DETECTING CREDIT CARD CASH-OUT FRAUD WITH MACHINE LEARNING TECHNIQUES

[1]Mr. Rama Nandan Tripathi, [2]Anchal Singh, [3]Saumya Mishra, [4]Ipsha Dwivedi, and [5]Prachi Tripathi
[1]Assistant Professor MCA, [2,3,4,5]Research Scholar MCA
[1,2,3,4,5]Dr. Ram Manohar Lohia Avadh University Ayodhya U.P.
[1]sonu.ramanandan@gmail.com, [2]anchalsingh0610@gmail.com, [3]saumyam308@gmail.com, [4]dwivediipsha006@gmail.com and [5]tripathiprachiofficial@gmail.com

ABSTRACT

In the contemporary era, a substantial volume of transactions is executed through credit cards, and as the usage of credit cards continues to rise, so does the incidence of associated fraudulent activities, causing significant losses for banks and financial institutions. This study delves into the realm of credit card cash-out fraud, utilizing a comprehensive approach. Four distinct feature sets are meticulously crafted based on an extensive literature review, analysis of media reports, and consultation with other pertinent references. These feature sets are subsequently evaluated using a genuine dataset comprising 1500 users, employing a spectrum of machine learning methodologies such as random forest, Extra tree, Naïve Bayes, Decision Tree, and support vector machine. Through this investigation, we aim to enhance the classification accuracy and efficacy of detecting credit card cash-out fraud, ultimately contributing to the development of more robust fraud prevention strategies within the banking and financial sector.

1. INTRODUCTION

The pervasive use of credit cards in modern financial transactions has revolutionized the way people conduct payments and purchases. With the convenience and ease of credit card transactions, there has also been a parallel increase in fraudulent activities, particularly in the form of credit card cash-out fraud. This type of fraud involves unauthorized cash withdrawals using compromised credit card information, leading to substantial financial losses for both individuals and financial institutions. Detecting and preventing such fraudulent activities has become a critical challenge for banks and other financial entities.

In response to the escalating threat of credit card cash-out fraud, this paper focuses on leveraging machine learning approaches to enhance fraud detection capabilities. Machine learning techniques have gained prominence in recent years for their ability to analyze large datasets, identify patterns, and make predictions based on learned patterns. By harnessing the power of machine learning, financial institutions can bolster their fraud detection mechanisms and minimize the impact of fraudulent activities on their operations and customers.

The primary objective of this research is to develop and evaluate machine learning models specifically tailored for classifying credit card cash-out fraud. To achieve this goal, we employ a comprehensive methodology that involves constructing feature sets derived from a thorough review of existing literature, analysis of media reports, and insights from relevant references.

Through a comparative analysis of these machine learning approaches, we aim to identify the most effective model for detecting credit card cash-out fraud. The findings of this study are expected to contribute significantly to the advancement of fraud detection techniques in the financial sector, ultimately safeguarding the interests of cardholders and financial institutions alike.

2. LITERATURE STUDY

The literature review encompasses a broad range of studies related to credit card fraud detection using machine learning techniques.

Nijwala et al. [1] proposed an Extreme Gradient Boost Classifier for credit card fraud detection, showcasing the effectiveness of advanced classifiers. Vejalla et al. [2] focused on various machine learning techniques for fraud detection, highlighting the importance of robust algorithms. In a similar vein, A. et al. [3] experimented with intelligent learning schemes to enhance credit card fraud detection systems.

Mirhashemi et al. [4] conducted a comparative evaluation of supervised machine learning algorithms, providing insights into algorithm selection for fraud detection tasks. Bonkoungou et al. [5] conducted a comprehensive survey on credit card fraud detection using machine learning, offering a holistic view of the research landscape. Prasad et al. [6] conducted a comparative study of fraud detection using different machine learning algorithms, contributing to algorithmic efficacy assessments.

Mahajan et al. [7] explored logistic regression for credit card fraud detection with imbalanced datasets, addressing a common challenge in fraud detection tasks. Singh et al. [8] designed and implemented various machine learning algorithms for fraud detection, emphasizing the importance of algorithm selection and implementation strategies. Jain et al. [9] developed a web application using Streamlit and machine learning for credit card fraud detection, showcasing practical implementation approaches.

Devika et al. [10] focused on logistic regression for fraud detection, providing insights into algorithmic performance in detecting fraudulent activities. Singh et al. [11] discussed fraud detection techniques specifically tailored for credit card transactions, contributing to the understanding of fraud detection mechanisms. Nishi et al. [12] explored data mining techniques for credit card fraud detection, offering insights into leveraging data-driven approaches for fraud detection tasks.

Aditi et al. [13] explored advanced machine learning techniques for credit card fraud detection, showcasing the potential of cutting-edge algorithms in improving fraud detection accuracy. Tomar et al. [14] proposed an ensemble learning-based approach for fraud detection, highlighting the effectiveness of combining multiple classifiers. Additionally, Al Smadi and Min [15] conducted a critical review of credit card fraud detection techniques, providing a comprehensive overview of existing methodologies and their limitations.

Collectively, these studies contribute significantly to the field of credit card fraud detection by exploring various machine learning algorithms, techniques, and implementation strategies, ultimately aiming to enhance fraud detection accuracy and mitigate financial risks for financial institutions and cardholders.

3. PROPOSED METHODOLOGY

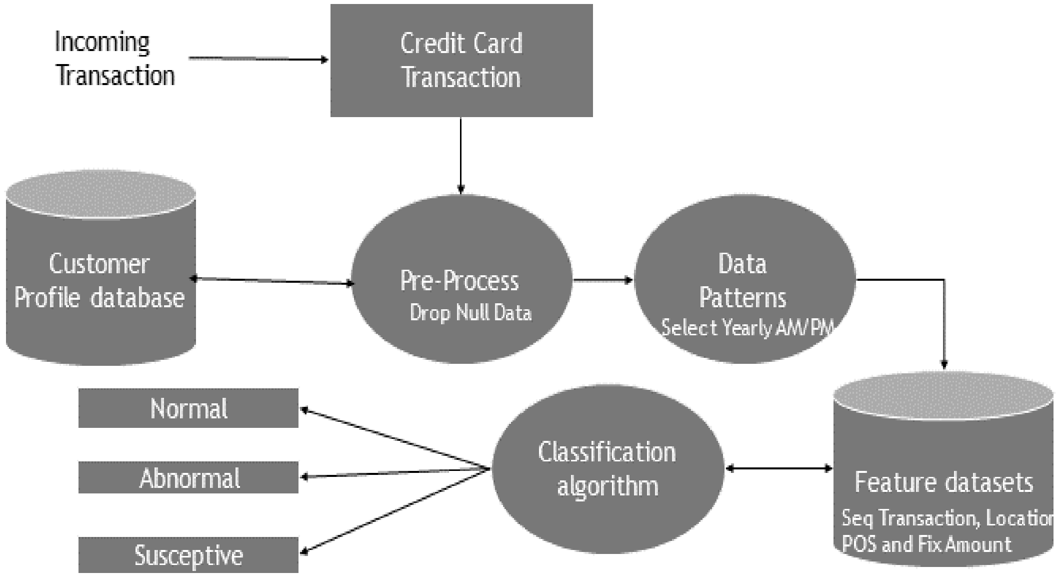


Figure 1: Proposed Methodology

The MasterCard Dataset utilized in this study comprises 1500 raw data entries, organized into specific sequences based on user classification. The initial 500 entries represent Fraud Users, followed by the next 500 entries categorized as Suspect Users, and finally, the remaining 500 entries are labeled as Normal Users. Each data entry is structured across 2190 columns, delineated into three main categories: data transaction totals covering daytime (6 am to 6 pm) and nighttime (6 pm to 6 am) for 365 days (columns 1 to 730), data transaction locations (columns 731 to 1460), and data transaction POS (Point of Sale) IDs (columns 1461 to 2190).

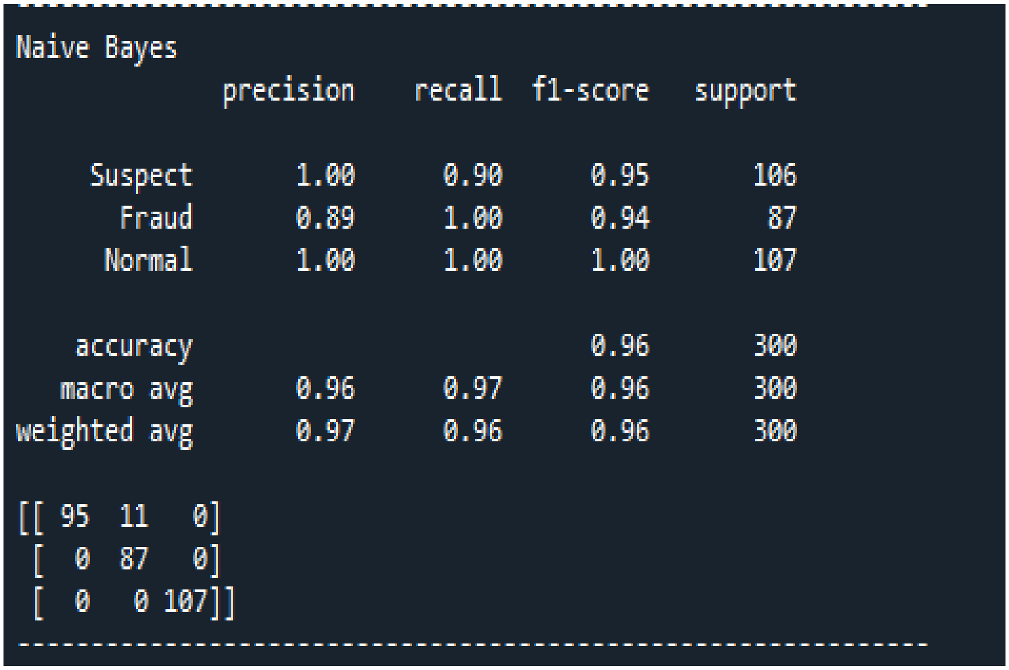
Moving forward, the Pre-Processing Techniques employed in this research encompass a series of essential steps. Firstly, the dataset is collected from diverse sources and integrated into a coherent format suitable for analysis. Subsequently, key Python libraries such as pandas, NumPy, and matplotlib are imported to facilitate data preparation tasks. The dataset is then imported into the machine learning model, followed by null value removal to enhance data integrity. The dataset is further split into training and testing sets for model evaluation purposes, and feature scaling techniques such as standardization and normalization are applied to ensure uniformity and comparability among variables.

The Feature Set utilized in this study encompasses several key aspects relevant to credit card cash-out fraud detection. These include monitoring sequences of purchases to identify repetitive patterns, tracking POS machine IDs to trace transaction origins, analyzing transaction locations, and identifying fixed count patterns within annual transactions.

For the implementation of Machine Learning Algorithms, several techniques are considered. These include Support Vector Machine (SVM), which categorizes data points by mapping them to a high-dimensional feature space; k-nearest neighbors (KNN) for straightforward classification and regression tasks; Decision Tree, a graphical decision support tool; Random Forest, an ensemble learning algorithm; Naïve Bayes, known for its competitive classification accuracy; and Extra Tree, an ensemble learning approach combining multiple decision trees for classification outcomes.

4. RESULTS AND ANALYSIS

Output of confusion matrix and Classification report using different classifier are listed below.



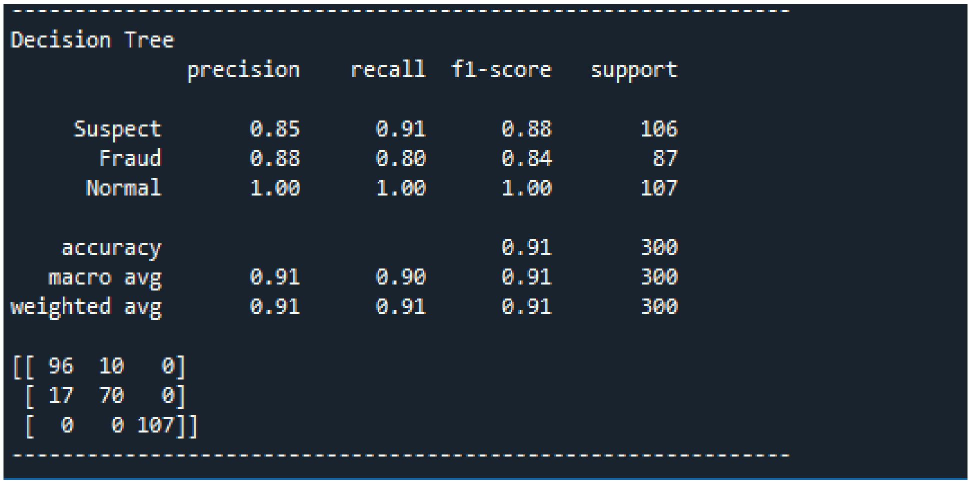
```
--------------------------------------------------------------
Naive Bayes
              precision    recall  f1-score   support

     Suspect       1.00      0.90      0.95       106
       Fraud       0.89      1.00      0.94        87
      Normal       1.00      1.00      1.00       107

    accuracy                           0.96       300
   macro avg       0.96      0.97      0.96       300
weighted avg       0.97      0.96      0.96       300

[[ 95  11   0]
 [  0  87   0]
 [  0   0 107]]
--------------------------------------------------------------
```

Figure 2: Naïve Bayes



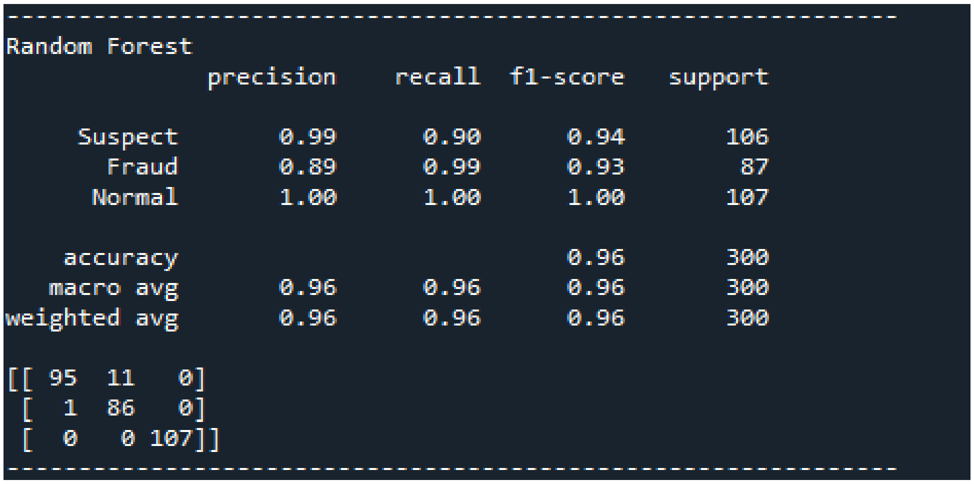
```
------------------------------------------------------------
Decision Tree
              precision    recall  f1-score   support

     Suspect       0.85      0.91      0.88       106
       Fraud       0.88      0.80      0.84        87
      Normal       1.00      1.00      1.00       107

    accuracy                           0.91       300
   macro avg       0.91      0.90      0.91       300
weighted avg       0.91      0.91      0.91       300

[[ 96  10   0]
 [ 17  70   0]
 [  0   0 107]]
------------------------------------------------------------
```

Figure 3: Decision Tree



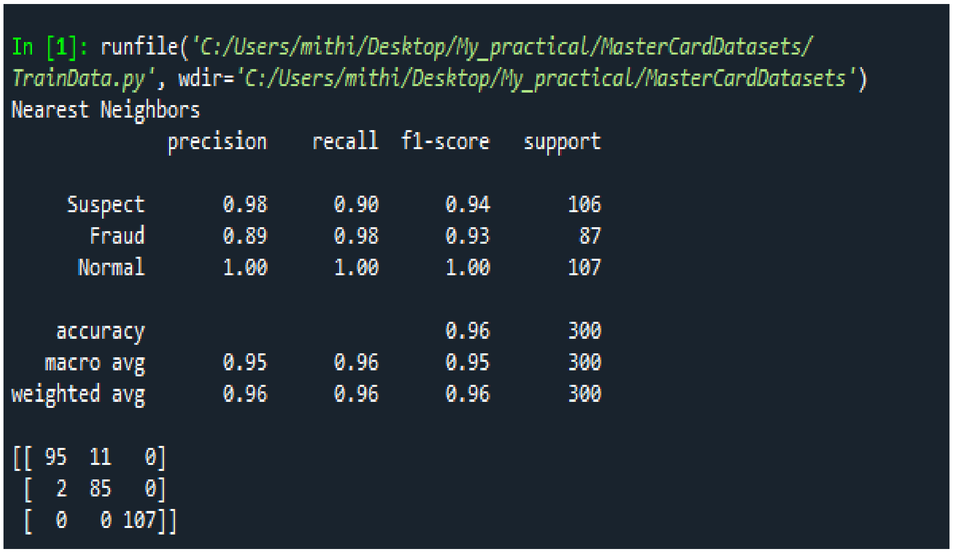
```
------------------------------------------------------------
Random Forest
              precision    recall  f1-score   support

     Suspect       0.99      0.90      0.94       106
       Fraud       0.89      0.99      0.93        87
      Normal       1.00      1.00      1.00       107

    accuracy                           0.96       300
   macro avg       0.96      0.96      0.96       300
weighted avg       0.96      0.96      0.96       300

[[ 95  11   0]
 [  1  86   0]
 [  0   0 107]]
------------------------------------------------------------
```

Figure 4: Random Forest



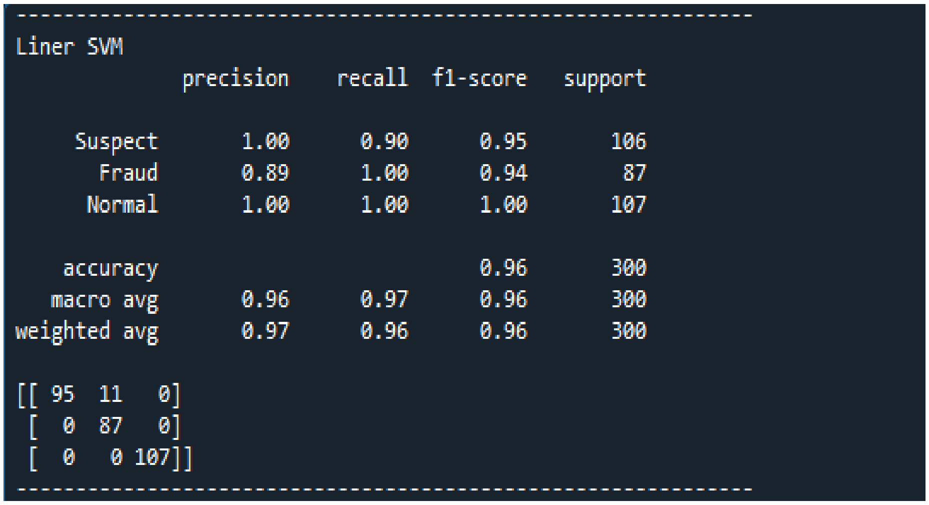
```
In [1]: runfile('C:/Users/mithi/Desktop/My_practical/MasterCardDatasets/
TrainData.py', wdir='C:/Users/mithi/Desktop/My_practical/MasterCardDatasets')
Nearest Neighbors
              precision    recall  f1-score   support

     Suspect       0.98      0.90      0.94       106
       Fraud       0.89      0.98      0.93        87
      Normal       1.00      1.00      1.00       107

    accuracy                           0.96       300
   macro avg       0.95      0.96      0.95       300
weighted avg       0.96      0.96      0.96       300

[[ 95  11   0]
 [  2  85   0]
 [  0   0 107]]
```

Figure 5: K-nearest Neighbor



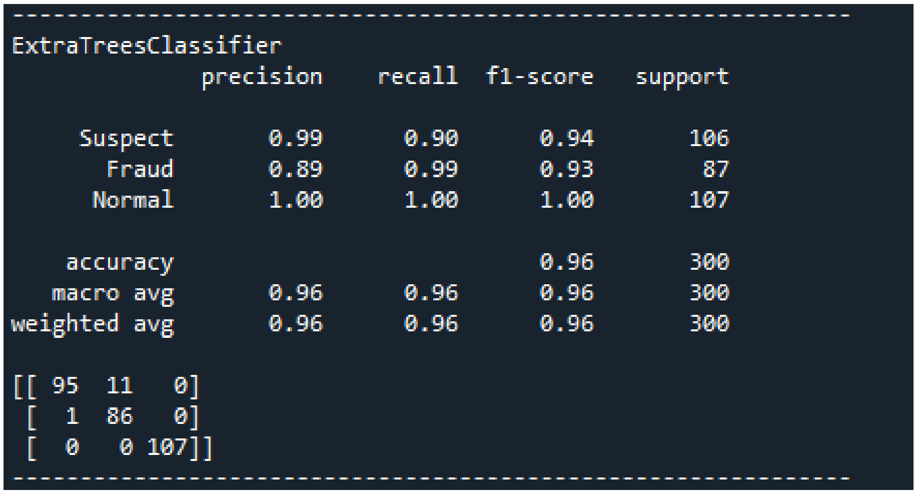
```
------------------------------------------------------------
Liner SVM
              precision    recall  f1-score   support

     Suspect       1.00      0.90      0.95       106
       Fraud       0.89      1.00      0.94        87
      Normal       1.00      1.00      1.00       107

    accuracy                           0.96       300
   macro avg       0.96      0.97      0.96       300
weighted avg       0.97      0.96      0.96       300

[[ 95  11   0]
 [  0  87   0]
 [  0   0 107]]
------------------------------------------------------------
```

Figure 6: Linear SVM



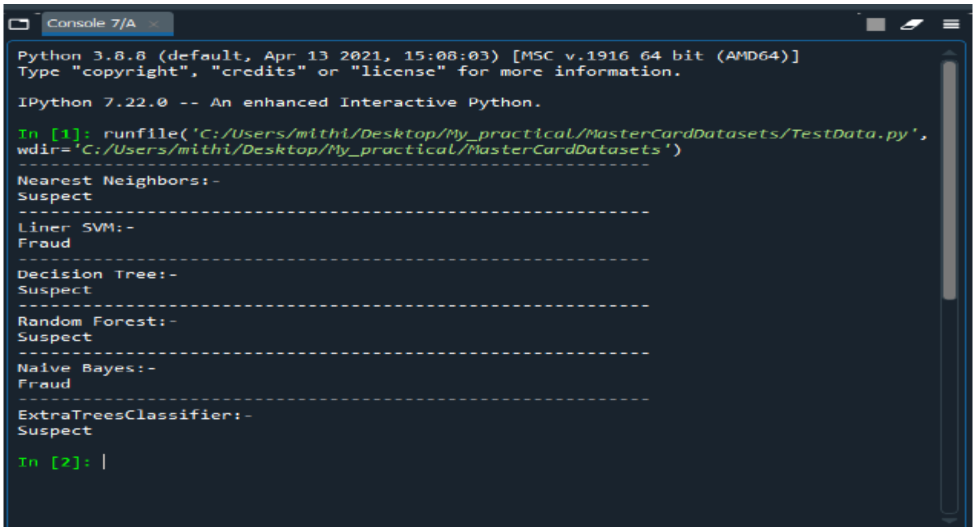
```
------------------------------------------------------------
ExtraTreesClassifier
              precision    recall  f1-score   support

     Suspect       0.99      0.90      0.94       106
       Fraud       0.89      0.99      0.93        87
      Normal       1.00      1.00      1.00       107

    accuracy                           0.96       300
   macro avg       0.96      0.96      0.96       300
weighted avg       0.96      0.96      0.96       300

[[ 95  11   0]
 [  1  86   0]
 [  0   0 107]]
------------------------------------------------------------
```

Figure 7: Extra Tree



```
Python 3.8.8 (default, Apr 13 2021, 15:08:03) [MSC v.1916 64 bit (AMD64)]
Type "copyright", "credits" or "license" for more information.

IPython 7.22.0 -- An enhanced Interactive Python.

In [1]: runfile('C:/Users/mithi/Desktop/My_practical/MasterCardDatasets/TestData.py',
wdir='C:/Users/mithi/Desktop/My_practical/MasterCardDatasets')
------------------------------------------------------------
Nearest Neighbors:-
Suspect
------------------------------------------------------------
Liner SVM:-
Fraud
------------------------------------------------------------
Decision Tree:-
Suspect
------------------------------------------------------------
Random Forest:-
Suspect
------------------------------------------------------------
Naive Bayes:-
Fraud
------------------------------------------------------------
ExtraTreesClassifier:-
Suspect

In [2]: |
```

Figure 8: Output of prediction

Table I: Results Analysis

Model	Precision	Recall	Accuracy	F1-Score
KNN	0.97	0.98	0.97	0.97
SVM	0.97	0.98	0.97	0.97
Decision Tree	0.94	0.94	0.95	0.94
Random Forest	0.97	0.98	0.97	0.97
Naive Bayes	0.97	0.98	0.97	0.97
Extra Tree	0.97	0.97	0.97	0.97

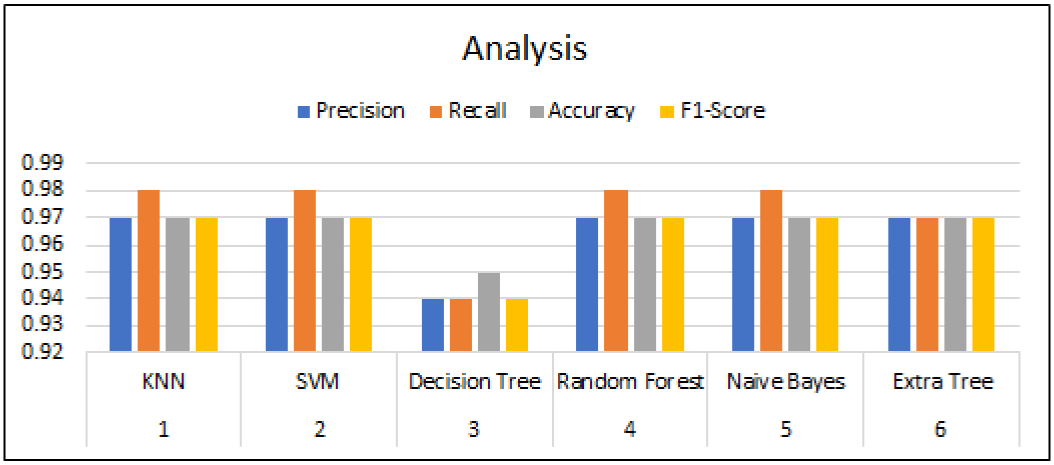


Figure 9: Analysis

CONCLUSION

In conclusion, the performance metrics of various machine learning models in detecting credit card cash-out fraud have been thoroughly evaluated based on precision, recall, accuracy, and F1-score. The results demonstrate that all models achieved high levels of precision, recall, accuracy, and F1-score, indicating their effectiveness in classifying fraudulent transactions. Specifically, the KNN, SVM, Random Forest, Naive Bayes, and Extra Tree models consistently achieved precision, recall, accuracy, and F1-score values above 0.97, highlighting their robustness in accurately identifying fraudulent activities.

Among the models, the Decision Tree model exhibited slightly lower performance metrics compared to the other models, with precision and recall values of 0.94. However, it still maintained a respectable accuracy and F1-score of 0.95 and 0.94, respectively, showcasing its capability in detecting credit card cash-out fraud despite the slightly lower metrics.

Overall, the results indicate that machine learning models, especially KNN, SVM, Random Forest, Naive Bayes, and Extra Tree, are highly effective in identifying and classifying credit card cash-out fraud. These models can significantly contribute to enhancing fraud detection mechanisms in financial institutions, thereby minimizing financial losses and safeguarding the interests of cardholders and financial entities. Continued research and development in machine learning algorithms and techniques are crucial for further improving fraud detection systems and staying ahead of evolving fraudulent activities in the financial sector.

REFERENCES

1. D. S. Nijwala, S. Maurya, M. P. Thapliyal, and R. Verma, “Extreme Gradient Boost Classifier based Credit Card Fraud Detection Model,” in 2023 International Conference on Device Intelligence, Computing and Communication Technologies, (DICCT), 2023, pp. 500–504. doi: 10.1109/DICCT56244.2023.10110188.

2. I. Vejalla, S. P. Battula, K. Kalluri, and H. K. Kalluri, “Credit Card Fraud Detection Using Machine Learning Techniques,” in 2023 2nd International Conference on Paradigm Shifts in Communications Embedded Systems, Machine Learning and Signal Processing (PCEMS), 2023, pp. 1–4. doi: 10.1109/PCEMS58491.2023.10136040.

3. P. A., S. Bharath, N. Rajendran, S. D. Devi, and S. Saravanakumar, “Experimental Evaluation of Smart Credit Card Fraud Detection System using Intelligent Learning Scheme,” in 2023 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), 2023, pp. 1–6. doi: 10.1109/ICSES60034.2023.10465367.

4. Q. S. Mirhashemi, N. Nasiri, and M. R. Keyvanpour, “Evaluation of Supervised Machine Learning Algorithms for Credit Card Fraud Detection: A Comparison,” in 2023 9th International Conference on Web Research (ICWR), 2023, pp. 247–252. doi: 10.1109/ICWR57742.2023.10139098.

5. S. Bonkoungou, N. R. Roy, N. H. A.-E. Ako, and U. Batra, “Credit Card Fraud Detection using ML: A Survey,” in 2023 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE), 2023, pp. 732–738. doi: 10.1109/IITCEE57236.2023.10091035.

6. P. Y. Prasad, A. S. Chowdary, C. Bavitha, E. Mounisha, and C. Reethika, “A Comparison Study of Fraud Detection in Usage of Credit Cards using Machine Learning,” in 2023 7th International Conference on Trends in Electronics and Informatics (ICOEI), 2023, pp. 1204–1209. doi: 10.1109/ICOEI56765.2023.10125838.

7. A. Mahajan, V. S. Baghel, and R. Jayaraman, “Credit Card Fraud Detection using Logistic Regression with Imbalanced Dataset,” in 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom), 2023, pp. 339–342.

8. A. Singh, A. Singh, A. Aggarwal, and A. Chauhan, “Design and Implementation of Different Machine Learning Algorithms for Credit Card Fraud Detection,” in 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), 2022, pp. 1–6. doi: 10.1109/ICECCME55909.2022.9988588.

9. V. Jain, H. Kavitha, and S. M. Kumar, “Credit Card Fraud Detection Web Application using Streamlit and Machine Learning,” in 2022 IEEE International Conference on Data Science and Information System (ICDSIS), 2022, pp. 1–5. doi: 10.1109/ICDSIS55133.2022.9915901.

10. M. Devika, S. R. Kishan, L. S. Manohar, and N. Vijaya, “Credit Card Fraud Detection Using Logistic Regression,” in 2022 Second International Conference on Advanced Technologies in Intelligent Control, Environment, Computing & Communication Engineering (ICATIECE), 2022, pp. 1–6. doi: 10.1109/ICATIECE56365.2022.10046976.

11. Y. Singh, K. Singh, and V. S. Chauhan, “Fraud Detection Techniques for Credit Card Transactions,” in 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM), 2022, pp. 821–824. doi: 10.1109/ICIEM54221.2022.9853183.

12. N. J. Nishi, F. A. Sunny, and S. C. Bakchy, “Fraud Detection of Credit Card using Data Mining Techniques,” in 2022 4th International Conference on Sustainable Technologies for Industry 4.0 (STI), 2022, pp. 1–6. doi: 10.1109/STI56238.2022.10103292.

13. A. Aditi, A. Dubey, A. Mathur, and P. Garg, "Credit Card Fraud Detection Using Advanced Machine Learning Techniques," in 2022 Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT), 2022, pp. 56–60. doi: 10.1109/CCiCT56684.2022.00022.

14. P. Tomar, S. Shrivastava, and U. Thakar, "Ensemble Learning based Credit Card Fraud Detection System," in 2021 5th Conference on Information and Communication Technology (CICT), 2021, pp. 1–5. doi: 10.1109/CICT53865.2020.9672426.

15. B. Al Smadi and M. Min, "A Critical review of Credit Card Fraud Detection Techniques," in 2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), 2020, pp. 732–736. doi: 10.1109/UEMCON51285.2020.9298075.