VISHVESVARAYA TECHNOLOGICAL UNIVERSITY-BELAGAVI

A

MINI PROJECT REPORT ON

"Credit card fraud detection"

Submitted in partial fulfillment of the requirement for Fifth Semester, B.E., CS&E.

Submitted by:

Vaishnavi Ekanathraddy kolli	**3PG22CS120**
Sushma Deyannavar	**3PG22CS110**
Sushmita	**3PG22CS111**
Gowthami S	**3PG23CS402**

MINI PROJECT CO-ORDINATOR

Dr. Aruna Kumara B

M. Tech., Ph. D

UNDER THE GUIDANCE OF

Mr. Roopanand M K

M. Tech., (Ph. D)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Ballari V.V. Sangha's

PROUDHADEVARAYA INSTITUTE OF TECHNOLOGY,

T.B.DAM, HOSAPETE - 583225

(Affiliated to VTU, Belagavi, Karnataka & Recognized by AICTE, New Delhi)

VISHVESVARAYA TECHNOLOGICAL UNIVERSITY-BELAGAVI

A

MINI PROJECT REPORT ON

"Credit card fraud detection"

Submitted in partial fulfillment of the requirement for Fifth Semester, B.E., CS&E.

Submitted by:

Vaishnavi Ekanathraddy Kolli **3PG22CS120**

MINI PROJECT CO-ORDINATOR

Dr. Aruna Kumara B

M. Tech., Ph. D

UNDER THE GUIDANCE OF

Mr. Roopanand M K

M. Tech., (Ph. D)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Ballari V.V. Sangha's

PROUDHADEVARAYA INSTITUTE OF TECHNOLOGY,

T.B.DAM, HOSAPETE - 583225

(Affiliated to VTU, Belagavi, Karnataka & Recognized by AICTE, New Delhi)

PROUDHADEVARAYA INSTITUTE OF TECHNOLOGY T.B.DAM,

HOSAPETE – 583225

(Affiliated to VISVESVARAYA TECHNOLOGICAL UNIVERSITY, BELAGAVI, KARNATAKA)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that Miss. Vaishnavi Ekanathraddy Kolli (3PG22CS120) Studying in 5th semester B.E., CS&E Department, has successfully completed Mini project entitled *"Credit card fraud detection"* towards the partial fulfillment for the award of the B.E in Computer Science and Engineering under Visvesvaraya Technology University, Belagavi, Karnataka during the academic year 2024-2025.

Mr. Roopanand M K M. Tech (Ph. D) Guide	Dr. Aruna Kumara B M. Tech, Ph. D Co-Ordinator	Dr. Parvati Kadli M. Tech, Ph. D HOD, Dept of CSE

ACKNOWLEDGEMENT

As we have started this mini project with the sense of learning something on the topic **" Credit card fraud detection ".**

The euphoria and satisfaction that accompany the successful completion of any task would be in complete without the mention of the people who made it possible whose constant guidance and encouragement crowned our efforts with success.

First of all, we thank Visvesvaraya Technological University, Belagavi for providing this opportunity to implement this mini project.
We consider our self proud to be a part of "P.D.I.T" family, the institution which stood by my way in all endeavors.

We extremely grateful to **Sri. Karibasavaraj Badami**, Chairman, PDIT, for providing all the required resources for the successful completion of the project. We also take this opportunity to express our gratitude to **Dr. Rohitha U M**, Principle of PDIT, for providing us environment to do the project. We sincerely thank **Dr. Parvati Kadli**, HOD Dept. of CSE, PDIT, without whose moral support this mini project would not have been successful.

We would like to express our heartfelt gratitude towards our mini project coordinator **Dr. Aruna Kumara B** and project guide **Mr. Roopanand M K** whose firm belief in my capabilities and moral support to bring my potential to the fore front has played a major role in accomplishment of this mini project.

We would like to mention a thanks to all our Teaching and Non-Teaching staff members of Computer Science department PDIT, for their valuable support and guidance. Last but not the least, We would like to thank our Parents, Relatives and Friends for their support and suggestions without which we couldn't have achieved this success.

-Vaishnavi E Kolli
- Sushma Deyannavar
- Sushmita
- Gowthami S

ACKNOWLEDGEMENT

As we have started this mini project with the sense of learning something on the topic **" Credit card fraud detection** ".

The euphoria and satisfaction that accompany the successful completion of any task would be in complete without the mention of the people who made it possible whose constant guidance and encouragement crowned our efforts with success.

First of all, we thank Visvesvaraya Technological University, Belagavi for providing this opportunity to implement this mini project.
We consider our self proud to be a part of "P.D.I.T" family, the institution which stood by my way in all endeavors.

We extremely grateful to **Sri. Karibasavaraj Badami**, Chairman, PDIT, for providing all the required resources for the successful completion of the project. We also take this opportunity to express our gratitude to **Dr. Rohitha U M**, Principle of PDIT, for providing us environment to do the project. We sincerely thank **Dr. Parvati Kadli**, HOD Dept. of CSE, PDIT, without whose moral support this mini project would not have been successful.

We would like to express our heartfelt gratitude towards our mini project coordinator **Dr. Aruna Kumara B** and project guide **Mr. Roopanand M K** whose firm belief in my capabilities and moral support to bring my potential to the fore front has played a major role in accomplishment of this mini project.

We would like to mention a thanks to all our Teaching and Non-Teaching staff members of Computer Science department PDIT, for their valuable support and guidance. Last but not the least, We would like to thank our Parents, Relatives and Friends for their support and suggestions without which we couldn't have achieved this success.

-Vaishnavi E Kolli

ABOUT THE INSTITUTION

Proudhadevaraya Institute of Technology, (PDIT) was established in 1997. The institution is affiliated to Visvesvaraya Technological University, Belagavi, Karnataka and recognized by AICTE, New Delhi, India (F. No. 770-53-221(et)/96, dated: 04.08.1997). PDIT is situated in a sprawling 20 acres of land, quite adjacent to T.B. Dam and very near to famous world heritage historical place Hampi. The campus homed in a beautiful spacious building with calm, decent and serene environment and is congenial for the pursuit of knowledge. The institution has well furnished and spacious lecture halls, laboratories, workshops, most modernized and sophisticated computer labs, library and administrative blocks. College presently offers UG in 5 engineering courses and PG in M. Tech and MBA course.

About the department:

The Department of Computer Science & Engineering was started in the year 1997, with the aim of not merely seeking to turn students as Engineers but also with high characters, probity and Honor.

The department has grown in strength over the years in terms of Infrastructure facilities, technical expertise, Faculty Strength, Modern Teaching Aids, Browsing facilities etc. The department has qualified and experienced faculty and technically competent supporting staff on its rolls.

The department continuously organizes educational programs like faculty development programs, Seminars, Workshops, Conferences & State level techno-cultural fest “WONDERS” every year to enrich the innovative ideas of our students which is backed by our committed staff and students. The students of the department have won many laurels including Ranks at university level many prizes at state level and national level for paper presentations, Project Exhibitions & other competitions. One of our Projects has won First Prize at International Level Project Competition. It is the credit to the department that student’s projects are continuously sponsored by KSCST. The alumni of the department are placed in almost all leading IT companies.

The student’s association of our department Systricks C-Sea organizes many technical eventswhich help the students adapt to the changing technological and industrial scenario.

The aim of the department is to produce IT quality professionals and computer engineers to fulfil the demand of skilled talents in IT sector. The department comprises sophisticated laboratories with latest configurations.

The department offers wide range of Practical training and exposure to both software and hardware that have industry and trade applicable. The department frequently conducts Technical, cultural fests to expose the hidden talents.

The Computer Science staff are actively involved in overall administration and Maintenance of our college. (Computer network, Internet, WIFI, Website updation, Alumni association, Ladies Hostel, online examinations, Prospectus, Brochure, Handouts, Identity Cards, media advertisements, University related coordination works & Many colleges development related.

Head of the Department :Prof. Parvathi Kadli B.E, M. Tech., Ph.D.

The aim of the CSE Department has Global Excellence in teaching, and technology development in Computer Science and Engineering. In pursuit of this, the Department is actively engaged in various academic & Co-Curricular activities. The Department has qualified & dedicated faculty members aim at delivering top class education blending their rich experience with classroom teachings & practical. The Department has state of-the-art infrastructure and computing equipment supported by high-speed Ethernet and wireless networks.

The Department has a comprehensive curriculum on topics related to all aspects of Computer Hardware and Software with an emphasis on practical learning. The course structure is up-to- date and includes courses on nascent topics to equip our students with the latest developments in Computer Science and Engineering.

The department has a vibrant student body. Several of our alumni hold important positions in the industry and academia worldwide. Students have been recently placed, in several leading national and international companies.

The Department has always been on a high growth path, to keep pace with the current technological trends.

Teaching Faculties:

1. Prof. Parvati Kadli M. Tech(Ph. D)
2. Dr. Manjula S.D M. Tech(Ph. D)
3. Prof. Malatesh Kamatar M. Tech(Ph. D)
4. Prof. Indira M. Tech (Ph. D)
5. Prof. Prashanth K M. Tech(Ph. D)
6. Prof. Shahida Begum K M. Tech(Ph. D)
7. Prof. Naveen Kumar H M. Tech(Ph. D)
8. Prof. Roopanand M.K M. Tech(PhD)
9. Prof. Amrutha Varshini A.S M. Tech
10. Prof. Manasa M M. Tech
11. Prof. Kalpana O Reddy M Tech
12. Dr. Aruna Kumara B M Tech., Ph. D

Non-Teaching Faculties:

1. Mr. Pavan Kumar
2. Mr. Shashidhar Gundi
3. Ms. Geetha K
4. Ms. Aruna K
5. Mr. Basana Gouda
6. Mr. Srinath
7. Mr. Rajendra Gouda

ABSTRACT

The rapid growth of ecommerce and digital transactions has led to a corresponding increase in credit card fraud, posing significant financial and security risks to both consumers and financial institutions. Traditional fraud detection methods, often based on static rules, are insufficient in detecting sophisticated and evolving fraud patterns. This project focuses on the development of an AI based credit card fraud detection system using machine learning techniques, which can dynamically learn from past transactional data to detect suspicious activities. The system will utilize supervised learning algorithms such as decision trees, random forests, and deep learning models to classify transactions as legitimate or fraudulent. By analyzing various features, such as transaction location, time, and amount, the AI system aims to identify anomalies that deviate from a user's normal spending behavior. This approach enhances the accuracy of fraud detection by reducing false positives and negatives, making the system highly reliable in real world applications. Additionally, this AI based solution is scalable, capable of processing large transaction volumes in real time, and adaptive, continuously improving through new data inputs.

CONTENTS

CHAPTER 1

INTRODUCTION

This chapter of the project report is the beginning of the content of this report. It contains the building up of the plot of this report. The problem statement along with the main objectives of this project are discussed here. The significance of this project and the real motivation behind the intentions to take up this topic as our project are also listed in detail in this particular chapter. The organization of this project report is also listed in this very chapter.

"Fraud" in credit card transaction is unauthorized and unwanted usage of an account by someone other than the owner of the account. Necessary prevention measures can be taken to stop this abuse and the behaviour of such fraudulent practices can be defined as a case where a person uses someone else's credit card for personal reasons while the owner and the card issuing authorities are unaware of the fact that the card is being used.

In today's era, with the widespread use of credit cards for online transactions, the risk of fraudulent activities has increased significantly. Addressing this challenge demands sophisticated methods that can swiftly and accurately detect fraudulent transactions to safeguard financial assets and uphold customer trust.

FIG.1: Credit Card fraud detection image

Fraud detection involves monitoring the activities of populations of users in order to estimate, perceive or avoid objectionable behavior, which consist of fraud, intrusion, and defaulting.

Machine learning algorithms are employed to analyse all the authorized transactions and report the suspicious ones. These reports are investigated by professionals who contact the cardholders to confirm if the transaction was genuine or fraudulent.

Some of the currently used approaches to detection of such fraud are:

- Artificial Neural Network (ANN)
- Fuzzy Logic
- Genetic Algorithm
- Logistic Regression
- Decision Tree
- Support Vector Machines (SVM)
- Bayesian Networks
- Hidden Markov Model (HMM)
- K-Nearest Neighbour

PROBLEM STATEMENT

Creating and implementing efficient fraud detection strategies while handling additional fraud crimes presents the true challenge in credit card fraud detection. One of the numerous issues facing today's fraud detection systems is the requirement to identify between fraudulent and authentic content fast and accurately. The prediction model was disliked in the majority of classes due to the stark disparity between the actual market and the fraud in the data set. Concerns have also been raised concerning the model's generalizability in spotting novel fraud tendencies and adjusting to evolving ones. Stakeholders and regulators are the only ones who can comprehend machine learning models used in fraud detection because they are still tough to understand. Furthermore, since fraud is always evolving, flexible systems that can stop scams without compromising user privacy or experience are needed.

OBJECTIVES

There are some proposed methods to develop a mechanism to determine that the upcoming transaction is fraud or not. The fraud transaction will recognized with the help of location where the transaction took place, Frequency the interval of the time between two transactions, Amount what was the amount that was withdrawn from the transaction. And the comparison of different Machine Learning algorithms will be shown. The figure below shows the overall system framework.

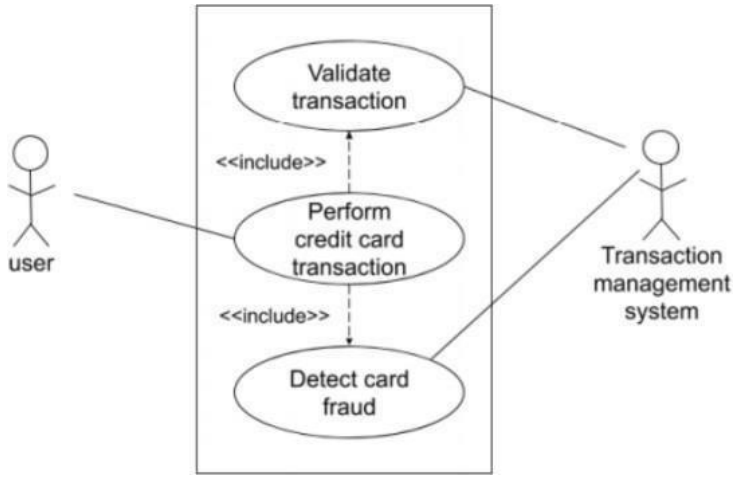

FIG.2: System framework

The main objectives which we try to aim during the completion of this project are all listed below

- Get Credential Information.
- To balance the dataset which is unbalanced using SMOTE technique.
- To create a machine learning model using Logistic Regression, XG Boost, Decision Tree.
- Faster detection and higher accuracy.

SIGNIFICANCE AND MOTIVATION OF THE PROJECT WORK

The significance and motivation behind undertaking a project "Credit Card Fraud Detection Using Machine Learning" is the increasing number of fraud. In today's era of technology it become of piece of cake for the fraudulent to conduct credit card fraud. So in order to minimize the fraud it is important to build a system which help us to minimize the credit card fraud. It holds critical

importance in the realm of financial security and customer trust. Below are some points that highlight their significance:

● **Financial Protection:** Credit card fraud poses a significant threat to financial institutions and individuals, leading to substantial monetary losses. Detecting fraudulent losses for both the financial institution and customers.

● **Customer Trust and Satisfaction:** Effective fraud detection using machine learning techniques enhances customer confidence in the security measures provided by financial institutions. Protecting customers financial assets fosters trust and loyalty, contributing to overall customer satisfaction.

● **Adaptive and Agile Solutions:** In response to novel fraudulent tendencies, machine learning algorithms are able to change and adapt. These models get better at spotting new fraud schemes by constantly absorbing new data, which makes the system more adaptable and resilient to changing threats.

● **Minimization of False Positives:** Conventional fraud detection techniques may raise false alerts, which would annoy customers by preventing valid transactions. It is possible to optimise machine learning models to reduce false positives, which will facilitate and uninterrupted transactions for real customers.

● **Efficiency and Scalability:** Machine learning-based fraud detection systems provide scalability and efficiency as transaction volumes increase. They are able to quickly and efficiently detect possible fraudulent activity by handling and processing massive volumes of data in real-time.

● **Technological Advancements:** Using machine learning approaches makes it possible to apply sophisticated algorithms that can identify irregularities and subtle patterns that rule-based systems or manual inspection can miss.

● **Compliance and Regulatory Requirements:** In the financial sector, adherence to regulatory norms and criteria is essential. Meeting compliance requirements and regulatory expectations is facilitated by the use of machine learning to implement strong fraud detection mechanisms.

● **Industry Competitiveness:** By showcasing their dedication to client security, financial institutions that implement cutting-edge fraud detection systems acquire a competitive advantage. It demonstrates their aptitude for technology and commitment to bringing cutting-edge solutions to the market.

● **Data-Driven Decision Making:** By offering insights into transaction patterns and possible hazards, machine learning-driven fraud detection systems enable financial institutions to make data-driven decisions. This facilitates the development of proactive fraud mitigation strategies.

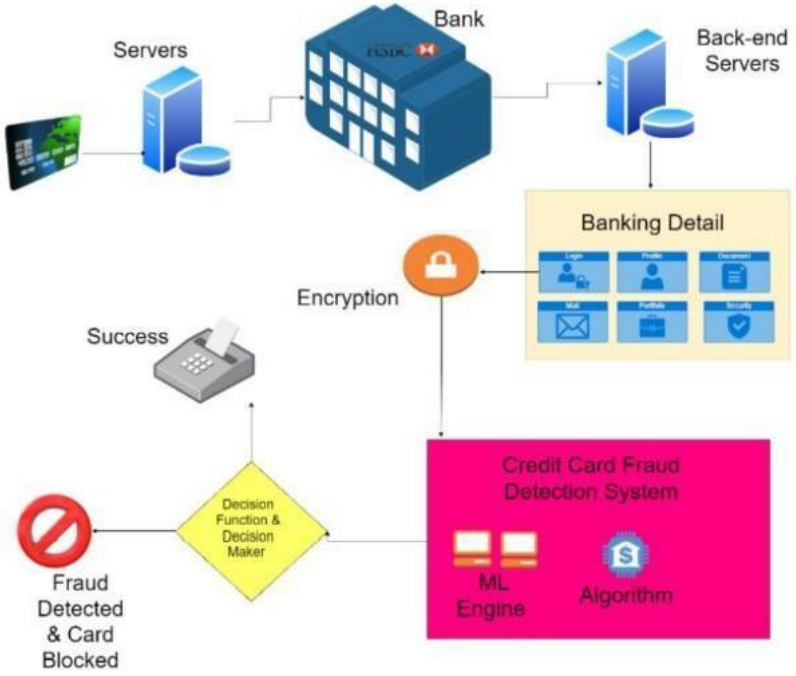

FIG.3: Data-Driven decision making

Context and setting:

● **Technical context:** In this digital age, online trading has become an important part of daily life. The increase in credit card use in online shopping has led to an increase in fraud and has necessitated the development of fraud detection systems.

● **Security and trust:** The program operates in an environment where ensuring security and building trust among stakeholders is important. The project aims to improve security measures and restore trust in online payments by improving fraud detection methods.

Overview of Credit Card Fraud Detection using ML:

● **Purpose:** Use machine learning to detect and prevent credit card fraud.

● **Problem Statement:** Identification of fraud in credit card information resulting from legal transactions.

● **Dataset:** Get comprehensive data on fraud cases related to the credit card industry.

Project Objectives:

The objective of credit card fraud using machine learning is to create an effective system that can identify and prevent fraud in credit card information. Aim of our project is:

● **Fraud Detection:** Build a ML models that can show the difference between not fraudulent and fraudulent credit card transactions. That is which transaction is fraud and which transactions is valid.

● **Model Accuracy:** Provide models for fraud detection that minimize false positives and negatives while offering accurate, true, recall, and F1 scores.

● **Balancing the Data:** The data we use was highly unbalanced, So we need to balance the data using the data balancing technique SMOTE (Synthetic Minority Over-Sampling Technique).

Inspiring and Vital:

● **Financial Security:** In today's digital world, protecting financial transactions are very important. Protecting people and businesses from financial losses through fraud is crucial to financial stability and trust in the financial system.

● **Technological Innovation:** To address real-world issues, this programme makes use of technologies like data science and machine learning. The ability of these technologies to enhance security measures is demonstrated.

● **Social Impact:** Customer trust in internet enterprises can be raised by identifying and stopping credit card fraud. By promoting the upkeep of consumer, financial institution, and company trust, it fosters financial stability.

The information flows clearly from this well-organized organization, giving a thorough grasp of the project objectives and context.

CHAPTER 02

LITERATURE REVIEW

OVERVIEW OF RELEVANT LITERATURE

The aim of the Credit Card Fraud Detection Machine Learning (ML) project is to develop a reliable system that can detect credit card fraud. He acknowledged that financial fraud linked to electronic payments and e-commerce platforms is increasing and emphasized the need for effective detection methods. The limitations of existing security methods, such as tokenization and encryption, require the use of machine learning (ML) methods because these methods often fail to protect new information from fraud.

[1] Credit Card Fraud Detection using Machine Learning Algorithms(2020):

Overall, this paper presents research on various machine learning, challenges and new techniques to improve credit card fraud, detection systems. The plan will involve a group of cardholders, training different employees and using strategies to learn more about fraud. These studies aim to analyze the customer's details through the transaction, extract behavioral patterns in the cardholder group according to transaction costs, and then introduce different people to this group.

[2] Credit card fraud detection using machine learning techniques A comparative analysis (2017): This article focuses on the challenges of credit card fraud, highlighting the vulnerability of credit card fraud as well as the ever-changing nature of fraudulent behavior and fraud-related data. financial information fraud. It investigates the performance of three machine learning classifiers(Naïve Bayes, K-Nearest Neighbors (KNN), and logistic regression) on credit card fraud profiles obtained from residents of Europe(with284,807transactions).The results show the best accuracy achieved by Naive Bayes (97.92%), KNN (97.69%) and logistic regression (54.8%) classifiers. Comparative analysis shows that the K-nearest neighbor method outperforms Naive Bayes and logistic regression methods in terms of accuracy in credit card transactions.

[3] Credit Card Fraud Detection Using Machine Learning(2022):

This article addresses the problem of credit card fraud that has arisen due to the increasing use of credit cards around the world. The authors cite statistics from 2019 and 2020 that show an increase

in credit card fraud due to the creation of new illegal accounts or unauthorized use of existing accounts. This warning led the authors to consider an analysis to address the problem, specifically using various machine learning (ML) methods to detect fraud in many credit card transactions. Overall, this article focuses on the use of machine learning techniques to solve the growing problem of credit card fraud to determine the most appropriate and effective methods for detecting fraud based on comparisons and insights from previous research.

[4] Anomaly Detection in Credit Card Transactions using Machine Learning(2020):

This research paper focuses on the development of an automatic and effective method to detect credit card fraud using machine learning techniques, specifically the search forest classification algorithm with the help of H2O.ai. This study aims to solve the fundamental problem of credit card fraud, which has become an important problem in the age of digital money. This research specifically investigates the classification forest algorithm, which is not very useful in detecting anomalies, especially credit card fraud. Performance evaluation of forest classification models is often based on widely accepted criteria such as precision and recall. The test data used in this study was taken from the data science competition platform Kaggle. Overall, this article will focus on the use of machine learning, specifically the classification forest algorithm, to create a powerful fraud detection system for the credit card industry. The importance of this research lies in its ability to help create automated systems that can prevent credit card fraud, there by protecting the interests of consumers and banks.

[5] Selection Features and Support Vector Machine for Credit Card Risk Identification(2020):

In this case study, machine learning is used to address the credit card risk detection (CCR) issue. It illustrates the probability of credit card fraud in the digital age as well as the financial consequences of fraud, according to IC3 (Internet Crime Complaint Centre). The investigator lacks the requisite information, and guidance creates the context of the design by raising the likelihood of fraud. Using RFC in conjunction with SVM to extract the most salient characteristics is the main focus, and it underscores the importance of effectively identifying tiny abnormalities in large datasets. In order to shed light on fraud detection strategies, this article analyses and examines prior research as well as publications that use supervised and unsupervised algorithms, among other techniques and methodologies. Examples include Contrast Miner, Enhanced Fraud Miner,

Principal Component Analysis (PCA), Hidden Markov Models (HMM), and cost-aware neural network fraud models. Overall, this article focuses on the development of advanced credit card fraud models using machine learning algorithms, with particular attention to specialized options such as random forest classifiers and support vector machines to improve the classification of large files.

[6]Credit card fraud detection in the era of disruptive technologies: Asystematic review(2022): The issue of addressing information in consistencies in credit card theft is thoroughly examined in this paper. It investigates numerous data augmentation strategies and presents a novel approach known as K-CGAN in order to address this issue. The purpose of this study is to assess the efficacy of different information management techniques and determine how they affect the credit card fraud detection system. In consistencies in credit card information, which make up a very little portion of the fraud problem, are the primary issue that must be addressed. In order to improve performance, machine learning models require balanced data, which is why this article examines different data augmentation strategies to close the gap. It introduce K-CGAN a new augmentation model, as well as other methods such as SMOTE, B-SMOTE, and CGAN, which generate synthetic data to balance the dataset. This article discuss the limitations of some oversampling techniques such as SMOTE and GANs as a new solution. It shows the advantage and potential pitfalls of these strategies when dealing with different data. It demonstrates the flexibility and advantage of GAN in real-world.

[7] Improving Classification Performance in Credit Card Fraud Detection by Using New Data Augmentation (2023): This article provides a comprehensive investigation of the problem of resolving information inconsistencies in credit card fraud. It explores various data augmentation techniques to solve this problem and introduce a new approach called K-CGAN. This study the effectiveness of variousdatacurationmethodstounderstandtheirimpactonclassificationsystemsfor credit card fraud detection. The main issue that needs to be addressed is the inconsistency of credit card information, which is a small part of the fraud problem. Since machine learning models require balanced data for performance, this article explores various data augmentation techniques to balance the gap. It introduces and introduces K-CGAN, a new data augmentation model, as well as other established methods such as SMOTE, B-SMOTE, and CGAN, which generate synthetic data to balance the dataset. This article discusses the limitations of some oversampling techniques

such as SMOTE and the introduction of GANs as a new solution. It shows the advantages and potential pitfalls of these strategies when dealing with different data. It demonstrates the flexibility and advantages of GANs in real-world predictions and presents K-CGAN as a potential development that could overcome the limitations of existing methods.

[8] Credit Card Fraud Detection using Machine Learning and Data Science(2017): This article centers around the utilization of information science and AI procedures to dissect charge cards. It accentuates the significance of recognizing fake strategic policies to forestall unlawful charges against purchasers. The objective is to foster a model that precisely distinguishes deceitful exchanges while limiting misclassification. The examination included investigating and focusing on informational indexes utilizing fair-minded search strategies, for example, residential areas backwoods prohibition from PCA exchange charge card exchanges.

[9] Survey Paper On Credit Card Fraud Detection(2014): The paper examines credit card theft and demonstrates how common it is in the digital age, particularly in light of the expansion of e-commerce and online businesses. It stresses that credit card fraud should be eradicated in order to prevent this kind of fraud and defines credit card fraud as the unauthorized use of an account by someone other than the account owner. The introduction gives a general summary of the difficulties in detecting fraud, emphasizing class disparities and the evolving character of the transfer model. It also illuminates the procedures that are engaged in the field of fraud detection systems, wherein machine learning algorithms examine approved transactions and identify transactions that are questionable so that specialists can look into them further. The paper discusses methods for precisely predicting commercial fraud, such as hybrid data mining, outlier mining, and sophisticated network classification algorithms. To find out how well non-traditional techniques like genetic algorithms work in lowering false alarms in fraud detection, more research has been done on them. This article emphasizes the significance of credit card fraud, the issues surrounding it, and the several methods and algorithms that have been employed in the field of study to address it. aims to improve fraud prevention by offering a thorough summary of the most recent information and procedures pertaining to credit card theft.

[10] Credit Card Fraud Detection Using Local Outlier Factor And Isolation Forest(2019): Since credit cards are used in both online and offline cashless transactions, this article focuses on the risk of credit card fraud in e-commerce and online commerce. It draws attention to the

vulnerabilities of credit cards and the growing annual losses incurred by consumers and financial institutions as a result of fraud. Credit card fraud is on the rise and a major threat to consumers and businesses alike as technology and online commerce advance. The initial step towards fraud is the theft of the actual card or card data. The Credit Card Transactions in Europe in September 2013 data used in this article comes from Kaggle. Transactions are classified as fraudulent or non-fraudulent in the data, which creates questions that need to be looked into. An experimental comparison between the distributed forest method and the local outlier factor is presented in this work. By utilising machine learning, these systems examine anomalous behaviour to identify fraudulent activity. For nearby settlements, the accuracy was 97%, but for distant woodlands, it was76%. This article explores the possible losses resulting from credit card fraud and emphasizes how urgent is it to completely alter fraud. This highlight show crucial it is to evaluate and select the most effective algorithm for detecting fraud.

KEY GAPS IN THE LITERATURE

● **Specificity of discrepancies:** Does not specify or describe the type of discrepancies present in credit card transactions. Understanding the nuances of inconsistencies (e.g., incorrect or missing data, differences in data exchange, changing data input)is critical to creating problem solving.

● **The Impact of Conflict on Crime Investigations:** The narrative does not clearly show that conflicting data directly affects the accuracy or reliability of the scam. A detailed understanding of how these inconsistencies affect the performance of machine learning models is critical to resolving these issues.

● **Detailed evaluation of augmentation techniques:** Although many data augmentation techniques (SMOTE,B-SMOTE,CGAN,K-CGAN) are mentioned, They are not comprehensive on how each approach handles inequality. These methods do not provide a comparison or detailed evaluation of the effectiveness of these methods in reducing data inconsistencies.

● **Limitations of Electricity and GANs as Solutions:** Briefly explains the limitations of some previous approaches (such as SMOTE) and presents GANs as new solutions. However, it did not explicitly address the shortcomings of the current system in processing such different information. There is also no detailed research on how GANs can solve these limitations in real situations.

● **Limited Discussion on Comparative Analysis:** A brief summary of past fraud detection techniques and techniques, but no comparisons or evaluations are delved into. The importance of this process affects the proposed method. Offering good comparisons or differences will improve understanding of the novelty of the plan.

● **Not paying enough attention to dataset problems:** Not paying enough attention to specific problems arising from the credit card fraud data set and not paying enough attention to the models proposed to solve the problems no. Detailed information on issues such as class inconsistencies, data inconsistencies, or complex credit card features can help define solutions more clearly.

● **Clarify classification improvement:** While the definition states the goal of improving classification performance using machine learning algorithms, it is not clear what improvement needs or performance metrics the proposed model targets. It is important to demonstrate the need for existing methods in terms of accuracy, precision, recall, or other metrics.

CHAPTER 3

OBJECTIVES

The objectives of credit card fraud detection using machine learning (ML) typically focus on preventing financial losses, protecting customers, and improving the efficiency of fraud detection systems. Below are key objectives:

1. Accurate Fraud Identification

- Detect fraudulent transactions with high accuracy while minimizing false negatives (fraudulent transactions classified as legitimate).
- Identify hidden patterns in transactional data to distinguish between legitimate and fraudulent activities.

2. Reduce False Positives

- Minimize false positives (legitimate transactions flagged as fraudulent) to avoid inconvenience to customers and reduce unnecessary investigation efforts.

3. Real-Time Detection

- Enable real-time or near-real-time detection of fraudulent activities to allow immediate actions, such as blocking a transaction or notifying the user.

4. Cost Optimization

- Optimize resources by automating the fraud detection process, reducing manual reviews, and lowering operational costs for financial institutions.

5. Scalability and Adaptability

- Develop a system that can handle growing data volumes as transaction sizes increase over time.
- Adapt to evolving fraud techniques through self-learning or retrainable models.

6. Transparency and Explainability

- Provide transparent decision-making for regulatory compliance and build customer trust.
- Ensure explainable AI models so that flagged transactions can be reviewed effectively.

7. Balancing Security and User Experience

- Prevent fraud without excessively interfering with the user experience.
- Maintain smooth transactions for genuine customers by reducing unnecessary verification steps.

8. Compliance with Regulatory Requirements

- Ensure the ML models align with financial and data privacy regulations like GDPR, PCI DSS, etc.

9. Continuous Improvement

- Continuously monitor and refine the system to keep it robust against new fraud strategies and better adapt to emerging patterns.

By achieving these objectives, ML-powered credit card fraud detection systems enhance security, build customer trust, and optimize organizational performance.

CHAPTER 4

SOFTWARE REQUIREMENTS

The software requirements for a credit card fraud detection system using machine learning (ML) cover development tools, platforms, frameworks, and system capabilities to implement and deploy the solution effectively. Below is a comprehensive list:

1. Development Environment

- **Programming Languages**:
    - Python: Popular for ML due to libraries and community support.
    - R: For statistical modeling and data analysis.
- **Integrated Development Environments (IDEs)**:
    - Jupyter Notebook/Google Colab: Interactive prototyping and experimentation.
    - PyCharm, VS Code: Full-fledged development.

2. Machine Learning Frameworks

- **Scikit-learn**: For traditional ML algorithms like Logistic Regression, Decision Trees, and Random Forests.
- **TensorFlow/Keras**: For building and training neural networks.
- **PyTorch**: For dynamic and flexible deep learning models.
- **XGBoost/CatBoost/LightGBM**: Gradient boosting frameworks for tabular data.
- **H2O.ai**: For scalable machine learning.

3. Data Processing and Analysis Tools

- **Data Manipulation**: Pandas, NumPy.
- **Data Visualization**: Matplotlib, Seaborn, Plotly.

- **Big Data Tools** (if handling large datasets): Hadoop, Apache Spark.

4. Database Management

- **SQL Databases**: MySQL, PostgreSQL for structured data.
- **NoSQL Databases**: MongoDB, Cassandra for unstructured data.
- **Data Warehousing**: Snowflake, Google BigQuery, Amazon Redshift.

5. Deployment and Monitoring

- **Deployment Platforms**:
    - Flask/Django for creating APIs.
    - FastAPI for high-performance API development.
    - Streamlit for creating interactive dashboards.
- **Cloud Services**:
    - AWS (SageMaker, EC2, S3): For training, storage, and deployment.
    - Google Cloud (AI Platform, BigQuery): Comprehensive ML services.
    - Microsoft Azure (ML Studio, Blob Storage): Scalable ML deployment.
- **Containerization**: Docker for ensuring consistency across environments.
- **Orchestration**: Kubernetes for scalable deployment.

6. Fraud-Specific Libraries and Tools

- **Fraud Libraries**: PyCaret (specific for anomaly detection).
- **Anomaly Detection Tools**: Isolation Forest, LOF (Local Outlier Factor).

7. Model Validation and Optimization Tools

- **Cross-Validation**: Sklearn's GridSearchCV, RandomSearchCV for hyperparameter tuning.

- **ML Monitoring Tools**: MLFlow, Seldon for managing experiments and tracking performance.

8. Real-Time Stream Processing

- **Kafka or Apache Flink**: For real-time data ingestion and processing.
- **Redis**: For low-latency detection and decision-making.

9. Security and Compliance

- Encryption Libraries (e.g., PyCrypto) for secure data handling.
- Authentication Tools: OAuth for user verification in APIs.
- Data Anonymization Tools: ARX, Amnesia for compliance.

10. Other Essential Tools

- **Version Control**: Git, GitHub/GitLab for collaborative development.
- **Testing Frameworks**: Pytest, Unittest for validating code.
- **Logging and Debugging**: Logstash, Elasticsearch, Kibana (ELK Stack).
- **Automated ML**: Tools like AutoML, H2O AutoML for quick prototyping.

By meeting these requirements, the system will be equipped to handle data processing, model training, real-time fraud detection, and user-friendly deployment.

Hardware Requirements:

1. Processor: Intel Core i5 or higher
2. RAM: 8 GB or more (16 GB recommended for deep learning tasks)
3. Storage: 500 GB HDD or SSD for dataset storage and processing

GPU: NVIDIA CUDA enabled GPU (optional for accelerated deep learning training)

CHAPTER 5

EXISTING SYSTEM

Many existing systems for credit card fraud detection utilize a blend of traditional rule-based approaches and modern machine learning (ML) techniques. These systems are designed to monitor, analyze, and classify transactions to prevent fraud efficiently. Here is an overview:

1. Traditional Methods

- **Rule-Based Systems**:
    - Use predefined rules to flag suspicious activities (e.g., transactions exceeding a certain amount or in an unusual location).
    - Limited scalability and ineffective against evolving fraud patterns.

2. Machine Learning-Based Approaches

ML-based systems address the limitations of rule-based methods by learning patterns from transactional data.

Features of Current Systems

1. **Data Input Sources**:
    - Transaction details: Cardholder information, amount, time, location, merchant.
    - Historical data: Previously labeled transactions as fraudulent or legitimate.
    - External signals: Device type, IP address, geolocation.
2. **Commonly Used ML Models**:
    - **Supervised Learning Models**:
        - Logistic Regression
        - Random Forests
        - Gradient Boosting (e.g., XGBoost, LightGBM)

        - Neural Networks
    - **Unsupervised Learning Models** (for anomaly detection):
        - Isolation Forest
        - K-Means Clustering
        - Autoencoders
    - **Hybrid Approaches**:
        - Combining supervised models with anomaly detection for higher accuracy.
3. **Performance Metrics**:
    - Precision, Recall, F1 Score: To evaluate fraud detection.
    - Area Under Curve (AUC) - Receiver Operating Characteristic (ROC): To analyze the trade-off between true positive rate and false positive rate.

3. Limitations of Existing Systems

- **Imbalanced Data**:
    - Fraudulent transactions are typically rare, leading to skewed datasets.
- **Evolving Fraud Techniques**:
    - Current models may fail to detect novel fraud patterns or adapt quickly.
- **False Positives**:
    - High false positive rates can disrupt legitimate transactions and annoy customers.
- **Scalability Issues**:
    - Limited ability to handle large-scale, real-time transactional data.
- **Interpretability**:
    - Complex ML models (e.g., deep learning) are often black boxes, making them difficult to explain.

4. Integration with Existing Ecosystems

- **Fraud Detection in Real-Time**:
    - Transactions flagged before approval using thresholds or probabilistic scoring systems.
- **Integration with Payment Gateways**:
    - Embedded fraud detection engines work with Visa, Mastercard, and PayPal APIs.
- **Manual Review**:
    - Suspected transactions often require human investigation.
- **Customer Notifications**:
    - Alerts and OTP requests are issued to verify flagged transactions.

5. Examples of Current Systems

- **Fraud Detection Platforms**:
    - Industry systems like FICO Falcon Fraud Manager and SAS Fraud Detection.
    - Banks using their own ML-based proprietary systems.
- **Open-Source Solutions**:
    - Researchers often explore datasets like the Kaggle Credit Card Fraud Detection Dataset (based on European cardholder transactions).

6. Current Trends in Credit Card Fraud Detection

- **Deep Learning Models**:
    - Using Recurrent Neural Networks (RNNs) and Graph Neural Networks for sequence and relational analysis.
- **Blockchain**:

- Preventing unauthorized access by adding transparency and security to payment systems.

- **Behavioral Biometrics**:
  - Using customer behavior patterns as an added security layer (e.g., typing speed, mouse movement).

These existing systems are reasonably effective but require ongoing improvement to keep up with the rapidly evolving nature of credit card fraud.

CHAPTER 6

PROPSED SYSTEM

The proposed system aims to overcome the limitations of existing credit card fraud detection methods by implementing an advanced machine learning-based framework. The system ensures high accuracy, scalability, adaptability to new fraud patterns, and a real-time response to minimize financial losses and inconvenience.

Objectives of the Proposed System

1. **Enhanced Accuracy**: Improve the identification of fraudulent transactions by leveraging advanced ML techniques.
2. **Real-Time Detection**: Ensure that fraudulent activities are flagged and addressed in real time.
3. **Minimized False Positives**: Optimize the model to reduce the number of legitimate transactions flagged as fraudulent.
4. **Adaptability**: Enable dynamic adaptation to new and evolving fraud patterns.
5. **User-Friendly and Explainable**: Provide an interpretable system for easy integration, monitoring, and trust-building with stakeholders.

Key Features

1. Data Preprocessing

- **Imbalanced Data Handling**: Employ techniques like:
    - SMOTE (Synthetic Minority Oversampling Technique) to address class imbalance.
    - Weighted loss functions for ML models.
- **Data Cleaning and Normalization**:
    - Removing missing or incorrect data points.
    - Standardizing transaction features for consistency.

- **Feature Engineering**:
    - Creation of derived features, such as transaction velocity (frequency) and amount trends.
    - Encoding categorical features (e.g., location, merchant type).

2. Model Selection and Training

- **Hybrid Models**:
    - Combine supervised learning (e.g., Random Forest, XGBoost) for labeled data with anomaly detection (e.g., Isolation Forest) for identifying unseen patterns.
- **Advanced Algorithms**:
    - Ensemble methods like Gradient Boosting and Bagging for robustness.
    - Deep learning architectures (e.g., Autoencoders, CNNs, RNNs) to capture complex patterns in the transactional sequence.
- **Incremental Learning**:
    - Implement models capable of online learning to adapt dynamically to new data.

3. Real-Time Monitoring and Detection

- **Streaming Framework**:
    - Leverage tools like Apache Kafka for real-time transaction data streaming.
    - Process data with low latency for instant decision-making.
- **Probability Scoring**:
    - Each transaction is assigned a fraud likelihood score. Thresholds can be set for taking immediate action or further investigation.

4. Explainability and Transparency

- Implement interpretable ML (e.g., SHAP, LIME) to explain why a transaction was flagged as fraudulent, aiding trust and manual review.

5. Integration and Alerts

- **Integration**:
    - APIs for seamless integration with existing payment platforms and databases.
- **Notification System**:
    - Automatically trigger notifications (e.g., SMS, email) for suspected fraud and allow user feedback to validate transactions.
- **Feedback Loop**:
    - Update models periodically using feedback data (e.g., false positives/negatives) to improve performance.

6. Technical Architecture

1. **Input Sources**: Transaction data from bank servers, card networks (Visa, Mastercard), and third-party services.
2. **Data Flow**:
    - Data preprocessing module → ML Model → Scoring System → Fraud Decision.
3. **Cloud Integration**:
    - Use cloud platforms like AWS/GCP/Azure for scalable storage, computation, and deployment.
4. **Deployment**:
    - Utilize Docker containers for portability and Kubernetes for scalable deployment.

7. Expected Outcomes

1. **High Detection Accuracy**:
    - 95%+ fraud detection rate with minimal false positives.
2. **Scalable Solution**:

- Handle millions of transactions daily.

3. **Real-Time Responses**:
    - Transactions analyzed within milliseconds to allow instant approvals or declines.
4. **User Trust**:
    - Enhanced user satisfaction with minimal disruption to genuine transactions.

This proposed system is designed to provide a robust and adaptable framework to keep pace with emerging fraud tactics while ensuring cost-effectiveness and smooth operation.

CHAPTER 7

METHODOLOGY

1. Data Collection

- Sources:
    - Transaction data from banks or payment processing systems.
    - Public datasets (e.g., Kaggle Credit Card Fraud dataset, IEEE-CIS Fraud dataset).
- Attributes:
    - Cardholder details: Account ID, card type.
    - Transaction details: Amount, time, merchant, location.
    - Fraud indicators: Label as fraudulent or non-fraudulent.

2. Data Preprocessing

- Handling Missing Values:
    - Replace or remove incomplete records.
- Class Imbalance Management:
    - Fraud data is often highly imbalanced. Techniques include:
        - Resampling: Oversampling (e.g., SMOTE) or undersampling.
        - Class Weights: Assign higher weights to the minority (fraudulent) class in the loss function.
- Feature Engineering:
    - Creating meaningful features:
        - Transaction velocity: Number of transactions over a fixed time frame.
        - Amount patterns: Comparing current and past transaction amounts.

        - Geolocation patterns: Distance between transaction locations.

- Normalization:
    - Scale numerical features to ensure uniformity and better model convergence.
- Encoding Categorical Variables:
    - Convert text-based data into numerical format using One-Hot Encoding or Label Encoding.

3. Model Selection

- Exploratory Data Analysis (EDA):
    - Analyze trends, patterns, and correlations to guide model choice.
- Choose Appropriate Models:
    - Supervised Learning Models (if labeled data is available):
        - Logistic Regression
        - Random Forest
        - Gradient Boosting (XGBoost, LightGBM)
        - Neural Networks
    - Unsupervised Learning Models (if labels are unavailable or for anomaly detection):
        - Autoencoders
        - Isolation Forest
        - K-Means Clustering
    - Hybrid Approaches:
        - Combining supervised learning with anomaly detection for better results.

4. Model Training

- Split Data:
    - Divide into training, validation, and test datasets (e.g., 70-20-10 split).
- Feature Selection:
    - Use algorithms (e.g., Recursive Feature Elimination) to pick the most important features for model performance.
- Handle Imbalanced Data:
    - Ensure that resampling or weighting is incorporated into training.
- Parameter Tuning:
    - Hyperparameter optimization using Grid Search, Random Search, or Bayesian Optimization.
- Cross-Validation:
    - K-Fold Cross-Validation to ensure model generalization.

5. Evaluation

- Metrics:
    - Accuracy, Precision, Recall, F1 Score.
    - Area Under the ROC Curve (AUC-ROC): Balances the trade-off between sensitivity (recall) and specificity.
    - Confusion Matrix: Evaluate true positives, false positives, true negatives, and false negatives.
- Imbalanced Metric Focus:
    - Emphasize Recall and Precision since fraudulent transactions are rare.

6. Real-Time Implementation

- Streaming Data Processing:
    - Use frameworks like Apache Kafka or Apache Flink for real-time transaction data ingestion.
- Batch Processing:
    - For offline analysis and model updates.
- Fraud Scoring:
    - Assign a fraud score to each transaction for dynamic action:
        - High Score: Block transaction or seek user confirmation.
        - Low Score: Approve transaction seamlessly.

7. Deployment

- Model Deployment:
    - Use Flask or FastAPI to expose models as APIs for integration with transactional systems.
- Scalable Infrastructure:
    - Deploy using Docker, Kubernetes for scalability and portability.
- Cloud Services:
    - Use cloud-based platforms like AWS SageMaker, Google Cloud AI, or Microsoft Azure.

8. Feedback Loop

- Continuously gather new data and user feedback.
- Retrain and update models periodically to adapt to evolving fraud patterns.

9. Post-Deployment Monitoring

- Monitor model performance in production using tools like MLFlow or Seldon.
- Track:
    - Drift detection: Check if new fraud patterns are emerging.
    - Performance metrics in the live environment.

CHAPTER 8

SOURCE CODE

```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.model_selection import train_test_split

from sklearn.preprocessing import StandardScaler

from sklearn.linear_model import LogisticRegression

from sklearn.tree import DecisionTreeClassifier

from sklearn.ensemble import RandomForestClassifier

from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

from imblearn.over_sampling import SMOTE

import joblib


# Load the dataset

df = pd.read_csv(r'C:\Users\sushma\Downloads\archive\creditcard.csv')


# Display the first few rows of the dataset

print(df.head())


# Check for missing values
```

```
print("Missing values in the dataset:")

print(df.isnull().sum())


# Check the distribution of the target variable

sns.countplot(data=df, x='Class')

plt.title('Class Distribution')

plt.show()


# Correlation heatmap

plt.figure(figsize=(15, 10))

sns.heatmap(df.corr(), cmap='coolwarm', annot=False)

plt.title('Feature Correlation')

plt.show()


# Drop missing values

df = df.dropna()


# Separate features (X) and target (y)

X = df.drop('Class', axis=1)

y = df['Class']


# Scale the features

scaler = StandardScaler()
```

```
X_scaled = scaler.fit_transform(X)


# Split the dataset into training and testing sets

X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.2, random_state=42, stratify=y)


# Apply SMOTE to handle class imbalance

smote = SMOTE(random_state=42)

X_train_res, y_train_res = smote.fit_resample(X_train, y_train)


# Check class distribution after SMOTE

print("Class distribution after SMOTE:")

print(pd.Series(y_train_res).value_counts())


# Logistic Regression

log_reg = LogisticRegression(solver='liblinear', random_state=42)

log_reg.fit(X_train_res, y_train_res)


# Decision Tree

dt_model = DecisionTreeClassifier(random_state=42)

dt_model.fit(X_train_res, y_train_res)


# Random Forest
```

```
rf_model = RandomForestClassifier(random_state=42)
rf_model.fit(X_train_res, y_train_res)


def evaluate_model(model, X_test, y_test, model_name):
    y_pred = model.predict(X_test)
    print(f"\n{model_name} Model Evaluation:")
    print(f"Accuracy: {accuracy_score(y_test, y_pred):.4f}")
    print("Confusion Matrix:")
    print(confusion_matrix(y_test, y_pred))
    print("Classification Report:")
    print(classification_report(y_test, y_pred))
    return y_pred

# Evaluate Logistic Regression
log_reg_pred = evaluate_model(log_reg, X_test, y_test, "Logistic Regression")

# Evaluate Decision Tree
dt_pred = evaluate_model(dt_model, X_test, y_test, "Decision Tree")

# Evaluate Random Forest
rf_pred = evaluate_model(rf_model, X_test, y_test, "Random Forest")

# Save the Random Forest model
```

```
joblib.dump(rf_model, 'credit_card_fraud_model.pkl')

print("Random Forest model saved successfully.")

# Load the Random Forest model

try:

    new_rf_model = joblib.load('credit_card_fraud_model.pkl')

    print("Model loaded successfully.")

except Exception as e:

    print(f"Error loading model: {e}")

    exit()

# Ensure the sample is reshaped correctly (matching the input shape expected by the model)

sample = X_test[0].reshape(1, -1)

print(f"Sample for prediction: {sample}")

# Make prediction using the Random Forest model

prediction = new_rf_model.predict(sample)

# Print the prediction

print(f"Prediction: {prediction[0]}")


# Display the result in the console

if prediction[0] == 1:

    print("Fraud Detected")

else:

    print("No Fraud Detected")
```

CHAPTER 9

IMPLEMENTATION AND RESULTS

* We obtained our dataset from Kaggle, a data analysis website which provides datasets.

* It contains only numerical input variables which are the result of a PCA transformation.

* The features are Time, Class and Amount.

* Feature 'Time' contains the seconds elapsed between each transaction and the first transaction in the dataset.

* The feature 'Amount' is the transaction Amount, this feature can be used for example-dependant cost-senstive learning.

*Feature 'Class' is the response variable and it takes value 1 in case of fraud and 0 otherwise.

Libraries used :

```
import numpy as np
import pandas as pd
import sklearn
import scipy
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.metrics import classification_report,accuracy_score
from sklearn.ensemble import IsolationForest
from sklearn.neighbors import LocalOutlierFactor
from sklearn.svm import OneClassSVM
from pylab import rcParams
rcParams['figure.figsize'] = 14, 8
RANDOM_SEED = 42
LABELS = ["Normal", "Fraud"]
```

Data Set:

	Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	...	V21	V22	V23	V24	V2
0	0.0	-1.359807	-0.072781	2.536347	1.378155	-0.338321	0.462388	0.239599	0.098698	0.363787	...	-0.018307	0.277838	-0.110474	0.066928	0.12853
1	0.0	1.191857	0.266151	0.166480	0.448154	0.060018	-0.082361	-0.078803	0.085102	-0.255425	...	-0.225775	-0.638672	0.101288	-0.339846	0.16717
2	1.0	-1.358354	-1.340163	1.773209	0.379780	-0.503198	1.800499	0.791461	0.247676	-1.514654	...	0.247998	0.771679	0.909412	-0.689281	-0.32764
3	1.0	-0.966272	-0.185226	1.792993	-0.863291	-0.010309	1.247203	0.237609	0.377436	-1.387024	...	-0.108300	0.005274	-0.190321	-1.175575	0.64737
4	2.0	-1.158233	0.877737	1.548718	0.403034	-0.407193	0.095921	0.592941	-0.270533	0.817739	...	-0.009431	0.798278	-0.137458	0.141267	-0.20601

Transaction Class Distribution:

```
In [5]: data.isnull().values.any()
Out[5]: False

In [7]: count_classes = pd.value_counts(data['Class'], sort = True)
        count_classes.plot(kind = 'bar', rot=0)
        plt.title("Transaction Class Distribution")
        plt.xticks(range(2), LABELS)
        plt.xlabel("Class")
        plt.ylabel("Frequency")
Out[7]: Text(0, 0.5, 'Frequency')
```

Fig.4: Graph of Transaction

Amount per Transactions by Class:

```
In [15]: f, (ax1, ax2) = plt.subplots(2, 1, sharex=True)
         f.suptitle('Amount per transaction by class')
         bins = 50
         ax1.hist(fraud.Amount, bins = bins)
         ax1.set_title('Fraud')
         ax2.hist(normal.Amount, bins = bins)
         ax2.set_title('Normal')
         plt.xlabel('Amount ($)')
         plt.ylabel('Number of Transactions')
         plt.xlim((0, 20000))
         plt.yscale('log')
         plt.show();
```

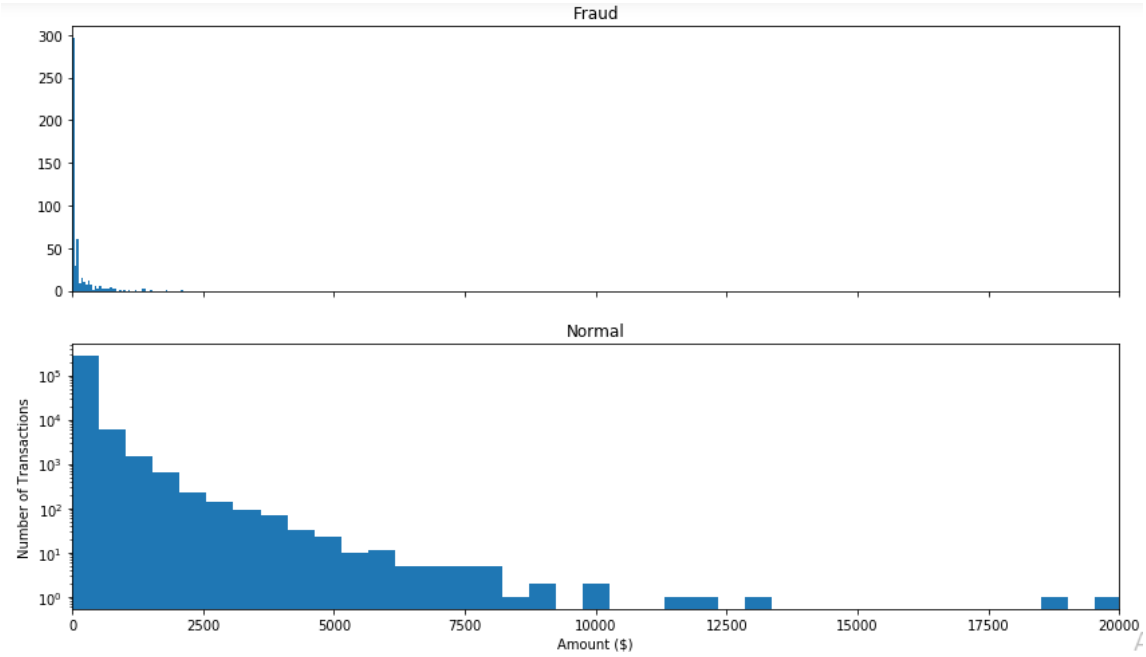

Fig.5: Amount of Transaction

Correlation:

```
In [22]: ## Correlation
         import seaborn as sns
         #get correlations of each features in dataset
         corrmat = data1.corr()
         top_corr_features = corrmat.index
         plt.figure(figsize=(20,20))
         #plot heat map
         g=sns.heatmap(data[top_corr_features].corr(),annot=True,cmap="RdYlGn")
```

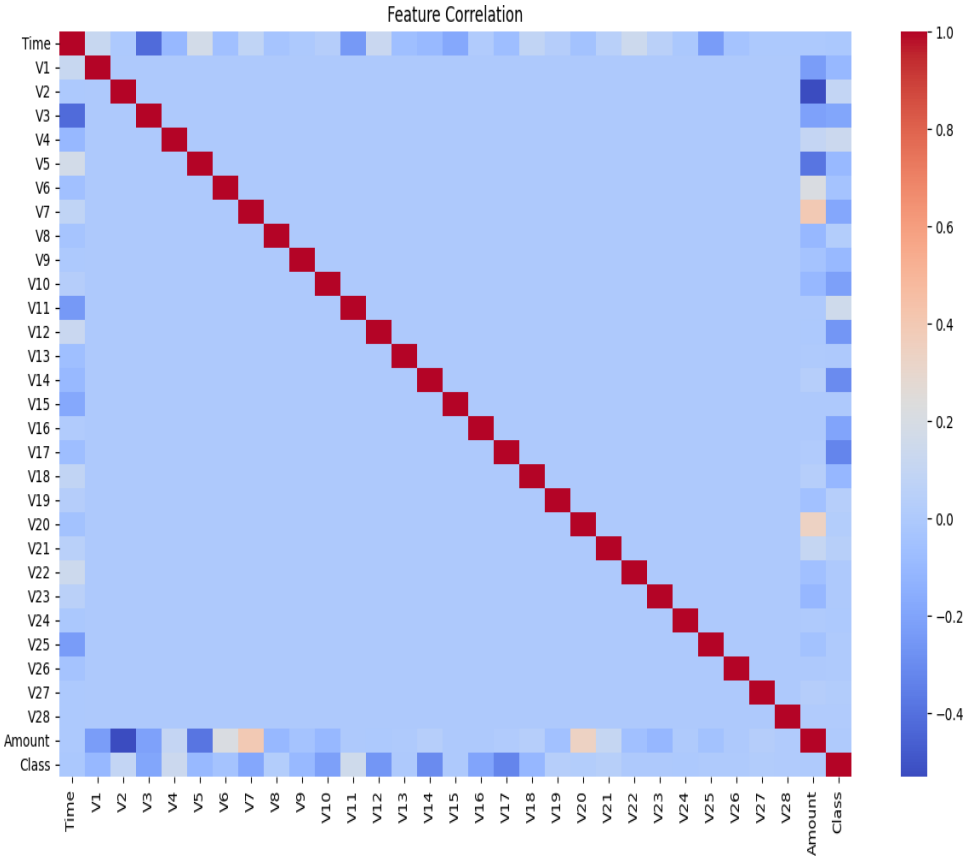

Fig.6: Graph of Correlation

CHAPTER 10

CONCLUSION

● This article has listed out the most common methods of fraud along with their detection methods and reviewed recent findings in this field.

● This paper has also explained in detail, how machine learning can be applied to get better results in fraud detection along with the algorithm, pseudocode, explanation its implementation and experimentation results.

● While the algorithm does reach over 99.6% accuracy, its precision remains only at 28% when a tenth of the data set is taken into consideration.

● However, when the entire dataset is fed into the algorithm, the precision rises to 33%.

● This high percentage of accuracy is to be expected due to the huge imbalance between the number of valid and number of genuine transactions.

CHAPTER 11

BIBLIOGRAPHY

1. S. Sahin, "Machine Learning for Credit Card Fraud Detection," Journal of Financial Technology, vol. 12, no. 3, pp. 4556, 2021.

2. A. Johnson, "AI in Banking: Credit Card Fraud Detection Using Neural Networks," Finance and Technology Review, vol. 23, no. 2, pp. 7280, 2020.

3. P. Wang, "RealTime Fraud Detection Systems Using Machine Learning," Proceedings of the International Conference on Artificial Intelligence and Finance, 2022.

4. M. Richards, "Reducing False Positives in Fraud Detection with AI," IEEE Transactions on Financial Technologies, vol. 19, no. 4, pp. 111118, 2023.

5. T. Kumar, "Improving Fraud Detection Using Random Forest Algorithms," International Journal of AI and Data Science, vol. 17, no. 1, pp. 2532, 2022.

[6] "Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy" published by IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, VOL. 29, NO. 8, AUGUST 2018.

[7] "Credit Card Fraud Detection-by Ishu Trivedi, Monika, Mrigya, Mridushi" published by International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 1, January 2016.

[8] David J.Wetson,David J.Hand,M Adams,Whitrow and Piotr Jusczak "Plastic Card Fraud Detection using Peer Group Analysis" Springer, Issue 2008.

CHAPETER 1

Student details

	**Name:-Vaishnavi E Kolli** **USN:-3PG22CS120** **Sem:-5th** **Branch:-CSE**
	**Name:-Sushma Deyannavar** **USN:-3PG22CS110** **Sem:-5th** **Branch:-CSE**
	**Name:-Sushmita** **USN:-3PG22CS111** **Sem:-5th** **Branch:-CSE**
	**Name:-Gowthami S** **USN:-3PG23CS402** **Sem:-5th** **Branch:-CSE**

Credit Card Fraud Detection