VISVESVARAYA TECHNOLOGICAL UNIVERSITY

"Jnana Sangama", Belagavi: 590018 , Karnataka , India

A Mini Project Report On

"Credit Card Fraud Detection Using ML Algorithms"

Submitted in partial fulfillment of the requirement for the award of Degree of Bachelor of Engineering in Computer Science and Engineering

Submitted by

SINDHUSHREE H R
(1VE21CS166)

NEELAMPRITANAK
(1VE21CS114)

BHAVANA D K
(1VE22CS401)

DINAKAR
(1VE22CS404)

Under the Guidance of
Mrs. DIVYARAJ G N
Assistant Professor
Department of CSE

Accredited by NAAC & NBA*
SVCE BENGALURU
SRI VENKATESHWARA COLLEGE OF ENGINEERING
— Affiliated to VTU, Approved by AICTE, Recognised by UGC u/s 2(f) & 12(B)—

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SRI VENKATESHWARA COLLEGE OF ENGINEERING

Vidyanagar, Bengaluru, Karnataka, India-562157

2023-2024

Accredited by NBA*

SVCE BENGALURU

SRI VENKATESHWARA COLLEGE OF ENGINEERING

Affiliated to VTU, Approved by AICTE

CERTIFICATE

This is to certify that Data science and Visualization with Mini project work entitled **“CREDIT CARD FRAUD DETECTION USING ML ALGORITHMS”** submitted in partial fulfilment of the requirement for VI semester Bachelor of Engineering in Computer Science and Engineering prescribed by the Visvesvaraya Technological University, Belgaum is a result of the Bonafede work carried out by **SINDHU SHREE H R[1VE21CS166], BHAVANA D K[1VE22CS401], NEELAM PRIYANKA[1VE21CS114], DINAKAR[1VE22CS404]** during the academic year 2023-24. It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the Report deposited in the departmental library. The project report has been approved as it satisfies the academic requirements in respect of Project work prescribed for the said Degree.

....

Signature of Guide
Mrs. Divyaraj G N
Asst. Prof, Dept of CSE,
SVCE, Bengaluru.

....

Signature of the HOD
Dr. Hema MS
Professor and HOD,
Dept of CSE, SVCE,
Bengaluru.

ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without complementing those who made it possible, whose guidance and encouragement made our efforts successful.

My sincere thanks to highly esteemed institution **SRI VENKATESHWARA COLLEGE OF ENGINEERING** for grooming me to be a Software Engineer.

We express our sincere gratitude to **Dr. NAGESWARA GUPTHA M**, Principal , SVCE , Bengaluru for providing required faculty

We would like to extend our sincere thanks to **Dr. HEMA M S**, HOD, Dept. of CSE, SVCE, Bengaluru for providing support and encouragement

We would like to express our sincere thanks to **Mrs. DIVYARAJ G N**, Asst. Prof, Dept. of CSE, SVCE, Bengaluru, for guidance and support in bringing this project to completion.

We am thankful to one and all who have been involved in this work directly or indirectly for the successful completion of this project.

Finally We am grateful to my parents and friends for their invaluable support, guidance and encouragement.

SINDHUSHREE H R (1VE21CS166)
BHAVANA D K (1VE22CS401)
NEELAM PRIYANAK (1VE21CS114)
DINAKAR (1VE22CS404)

DEPARTMENT VISION

Global Excellence with Local relevance in Information Science and Engineering Education, Research and Development

DEPARTMNET MISION

M1. Strive for academic excellence in Information Science and Engineering through student centric innovative teaching-learning process, competent faculty members, efficient assessment and use of ICT.

M2. Establish Centre for Excellence in various vertical of Information Science and Engineering to promote collaborative research and Industry Institute Interaction.

M3. Transform the engineering aspirants to socially responsible, ethical, technically competent and value added professional or entrepreneur.

PROGRAM OUTCOMES

1. Engineering Knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

2. Problem Analysis: Identify, formulate, research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

3. Design/development of Solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

4. Conduct Investigations of Complex Problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

5. Modern Tool usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modelling to complex engineering activities with an understanding of the limitations.

6. The Engineer and Society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

7. Environment and Sustainability: Understand the impact of the professional

engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

8. Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

9. Individual and Team Work: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

10. Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

11. Project Management and Finance: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

12. Life-long Learning: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

PROGRAM EDUCATIONAL OBJECTIVES

Knowledge:

Computer Science and Engineering Graduates will have professional technical career in inter disciplinary domains providing innovative and sustainable solutions using modern tools.

Skills:

Computer Science and Engineering Graduates will have effective communication, leadership, team building, problem solving, decision making and creative skills.

Attitude:

Computer Science and Engineering Graduates will practice ethical responsibilities towards their peers, employers and society.

PROGRAM SPECIFIC OUTCOMES

PSO1:

Ability to adopt quickly for any domain, interact with diverse group of individuals and be an entrepreneur in a societal and global setting.

PSO2:

Ability to visualize the operations of existing and future software Applications.

ABSTRACT

It examines the increasing problem on credit card(CD) fraud as evolving within the rapidly developing world of electronic commerce. With credit cards becoming more widespread as a payment method, the no of false transactions are also increasing at an alarming rate across the world and leading to significant monetary loss. In the face of such challenge, a no of sophisticated approaches rooted in artificial intelligence, It gives an overview of the techniques, stressing the necessity for effective fraud detection systems to protect damage from credit card issued by banks. Machine learning(ML) with approaches that includes DT, LR, RF, Genetic Algorithms and ensemble techniques is then considered to demonstrate the efficacy of finding out genuine transactions from all fraudulent ones. Real-world datasets are utilized in the research to show how competitive this work against existing systems. This last section concludes on the ever evolving fraud detection ideas and a broad understanding of different counter–measures across domains with cases in credit card.

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

Credit card technology, first developed by Bank of America in 1958, has evolved significantly, and with it, the sophistication of fraudulent activities. By 2020, there were 1.4 million allegations of identity theft, with 393,207 cases involving Credit Conversion Factor (CCF) fraud, making it the second most prevalent form of identity theft. The CCF is a critical measure used in banking and finance to assess the credit risk associated with various financial instruments and is pivotal in calculating credit exposure for regulatory purposes.

The Annual Fraud Report on United Kingdom Finance for 2022 highlighted that over £1.2 billion was stolen through both authorized and unauthorized criminal activity, equating to £2,300 lost every minute. Notably, 18% of Authorized Fraud involving Push Payment (APP) incidents occurred via phone lines, while 78% originated online. Credit card fraud remains a significant global issue, with 45,120 scam cases reported in 2020 alone.

The past decade has witnessed exponential growth in internet usage, leading to increased adoption of electronic bill payments, tap-and-pay services, and e-commerce. Consequently, fraudsters have intensified their efforts to exploit credit card transactions. Fraud detection now involves monitoring user activities to predict, understand, and prevent illicit behavior, such as fraud, intrusion, and defaults. However, the challenge lies in distinguishing a small number of fraudulent transactions from a vast number of legitimate ones, as transaction patterns can evolve over time.

1.1. BACKGROUND

Credit card fraud has become increasingly prevalent with the rise of online shopping, electronic payments, and global financial transactions. Fraudulent activities can take various forms, such as unauthorized transactions, identity theft, and account takeovers. As digital payment systems continue to expand, so does the complexity and frequency of credit card fraud. Financial institutions, e-commerce platforms, and payment processors face significant challenges in detecting and preventing fraud while maintaining a seamless user experience. The development and deployment of effective fraud detection systems are essential to mitigate financial losses, protect consumers, and maintain trust in the financial ecosystem.

Advanced fraud detection systems leverage a combination of rule-based approaches, statistical methods, and machine learning techniques to identify and combat fraudulent activities. Rule-based systems use predefined criteria to flag suspicious transactions, while machine learning models analyze large datasets to uncover patterns indicative of fraud. These systems must continually evolve to adapt to new fraud tactics, ensuring they remain effective against increasingly sophisticated schemes. Collaboration between financial institutions, technology providers, and regulatory bodies is crucial in developing comprehensive strategies to detect and prevent fraud. This collaborative approach not only enhances the accuracy and efficiency of fraud detection but also fosters a unified effort in safeguarding the financial ecosystem against the growing threat of credit card fraud.

1.2. PURPOSE

The primary purpose of credit card fraud detection is to identify and prevent unauthorized and fraudulent transactions in real-time. This involves:

- **Protecting Consumers:** Ensuring cardholders are protected from fraud is paramount. Effective fraud detection systems help in promptly identifying and halting unauthorized transactions, thereby safeguarding consumers' financial assets and personal information. This protection extends to preventing identity theft, where fraudsters use stolen personal details to carry out fraudulent activities.
- **Reducing Financial Losses:** Fraudulent activities lead to significant monetary losses for banks, merchants, and payment processors. By detecting and preventing fraud in real-time, these financial entities can minimize their losses. This involves not only stopping unauthorized transactions but also recovering funds when possible and preventing future fraud attempts.
- **Maintaining Trust:** Customer confidence in the security of credit card transactions is vital for the sustained growth of digital payment systems. When consumers feel secure, they are more likely to engage in online shopping and use electronic payments. Maintaining trust involves ensuring that legitimate transactions are processed smoothly while effectively blocking fraudulent ones without causing inconvenience to the cardholder.
- **Compliance:** Financial institutions and payment processors must adhere to various regulatory requirements and industry standards designed to prevent financial crimes and protect consumers. Compliance with regulations such as the Payment Card Industry Data Security Standard (PCI DSS) and anti-money laundering (AML) laws is essentialSCOPE

1.3. SCOPE

The scope of credit card fraud detection involves various techniques and challenges essential for identifying and preventing fraud. Detection methods include rule-based systems with predefined thresholds and machine learning models that analyze historical data for fraud patterns. Hybrid approaches combine both to enhance accuracy. The types of fraud addressed are Card-Not-Present (CNP) fraud, card-present fraud, account takeover, and application fraud. Challenges include managing imbalanced data, ensuring real-time processing, adapting to evolving fraud tactics, and minimizing false positives to maintain a seamless user experience. Key evaluation metrics are accuracy, precision, recall, and the F1 score, which collectively measure the system's performance in detecting fraud while balancing false positives and false negatives. This comprehensive scope aims to protect consumers, reduce financial losses, and maintain trust in the financial system.

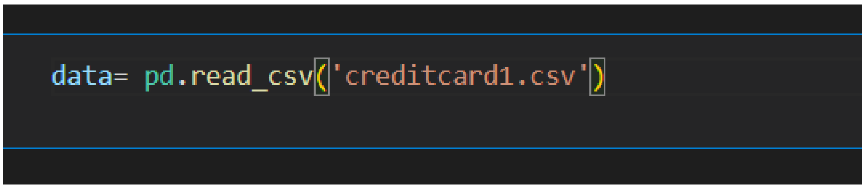
Chapter 2

METHODOLOGY

The methodology for credit card fraud detection using machine learning algorithms involves several key steps: data collection, preprocessing, model selection, training, evaluation, and deployment. This section details the application of logistic regression, random forest, and decision tree algorithms.

2.1. Data Collection

The dataset used for this analysis was obtained from Kaggle and includes credit card transactions performed by cardholders throughout Europe in September 2013. The dataset contains features such as transaction amount, time, and anonymized features (V1 to V28) resulting from a PCA transformation, along with a label indicating whether a transaction is fraudulent.



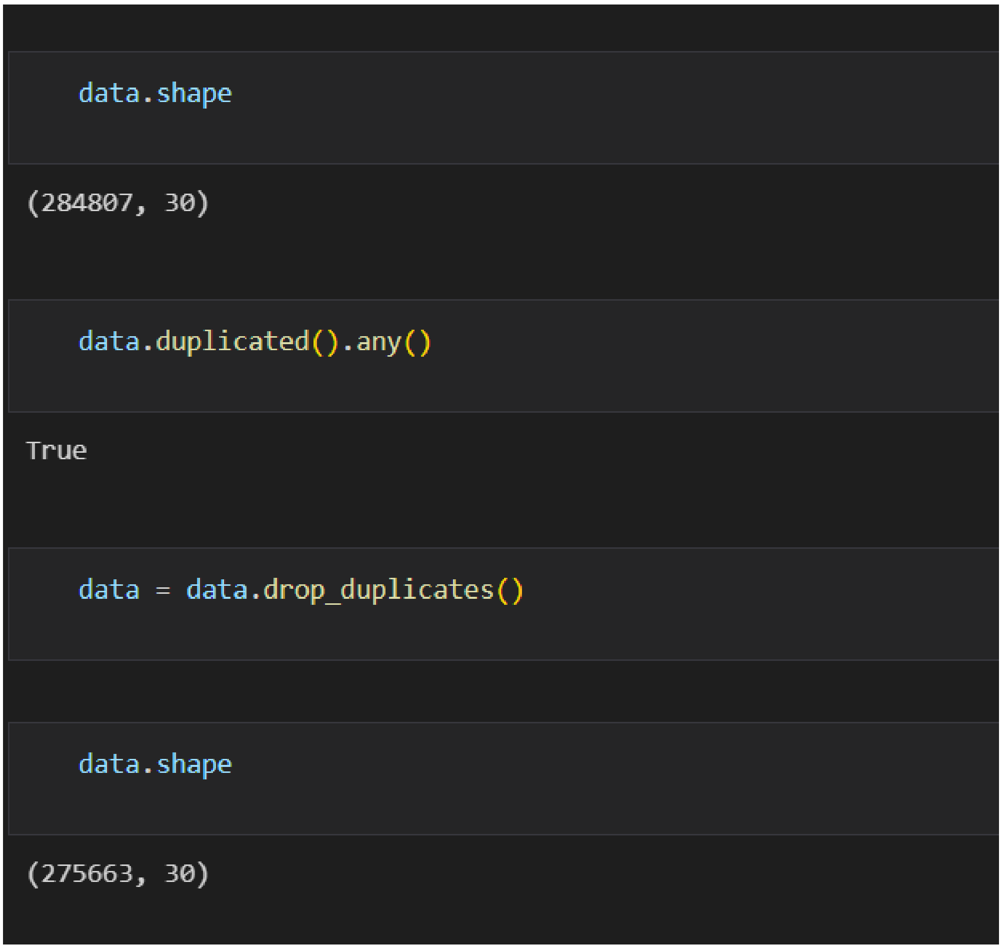
```
data= pd.read_csv('creditcard1.csv')
```

2.2. Data Preprocessing

Data preprocessing is crucial to ensure the quality and consistency of the dataset. This includes:

- **Handling Missing Values:** Imputing or removing missing data to maintain dataset integrity.
- **Feature Scaling:** Normalizing numerical features to ensure they contribute equally to the model.
- **Encoding Categorical Variables:** Converting categorical data into numerical format using techniques such as one-hot encoding.

- **Balancing the Dataset:** Addressing the imbalance between fraudulent and non-fraudulent transactions using techniques like oversampling, undersampling, or synthetic data generation (e.g., SMOTE).



```
data.shape
```

```
(284807, 30)
```

```
data.duplicated().any()
```

```
True
```

```
data = data.drop_duplicates()
```

```
data.shape
```

```
(275663, 30)
```

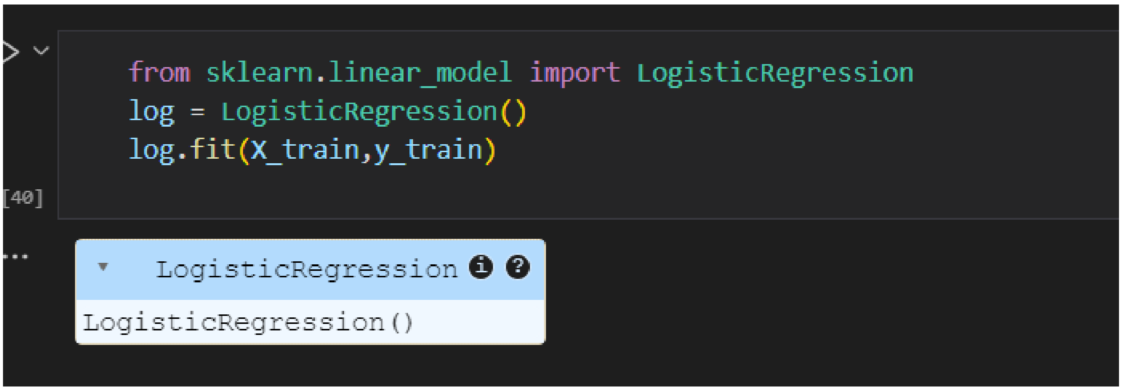
2.3. Model Selection

The following machine learning algorithms are selected for fraud detection:

- **Logistic Regression:** A linear model that predicts the probability of a transaction being fraudulent based on the input features. It is simple, interpretable, and effective for binary classification problems.

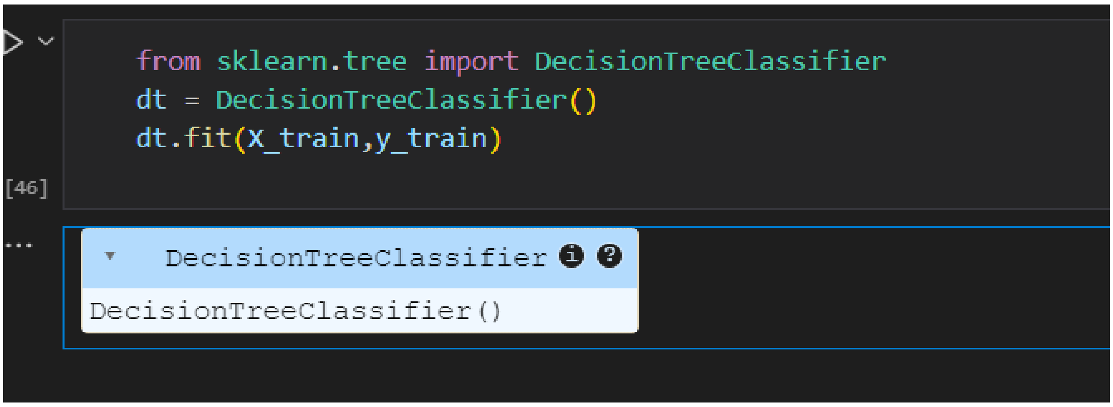
- **Decision Tree:** A non-linear model that splits the data into branches based on feature values, leading to a decision on whether a transaction is fraudulent. It is easy to visualize and interpret.

- **Random Forest:** An ensemble method that builds multiple decision trees and combines their outputs to improve prediction accuracy and reduce overfitting. It is robust and handles large datasets with high dimensionality well.



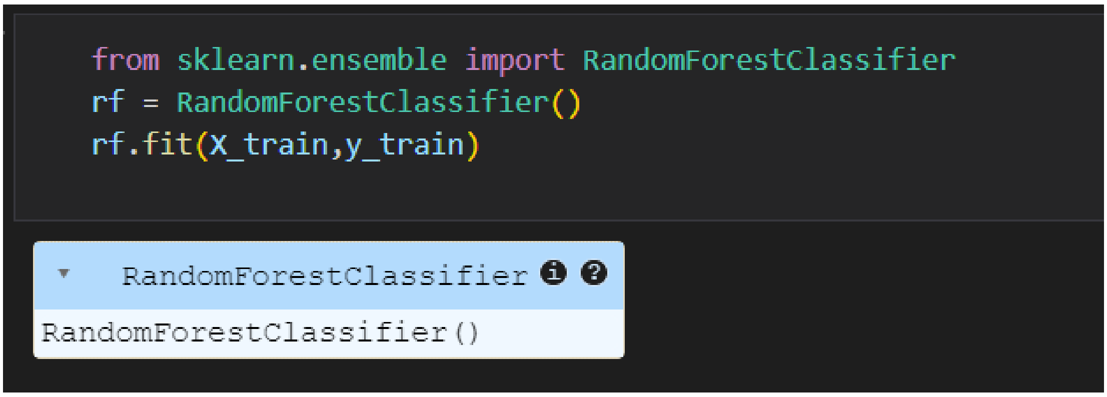
```
from sklearn.linear_model import LogisticRegression
log = LogisticRegression()
log.fit(X_train,y_train)
```

```
LogisticRegression
LogisticRegression()
```



```
from sklearn.tree import DecisionTreeClassifier
dt = DecisionTreeClassifier()
dt.fit(X_train,y_train)
```

```
DecisionTreeClassifier
DecisionTreeClassifier()
```



```
from sklearn.ensemble import RandomForestClassifier
rf = RandomForestClassifier()
rf.fit(X_train,y_train)
```

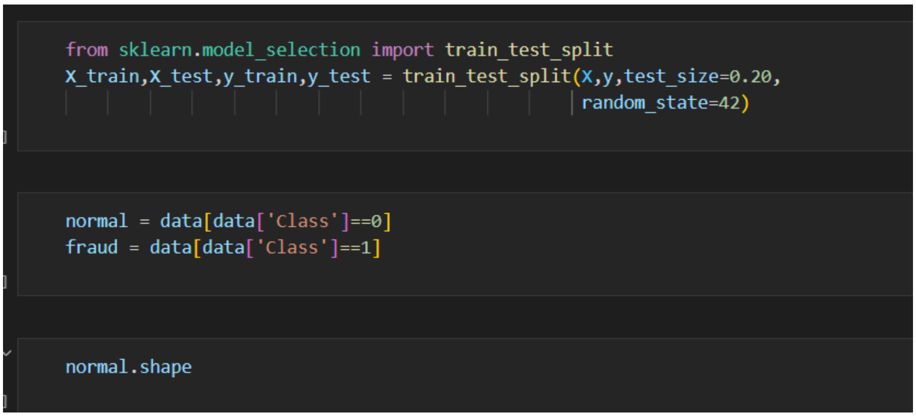
```
RandomForestClassifier
RandomForestClassifier()
```

2.4. Model Training

The models are trained on the preprocessed dataset:

- **Train-Test Split:** Dividing the dataset into training and testing subsets (e.g., 70% training, 30% testing) to evaluate model performance.

- **Training the Models:** Using the training data to fit the logistic regression, decision tree, and random forest models. Hyperparameters are tuned using techniques such as cross-validation.



```
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test = train_test_split(X,y,test_size=0.20,
                                                 random_state=42)
```

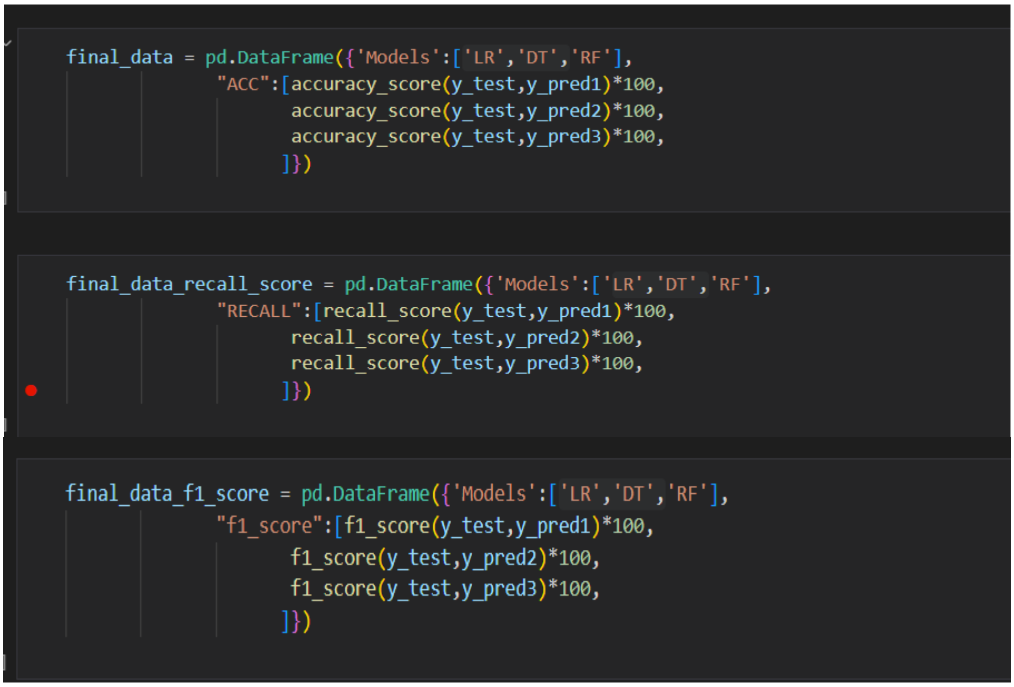
```
normal = data[data['Class']==0]
fraud = data[data['Class']==1]
```

```
normal.shape
```

2.5. Model Evaluation

The models are evaluated using the testing data with the following metrics:

- **Accuracy:** The proportion of correctly identified transactions (both fraudulent and legitimate).
- **Precision:** The proportion of true positive frauds among all transactions flagged as fraud.
- **Recall (Sensitivity):** The proportion of actual frauds correctly identified.
- **F1 Score:** The harmonic mean of precision and recall, providing a balanced measure.
- **Area Under the Curve (AUC):** The area under the ROC curve, indicating the model's ability to distinguish between fraudulent and legitimate transactions.



```
final_data = pd.DataFrame({'Models':['LR','DT','RF'],
                    "ACC":[accuracy_score(y_test,y_pred1)*100,
                          accuracy_score(y_test,y_pred2)*100,
                          accuracy_score(y_test,y_pred3)*100,
                         ]})
```

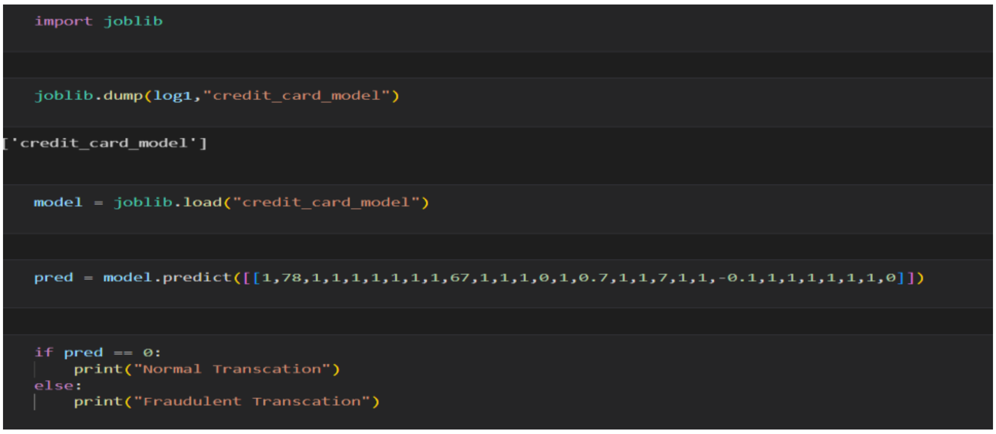
```
final_data_recall_score = pd.DataFrame({'Models':['LR','DT','RF'],
                    "RECALL":[recall_score(y_test,y_pred1)*100,
                          recall_score(y_test,y_pred2)*100,
                          recall_score(y_test,y_pred3)*100,
                         ]})
```

```
final_data_f1_score = pd.DataFrame({'Models':['LR','DT','RF'],
                    "f1_score":[f1_score(y_test,y_pred1)*100,
                          f1_score(y_test,y_pred2)*100,
                          f1_score(y_test,y_pred3)*100,
                         ]})
```

2.6. Model Deployment

The best-performing model is deployed into the fraud detection system:

- **Real-Time Processing:** Integrating the model with transaction processing systems to detect fraud in real-time.
- **Continuous Monitoring:** Regularly updating the model with new transaction data to maintain its effectiveness against evolving fraud tactics.
- **Feedback Loop:** Incorporating feedback from flagged transactions to refine and improve the model over time.



```
import joblib
```

```
joblib.dump(log1,"credit_card_model")
```

```
['credit_card_model']
```

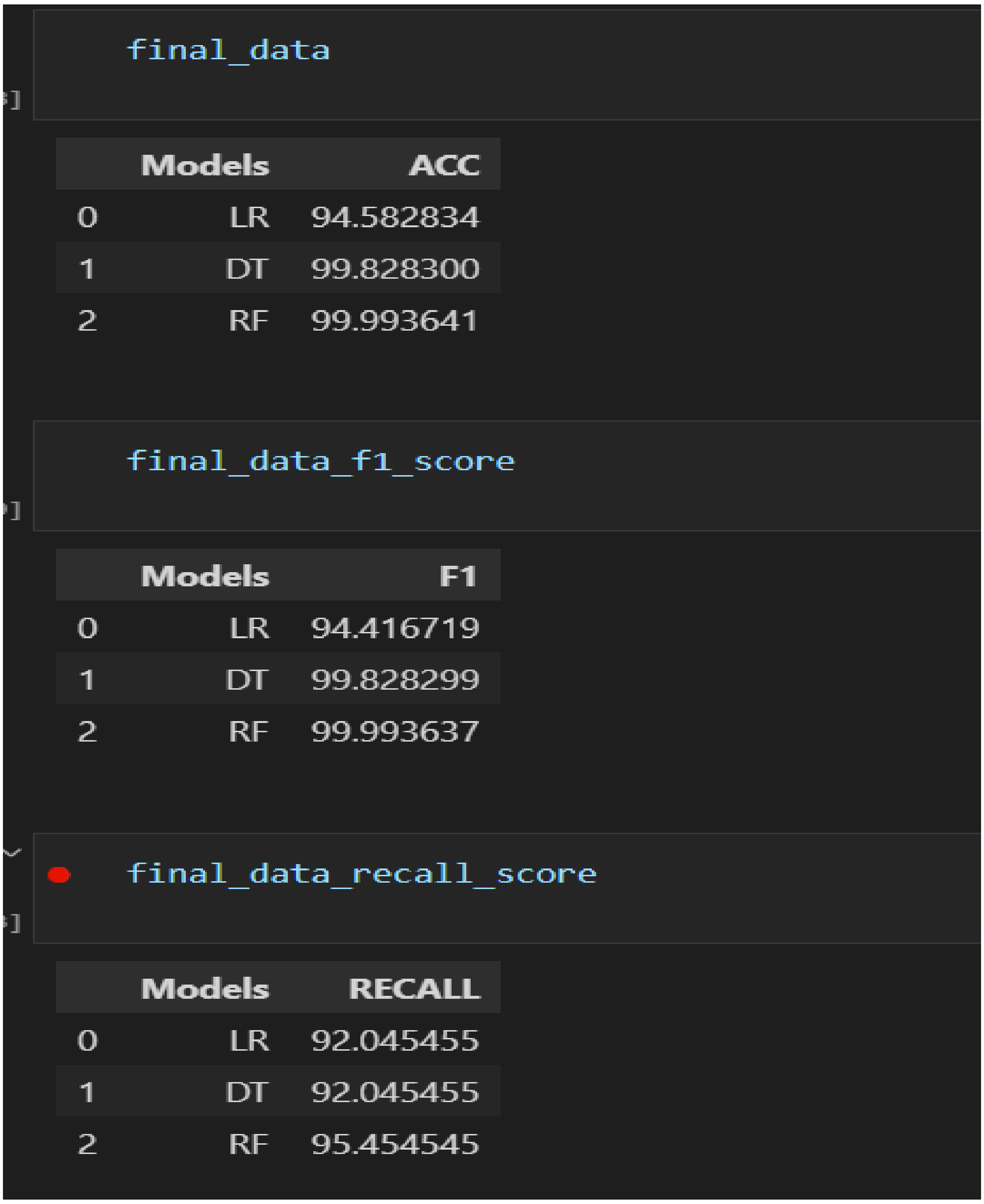
```
model = joblib.load("credit_card_model")
```

```
pred = model.predict([[1,78,1,1,1,1,1,1,1,67,1,1,1,0,1,0.7,1,1,7,1,1,-0.1,1,1,1,1,1,1,0]])
```

```
if pred == 0:
    print("Normal Transcation")
else:
    print("Fraudulent Transcation")
```

Chapter 3

RESULT



```
final_data
```

	Models	ACC
0	LR	94.582834
1	DT	99.828300
2	RF	99.993641

```
final_data_f1_score
```

	Models	F1
0	LR	94.416719
1	DT	99.828299
2	RF	99.993637

```
final_data_recall_score
```

	Models	RECALL
0	LR	92.045455
1	DT	92.045455
2	RF	95.454545

CONCLUSION

The significance of applying techniques for machine learning to improve security measures in financial transactions is highlighted by the study on identifying creditcard theft. The review of several algorithms, such as Random Forest, decision trees, and LR, emphasizes how crucial it is to choose the right model for fraud_ detection. The evaluation's measurements of accuracy, recall, and precision demonstrate how well these models distinguish between authentic and fraudulent transactions.. The results highlight the necessity of an all-encompassing strategy that takes into account variables like dataset imbalance, algorithm efficiency, When selecting a machine_learning model for creditcard fraud_detection . We determined the precision, recall (sensitivity), F1-score, and other performance indicators for each method, which allowed us choose the best one out of the few that were employed. Among the algorithms for machine learning, random forest produced the best results. Having a 100% recall score, 99% f1 score, and 99% accuracy.

FUTURE ENHANCEMENT

Future enhancements in credit card fraud detection will include advanced machine learning techniques like deep learning and ensemble methods to improve accuracy and robustness. Real-time processing frameworks and scalable cloud platforms will handle large transaction volumes quickly. Anomaly detection with unsupervised learning and behavioural analysis will identify new fraud patterns. Building interpretable models will ensure regulatory compliance and improve trust. Data augmentation and additional data sources will balance and enrich transaction data. Online learning and continuous feedback will keep models updated with new fraud tactics. Collaboration in data sharing and developing collective defences will enhance detection. Multi-factor and biometric authentication will add security, while privacy regulations and ethical AI practices will protect user privacy and ensure fairness. These enhancements will help secure financial transactions against evolving fraud tactics.

REFERENCES

[1] Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. “A comprehensive study of credit card fraud detection based on machine learning”. Published - 2011

[2] Dal Pozzolo, A., Caelen, O., Johnson, R. A., & Bontempi, G. “Machine Learning Techniques for Credit Card Fraud Detection”, (2015). IEEE Symposium Series on Computational Intelligence.

[3] Whitrow C, Hand, D. J, Juszczak, P, Weston, D, & Adams, N. M, “Evaluating the Performance of Different Machine Learning Algorithms for Credit Card Fraud Detection”, Published – 2009 , Data Mining and Knowledge Discovery, 18(1), 30-55.

[4] Patil, S. V., & Kulkarni, R. V. (2016). "Credit card fraud detection using decision tree induction algorithm." International Journal of Innovative Research in Computer and Communication Engineering, 4(5), 944-948.

[5] Abdallah, A., Maarof, M. A., & Zainal, A. “A Survey on Credit Card Fraud Detection Using Machine Learning” (2016). Journal of Network and Computer Applications, 68, 90-113