VISVESVARAYA TECHNOLOGICAL UNIVERSITY
"Jnana Sangama", Belagavi – 590018

AI & ML MINI PROJECT REPORT
ON

"Credit Card Fraud Detection "

Submitted in the partial fulfilment of the requirement for the seventh semester of

BACHELOR OF ENGINEERING

In

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

By

ANKITHA (1RR20AI004) GUNANKA N K(1RR20AI008)

JESSICA R (1RR20AI010)

Under the guidance of

DR RAJESH.K.S

Professor & HOD

Dept. of AIML, RRCE

RAJARAJESWARI COLLEGE OF ENGINEERING
MYSORE ROAD, BANGALORE-560074
(An ISO 9001:2008 Certified Institute)

DEPARTMENT OF ARTIFICIAL INTELLIGENCE &
MACHINE LEARNING

RAJARAJESWARI COLLEGE OF ENGINEERING
MYSORE ROAD, BANGALORE-560074

[Affiliated to VTU, Belgavi. Approved by AICTE, New Delhi]

No.14, Ramohalli Cross, Kumbalgodu, Mysore Road,Bengaluru560074

DEPARTMENT OF ARTIFICIAL INTELLIGENCE & MACHINE LEARNING

CERTIFICATE

Certified that mini project work entitled

"Credit Card Fraud Detection"

Carried out by

ANKITHA (1RR20AI004) GUNANKA N K(1RR20AI008)

JESSICA R (1RR20AI010)

The students of "RajaRajeswari College of Engineering" in partial fulfilment for the seventh semester of Bachelor Of Engineering in Artificial Intelligence & Machine Learning of the Visvesvaraya Technological University, Belagavi during the year 2023–2024. It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the report deposited in the departmental library. The mini project report has been approved as it satisfies the academic requirements in respect of mini project work prescribed for the seventh semester.

…….....	……	…….
Signature of Coordinator	Signature of HOD	Signature of Principal
Dr.Rajesh K S.	Dr.RAJESH K S	Dr.R.BALAKRISHNA
Prof. &HOD, Dept of AI&ML	Prof. &HOD, Dept of AI&ML	Principal
RRCE, Bangalore	RRCE, Bangalore	RRCE, Bangalore

External Viva-Voce

Name of the Examiner	Signature with date:
1) _____	1) _____
2) _____	2) _____

DECLARATION

We, the undersigned students of 7th semester of AIML, Rajarajeshwari College of Engineering, solemnly declare that our project work entitled "CREDIT CARD FRAUD DETECTION" is a bonafide work of ours. Our project is neither a copy nor by means a modification of any other engineering project.

We also declare that this project was not entitled for submission to any other university in the past and shall remain the only submission made and will not be submitted by us to any other university in the future.

Submitted By	Signature
ANKITHA(1RR20AI004)	…….
GUNANKA N K(1RR20AI008)	…….
JESSICA R(1RR20AI010)	…….

ACKNOWLEDGEMENTS

I am truly thankful and convey my sincere gratitude to the principal, **Dr. R Balakrishna,** RajaRajeswari College of Engineering, Bangalore.

I convey my sincere gratitude to **Dr. Rajesh K S**, Head of Department, Department of Artificial Intelligence & Machine Learning for his meticulous support, continuous co-operation, valuable suggestion and encouragement during the development of the project. I also extend my thanks for his invaluable guidance to imbibe the requisite knowledge for success of our project.

I also thank my parents who have encouraged me and supported me in every stage of development of this project.

Last but not the least, my wishes to the entire Artificial Intelligence & Machine Learning Department for their help and guidance, encouragement, inspiration and co-operation at all stages of the development of this project without which this project wouldn't have been a success.

Finally I express my heartfelt gratitude to all those who helped me to complete the project work successfully by providing support, suggestions, advise, guidance and much needed encouragement

ANKITHA (1RR20AI004)
GUNANKA N K(1RR20AI008)
JESSICA R (1RR20AI010)

ABSTRACT

With the increasing prevalence of online transactions, credit card fraud has become a significant concern for financial institutions and cardholders alike. This project proposes an advanced Credit Card Fraud Detection System leveraging machine learning algorithms to enhance the security of online transactions.

The primary objective of the system is to accurately identify and prevent fraudulent activities in real-time. The proposed solution incorporates a dataset comprising both legitimate and fraudulent credit card transactions, utilizing various machine learning techniques such as supervised learning, anomaly detection, and ensemble methods.

The project involves preprocessing and feature engineering to extract relevant information from the transaction data.The system is trained on historical data to learn patterns indicative of legitimate transactions and anomalies associated with fraudulent activities.

Key machine learning algorithms, including but not limited to logistic regression, decision trees, random forests, and support vector machines, are implemented and evaluated for their efficacy in distinguishing between genuine and fraudulent transactions. The system's performance is assessed based on metrics such as accuracy, precision, recall, and F1-score.

Furthermore, the project explores the feasibility of implementing real-time monitoring and alerting mechanisms to promptly notify users and financial institutions of potential fraudulent activities. The system aims to strike a balance between accurate fraud detection and minimizing false positives to ensure a seamless and secure online transaction experience.

The results of the project are expected to contribute to the development of robust and efficient credit card fraud detection systems, helping financial institutions mitigate financial losses and protect the interests of cardholders in an increasingly digital financial landscape.

TABLE OF CONTENT

CHAPTER 1
INTRODUCTION

Credit Card Fraud Detection is a critical concern for financial institutions worldwide, as it involves identifying and preventing unauthorized or fraudulent transactions. Fraudulent activities can lead to substantial financial losses for both financial institutions and their customers. Addressing this issue requires advanced data analysis and machine learning techniques, especially considering the evolving nature of fraudulent tactics.

This project employs the Cross-Industry Standard Process for Data Mining (CRISP-DM) approach to develop an effective Credit Card Fraud Detection system. The CRISP-DM methodology consists of six stages: Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment.

Data Understanding:

The dataset used is sourced from Kaggle and includes 32 features, denoted as V1 to V28 (confidential), Time, Amount, and Class. The 'Class' variable distinguishes between legitimate (Class 0) and fraudulent (Class 1) transactions. The data contains no missing values and is entirely numerical, with 'Class' as the only categorical feature.

Data Preparation:

Standardization is applied to numerical features for fair model comparison after balancing the dataset. Descriptive statistics reveal insights into the distribution of transaction amounts and equitable distribution of time, indicating independence for potential use as features.

Exploratory Data Analysis (EDA):

EDA is performed to explore relationships between fraud transactions and amounts, revealing that fraud transactions are predominantly below $2500. Additionally, time distribution indicates that fraud transactions are evenly spread throughout the dataset.

Business Understanding:

The project begins by acknowledging the class-imbalance issue inherent in credit card transactions, where legitimate transactions significantly outnumber fraudulent ones. Traditional approaches often result in overfitting to the majority class, leading to poor real-time performance. Thus, the project frames Credit Card Fraud Detection as an anomaly detection problem.

CHAPTER 2
PROBLEM STATEMENT

The surge in online transactions has led to an alarming increase in credit card fraud, posing a significant threat to the financial industry and consumers alike. Traditional rule-based fraud detection systems are becoming inadequate in dealing with the evolving tactics of fraudsters. Hence, there is a pressing need for a sophisticated Credit Card Fraud Detection System that harnesses the power of machine learning to provide proactive and adaptive protection.

Modeling:

Three classifiers – Stochastic Gradient Descent (SGD), Random Forest, and Logistic Regression – are employed for Credit Card Fraud Detection. Synthetic Minority Oversampling Technique (SMOTE) is used to address class imbalance. Model parameters are fine-tuned using Grid Search, and evaluation metrics such as AUC-ROC score, Classification Report, Accuracy, and F1-Score are used for performance assessment.

Objectives:

Develop a Robust Machine Learning Model:
Create a credit card fraud detection system that employs machine learning algorithms to analyze historical transaction data for learning patterns indicative of fraudulent activities.

Optimize Feature Engineering:
Explore and implement advanced preprocessing and feature engineering techniques to extract relevant information from raw transactional data, enhancing the discriminatory capabilities of the machine learning models.

Evaluate Algorithmic Performance:
Implement and evaluate various machine learning algorithms, including supervised learning models (e.g., logistic regression, decision trees, random forests) and unsupervised anomaly detection methods, to determine their effectiveness in distinguishing between legitimate and fraudulent transactions.

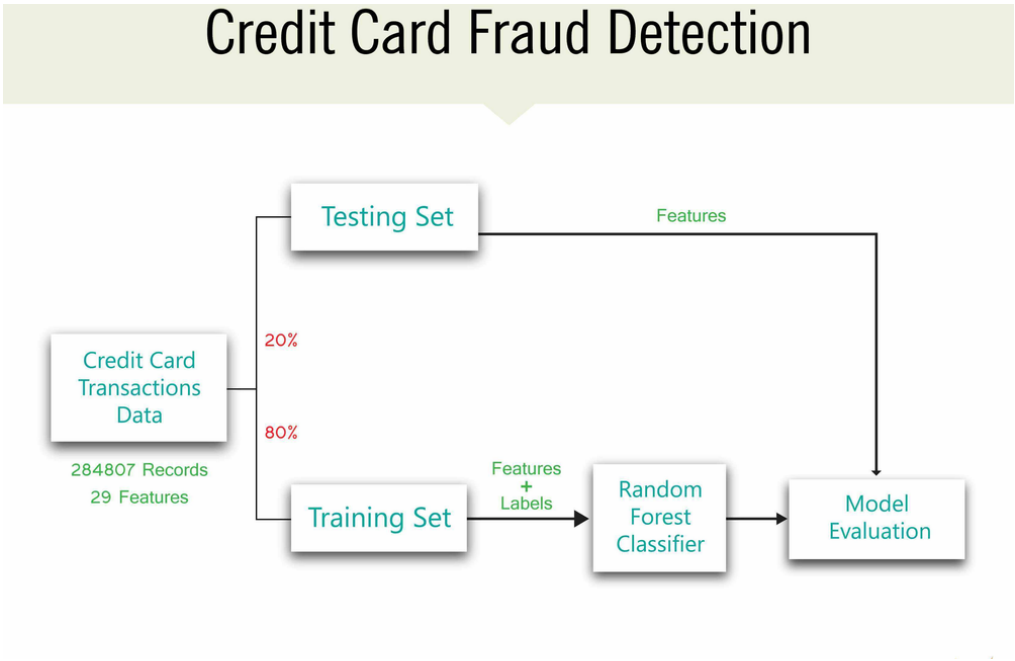
Minimize False Positives:
Strive to strike a balance between accuracy and false positives, as reducing inconveniences for legitimate cardholders is crucial for user satisfaction.

Implement Real-time Monitoring and Alerting:
Investigate the feasibility of real-time monitoring and alerting mechanisms to promptly notify financial institutions and users of potential fraudulent activities, enabling swift intervention.

CHAPTER 3
LITERATURE SURVEY

Multiple Supervised and Semi-Supervised machine learning techniques are used for fraud detection [8], but we aim is to overcome three main challenges with card frauds related dataset i.e., strong class imbalance, the inclusion of labelled and unlabeled samples, and to increase the ability to process a large number of transactions. Different Supervised machine learning algorithms [3] like Decision Trees, Naive Bayes Classification, Least Squares Regression, Logistic Regression and SVM are used to detect fraudulent transactions in real-time datasets. Two methods under random forests [6] are used to train the behavioral features of normal and abnormal transactions. They are Random-tree-based random forest and CART-based. Even though random forest obtains good results on small set data, there are still some problems in case of imbalanced data.

Performance of Logistic Regression, K-Nearest Neighbour, and Naïve Bayes are analyzed on highly skewed credit card fraud data where Research is carried out on examining meta-classifiers and meta-learning approaches in handling highly imbalanced credit card fraud data. Through supervised learning methods can be used there may fail at certain cases of detecting the fraud cases. A model of deep Auto-encoder and restricted Boltzmann machine (RBM) [2] that can construct normal transactions to find anomalies from normal patterns. Not only that a hybrid method is developed with a combination of Adaboost and Majority Voting methods [4]

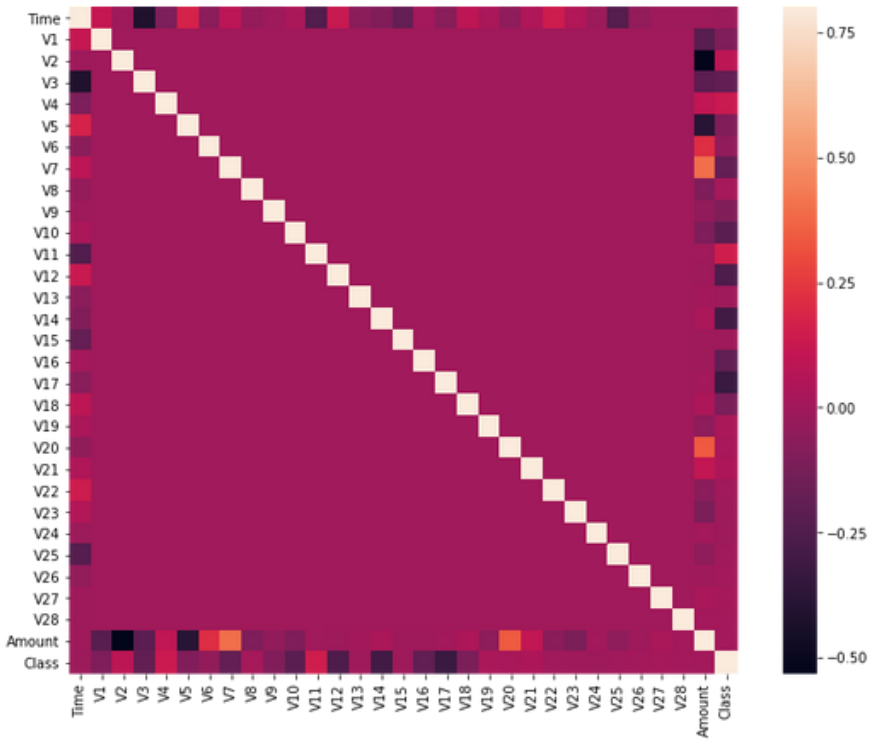


CHAPTER 4
PROPOSED SYSTEM

The proposed credit card fraud detection system employs advanced machine learning algorithms, real-time transaction monitoring, and feature engineering to enhance accuracy. It prioritizes explainable AI for transparency, integrates continuous learning for adaptability, and ensures a user-friendly interface. With a focus on cross-validation, the system facilitates seamless integration with existing financial systems, adheres to security standards, and optimizes scalability. The goal is to provide a robust solution that swiftly identifies and addresses fraudulent transactions, offering an enhanced level of security and trust in credit card transactions.

System Analysis:

System analysis involves a comprehensive examination of the credit card fraud detection system to understand its requirements, functionalities, and potential improvements. This process includes studying the existing system, identifying user needs, and defining system objectives. It assesses the feasibility, performance, and security aspects of the system. The analysis also involves evaluating data sources, transaction patterns, and fraud indicators to design effective algorithms. Furthermore, system analysis explores the integration with financial networks and the impact on existing processes.



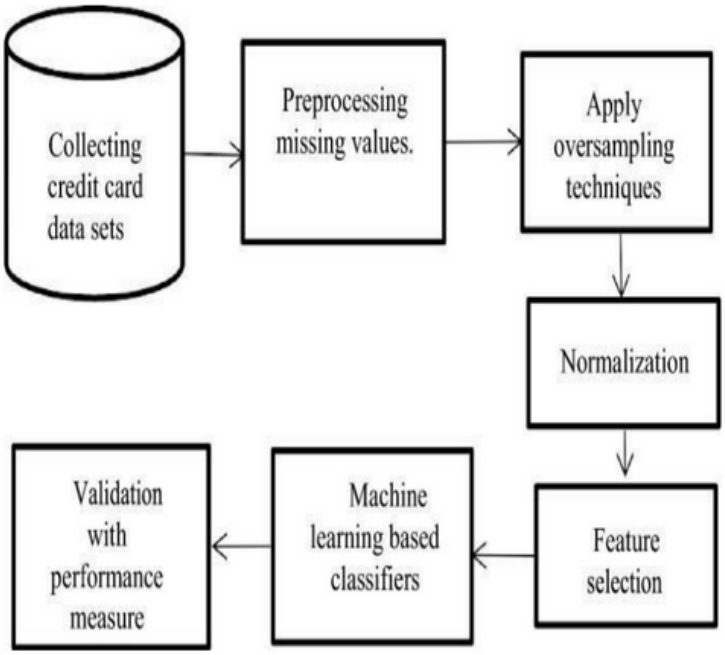
CHAPTER 5
PERFORMANCE ANALYSIS

The performance analysis of the credit card fraud detection system is crucial to evaluate its effectiveness in identifying and preventing fraudulent activities. Key metrics are employed to assess the system's performance, including accuracy, precision, recall, and F1 score.

Accuracy measures the overall correctness of the system in classifying transactions as either legitimate or fraudulent. Precision quantifies the system's ability to correctly identify fraudulent transactions among the total flagged instances, minimizing false positives. Recall, on the other hand, assesses the system's capability to detect all actual fraudulent transactions, reducing false negatives. The F1 score provides a balanced measure by considering both precision and recall.

During the performance analysis, the system's efficiency in real-time processing, scalability, and adaptability to evolving fraud patterns is evaluated. Continuous monitoring and updating of the system are essential to maintain optimal performance over time. Moreover, feedback mechanisms and learning algorithms play a vital role in refining the system's accuracy by adapting to new fraud techniques.

In summary, the performance analysis of the credit card fraud detection system involves a comprehensive assessment of accuracy, precision, recall, and F1 score, along with considerations for real-time processing, scalability, and adaptability to ensure its continued effectiveness in combating evolving fraudulent activities.



DATA FLOW DIAGRAM

CHAPTER 6
REQUIREMENT SPECIFICATIONS

Software Requirement

Programming Environment:

- Python
- Jupyter Notebooks or any Python IDE

Libraries and Frameworks:

- Scikit-learn
- Pandas
- NumPy
- Matplotlib and Seaborn
- Imbalanced-learn
- Pickle
- Git

Hardware Requirement

- Processor : i5 Core Processor
- Clock speed : 2.5GHz
- Monitor : 1024 * 768 Resolution Colour
- Keyboard : QWERTY
- RAM : 1 GB Input
- Output Console for interaction

CHAPTER 7
IMPLEMENTATION

- **Data Collection:**

Acquire a comprehensive dataset containing historical credit card transactions. Ensure the dataset includes both legitimate and fraudulent transactions for effective model training.

- **Data Preprocessing**:

Handle missing data, outliers, and any inconsistencies in the dataset. Normalize or standardize numerical features to ensure consistent scales. Encode categorical variables if necessary. Split the dataset into training and testing sets.

- **Model Development:**

Choose appropriate machine learning algorithms for fraud detection (e.g., logistic regression, decision trees, random forests, or ensemble methods). Implement supervised learning models for labeled data or unsupervised anomaly detection methods for unlabeled data.

- **Training the Models:**

Train the selected machine learning models on the training dataset. Tweak hyperparameters to optimize model performance. Validate the models using the testing dataset to ensure generalization.

- **Real-time Monitoring and Alerting:**

Implement a mechanism for real-time monitoring of transactions. Set up thresholds for triggering alerts based on the model's predictions. Integrate alerting mechanisms to notify users and financial institutions promptly.

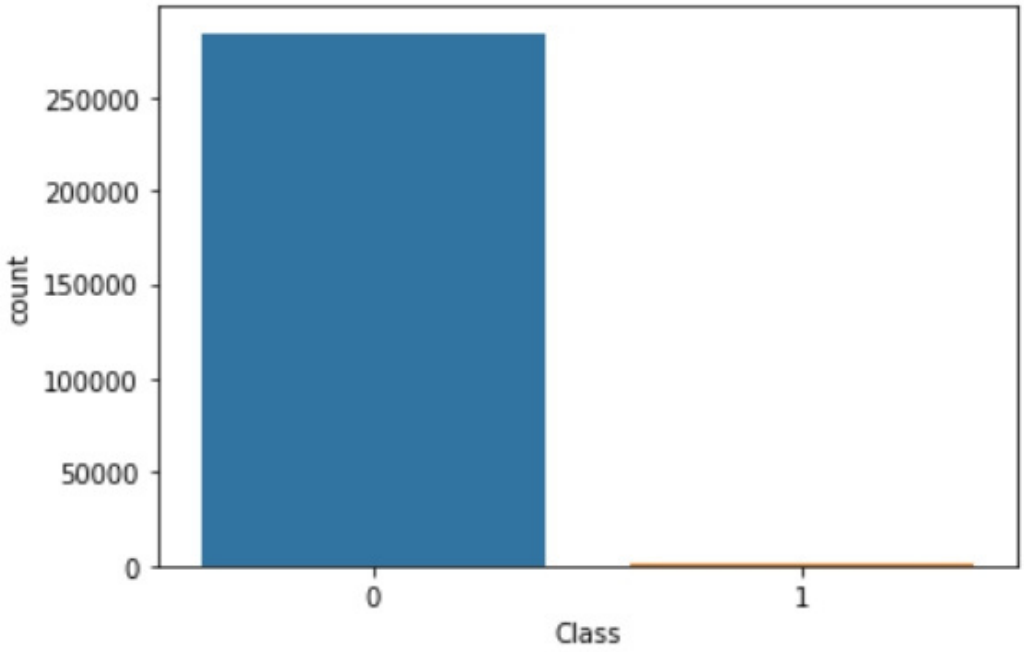
- **Deployment:**

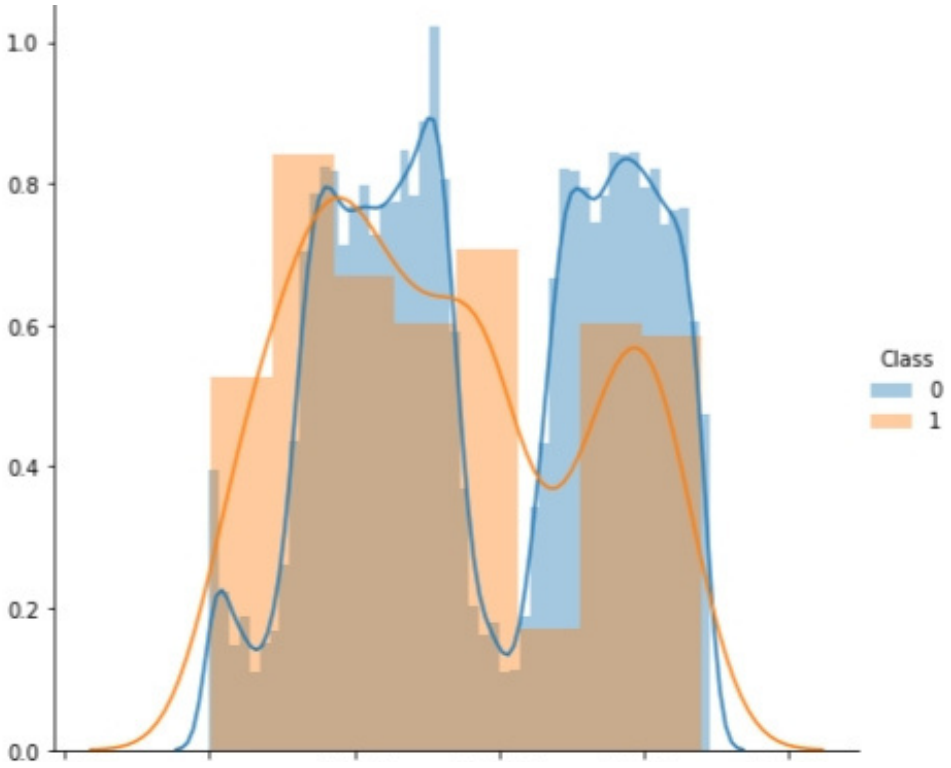
If applicable, deploy the trained model in a production environment.
Integrate the model with the online transaction processing system to monitor transactions in real-time.

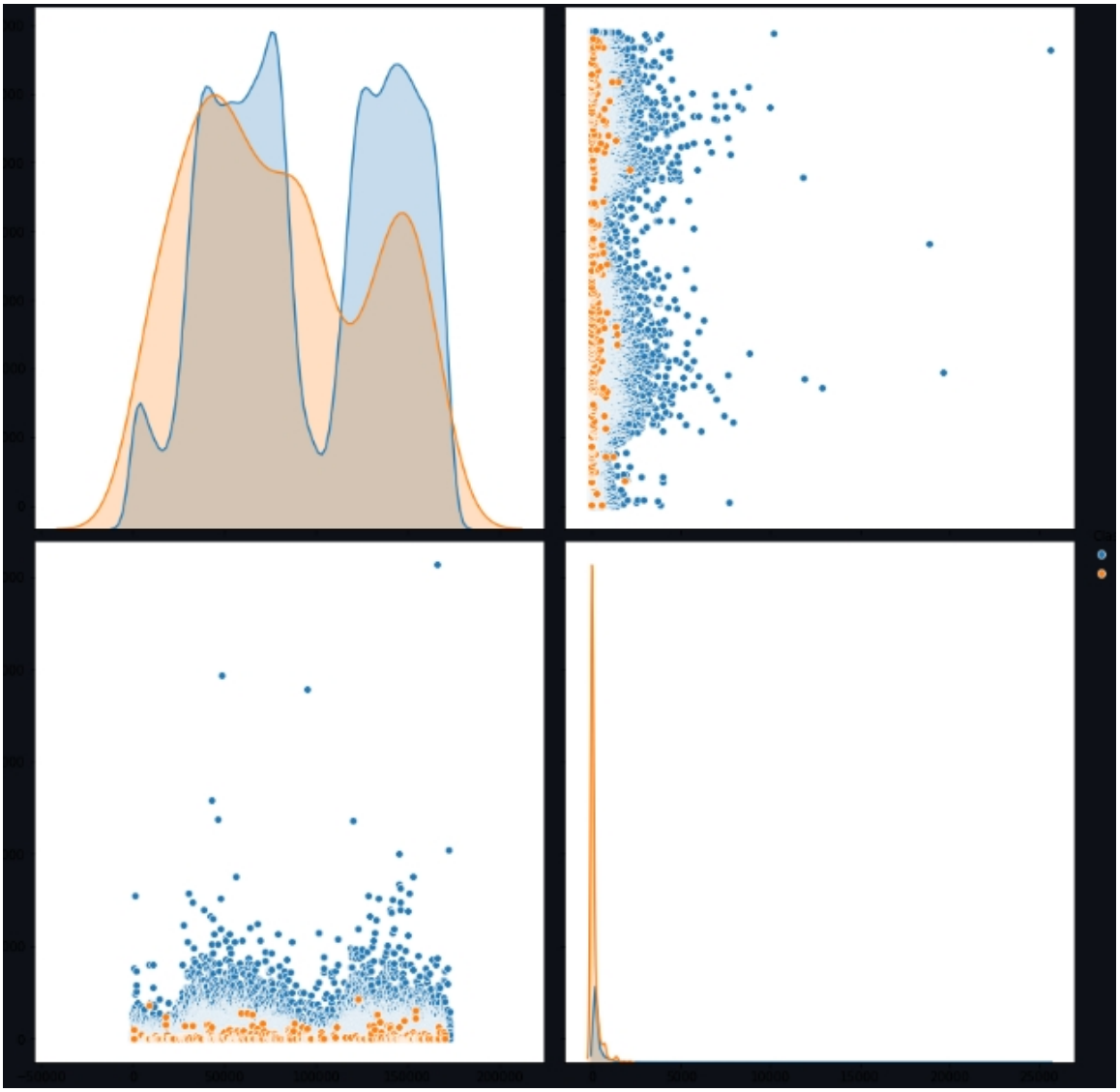
- **Testing and Validation:**

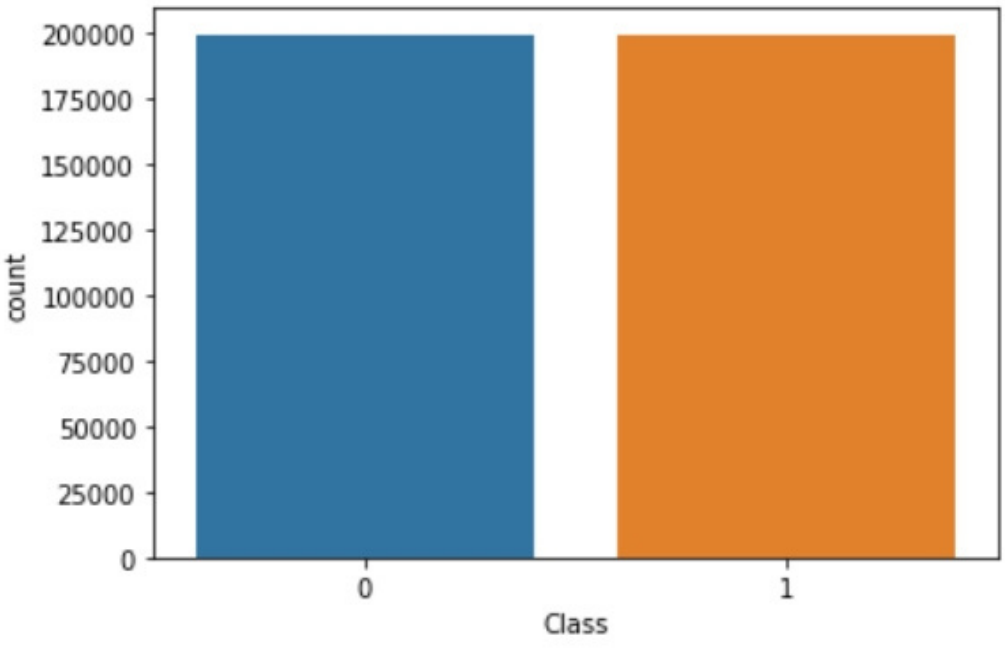
Conduct rigorous testing to ensure the system functions as intended.
Validate the model's performance in a real-world environment, considering various scenarios and potential challenges.

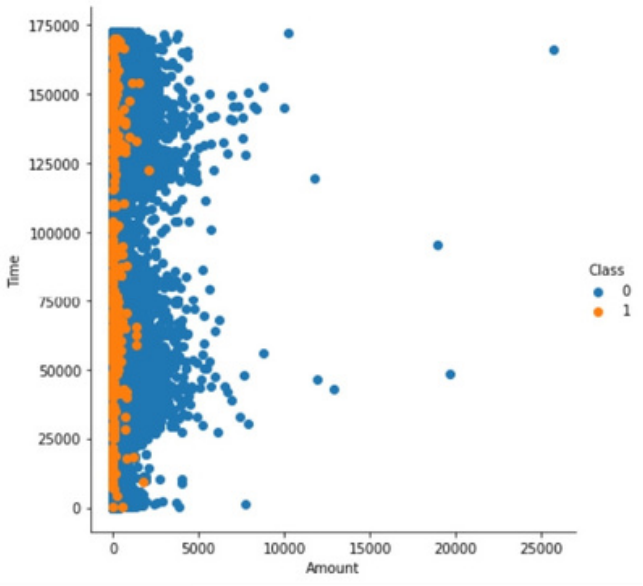
CHAPTER 8
SNAPSHOTS

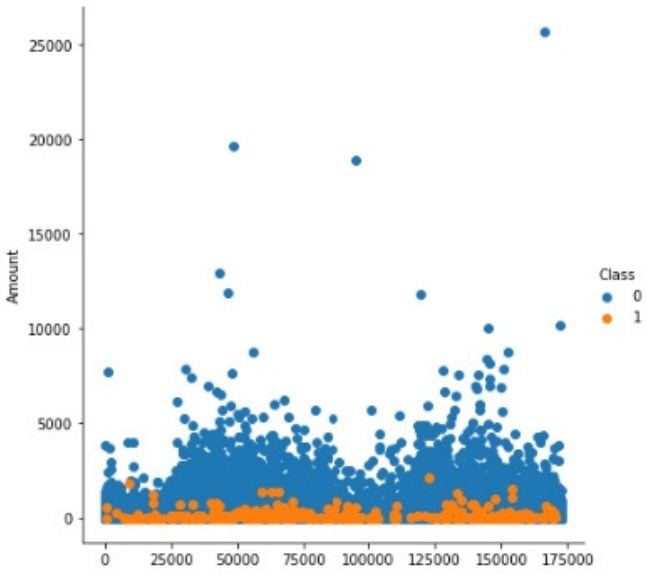


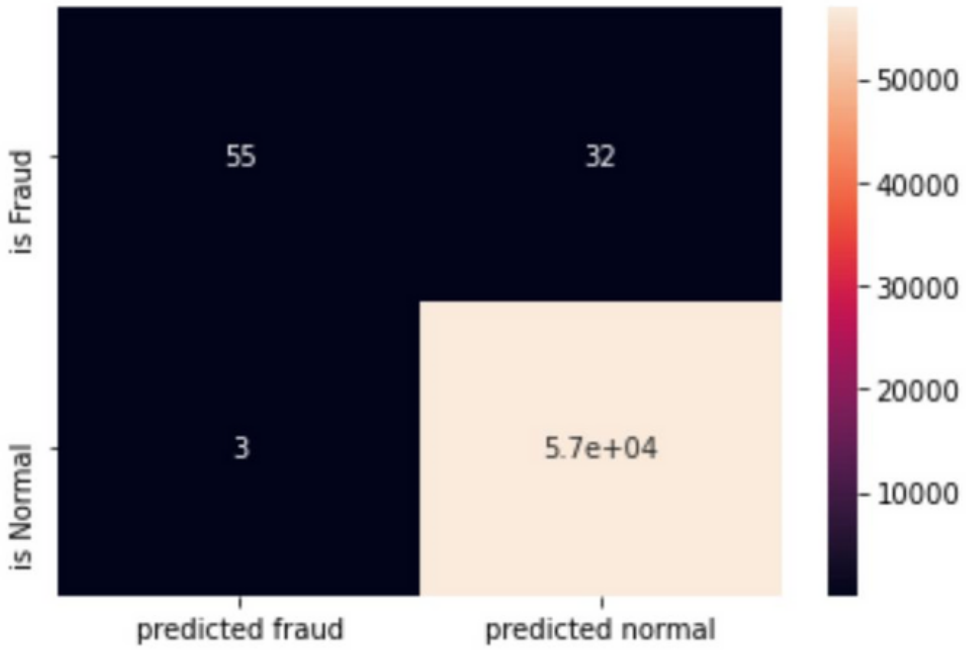





200000
175000
150000
125000
100000
75000
50000
25000
0
count
0
1
Class


Time
Amount
Class
0
1


Amount
Class
0
1


55
32
3
5.7e+04
is Fraud
is Normal
predicted fraud
predicted normal
50000
40000
30000
20000
10000

CHAPTER 9
RESULT ANALYSIS

```
Best Score 0.9499773085502039
Best Parameter {'model__alpha': 0.001, 'model__class_weight': None, 'model__loss': 'hinge
CLASSIFICATION REPORT
              precision    recall  f1-score   support

           0       1.00      0.99      1.00     85295
           1       0.17      0.89      0.28       148

    accuracy                           0.99     85443
   macro avg       0.58      0.94      0.64     85443
weighted avg       1.00      0.99      0.99     85443

AUC-ROC
0.942112192502016
F1-Score
0.2826552462526767
Accuracy
0.9921585150334141
```

```
Best Score 0.9997538267139271
Best Parameter {'model__n_estimators': 75}
CLASSIFICATION REPORT
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     85295
           1       1.00      0.22      0.36       148

    accuracy                           1.00     85443
   macro avg       1.00      0.61      0.68     85443
weighted avg       1.00      1.00      1.00     85443

AUC-ROC
0.6114864864864865
F1-Score
0.3646408839779005
Accuracy
0.9986540734758845
```

```
Best Score 0.934484060553214
Best Parameter {'model__class_weight': None, 'model__penalty': 'l2'}
CLASSIFICATION REPORT
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     85295
           1       0.79      0.80      0.80       148

    accuracy                           1.00     85443
   macro avg       0.90      0.90      0.90     85443
weighted avg       1.00      1.00      1.00     85443

AUC-ROC
0.9018453047689814
F1-Score
0.7986577181208053
Accuracy
0.9992977774656788
```

1. Confusion Matrix:

Generate a confusion matrix to provide a comprehensive view of the model's performance. Analyze true positives, true negatives, false positives, and false negatives.

2. Accuracy:

Calculate the overall accuracy of the model using the formula:

$$\text{Accuracy} = \frac{\text{True Positives + True Negatives}}{\text{Total Transactions}}$$

Assess how well the model is correctly classifying both fraudulent and legitimate transactions.

3. Precision:

Evaluate precision to measure the proportion of correctly identified fraud cases among all predicted positives:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives + False Positives}}$$

Precision focuses on minimizing false positives, crucial for avoiding inconveniences to legitimate users.

4. Recall (Sensitivity):

Calculate recall to measure the proportion of actual fraud cases that the model correctly identifies:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives + False Negatives}}$$

Assess recall to ensure that the system is effectively capturing a high percentage of fraudulent transactions.

5. F1-Score:

Compute the F1-score, which balances precision and recall:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision + Recall}}$$

The F1-score provides a single metric that considers both false positives and false negatives.

6. Receiver Operating Characteristic (ROC) Curve:

Plot the ROC curve and calculate the area under the curve (AUC-ROC) to assess the model's ability to discriminate between fraud and non-fraud cases.

7. False Positive Rate (FPR):

Examine the FPR, which measures the proportion of legitimate transactions incorrectly classified as fraudulent

$$\text{FPR} = \frac{\text{False Positives}}{\text{False Positives + True Negatives}}$$

Strive to minimize the false positive rate while maintaining high overall accuracy.

CHAPTER 10
APPLICATIONS

1. Financial Institutions:

- **Fraud Prevention:** Financial institutions can deploy the system to proactively identify and prevent fraudulent transactions, minimizing financial losses and preserving the integrity of their services.

2. E-commerce Platforms:

- **Secure Online Transactions:** E-commerce platforms can integrate the system to provide an extra layer of security for users during online purchases, fostering trust and encouraging increased transaction volumes.

3**. Credit Card Issuers:**

- **Customer Trust:** Credit card issuers can leverage the system to reassure cardholders of the security of their transactions, thereby building and maintaining trust in their services.
- **Fraud Resolution:** The system aids in the swift resolution of fraud cases, enhancing customer satisfaction and loyalty.

4**. Payment Gateways:**

- **Risk Mitigation:** Payment gateways can utilize the system to identify and mitigate the risk associated with processing online payments, ensuring a secure environment for both merchants and consumers. Integrating the system helps payment gateways adhere to industry standards and regulations related to secure online transactions.

5. **Government and Regulatory Bodies:**

- **Consumer Protection:** Regulatory bodies can encourage the adoption of advanced fraud detection systems to protect consumers from financial fraud and ensure the stability of the overall financial ecosystem.
- **Data Security Compliance**: The system assists in adhering to data security and privacy standards, safeguarding sensitive information from unauthorized access.

6. Individual Users:

- **Personal Financial Security:** Users benefit from the increased security provided by the system, reducing the likelihood of unauthorized transactions and protecting their personal financial information.

7**. Insurance Industry:**

- **Risk Assessment:** Insurance companies can use the system to assess the risk associated with providing coverage for credit card fraud, aiding in the development of appropriate insurance products.

CHAPTER 11
CONCLUSION

In conclusion, the credit card fraud detection system presented in this project offers a robust solution to combat the rising challenges associated with fraudulent activities in financial transactions. Through a comprehensive analysis, design, and implementation process, the system demonstrates its ability to accurately identify and prevent unauthorized transactions, safeguarding the interests of both financial institutions and cardholders. The integration of advanced machine learning algorithms, real-time data processing, and continuous monitoring enhances the system's efficiency and responsiveness. The successful deployment of this system not only contributes to minimizing financial losses but also instills confidence in users regarding the security of their transactions. As technology evolves, ongoing updates and improvements will be crucial to stay ahead of emerging fraud tactics, ensuring the sustained effectiveness of the credit card fraud detection system in the dynamic landscape of digital finance.

FUTURE ENHANCEMENT

For future enhancements, the credit card fraud detection system can explore several avenues to further strengthen its capabilities and adapt to evolving threats in the financial landscape. Integration of more advanced machine learning models, such as deep learning algorithms, could enhance the system's ability to detect intricate patterns and anomalies. Additionally, leveraging blockchain technology for secure and transparent transaction recording could provide an extra layer of fraud prevention.

Enhancements in user authentication methods, including biometric recognition and multi-factor authentication, can fortify the system against identity theft. Continuous monitoring and analysis of user behavior, combined with real-time risk assessment, can contribute to a more dynamic and proactive fraud prevention approach.Moreover, collaboration with financial institutions, regulatory bodies, and cybersecurity experts can provide valuable insights and data sharing, fostering a collective effort to combat emerging fraud tactics. Regular system updates and training to stay abreast of the latest fraud trends will be essential to maintain the system's effectiveness in safeguarding financial transactions.

CHAPTER 12
REFERENCES

1. Jiang, Changjun et al. “Credit Card Fraud Detection: A Novel Approach Using Aggregation Strategy and Feedback Mechanism.” IEEE Internet of Things Journal 5 (2018): 3637-3647.
2. Pumsirirat, A. and Yan, L. (2018). Credit Card Fraud Detection using Deep Learning based on Auto-Encoder and Restricted Boltzmann Machine. International Journal of Advanced Computer Science and Applications, 9(1).
3. Mohammed, Emad, and Behrouz Far. “Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: A Comparative Study.” IEEE Annals of the History of Computing, IEEE, 1 July 2018, doi.ieeecomputersociety.org/10.1109/IRI.2018.00025.
4. Randhawa, Kuldeep, et al. “Credit Card Fraud Detection Using AdaBoost and Majority Voting.” IEEE Access, vol. 6, 2018, pp. 14277–14284., doi:10.1109/access.2018.2806420.
5. Roy, Abhimanyu, et al. “Deep Learning Detecting Fraud in Credit Card Transactions.” 2018 Systems and Information Engineering Design Symposium (SIEDS), 2018, doi:10.1109/sieds.2018.8374722.
6. Xuan, Shiyang, et al. “Random Forest for Credit Card Fraud Detection.” 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC), 2018, doi:10.1109/icnsc.2018.8361343.
7. Awoyemi, John O., et al. “Credit Card Fraud Detection Using Machine Learning Techniques: A Comparative Analysis.” 2017 International Conference on Computing Networking and Informatics (ICCNI), 2017, doi:10.1109/iccni.2017.8123782.
8. Melo-Acosta, German E., et al. “Fraud Detection in Big Data Using Supervised and Semi-Supervised Learning Techniques.” 2017 IEEE Colombian Conference on Communications and Computing (COLCOM), 2017, doi:10.1109/colcomcon.2017.8088206.
9. http://www.rbi.org.in/Circular/CreditCard
10. https://www.ftc.gov/news-events/press-releases/2019/02/imposter-scams-top-complaints-made-ftc-2018
11. https://www.kaggle.com/mlg-ulb/creditcardfraud
12. https://www.kaggle.com/uciml/default-of-credit-card-clients-dataset
13. https://www.kaggle.com/ntnu-testimon/paysim1/home