# **AUTOMATED FINANCIAL FRAUD DETECTION**

**MACHINE LEARNING PROJECT REPORT**

Submitted by

M. Lakshmi Prasanna (99220040304)

A. Chermila (99220040010)

U. Harshitha (99220040216)

B. Gayatri (99220040461)

**B. Tech - Computer Science and Engineering, Artificial Intelligence and Machine Learning**

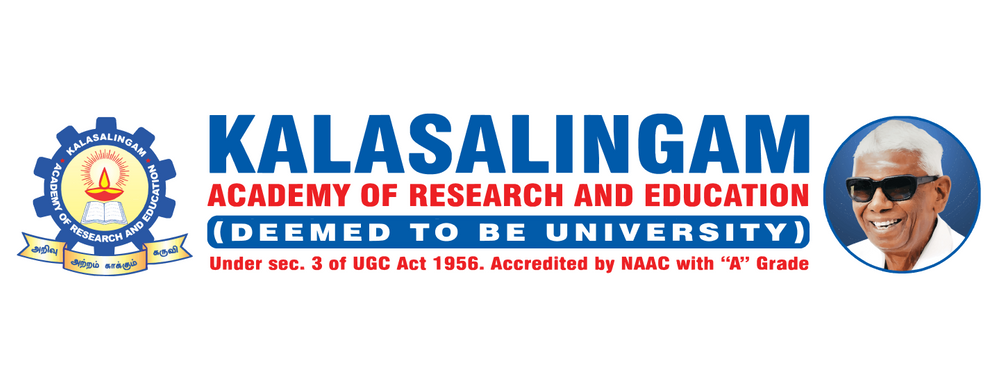


**Kalasalingam Academy of Research and Education**

(Deemed to be University)

Anand Nagar, krishnankoil-626126

March 2024



SCHOOL OF COMPUTING

DEPARTMENT OF SCIENCE AND ENGINEERING

**DECLARATION BY THE STUDENT**

We here by declare that this project **“AUTOMATED FINANCIAL FRAUD DETECTION”** is our genuine work and no part of it has been reproduced from any other works.

**Register No. Student Name Sign**

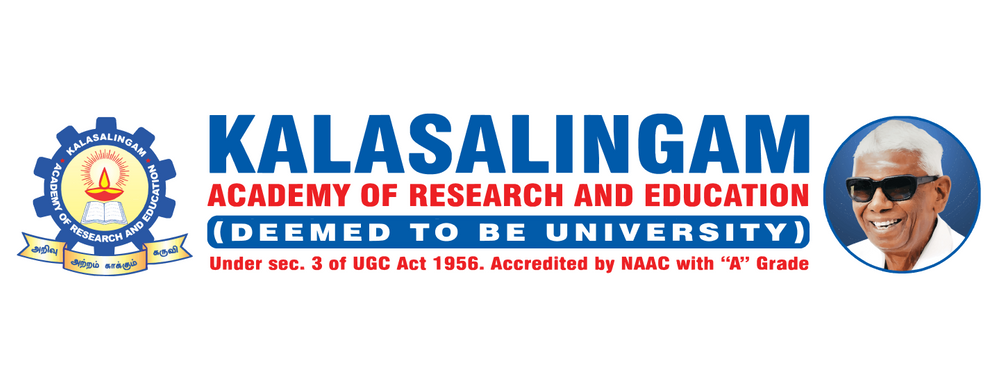
9220040304 M. Lakshmi Prasanna

99220040010 A. Chermila

99220040216 U. Harshitha

99220040461 B. Gayatri

Date:

BONAFIDE CERTIFICATE

Certified that this project report **“AUTOMATED FINANCIAL FRAUD DETECTION”** is the bonafide work of“**M. Lakshmi Prasanna (99220040304), A. Chermila (99220040010), U. Harshitha (99220040216), B. Gayatri (99220040461)”** who carried out the project work under my supervision.

Dr. N. Suresh Kumar Mr. R. Mari Selvan

Head of the department Supervisor

Professor Assistant Professor

Department of CSE Department of CSE

Kalasalingam Academy of Kalasalingam Academy

Research and Education of Research and Education

Krishnankoil-626126 Krishnankoil-626126

Project Final Review Viva-voice held on \_\_\_\_\_\_\_\_\_\_\_

Internal Examiner External Examiner

**ACKNOWLEDGEMENT**

We would like to express our sincere gratitude to all the researchers, data scientists, and developers who have contributed to the field of automated fraud detection using machine learning. Their innovative work and dedication have paved the way for advancements in fraud prevention technology, helping businesses and organizations combat financial crimes more effectively. we also acknowledge the support and collaboration of industry partners, academic institutions, and regulatory bodies who have provided valuable insights and resources to enhance our understanding of fraud detection techniques. Furthermore, we extend our appreciation to the open-source community for sharing tools, libraries, and datasets that have enabled us to experiment and develop machine learning models for fraud detection. Lastly, we thank our colleagues and team members for their hard work and collaboration in researching, designing, and implementing automated fraud detection systems. Their expertise and commitment have been instrumental in the success of our projects. Together, we are working towards a safer and more secure digital environment by leveraging the power of machine learning for fraud prevention.

**Table of Contents**

**Chapter No. Title Page No.**

Abstract 6

1 Introduction 7

2 Literature Survey 8-10

2.1 Features and advantages 11-12

2.2 Limitations and challenges 13

3 Methodology 14-15

3.1 Packages used 15

3.2 Data collection methods, sources 16

4 Proposed works

4.1 Flowchart 16

4.2 Code Implementation 17-20

5 Reference papers 21

**ABSTRACT**

Automated fraud detection using machine learning techniques has become a critical component in various industries to combat fraudulent activities. This project report presents a detailed analysis of implementing machine learning algorithms for fraud detection, including logistic regression, decision trees, and random forests. The report covers data preprocessing, feature engineering, and model evaluation techniques to enhance the accuracy. The report emphasizes the significance of automated fraud detection in industries like finance, e-commerce, and healthcare, highlighting the benefits of using machine learning models for real-time fraud prevention. It also addresses the challenges associated with implementing machine learning for fraud detection, such as data imbalance and model interpretability**.** Through this project report, the effectiveness of machine learning algorithms in detecting and preventing fraudulent activities is demonstrated, showcasing the potential for automated fraud detection systems to enhance security and protect businesses from financial losses.

**CHAPTER 1**:

**INTRODUCTION**

Automated fraud detection using machine learning is a powerful approach to identify and prevent fraudulent activities across various industries, including finance, insurance, etc. With the increasing volume and complexity of fraudulent activities, traditional methods of fraud detection are no longer sufficient. As a result, many organizations are turning to automated fraud detection systems powered by machine learning algorithms.

Machine learning algorithms have the ability to analyse large volumes of data in real-time, identify patterns and anomalies, and make accurate predictions about potentially fraudulent activities. By leveraging historical data and continuously learning from new data, these systems can adapt and improve their detection capabilities over time. Machine learning models are trained on historical data to learn patterns and anomalies associated with fraudulent behaviour. These models can then be deployed to automatically detect and flag suspicious transactions or activities in real-time.

One key advantage of using machine learning for fraud detection is its ability to adapt to evolving fraud patterns. As fraudsters change their tactics, machine learning models can be retrained with new data to stay effective.

**CHAPTER 2**

**LITERATURE SURVEY**

A literature survey on automated financial credit card fraud detection would typically cover a wide range of approaches, methodologies, and technologies employed in detecting fraudulent activities in credit card transactions.

Here is a structured breakdown of what such a literature survey might include:

Introduction to Credit Card Fraud Detection:

Overview of the importance of detecting credit card fraud. Explanation of the types of credit card fraud (e.g., identity theft, card skimming, account takeover).

Brief history and evolution of credit card fraud detection techniques.

Traditional Approaches:

Rule-based systems:

This includes systems that rely on predefined rules to flag potentially fraudulent transactions based on criteria such as transaction amount, frequency, location, etc.

Signature-based systems:

These systems compare current transactions with known patterns of fraudulent transactions.

Machine Learning and Data Mining Techniques:

Supervised learning algorithms:

Classification techniques such as logistic regression, decision trees, random forests, support vector machines (SVM), and neural networks have been applied to detect fraudulent transactions.

Unsupervised learning algorithms:

Clustering algorithms like K-means and anomaly detection techniques such as Isolation Forest, One-Class SVM, and autoencoders have been used for detecting unusual patterns indicative of fraud.

Ensemble methods:

Techniques such as boosting and bagging have been employed to improve the performance of individual classifiers.

Deep learning:

Deep neural networks, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have shown promise in detecting complex patterns in credit card transactions.

Feature Engineering and Selection:

Feature extraction:

Techniques for extracting relevant features from credit card transaction data, such as transaction amount, merchant category, time of transaction, geographical location, etc.

Feature selection:

Methods for selecting the most informative features to improve model performance and reduce computational overhead.

Evaluation Metrics:

Common metrics used to evaluate the performance of fraud detection systems, such as accuracy, precision, recall, F1-score, ROC curve, and AUC-ROC. Emerging technologies and methodologies in credit card fraud detection, such as blockchain-based solutions, explainable AI, and federated learning.

Potential research directions and areas for improvement in terms of accuracy, efficiency, and scalability. When conducting a literature survey, it is essential to review recent papers from reputable journals, conferences, and academic repositories. Additionally, it is helpful to analyse both theoretical frameworks and empirical studies to gain a comprehensive understanding of the subject matter.

**2.1 FEATURES AND ADVANTAGES**

**Advantages:**

* **Improved Detection Accuracy:** ML algorithms can analyse vast amounts of data to identify patterns and anomalies that might indicate fraudulent activities.
* ML models continuously monitor transactions and activities.
* ML models can easily scale to handle large volumes of transactions.
* ML models can adapt to new fraud patterns and techniques, making them more effective at detecting emerging threats compared to static rule-based systems.
* It can reduce false positives by learning from past data and

Refining their detection criteria, leading t more accurate identification of fraudulent activities.

**FEATURES:**

* Machine Learning Algorithms: Employing models to detect fraudulent activities.
* Real-time Monitoring: Constantly monitoring transactions for suspicious behaviour**.**
* Pattern Recognition: Identifying common fraud patterns using data analytics**.**
* Geolocation Tracking: Verifying the location of the transaction.
* Transaction Limits: Setting limits to minimize potential losses.
* Two-Factor Authentication: Adding an extra layer of security for transactions**.**
* Card Verification Value (CVV) Checks: Verifying the CVV for online transactions**.**
* Collaboration: Sharing information and patterns with other financial institutions to enhance detection capabilities.
* Anomaly Detection: Identifying unusual patterns in transactions.
* Behavioural Analysis: Monitoring spending habits for deviations.

**2.2 LIMITATIONS AND CHALLENGES**

**LIMITATIONS:**

* **Data quality:** ML models heavily rely on quality of input data. If the data is incomplete or biased it can lead to erroneous results.
* **Overfitting:** ML modelstrained on historical data may become too specialized to that data and fail to generalize well to new, unseen data, leading to poor performance.
* Some models can be complex and difficultto interpret.
* Using sensitive customer data for training ML models raise to privacy concerns.
* Implementing and maintaining an ML-based fraud detection system can be costly and required specialized expertise in ML, cyber security, data engineering.

**CHALLENGES:**

* Data Quality and quantity
* Imbalanced Data
* Adaptability to new fraud patterns
* Model Interpretability
* Overfitting
* Adversarial Attacks
* Privacy and Compliance
* Model Explainability
* Scalability

**CHAPTER 3**

**METHODOLOGY**

The methodology for automated financial fraud detection using machine learning typically involves several key steps:

Data Collection: Gather transactional and historical data from various sources, including transaction logs, customer information, and previous fraud cases. Ensure the data is comprehensive and representative of the fraud patterns you aim to detect.

Data Preprocessing: Clean the data to remove duplicates, fill in missing values, and standardize formats. Perform feature engineering to extract relevant features that can help the ML model identify fraud patterns**.**

Feature Selection: Select the most relevant features for training the ML model. Use techniques such as correlation analysis, feature importance, and domain knowledge to identify the most predictive features.

Model Selection: Choose the appropriate ML model(s) for fraud detection. Commonly used models include logistic regression, decision trees, random forests, and neural networks. Consider factors such as model complexity, interpretability, and scalability.

Model Training: Split the data into training and validation sets. Train the selected ML model(s) on the training set using techniques like cross-validation to optimize performance and prevent overfitting**.**

Model Evaluation: Evaluate the trained model(s) using the validation set to assess performance metrics such as accuracy, precision, recall, and F1 score. Adjust the model parameters as needed to improve performance**.**

Model Deployment: Deploy the trained model(s) into a production environment for real-time fraud detection. Implement monitoring mechanisms to ensure the model performs effectively and efficiently**.**

**3.1. PACKAGES USED**

* Pandas
* NumPy
* TensorFlow
* OS
* Datetime
* Sklearn.metrics
* Seaborn
* matplotlib. Pyplot
* matplotlib. grid spec.

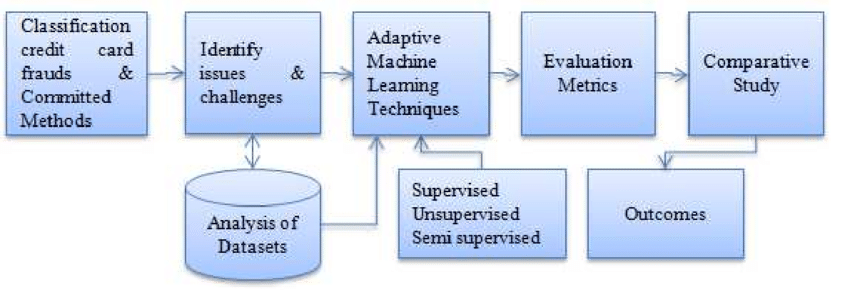
**3.2. DATA COLLECTION METHODS AND SOURCES**

* Transaction data
* Customer data
* Merchant data
* Device and IAP address data
* External data sources
* Behavioural Biometrics
* Machine Learning models
* API and Web scraping
* Fraudulent Activity reports

**CHAPTER 4**

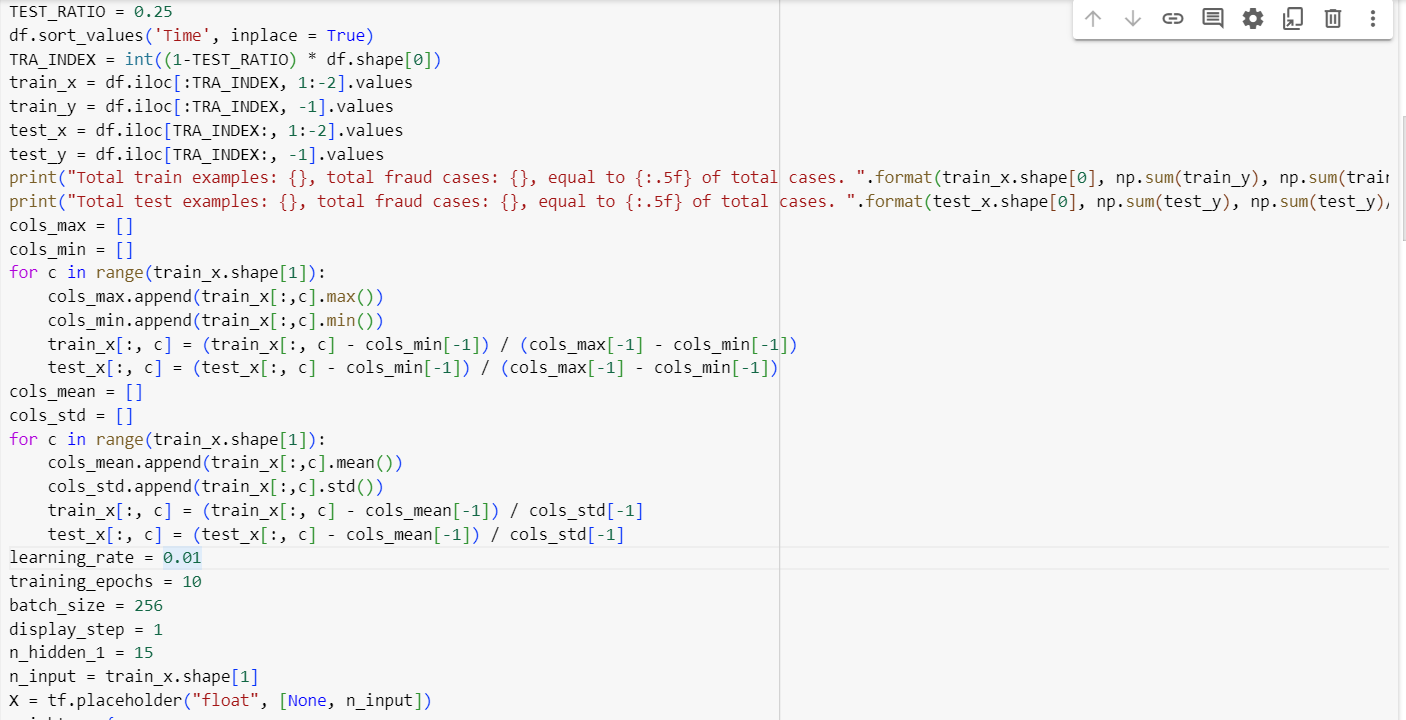
**PROPOSED WORKS**

**4.1. Flow Chart**



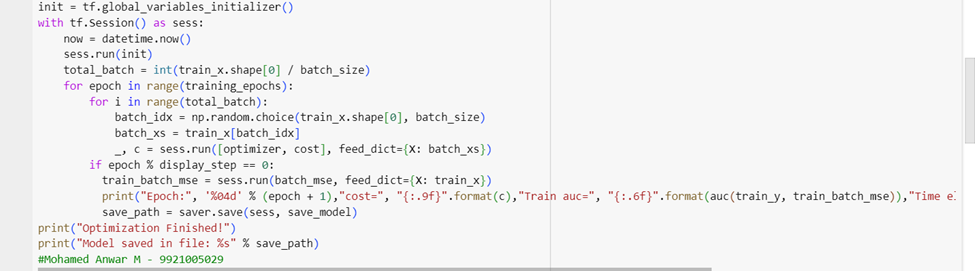
**4.2 CODE Implementation**

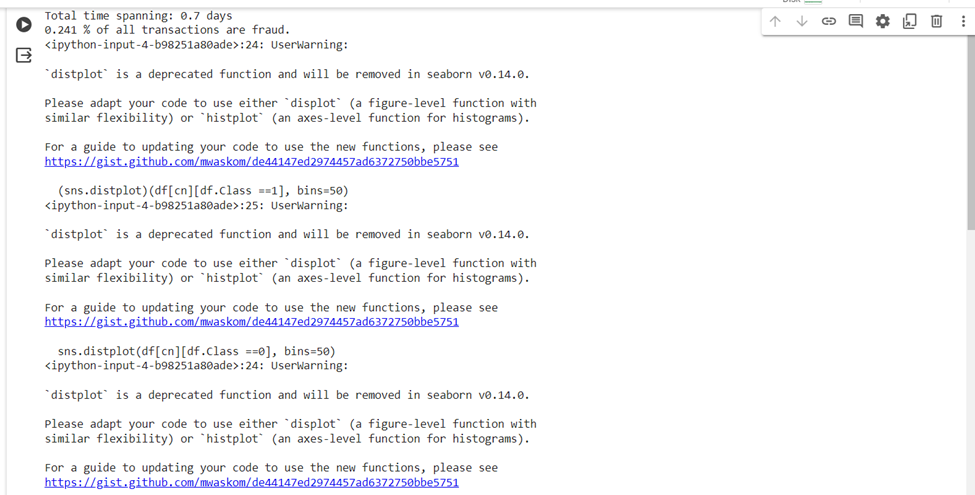
****

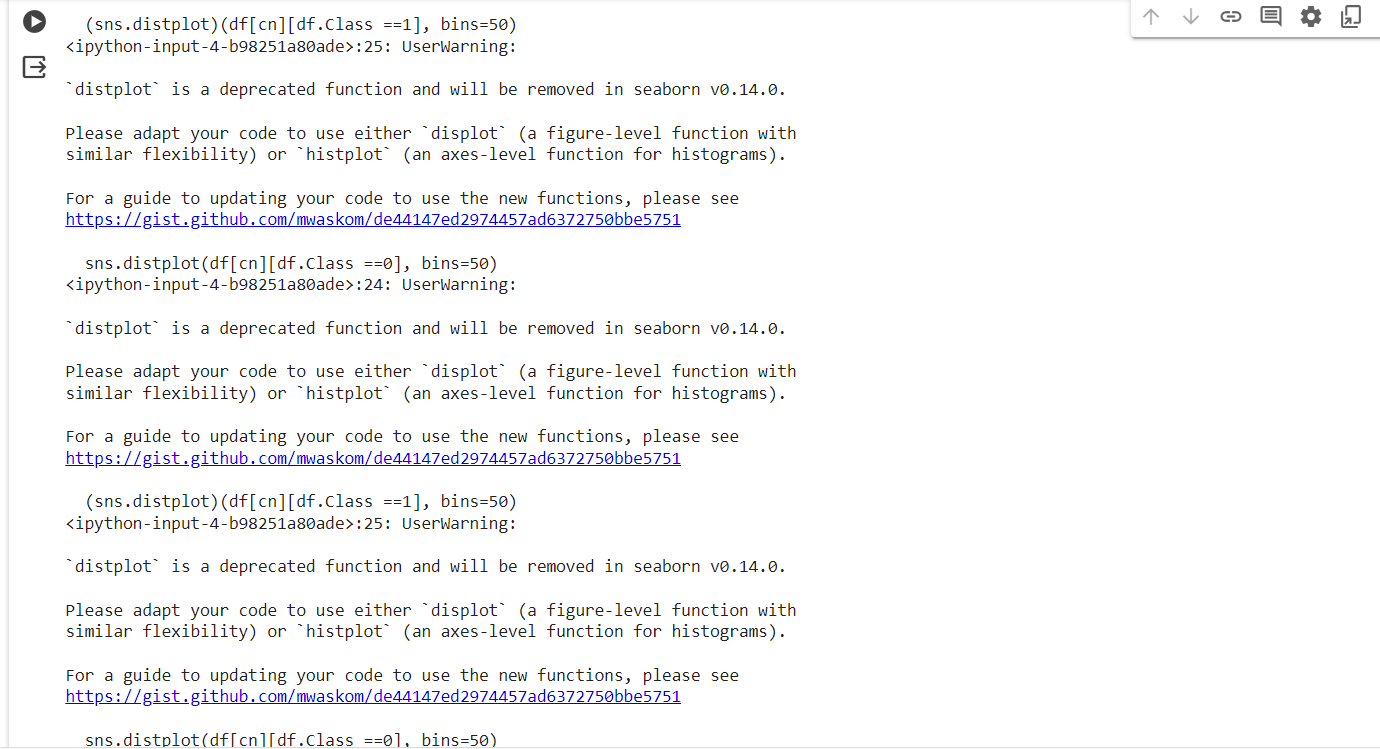
****

**A screenshot of a computer

Description automatically generated**

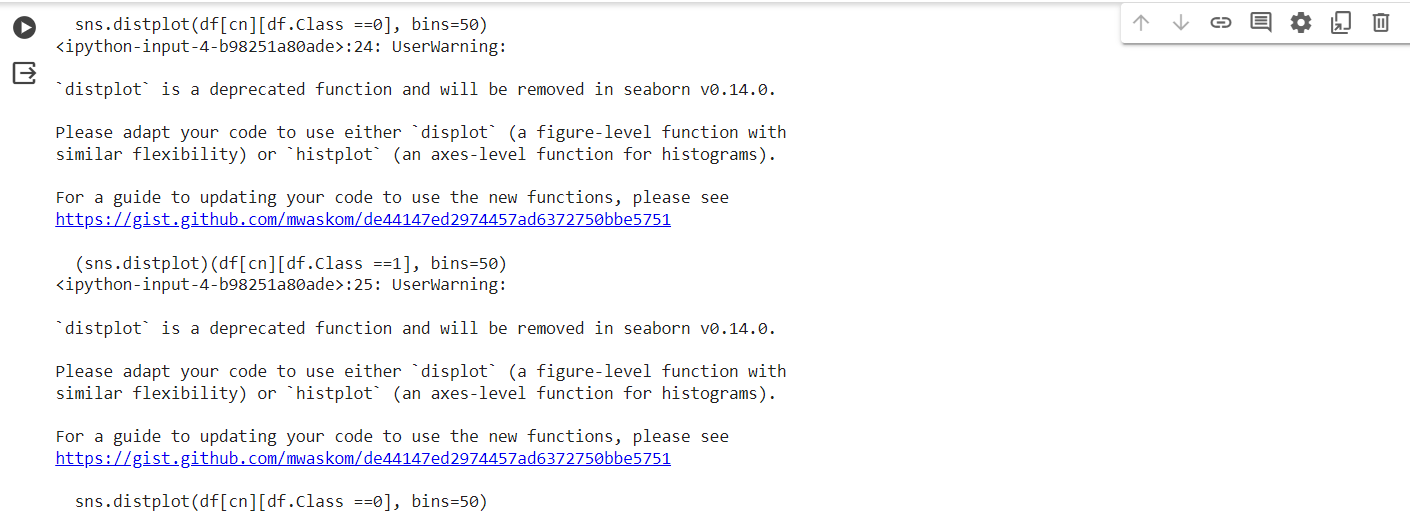


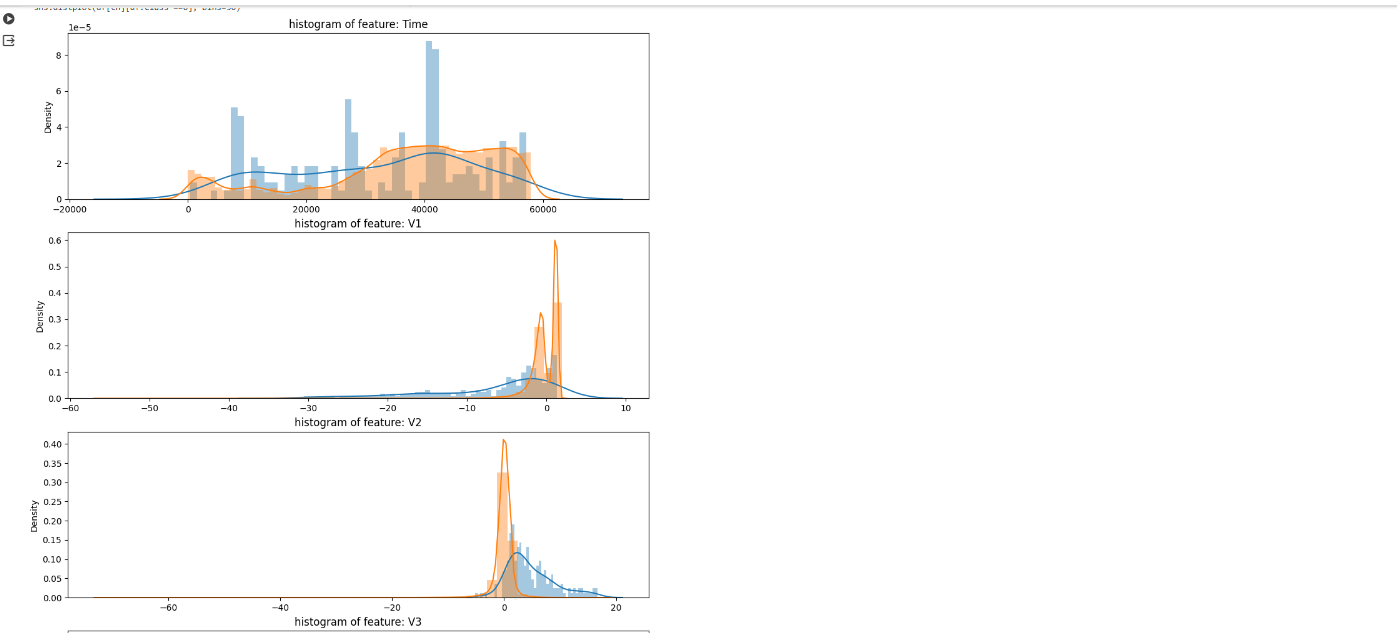
****

****

**A screenshot of a computer

Description automatically generated**

****

****

**A graph with a line and numbers

Description automatically generated with medium confidence**

**Chapter 5**

**REFERENCE PAPERS**

<https://scholar.google.co.in/scholar?as_ylo=2024&q=reference+paper+links+on+automated+financial+fraud+detection+using+machine+learning&hl=en&as_sdt=0,5&as_vis=1_>

1. "Fraud Detection Using Machine Learning: A Review" by Ronald J. van der Aalst, Rob J.M. Vaessen, and Marc J.M. van Kervel.
2. "Detecting Financial Statement Fraud: Three Essays on Fraud Predictors, Multi-Classifier Combination, and Fraud Detection Systems" by Eunjoo Cho.
3. "A Survey of Credit Card Fraud Detection Techniques: Data and Technique-Oriented Perspective" by Shehab Gamalel-Din and Waleed G. Aref.
4. "Anomaly Detection: A Survey" by Varun Chandola, Arindam Banerjee, and Vipin Kumar. While not specific to financial fraud, this survey paper covers various anomaly detection techniques, which are often used in financial fraud detection systems.
5. "Deep Learning for Anomaly Detection: A Survey" by Chong Zhou and Randy C. Paffenroth.

1. "Fraud Analytics: Strategies and Methods for Detection and Prevention" by Delena D. Spann.
2. "Machine Learning for Financial Fraud Detection: A Systematic Literature Review" by Andrea Dal Pozzolo, Olivier Caelen, Reid A. Johnson, and Gianluca Bontempi.
3. "Credit Scoring and Its Applications" edited by Lyn C. Thomas and David B. Edelman.
4. "Handbook of Financial Cryptography and Security" edited by Burton Rosenberg.
5. "The Data Science Handbook" edited by Field Cady. This handbook offers insights into various aspects of data science, including machine learning and data mining techniques used in fraud detection and prevention in finance and other industries.