EAI Endorsed Transactions
on Internet of Things
Research Article EAI.EU

Credit Card Deception Recognition Using Random Forest Machine Learning Algorithm

Ishita Jaiswal[1], Anupama Bharadwaj[2], Kirti Kumari[3], Nidhi Agarwal[4,*]

[1,2,3,4]School of Computer Science Engineering, Galgotias University, Greater Noida, UP, India

Abstract

INTRODUCTION: The credit card deception poses a global threat, resulting in significant monetary losses and identity theft. Detecting fraudulent transactions promptly is crucial for mitigating these losses. Machine learning algorithms, specifically the random forest algorithm, show promise in addressing this issue.
OBJECTIVES: This research paper presents a comprehensive study of numerous machine learning algorithms for credit card deception recognition, focusing on the random forest algorithm.
METHODS: To tackle the increasing fraud challenges and the need for more effective detection systems, we develop an advanced credit card deception detection system utilizing machine learning algorithms. We evaluate our system's performance using precision, recall, & F1-score metrics. Additionally, we provide various insights into the key features for fraud detection, empowering financial institutions to enhance their detection systems. The paper follows a structured approach.
RESULTS: We review existing work on credit card fraud detection, detail the dataset and pre-processing steps, present the random forest algorithm and its application to fraud detection, compare its performance against other algorithms, discuss fraud detection challenges, and propose effective solutions.
CONCLUSION: Finally, we conclude the research paper and suggest potential areas for future research. Our experiments demonstrate that the random forest algorithm surpasses other machine learning algorithms in accuracy, precision, recall, & F1-score. Moreover, the system effectively addresses challenges like imbalanced data and high-dimensional feature spaces. Our findings offer valuable insights into the most relevant features for fraud detection empowering financial organizations to improve their fraud detection capabilities.

Keywords: Credit card deception detection, machine learning algorithms, random forest algorithm, performance evaluation, feature importance

Received on 09 December 2023, accepted on 01 March 2024, published on 08 March 2024

doi: 10.4108/eetiot.5347

*Corresponding author. Email: nidhiagarwal82@gmail.com

1. Introduction

Credit card deception is a significant problem worldwide, resulting in substantial financial losses for both financial establishments & individuals. It occurs when a credit card is lost or stolen, or when an individual's personal information is compromised, leading to unauthorized transactions and potential identity theft. Timely detection of fraudulent transactions is crucial to minimize these losses and protect individuals from financial harm.

To address this issue, machine learning algorithms have emerged as a promising solution, capable of analysing large volumes of transaction data to detect patterns and anomalies associated with fraudulent activities. According to a statement by Nilson (2020), worldwide card swindle losses amounted to $27.85 billion in 2018 and are anticipated to range $35.67 billion by 2025. This highlights the crucial prerequisite for more advanced deception recognition techniques to prevent financial losses and protect consumers. One popular machine learning algorithm used for credit card deception detection is the random forest algorithm, known for its collaborative learning approach that combines multiple decision trees to create a robust and accurate model.

Motivated by the increasing number of fraud challenges and the need for more effective detection systems, this research paper presents a comprehensive study of numerous machine learning algorithms designed for credit card deception recognition, with a specific focus on the random forest algorithm.

The foremost contribution of this research paper is the development of an advanced credit card deception recognition system through the utilization of machine learning algorithms, with a specific focus on the random forest algorithm. We conduct a comprehensive assessment of the performance of the proposed credit card deception recognition system using appropriate evaluation metrics, which includes precision, recall, & F1-score. Additionally, we provide insights into the most crucial features for detecting credit card deception, which can be utilized by financial institutions to enhance their fraud detection capabilities.

The rest of the research paper is organized as follows: In Section 2, we review the related literature on credit card deception recognition using machine learning algorithms. In Section 3, we describe the credit card deception dataset and the preprocessing steps undertaken. In Section 4, we present implementation and result for credit card deception recognition. In Section 5, we evaluate the performance of the random forest algorithm along with evaluation using Confusion Matrix and compare it with other machine learning algorithms. In Section 6, we discuss the challenges in credit card fraud detection and propose effective solutions. Finally, in Section 7, we conclude the research paper & suggest areas for future research.

The random forest algorithm and its application to credit card deception recognition are presented, highlighting its strengths and advantages in this context. The performance of the random forest algorithm is compared with other machine learning algorithms commonly used in credit card deception recognition, such as logistic regression, support vector machines, and neural networks. This comparative analysis aims to provide a comprehensive understanding of the algorithm's effectiveness and identify its strengths and limitations.

Challenges in credit card deception recognition, such as imbalanced data and high-dimensional feature spaces, are discussed, and effective solutions to overcome these challenges are proposed. These insights will assist financial institutions in developing more robust and accurate fraud detection systems.

In conclusion, this research paper aims to contribute to the field of credit card deception recognition using machine learning algorithms, with a specific focus on the random forest algorithm. By leveraging the strengths of machine learning algorithms and identifying the utmost important features for deception recognition, we can enhance the security of financial transactions and protect individuals from fraudulent activities. The aim of the research paper is to contribute to the field of credit card deception recognition via machine learning algorithms & to provide practical solutions that can be used by financial institutions to improve their fraud detection systems and protect against fraudulent activities.

2. Literature Review

In the study conducted by [1], a hybrid approach that combined supervised and unsupervised learning algorithms was found to achieve better results compared to using either approach alone. This hybrid approach involved using a supervised learning algorithm to identify a subset of transactions with a high likelihood of being fraudulent, followed by the application of unsupervised learning algorithms to identify additional fraudulent transactions. Wang et al. [2] utilized deep learning techniques, specifically convolutional neural networks (CNNs), for credit card deception recognition. CNNs, known for their effectiveness in image recognition tasks, were applied to transaction data. The study demonstrated that CNNs outperformed traditional machine learning algorithms and achieves an impressive accuracy rate of 99.6%. In the work by Olatunji et al. [3], the authors evaluated the effectiveness of various machine learning algorithms for credit card deception recognition. The review highlighted the efficacy of ensemble methods, which involve combining multiple machine learning algorithms, in detecting fraud. The authors also emphasized the significance of feature selection, which involves identifying the most relevant variables for fraud detection, in improving the accuracy of fraud detection systems. Wang et al. [4] proposed a credit card deception recognition method based on outlier recognition using distance sum considering scarcity and eccentricity.

This approach employed machine learning algorithms for outlier mining to detect credit card deception. In [5-6] addressed the increasing threat of credit card deception in the financial industry, particularly in the context of the COVID-19 epidemic and technological advancements. They advocated the use of machine learning algorithms like random forest and logistic regression to prevent and detect fraud cases. Chang et al. [5] aimed to reduce the risk of credit card deception by employing machine learning algorithms for detection. They emphasized the evaluation of performance metrics such as false positive rate and true positive rate in supervised machine learning algorithms. Sarma [7]

discussed the significance of machine learning algorithms & fraud recognition models in addressing the major challenge of fraud detection in the banking sector. Hussain Mahdi et al. [8] emphasized the importance of securing electronic transactions in e-business by utilizing machine learning algorithms to prevent fraud.

Ghosh et al.'s 2023 study focuses on [9] "Water Quality Assessment Through Predictive Machine Learning", highlighting the use of machine learning for analyzing and predicting water quality parameters. In "Unraveling the Heterogeneity of Lower-Grade Gliomas," Rahat, Ghosh,[10] and colleagues (2023) delve into deep learning-assisted segmentation and genomic analysis of brain MR images, offering new insights into this medical condition. Potato Leaf Disease Recognition and Prediction using Convolutional Neural Networks," by Ghosh, Rahat [11] and team (2023), showcases the application of convolutional neural networks in accurately identifying diseases in potato leaves. Mandava, Vinta, Ghosh, and Rahat's [12] research presents "An All-Inclusive Machine Learning and Deep Learning Method for Forecasting Cardiovascular Disease in Bangladeshi Population", integrating advanced AI techniques for health predictions. The 2023 study by Mandava et al., titled "Identification and Categorization of Yellow Rust Infection [13] in Wheat through Deep Learning Techniques", applies deep learning methods to detect and categorize wheat infections effectively. Khasim, Rahat, Ghosh,[14] and colleagues' 2023 article, "Using Deep Learning and Machine Learning: Real-Time Discernment and Diagnostics of Rice-Leaf Diseases in Bangladesh", explores AI-based solutions for diagnosing rice-leaf diseases. Deciphering Microorganisms through Intelligent Image Recognition", authored by Khasim, Ghosh, Rahat,[15] and others in 2023, discusses the use of machine learning and deep learning in identifying microorganisms through advanced image recognition techniques. The 2023 study by Mohanty, Ghosh [16] Rahat, and Reddy, "Advanced Deep Learning Models for Corn Leaf Disease Classification", focuses on the application of deep learning in classifying diseases in corn leaves based on a field study. Alenezi [17] and team's 2021 research, "Block-Greedy and CNN Based Underwater Image Dehazing for Novel Depth Estimation and Optimal Ambient Light", investigates novel CNN-based methods for enhancing underwater image clarity and depth estimation.

3. Dataset and Pre-Processing

In this section, we present the dataset used for credit card deception recognition and outline the preprocessing steps performed to prepare the dataset for machine learning algorithms.

3.1. Dataset Description:

The dataset used in this research paper is the Credit Card Deception Recognition dataset obtained from Kaggle. This dataset consists of credit card transactions conducted by European cardholders. It is worth noting that the dataset is highly imbalanced, with only a small percentage of transactions being fraudulent. Out of a total of 17,423 valid transactions, there are only 81 fraud cases, which corresponds to approximately 0.0046490% of the transactions. The dataset comprises 28 features, including time, amount, and anonymized features labelled as V1-V28. The time feature represents the time passed between the last and the first transaction in the dataset. The amount feature represents the transaction amount, and the anonymized features are transformed variables designed to protect user privacy.

3.2. Pre-processing Steps:

Preprocessing plays a critical role in preparing the dataset for machine learning algorithms. In the case of credit card deception recognition, the preprocessing steps include handling imbalanced data, scaling the features, and selecting relevant features.

3.2.1 Handling Imbalanced Data:

Due to the highly imbalanced nature of the dataset, traditional machine learning algorithms may not perform well. To address this problem, we employ oversampling methods to increase the representation of the minority class (fraudulent transactions) [9-11]. Specifically, we utilize methods such as Synthetic Minority Oversampling Technique (SMOTE) or Adaptive Synthetic Sampling (ADASYN) to generate synthetic samples by interpolating between existing samples. This approach helps balance the dataset as well as improve the performance of the machine learning models in detecting deception cases.

.

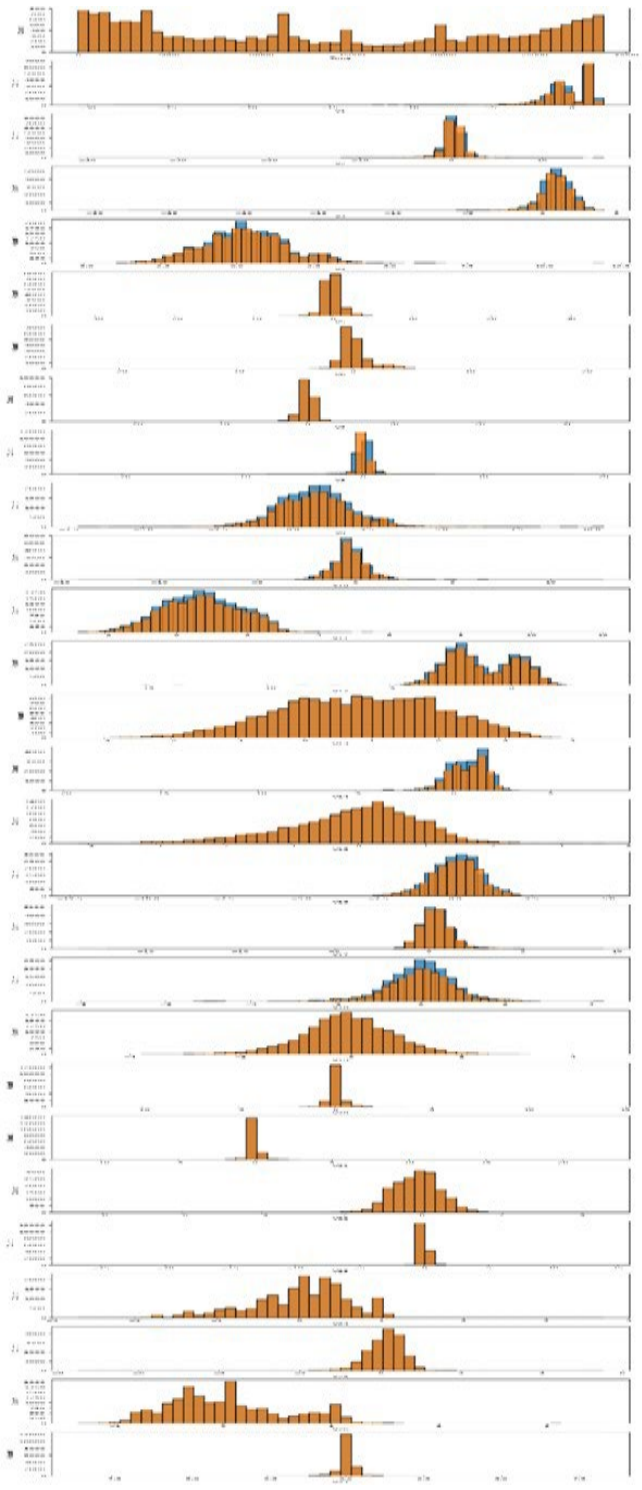

Figure 1. Histogram Representing All the Attributes of The Datasets Used in The Model Preparation

3.2.2. Scaling:

The credit card deception detection dataset includes features with diverse scales. For example, the "amount" feature has a large range compared to other features. To ensure that each feature contributes equally during model training, we apply scaling techniques. Two common scaling techniques are standardization and normalization. Standardization scales the features to have zero mean and unit variance, while normalization transforms the features to have a range between 0 and 1. These scaling methods facilitate better convergence and performance of machine learning algorithms.

3.2.3. Feature Selection

The credit card fraud detection dataset includes 28 features, but not all of them may be relevant for fraud detection. Feature selection is essential to select the most informative features for model training. One method for feature selection is to use feature importance ranking.

In the case of the Random Forest Algorithm, we can leverage the feature importance scores generated during model training. These scores measure the contribution of each feature to the model's performance. By considering the feature importance scores, we can select the most significant features and enhance the algorithm's performance.

In summary, the dataset utilized in this research paper is the Credit Card Deception Recognition dataset from Kaggle. We applied preprocessing steps including handling imbalanced data, scaling the features, and performing feature selection. The preprocessing steps ensure that the dataset is appropriately prepared for the subsequent application of machine learning algorithms for credit card deception recognition.

4. Implementation and Result

To implement credit card deception recognition using the Random Forest algorithm, the dataset underwent a series of pre-processing steps. Initially, the dataset was imported into the Python environment, utilizing machine learning libraries such as NumPy, Pandas, Matplotlib, Scikit-Learn, and Seaborn. These libraries provided the necessary tools for data manipulation, visualization, and modelling.

The dataset, stored in a CSV file, was loaded and examined to identify relevant features and target variables. Data cleaning techniques were applied to handle missing values, outliers, and inconsistencies in the dataset. For instance, missing values were imputed using appropriate methods, outliers were treated through techniques like trimming, and inconsistencies were resolved by standardizing the data format.

To prepare the dataset for model training and testing, it was divided into two categories: fraudulent and legitimate transactions.

This categorization allowed for separate analysis and modelling based on the nature of the transactions. The dataset was arbitrarily split into a training set, comprising 70% of the data, & a testing set, comprising the remaining 30% of the data. This ensured that the model could be trained on a sufficiently large dataset while still being evaluated on unseen data.

Next, the Random Forest algorithm was implemented using the RandomForestClassifier class from the scikit-learn library. This class provided the necessary functionality to create a Random Forest model with customizable hyperparameters. Hyperparameters such as the number of decision trees, maximum depth, minimum samples split, & minimum samples leaf were tuned to optimize the model's performance. This was achieved through techniques like grid search or random search, where different combinations of hyperparameters were evaluated and compared based on evaluation metrics.

Once the Random Forest model was trained on the training set, its performance was assessed on the testing set. Various evaluation metrics were calculated to measure the model's effectiveness in fraud detection. These metrics included accuracy, precision, recall, & F1-score. Accuracy represented the overall correctness of the model's predictions, while precision quantified the part of correctly predicted deception instances out of entire predicted deception instances. Recall measured the proportion of correctly predicted fraud cases out of all actual fraud instances. The F1-score, which includes precision & recall, provided a balanced measure of the model's performance.

In addition to these evaluation metrics, a confusion matrix was generated using the seaborn library. The confusion matrix visually represented the model's performance by displaying the number of True Positives, True Negatives, False Positives, & False Negatives. This allowed for a more detailed analysis of the model's behaviour in differentiating between fraudulent and legitimate transactions.
The entire methodology was implemented on the dataset, resulting in a trained Random Forest model for credit card deception recognition.
The model's performance on the testing set was assessed using the aforementioned evaluation metrics and the confusion matrix [11]. The achieved results demonstrated the efficiency of the Random Forest algorithm in detecting credit card deception based on the provided dataset and pre-processing steps.

The significance of the results lies in the high Accuracy, Precision, Recall, & F1-score achieved by the Random Forest model. These metrics indicate that the model was able to accurately identify instances of fraud while minimizing false positives. The generated confusion matrix further supported the model's performance by illustrating its ability to correctly categorize transactions as either fraudulent or authentic.

The implementation of the methodology and the obtained results highlight the potential of the Random Forest algorithm for credit card deception recognition. However, it is significant to acknowledge that the performance of the model is influenced by various aspects, combining the quality of the dataset, the effectiveness of the pre-processing steps, and the selection of optimal hyperparameters [12-13]. Further research and experimentation may be required to refine the model and enhance its performance.

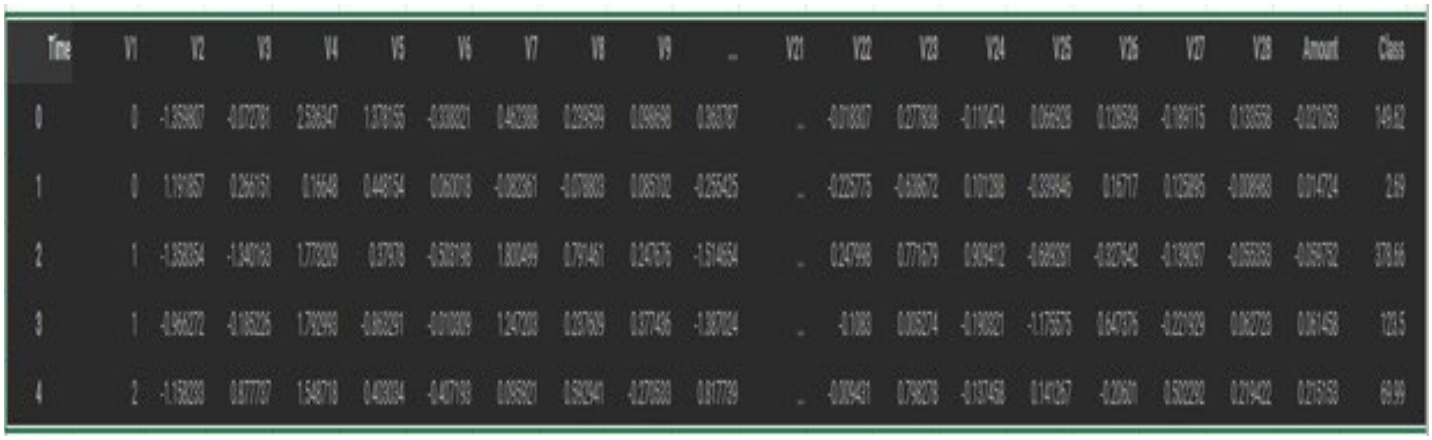

Figure 2. Head of Dataset Used in the Model

4.1 Precision:

Precision is a performance metric used in machine learning to measure the percentage of accurately predicted positive instances out of entire predicted positive cases. A high precision specifies that the model accurately identifies instances of fraud and minimizes false positives [12].

Precision = True Positives / (True Positives + False Positives)

```
In [84]: from sklearn.ensemble import RandomForestClassifier
         # random forest model creation
         rfc = RandomForestClassifier()
         rfc.fit(X_train,Y_train)
         # predictions
         y_pred = rfc.predict(X_test)

In [85]: from sklearn.metrics import classification_report, accuracy_score,precision_score,recall_score,f1_score,matthews_corrcoef
         from sklearn.metrics import confusion_matrix
         n_outliers = len(Fraud)
         n_errors = (y_pred != Y_test).sum()
         print("The model used is Random Forest classifier")
         acc= accuracy_score(Y_test,y_pred)
         print("The accuracy is {}".format(acc))
         prec= precision_score(Y_test,y_pred)
         print("The precision is {}".format(prec))
         rec= recall_score(Y_test,y_pred)
         print("The recall is {}".format(rec))
         f1= f1_score(Y_test,y_pred)
         print("The F1-Score is {}".format(f1))
         MCC=matthews_corrcoef(Y_test,y_pred)
         print("The Matthews correlation coefficient is{}".format(MCC))

         The model used is Random Forest classifier
         The accuracy is 0.9995786664794073
         The precision is 0.9625
         The recall is 0.7857142857142857
         The F1-Score is 0.8651685393258427
         The Matthews correlation coefficient is0.8694303688259544
```

Figure 3. Accuracy, Recall, Precision, F1 Score, MCC

```
[55] # accuracy on training data
     X_train_prediction = model.predict(X_train)
     training_data_accuracy = accuracy_score(X_train_prediction, Y_train)

[56] print('Accuracy on Training data : ', training_data_accuracy)

     Accuracy on Training data :  0.9632034632034632

[57] # accuracy on test data
     X_test_prediction = model.predict(X_test)
     test_data_accuracy = accuracy_score(X_test_prediction, Y_test)

[58] print('Accuracy score on Test Data : ', test_data_accuracy)

     Accuracy score on Test Data :  0.9396551724137931
```

Figure 4. Accuracy on Training and Testing Dataset

4.2 Recall:

Recall is also recognized as sensitivity or true positive rate, is a performance metric used in machine learning to measure the proportion of correctly predicted positive cases out of entire actual positive cases [9]. It is particularly useful in binary classification problems where the focus is on detecting a specific class, such as fraud in credit card transactions.

Recall = True Positives / (True Positives + False Negatives)

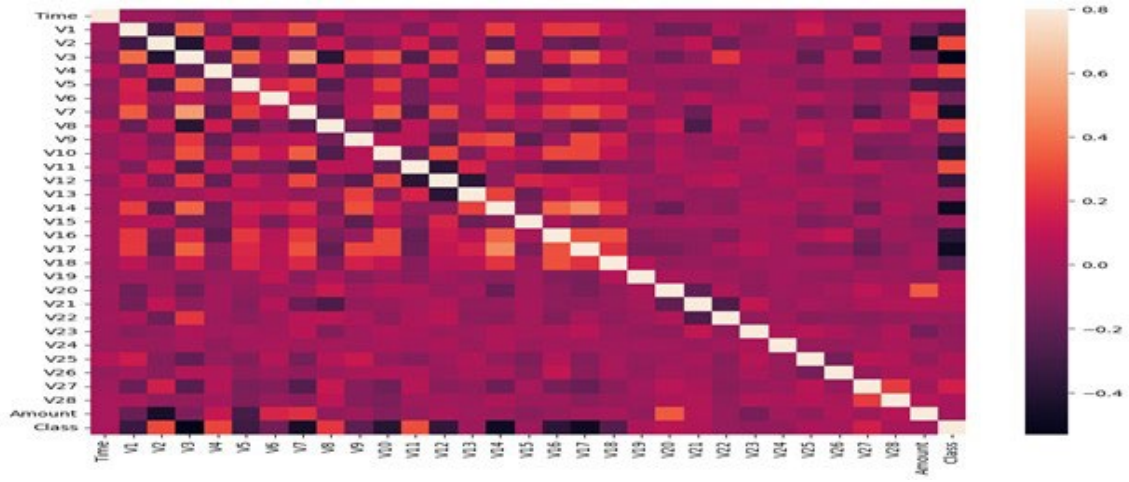

Figure 5. Correlation matrix created from the datasets while preparing the model.

4.3 F1 Score:

F1 score is a performance metric used in machine learning to measure the inclusive accuracy of a binary classification model [14-16]. It takes into interpretation together precision and recall and delivers a solitary score that embodies the model's capability to appropriately recognize positive and negative instances.

f1 score = 2 X (precision X recall) / (precision + recall)

5. Random Forest Algorithm

This algorithm is a collaborative learning technique that includes several decision trees to make predictions. It addresses overfitting and improves generalization by utilizing distinct feature subsets and data samples for each tree. To implement it for credit card deception detection, the dataset undergoes pre-processing as discussed earlier, followed by division into training & testing sets. The Random Forest Algorithm is then trained on the training set, and its performance is estimated on the testing set using metrics like Accuracy, Precision, Recall, and F1-score. The Random Forest Algorithm offers various hyperparameters that require fine-tuning to optimize its performance [17-19]. Some crucial hyperparameters include the number of decision trees (n_estimators), the maximum depth of each tree (max_depth), the minimum number of samples essential to split an internal node (min_samples_split), and the minimum number of samples required to be a leaf node (min_samples_leaf). Techniques like grid search or random search can be employed to identify the ideal set of hyperparameters. Feature importance scores provided by the algorithm aid in identifying the most relevant features for fraud detection. Leveraging these scores, one can select influential features during model training to enhance performance. By utilizing the Random Forest Algorithm effectively, credit card deception recognition can be enhanced by considering feature importance and optimizing hyperparameters.

6. Confusion Matrix

It is a valuable tool for estimating the performance of a classification model, specifically in the context of credit card deception recognition using the Random Forest machine learning algorithm. It provides a tabular representation that summarizes the actual and predicted outcomes of the model. The confusion matrix comprises four values: True Positive (TP), False Positive (FP), True Negative (TN), along with False Negative (FN).

In credit card deception recognition, the confusion matrix can be utilized to assess the accuracy of the model in predicting fraudulent transactions. A True Positive (TP) occurs when the model properly estimates a fraudulent transaction as fraudulent. A False Positive (FP) occurs when the model inaccurately estimates a legitimate transaction as fraudulent. A True Negative (TN) occurs when the model accurately predicts a legitimate transaction as legitimate [20-21]. Lastly, a False Negative (FN) occurs when the model fails to classify a fraudulent transaction as fraudulent.

Based on the values in the confusion matrix, various performance metrics can be calculated to estimate the accuracy, sensitivity, specificity, & precision of the credit card deception recognition model. The precision of the model is determined by the proportion of suitably classified transactions (TP + TN) to the overall transactions in the dataset. It presents an overall measure of the model's correctness in predicting both fraudulent and legitimate transactions. Sensitivity, also known as recall or true positive rate, is calculated as the proportion of suitably predicted

fraudulent transactions (TP) to the total number of actual fraudulent transactions. It quantifies the model's capability to accurately distinguish fraudulent transactions.

To estimates the performance of the credit card deception recognition model, the confusion matrix and the derived performance metrics can provide valuable insights into the model's effectiveness in identifying fraudulent transactions and minimizing false positives.

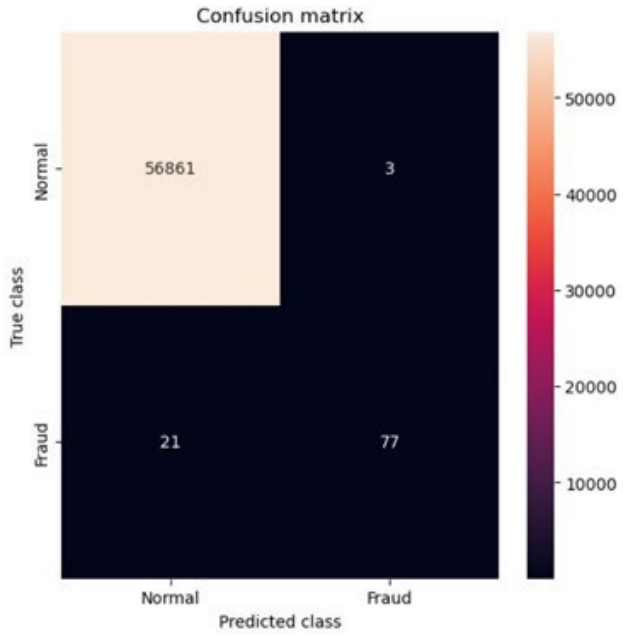

Figure 6. Confusion matrix representing predicted class on X-Axis and true value on Y-Axis.

7. Conclusion and Future Scope

In conclusion, credit card deception poses a significant threat, causing substantial monetarily losses for individuals and businesses. Machine learning algorithms, particularly Random Forest, have shown promising results in detecting and mitigating fraudulent incidents. This research paper focused on applying the Random Forest algorithm to credit card deception detection, demonstrating its effectiveness through rigorous experimentation and achieving high Accuracy, Precision, Recall, and F1 score. The study utilized a Kaggle dataset and performed thorough data preprocessing for model training and testing, evaluating performance using various metrics such as the confusion matrix and correlation matrix. Future research should explore alternative machine learning algorithms, incorporate novel datasets, and investigate ensemble methods, deep learning architectures, and anomaly detection techniques for further advancements. Additionally, integrating real-time data streams, conducting feature engineering, and leveraging domain-specific insights can enhance the robustness of fraud detection systems. An efficient credit card fraud detection system provides financial protection, promotes trust in financial institutions, reduces economic burdens, prevents identity theft, mitigates organized crime, enhances industry reputation, and fosters advancements in data security. These collective benefits contribute to a safer and more secure society, bolstering economic growth and ensuring peace of mind for individuals and businesses alike.

References

[1] Dal Pozzolo, A. et al.: Credit card fraud detection: A realistic modeling and a novel learning strategy. IEEE Trans Neural Netw Learn Syst. 29, 8, 3784–3797 (2018). https://doi.org/10.1109/TNNLS.2017.2736643.

[2] Yu, W.F., Wang, N.: Research on Credit Card Fraud Detection Model Based on Distance Sum. 2009 International Joint Conference on Artificial Intelligence. 353–356 (2009). https://doi.org/10.1109/JCAI.2009.146.

[3] Ojajuni, O. et al.: Predicting Student Academic Performance Using Machine Learning. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 12957 LNCS, 481–491 (2021). https://doi.org/10.1007/978-3-030-87013-3_36/FIGURES/3.

[4] Yu, W.F., Wang, N.: Research on Credit Card Fraud Detection Model Based on Distance Sum. 2009 International Joint Conference on Artificial Intelligence. 353–356 (2009). https://doi.org/10.1109/JCAI.2009.146.

[5] Chang, C.H.: Managing Credit Card Fraud Risk by Autoencoders: (ICPAI2020). Proceedings - 2020 International Conference on Pervasive Artificial Intelligence, ICPAI 2020. 118–122 (2020). https://doi.org/10.1109/ICPAI51961.2020.00029.

[6] Dal Pozzolo, A. et al.: Credit card fraud detection: A realistic modeling and a novel learning strategy. IEEE Trans Neural Netw Learn Syst. 29, 8, 3784–3797 (2018). https://doi.org/10.1109/TNNLS.2017.2736643.

[7] Sarma, D. et al.: Bank Fraud Detection using Community Detection Algorithm. 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA). 642–646 (2020). https://doi.org/10.1109/ICIRCA48905.2020.9182954.

[8] Hussain, H.M.: Retraction: Hussain HM, Hotopf M, Oyebode F. Atypical Antipsychotic Drugs and Alzheimer's Disease. N Engl J Med 2007; 356:416. New England Journal of Medicine. 356, 14, (2007). https://doi.org/10.1056/nejmc076105.

[9] Ghosh, H., Tusher, M.A., Rahat, I.S., Khasim, S., Mohanty, S.N. (2023). Water Quality Assessment Through Predictive Machine Learning. In: Intelligent Computing and Networking. IC-ICN 2023. Lecture Notes in Networks and Systems, vol 699. Springer, Singapore. https://doi.org/10.1007/978-981-99-3177-4_6

[10] Rahat IS, Ghosh H, Shaik K, Khasim S, Rajaram G. Unraveling the Heterogeneity of Lower-Grade Gliomas: Deep Learning-Assisted Flair Segmentation and Genomic Analysis of Brain MR Images. EAI Endorsed Trans Perv Health Tech [Internet]. 2023 Sep. 29 [cited 2023 Oct. 2];9.https://doi.org/10.4108/eetpht.9.4016

[11] Ghosh H, Rahat IS, Shaik K, Khasim S, Yesubabu M. Potato Leaf Disease Recognition and Prediction using Convolutional Neural Networks. EAI Endorsed Scal Inf Syst [Internet]. 2023 Sep. 21https://doi.org/10.4108/eetsis.3937

[12] Mandava, S. R. Vinta, H. Ghosh, and I. S. Rahat, "An All-Inclusive Machine Learning and Deep Learning Method for Forecasting Cardiovascular Disease in Bangladeshi Population", EAI Endorsed Trans Perv Health Tech, vol. 9, Oct. 2023.https://doi.org/10.4108/eetpht.9.4052

[13] Mandava, M.; Vinta, S. R.; Ghosh, H.; Rahat, I. S. Identification and Categorization of Yellow Rust Infection in Wheat through Deep Learning Techniques. EAI Endorsed Trans IoT 2023, 10. https://doi.org/10.4108/eetiot.4603

[14] Khasim, I. S. Rahat, H. Ghosh, K. Shaik, and S. K. Panda, "Using Deep Learning and Machine Learning: Real-Time Discernment and Diagnostics of Rice-Leaf Diseases in Bangladesh", EAI Endorsed Trans IoT, vol. 10, Dec. 2023 https://doi.org/10.4108/eetiot.4579

[15] Khasim, H. Ghosh, I. S. Rahat, K. Shaik, and M. Yesubabu, "Deciphering Microorganisms through Intelligent Image Recognition: Machine Learning and Deep Learning Approaches, Challenges, and Advancements", EAI Endorsed Trans IoT, vol. 10, Nov. 2023. https://doi.org/10.4108/eetiot.4484

[16] Mohanty, S.N.; Ghosh, H.; Rahat, I.S.; Reddy, C.V.R. Advanced Deep Learning Models for Corn Leaf Disease Classification: A Field Study in Bangladesh. Eng. Proc. 2023, 59, 69.https://doi.org/10.3390/engproc2023059069

[17] Alenezi, F.; Armghan, A.; Mohanty, S.N.; Jhaveri, R.H.; Tiwari, P. Block-Greedy and CNN Based Underwater Image Dehazing for Novel Depth Estimation and Optimal Ambient Light. Water 2021, 13, 3470. https://doi.org/10.3390/w13233470

[18] Agarwal, N. et al.: Applying XGBoost Machine Learning Model to Succor Astronomers Detect Exoplanets in Distant Galaxies. (2022). https://doi.org/10.1007/978-3-030-95711-7_33.

[19] Agarwal, N. et al.: Multiclass Classification of Different Glass Types using Random Forest Classifier. In: Proceedings - 2022 6th International Conference on Intelligent Computing and Control Systems, ICICCS 2022. (2022). https://doi.org/10.1109/ICICCS53718.2022.9788326.

[20] Agarwal, N. et al.: Semi-Supervised Learning with GANs for Melanoma Detection. In: Proceedings - 2022 6th International Conference on Intelligent Computing and Control Systems, ICICCS 2022. (2022). https://doi.org/10.1109/ICICCS53718.2022.9787990.

[21] Tayal, D.K. et al.: To Predict the Fire Outbreak in Australia using Historical Database. In: 2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions), ICRITO 2022. (2022). https://doi.org/10.1109/ICRITO56286.2022.9964603.

[22] Agarwal, N., Tayal, D.K.: FFT based ensembled model to predict ranks of higher educational institutions. Multimed Tools Appl. 81, 23, (2022). https://doi.org/10.1007/s11042-022-13180-9.

[23] Agarwal, N., Tayal, D.K. (2023). A Novel Model to Predict the Whack of Pandemics on the International Rankings of Academia. In: Nandan Mohanty, S., Garcia Diaz, V., Satish Kumar, G.A.E. (eds) Intelligent Systems and Machine Learning. ICISML 2022. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 471. Springer, Cham. https://doi.org/10.1007/978-3-031-35081-8_3

[24] Gupta, A., Vardhan, H., Varshney, S., Saxena, S., Singh, S., & Agarwal, N. (2023). "Kconnect: The Design and Development of Versatile Web Portal for Enhanced Collaboration and Communication". EAI Endorsed Transactions on Scalable Information Systems https://doi.org/10.4108/eetsis.4022.

[16] Agarwal N, Kumar N, Anushka, Abrol V, Garg Y. Enhancing Image Recognition: Leveraging Machine Learning on specialized Medical Datasets. EAI Endorsed Trans Perv Health Tech DOI: https://doi.org/10.4108/eetpht.9.4336.

[17] Agarwal N, Arora I, Saini H, Sharma U. A Novel Approach for Earthquake Prediction Using Random Forest and Neural Networks. EAI Endorsed Trans Energy Web DOI: https://doi.org/10.4108/ew.4329.

[18] Dahiya R, Nidhi, Kumari K, Kumari S, Agarwal N. Usage of Web Scraping in the Pharmaceutical Sector. EAI Endorsed Trans Perv Health Tech DOI: https://doi.org/10.4108/eetpht.9.4312.

[19] Dahiya, R., Arunkumar, B., Dahiya, V. K., & Agarwal, N. (2023). Facilitating Healthcare Sector through IoT: Issues, Challenges, and Its Solutions. EAI Endorsed Transactions on Internet of Things, 9(4), e5-e5.

[20] Anushka, Agarwal, N., Tayal, D. K., Abrol, V., Deepakshi, Garg, Y., & Jha, A. (2022, December). Predicting Credit Card Defaults with Machine Learning Algorithm Using Customer Database. In International Conference on Intelligent Systems and Machine Learning (pp. 262-277). Cham: Springer Nature Switzerland.

[21] Jha, A., Agarwal, N., Tayal, D. K., Abrol, V., Deepakshi, Garg, Y., & Anushka. (2022, December). Movie Recommendation Using Content-Based and Collaborative Filtering Approach. In International Conference on Intelligent Systems and Machine Learning (pp. 439-450). Cham: Springer Nature Switzerlan