths is its precision of 0.9980, as the model only produced 624 false positives, a higher rate than SVM and Logistic Regression but is much better than having the equivalent number of false negatives. This high precision means that when Random Forest predicts a positive case, it is almost always correct, which is beneficial in other applications where false positives must be kept to a minimum but not as important as minimizing false negatives for us. The recall of 0.9891 for Random Forest, however, reflects a minor limitation. With 115 false negatives, the model failed to identify some positive cases. While this number is much smaller than Linear Regression's 1,100 false negatives, it is still higher than that of Logistic Regression and SVM. This indicates that while Random Forest is strong in precision, it may require further tuning or alternative configurations to improve its sensitivity to positive cases, making it more appropriate for applications where precision is valued over recall. In terms of true negative identification, Random Forest correctly predicted 56,352 true negatives. The relatively higher number of 624 false positives, compared to the other models, suggests that Random Forest might flag more negatives as positives, which is much better than the opposite but still not ideal as minimizing both false positives and negatives is the best situation.

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
Volume 12 Issue 12 – December 2024 - Pages 1-10
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395

The Support Vector Machine (SVM) outperformed all other models, with a testing accuracy of 0.9985, and an F1 score of 0.9985. This indicates that the model performs exceptionally well in balancing precision and recall, making it highly reliable for our classification task.

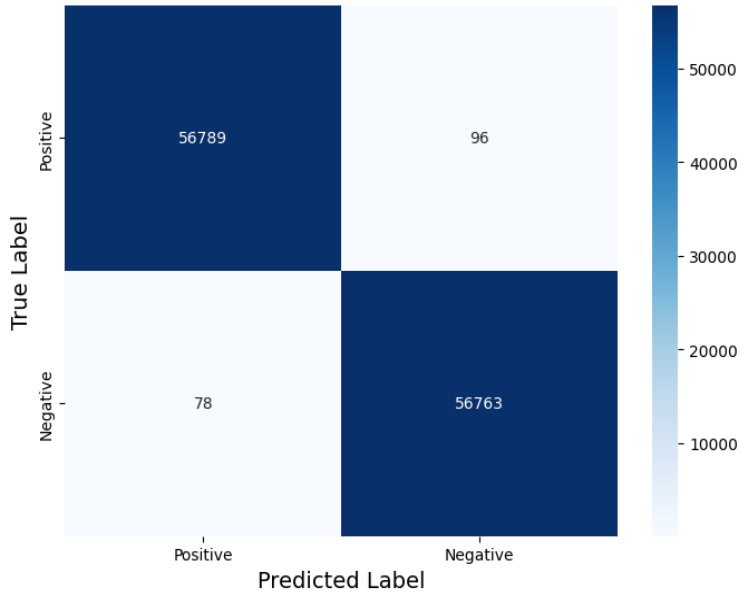

Figure 5. Confusion matrix of Support Vector Machine

The precision of 0.9986 is the highest among the models and figure 5 highlights it only had 96 false positives, showing that SVM is highly accurate in identifying true positives while minimizing false positives. SVM also achieved a recall of 0.9983, with only 78 false negatives, the lowest across all models. This demonstrates the model's ability to capture almost all positive cases, ensuring that very few true positives go undetected. This performance makes SVM ideal for our application where both false positive and false negatives should be ideally minimized. The confusion matrix for SVM reveals 56,789 true positives and 56,763 true negatives, indicating that the model is highly effective at correctly classifying both positive and negative cases. With its minimal number of misclassifications (both false positives and false negatives), SVM is clearly the best-performing model for this dataset.

In conclusion, SVM was the top-performing model with the highest testing accuracy (0.9985), F1 score (0.9985), precision (0.9986), and recall (0.9983), making it the best choice for our applications where, ideally, both false positives and false negatives would be minimized. Random Forest exhibited strong precision (0.9980) but had a noticeably higher number of false positives, making it less ideal for our application. Logistic Regression also performed well, showing a good balance of precision and recall, making it a strong alternative when model simplicity and efficiency are prioritized. Linear Regression, while exhibiting high precision (0.9978), struggled with recall (0.9807), making it less ideal for cases where capturing all positive instances is crucial. Overall, SVM is the most reliable choice for this dataset, but Logistic Regression offers a practical, interpretable alternative with slightly lower performance.

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
Volume 12 Issue 12 – December 2024 - Pages 1-10
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395

5. Future Works

The data from the experiment applies to the thesis/research question as it attempts to show the correlation between model complexity and effectiveness in detecting fraudulent transactions. What if we used more advanced AI models, such as Deep neural networks, could they better detect and predict fraudulent credit card transactions? It was hypothesized that this would be the case as one would reasonably think that more complex AI models are more accurate. In the experiment, ten AI models of varying complexities were developed to truly test this. There exist several limitations to the experiment. Mainly, computational and financial limitations. Since the more complex AI models require more computational power to train to their fullest potential, providing as much power as possible is highly advisable. However, due to being a student, the researcher was unable to provide their laptop's full computational power to train a recurrent neural network, for example, to near 100% accuracy. Additionally, Google Colab, the web-based software used for the second trial in the experiment, has its free-tier limits in place, so it made multiple attempts to fully train some AI models, while others had to be rewritten to be less computationally expensive. Moving on, from the experiment (data table available above), it can be noted that the more complex models did indeed achieve higher accuracy than the less complex models. Looking at Linear Regression (one of the most basic models made) in comparison to a Deep Neural Network (the second most complex AI model that was trained), there is about a 3% gap in accuracy. Though this seems negligible, it will scale to truly massive amounts of money saved if implemented in a real banking system.

6. Conclusion

Summarizing this entire article, the following can be said. Firstly, credit card fraud is a tremendous issue today with statistics such as nearly 12 billion dollars attempted to be stolen and 9.5 million people affected every year (security.org, 2023). Secondly, credit card fraud can happen in many ways, such as using details online or stealing a physical credit card. Lastly, the security measures of banks and credit card companies are inadequate, especially in comparison to the cybercriminals in today's ever-increasing digital world. To better these security systems, banks should begin the implementation and usage of more complex Deep Neural Networks (DNNs). The reason behind using a DNN is that they have been highlighted as the most accurate and precise out of the ten AI models used in my experiment as explained previously (this can be seen in Section 5, Figure 1). The key idea behind this research study is to better ensure the financial assets of both individuals and businesses with the use of Artificial Intelligence and Machine Learning.

References

[1]. Mienye, I. D., & Jere, N. (2024). Deep Learning for Credit Card Fraud Detection: A Review of Algorithms, Challenges, and Solutions. IEEE Access.
[2]. Chatterjee, P., Das, D., & Rawat, D. B. (2024). Digital twin for credit card fraud detection: Opportunities, challenges, and fraud detection advancements. Future Generation Computer Systems.
[3]. Ali, M. M., & Mohd Zaharon, N. F. (2024). Phishing—A cyber fraud: The types, implications and governance. International Journal of Educational Reform, 33(1), 101-121.
[4]. Ghosh, S., & Reilly, D. L. (1994, January). Credit card fraud detection with a neural network. In System Sciences, 1994. Proceedings of the Twenty-Seventh Hawaii International Conference on (Vol. 3, pp. 621-630). IEEE.

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
Volume 12 Issue 12 – December 2024 - Pages 1-10
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395

[5]. Olayinka, O., Ogundile., Oluwaseyi, P., Babalola., Afolakemi, Ogunbanwo., Olayinka, O., Ogundile., Vipin, Balyan. (2024). Credit Card Fraud: Analysis of Feature Extraction Techniques for Ensemble Hidden Markov Model Prediction Approach. doi: 10.20944/preprints202407.2157.v1

[6]. Tayo, P., Ogundunmade., Adedayo, A., Adepoju. (2024). Modelling Credit Card Fraud Data using Machine Learning Algorithms. 1(2):43-49. doi: 10.62527/comien.1.2.10

[7]. Rohit, Pandey., Rakesh, Panigrahi., Hritik, Mishra., Adarsh, Mishra., Vaishali, Bhusari. (2024). CredChecker:-Credit Card Fraud Detection WebApp. International Journal of Advanced Research in Science, Communication and Technology, doi: 10.48175/ijarsct-17839

[8]. Ahmad, Ahmad., Salah, N., Mjeat., Daniah, Abul, Qahar, Shakir., Mohammed, Awad, Alfwair. (2023). Credit Card Fraud Detection Model Based on Correlation Feature Selection. Journal of cybersecurity and information management, 14(2):334-342. doi: 10.54216/jcim.140224

[9]. Anuj, Raturi., Mayur, Pal., H., K., Narang., Satvik, Vats., Vikrant, Sharma., Satya, Prakash, Yadav. (2024). A Comparative Analysis of Machine Learning Algorithms for Credit Card Fraud Detection. doi: 10.1109/icecccc61767.2024.10593936

[10]. Credit Card Fraud Detection Dataset 2023. (2023, September 18). Kaggle. https://www.kaggle.com/datasets/nelgiriyewithana/credit-card-fraud-detection-dataset-2023/data

[11]. Xia, T., Xiao, J., Huang, Y., Hu, C., Song, S., Huang, X., & Wang, J. (2024). Time series data encoding in Apache IoTDB: comparative analysis and recommendation. The VLDB Journal, 33(3), 727-752.

[12]. Demircioğlu, A. (2024). The effect of normalization methods in radiomics. Insights into Imaging, 15(1), 2.

[13]. Baria, J. B., Baria, V. D., Bhimla, S. Y., Prajapati, R., Rathva, M., & Patel, S. (2024). Deep Learning based Improved Strategy for Credit Card Fraud Detection using Linear Regression. Journal of Electrical Systems, 20(10s), 1295-1301.

[14]. Saeed, V. A., & Abdulazeez, A. M. (2024). Credit Card Fraud Detection using KNN, Random Forest and Logistic Regression Algorithms: A Comparative Analysis. The Indonesian Journal of Computer Science, 13(1).

[15]. Chen, B. From Logistic Regression to Deep Learning: Machine Learning Advances for Credit Card Churn.

[16]. Sutedja, I., Lim, J., Setiawan, E., & Adiputra, F. R. (2024). Credit Card Approval Prediction: A Systematic. Journal Of Theoretical And Applied Information Technology, 102(3).

[17]. Kumar, A., Poojitha, M. V., Anuhya, T., Srinivas, K., & Bhargavi, M. (2024, July). Credit card fraud detection. In 2024 8th International Conference on Inventive Systems and Control (ICISC) (pp. 79-82). IEEE.

[18]. Raj, P. S., Murugan, S., & Kanipriya, M. (2024, July). Credit card fraud detection using machine learning. In AIP Conference Proceedings (Vol. 3075, No. 1). AIP Publishing.

[19]. Schwartz, D. (2024). The rise of a nudge: Field experiment and machine learning on minimum and full credit card payments. Working Paper.

[20]. Alshawi, B. (2024). Comparison of SVM kernels in Credit Card Fraud Detection using GANs. International Journal of Advanced Computer Science & Applications, 15(1).

[21]. Seera, M., Lim, C. P., Kumar, A., Dhamotharan, L., & Tan, K. H. (2024). An intelligent payment card fraud detection system. Annals of operations research, 334(1), 445-467.