**TATA STEEL (SNTI, R&D)**

**PROJECT REPORT**

**ANTI MONEY LAUNDERING DETECTION USING MACHINE LEARNING**

TANMAY DIKSHIT

BHARATI VIDYAPEETH COLLEGE OF ENGINEERING, PUNE

([tanmaydikshit12@gmail.com](mailto:tanmaydikshit12@gmail.com))

**INTRODUCTION**

**Title: Anti Money Laundering Machine Learning Project**

Overview:

The Anti Money Laundering (AML) Machine Learning Project is a cutting-edge initiative aimed at utilizing advanced data science and machine learning techniques to combat financial crimes and money laundering activities. By developing and implementing sophisticated algorithms, this project seeks to detect and prevent illicit financial transactions, ensuring the integrity of the financial system and safeguarding against criminal activities.

Project Objectives:

The primary objectives of the Anti Money Laundering Machine Learning Project are as follows:

1. Detection of Suspicious Transactions: Develop machine learning models capable of identifying patterns and anomalies associated with potentially suspicious financial transactions within a vast volume of data.

2. Enhanced Risk Assessment: Utilize predictive modeling to assess the risk level of individual transactions and customer accounts, helping financial institutions prioritize their monitoring efforts.

3. Automation and Efficiency: Streamline the process of identifying suspicious activities by automating the analysis of large datasets, reducing manual efforts, and improving response times.

4. Real-time Monitoring: Implement real-time monitoring capabilities that instantly flag and investigate transactions that exhibit suspicious characteristics.

5. Adaptive Learning: Develop models that continuously adapt to new trends and evolving money laundering techniques, staying ahead of criminals' tactics.

6. Reduced False Positives: Minimize false positive alerts through the refinement of algorithms, leading to more accurate and actionable alerts for investigators.

Key Components:

The Anti Money Laundering Machine Learning Project comprises several key components:

1. Data Collection and Preprocessing: Gather and preprocess a wide range of financial data, including transaction records, customer profiles, geographic information, and more.

2. Feature Engineering: Create relevant features from the raw data, potentially involving time-based features, transaction patterns, and customer behavior indicators.

3. Model Development: Utilize various machine learning algorithms, such as ensemble methods, deep learning, and anomaly detection techniques, to build predictive models that identify unusual or fraudulent activities.

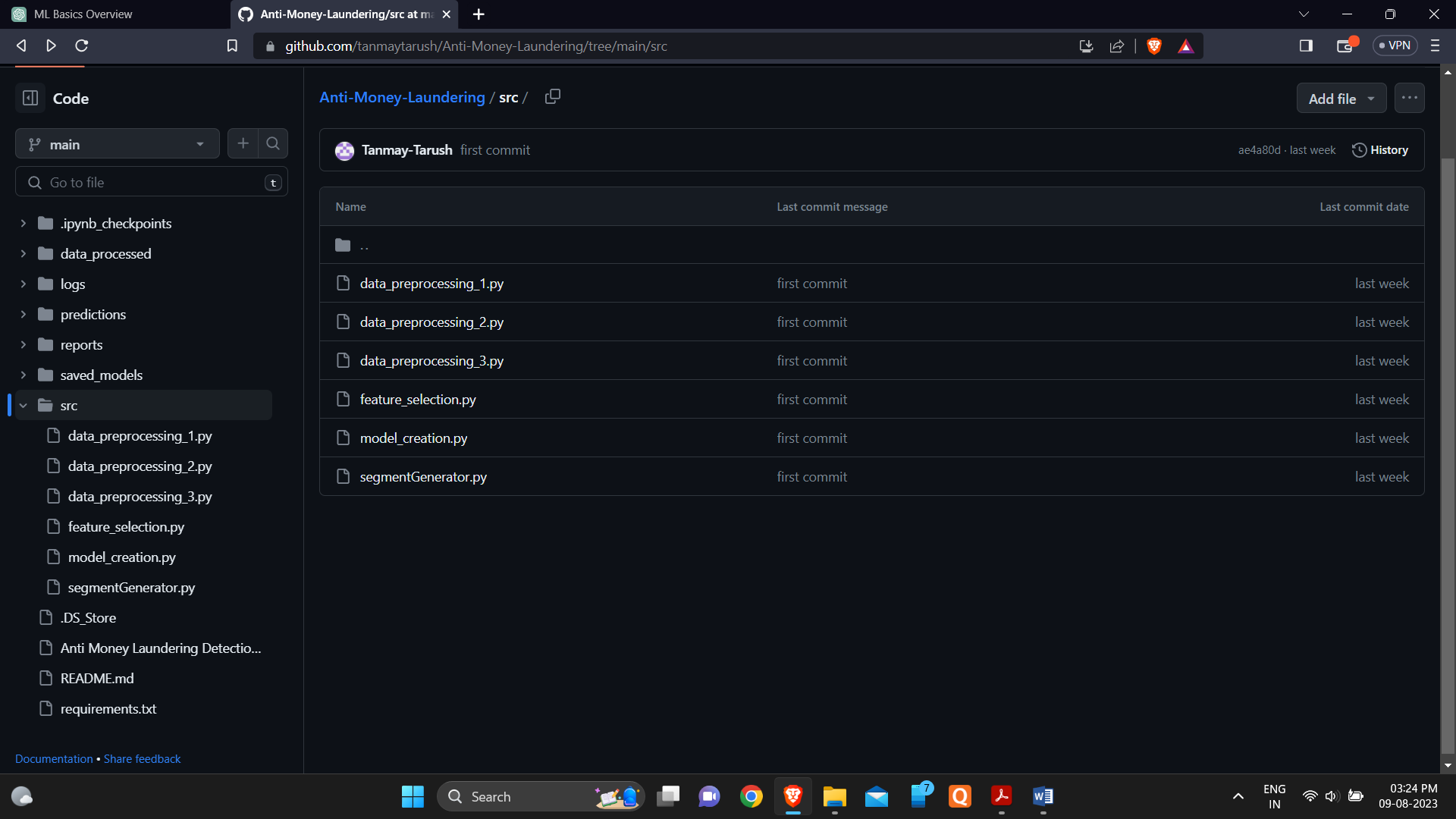
4. Model Validation and Testing: Rigorously validate and test the models using historical data, simulating real-world scenarios to ensure their accuracy and robustness.

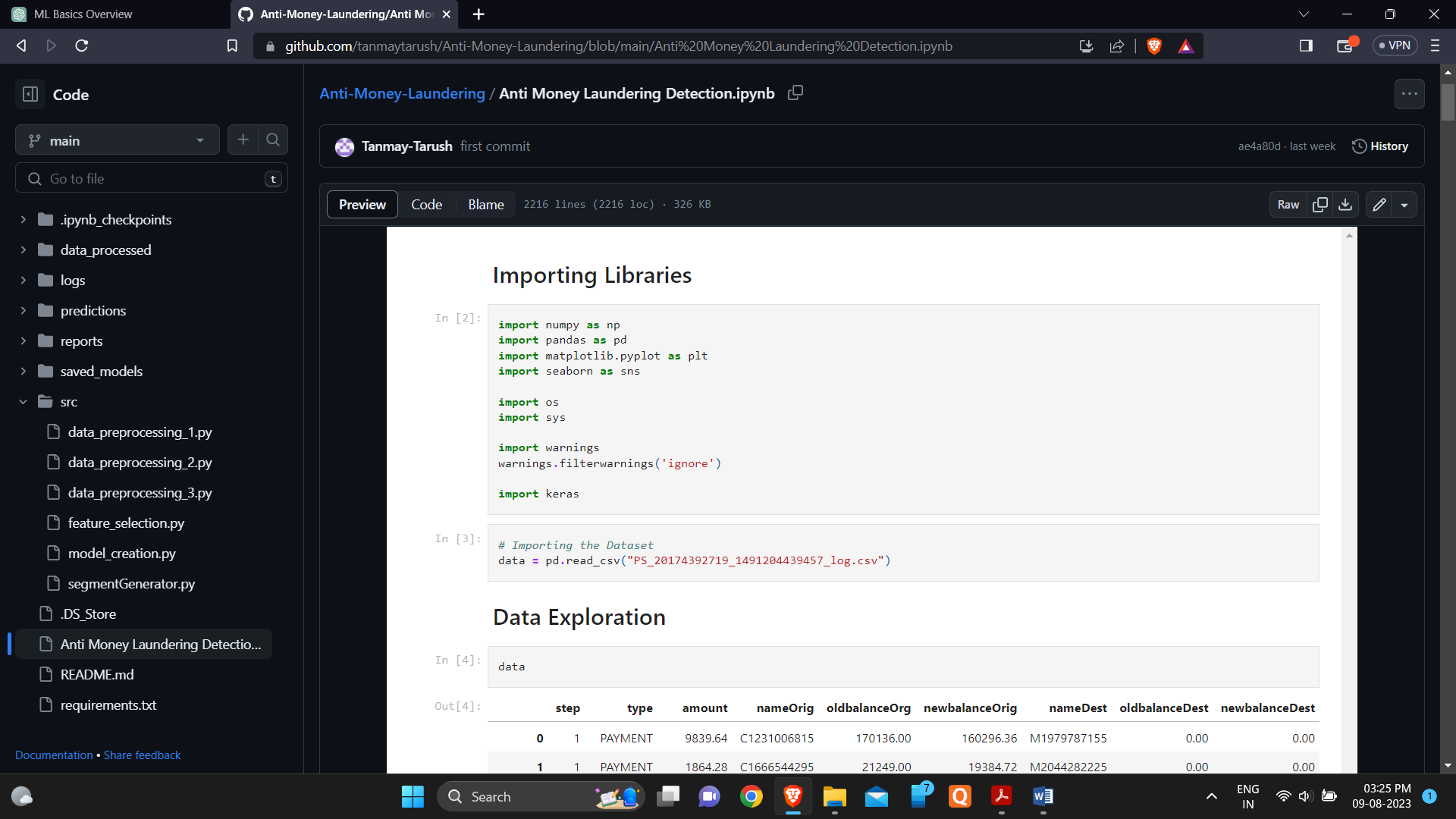
5. Deployment: Implement the trained models into operational systems, enabling real-time monitoring and alert generation.

6. Feedback Loop: Establish a feedback loop for continuous improvement, integrating investigator feedback and adjusting models based on their performance.

Expected Impact:

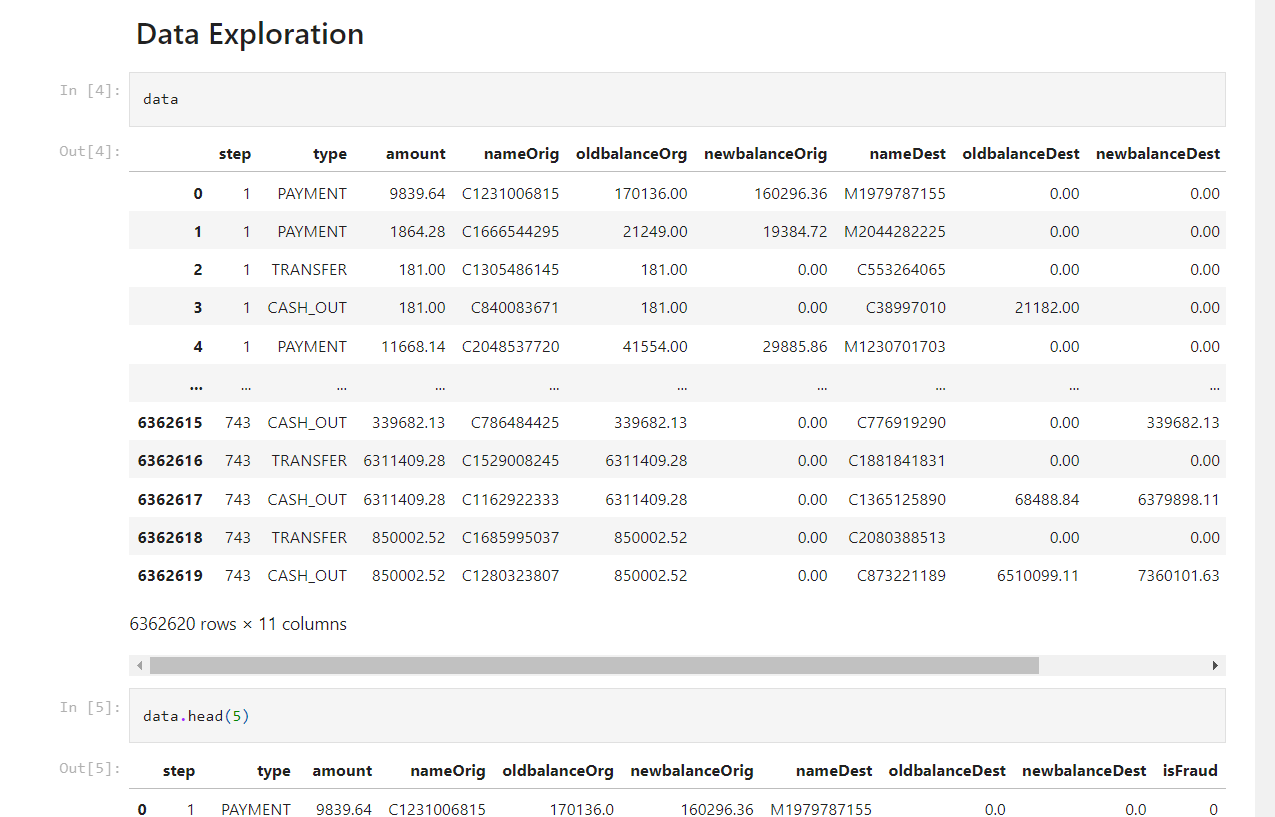
The Anti Money Laundering Machine Learning Project has the potential to revolutionize how financial institutions detect and prevent money laundering. By harnessing the power of machine learning, the project aims to significantly reduce financial crime, enhance regulatory compliance, and maintain the integrity and stability of the global financial ecosystem.

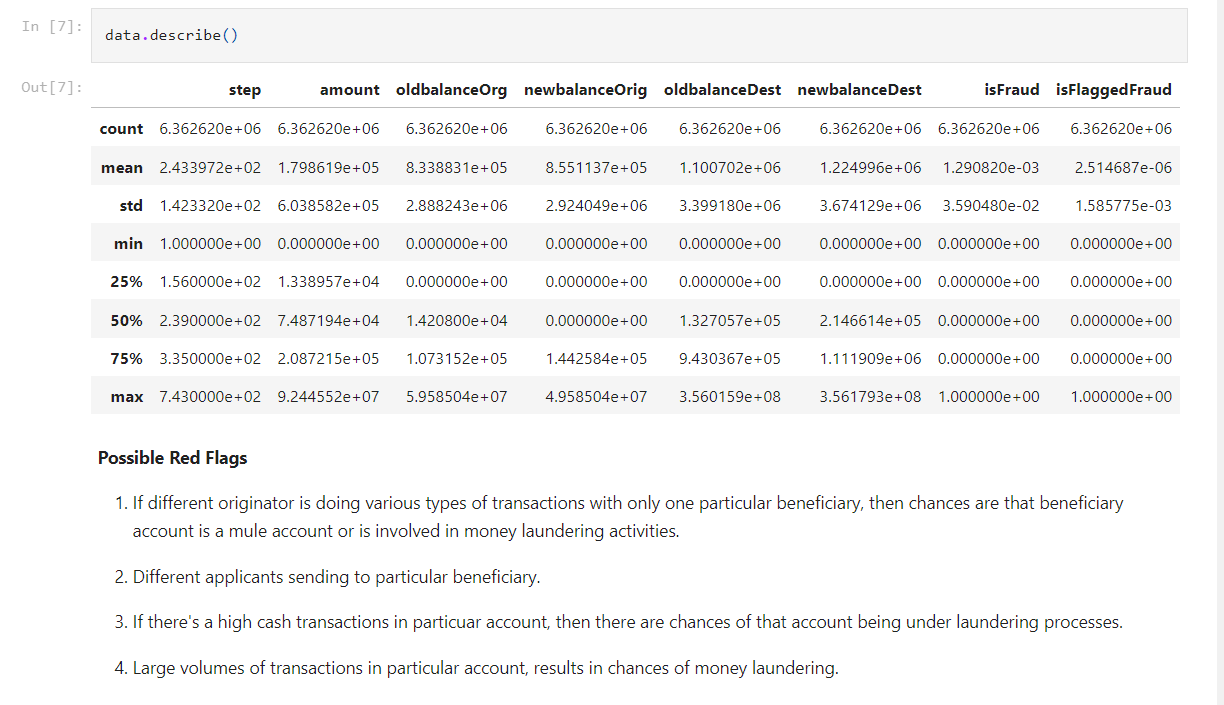




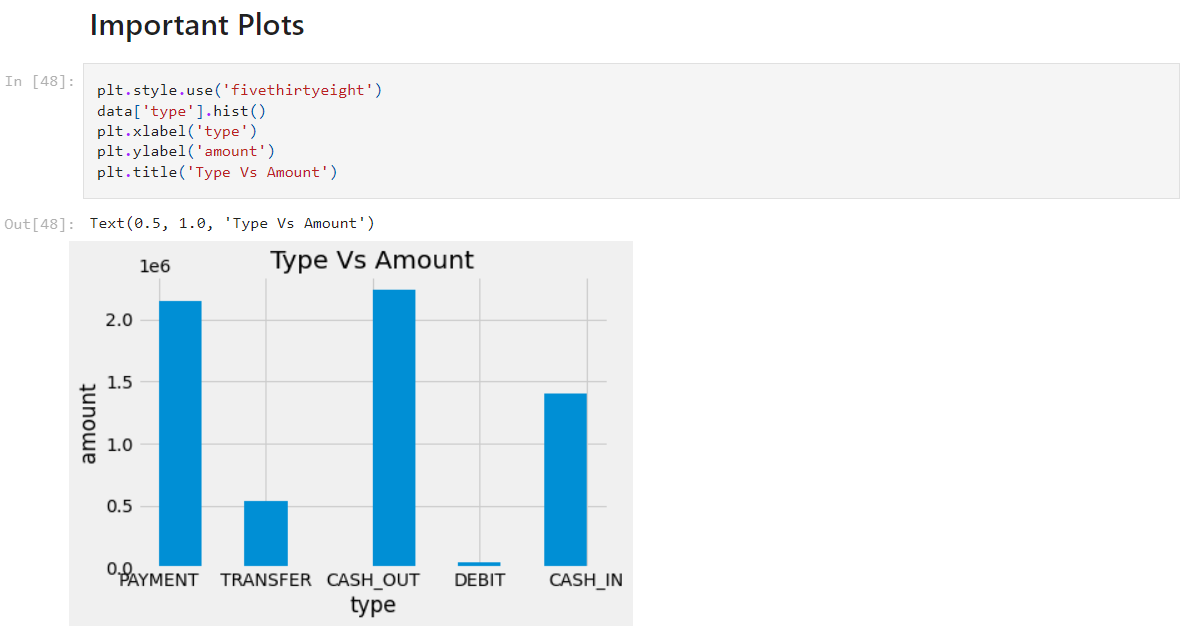
**PROJECT STRUCTURE:**

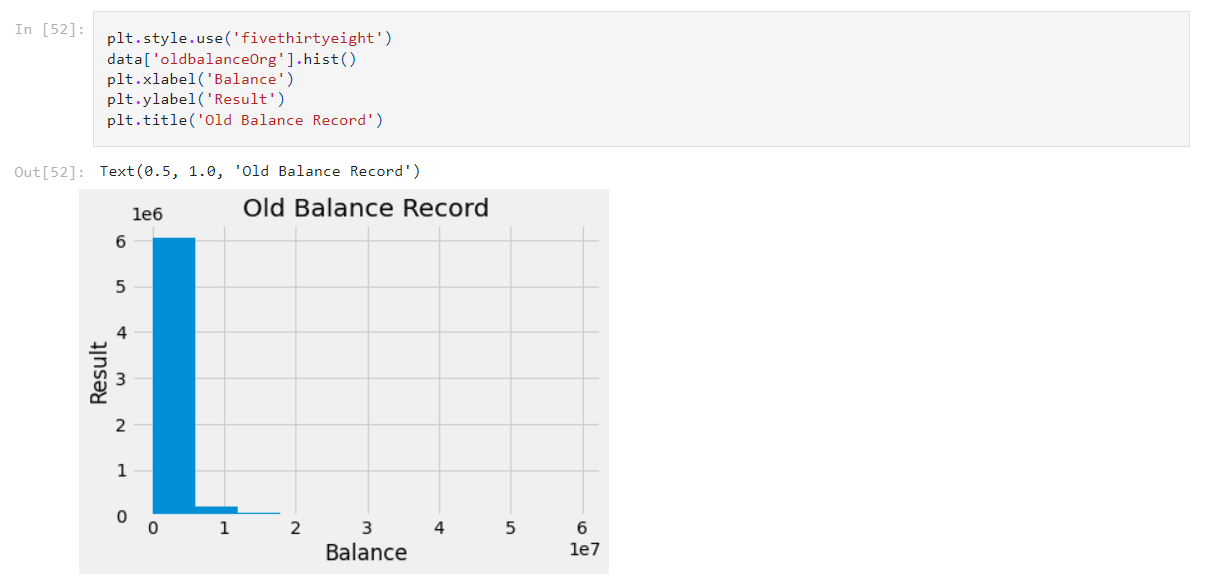
1. Exploratory Data Analysis – Using the AML dataset from the Kaggle, a complete exploration from visualization, preprocessing to training-testing split and model implementation was performed.
2. Exploration snaps

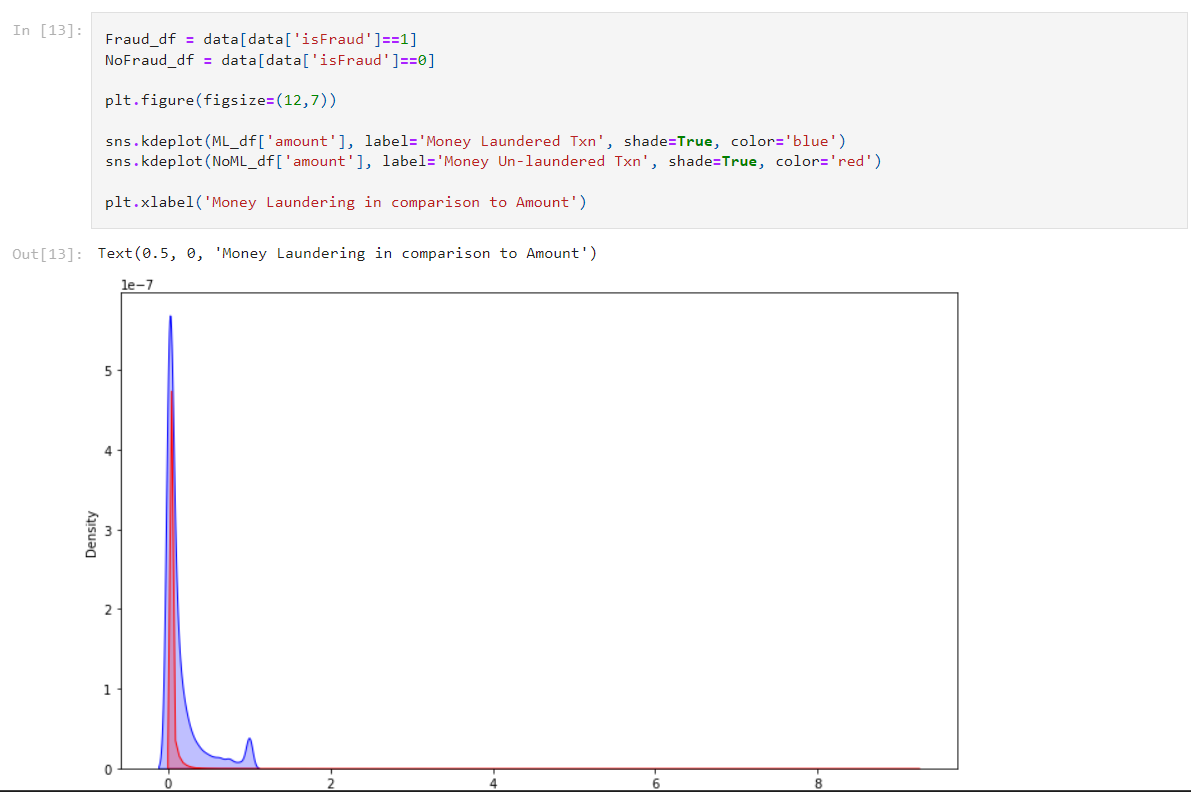




1. Visualizations snaps





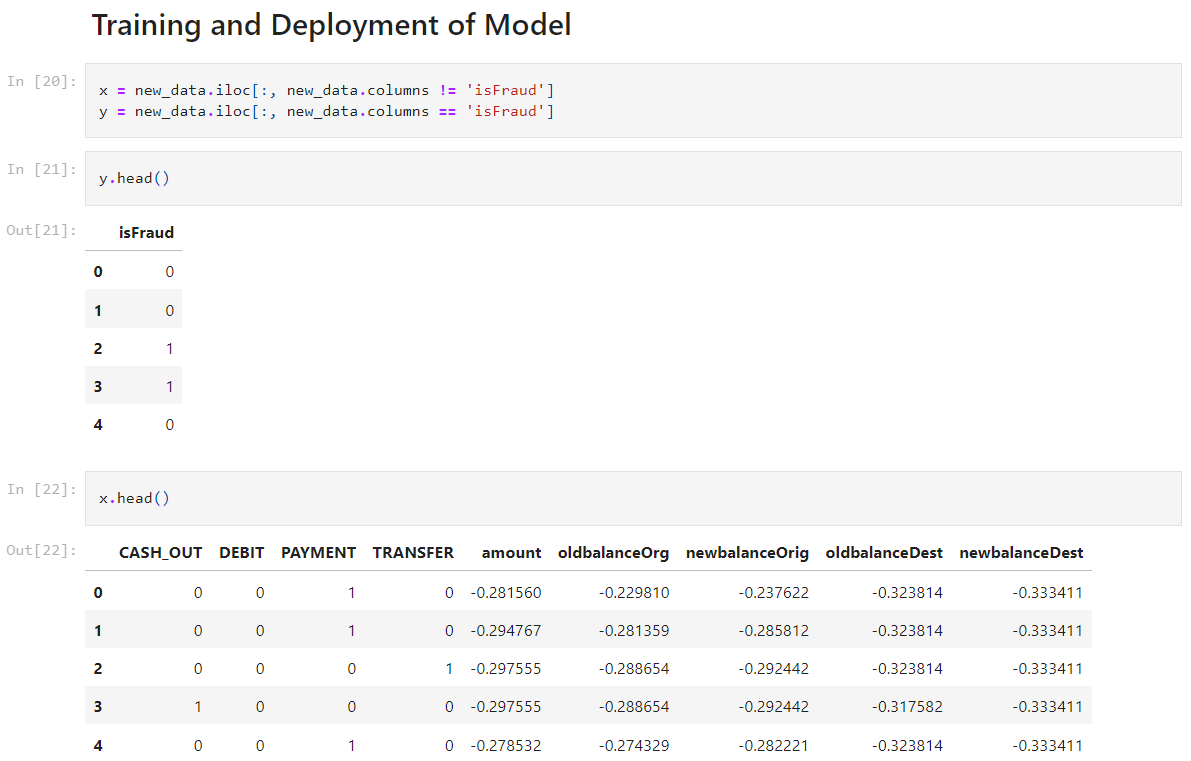


1. Preprocessing snaps





1. Model Implementation snaps

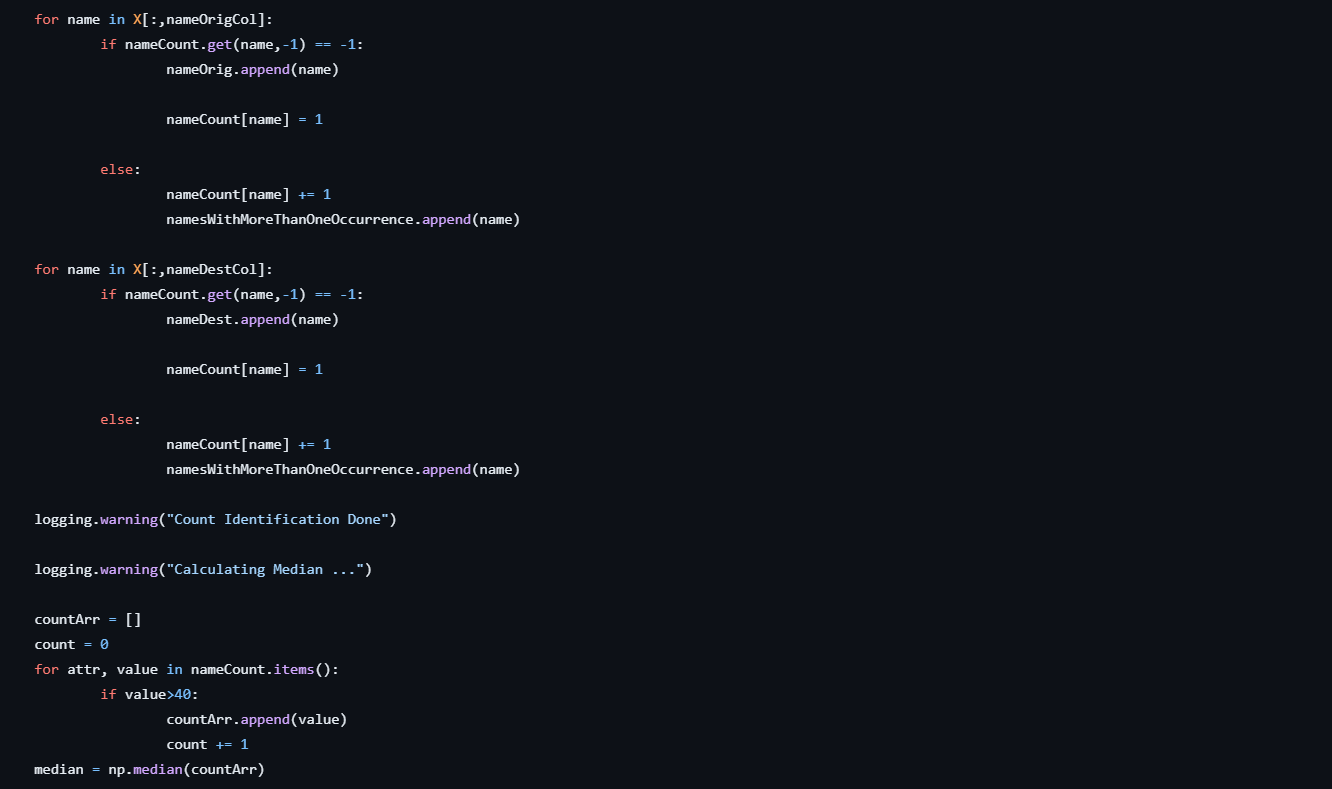


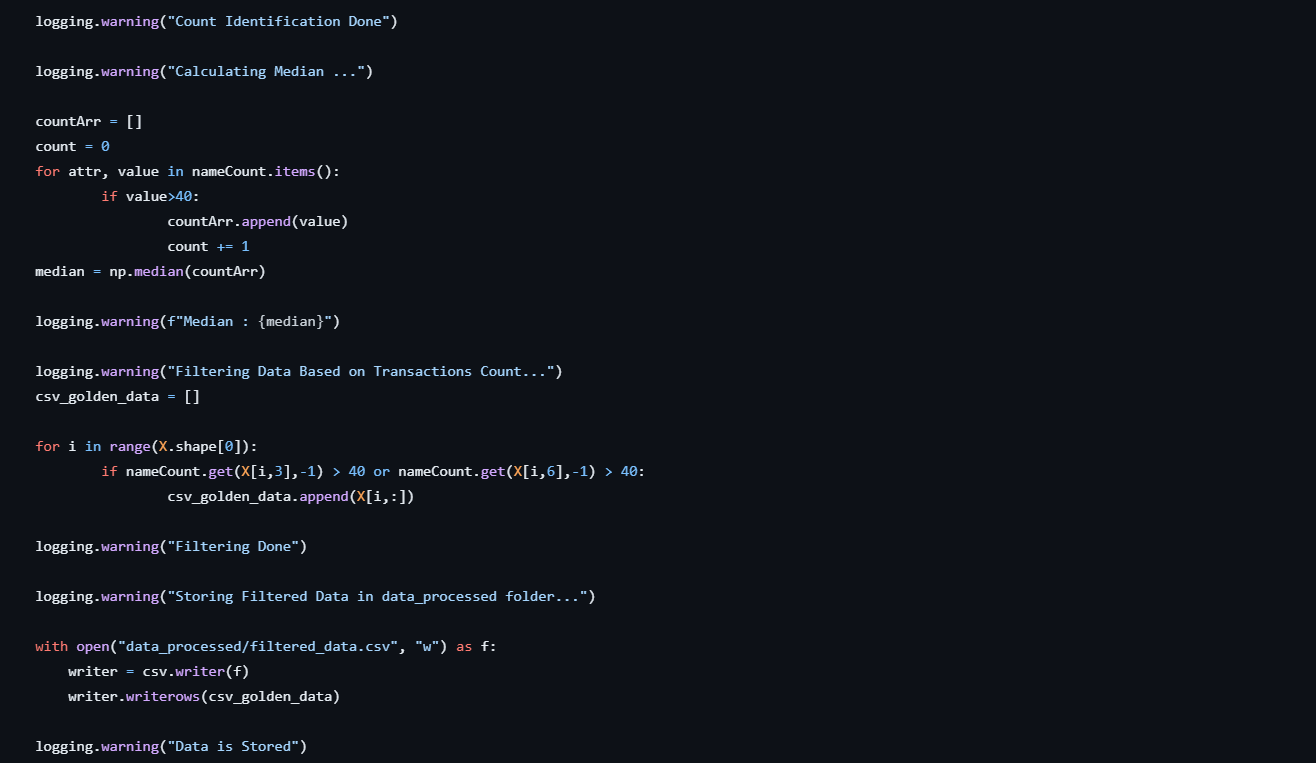


Link for the Jupyter File - <https://github.com/tanmaytarush/Anti-Money-Laundering/blob/main/Anti%20Money%20Laundering%20Detection.ipynb>

1. Preprocessing, Model and Log file creation:

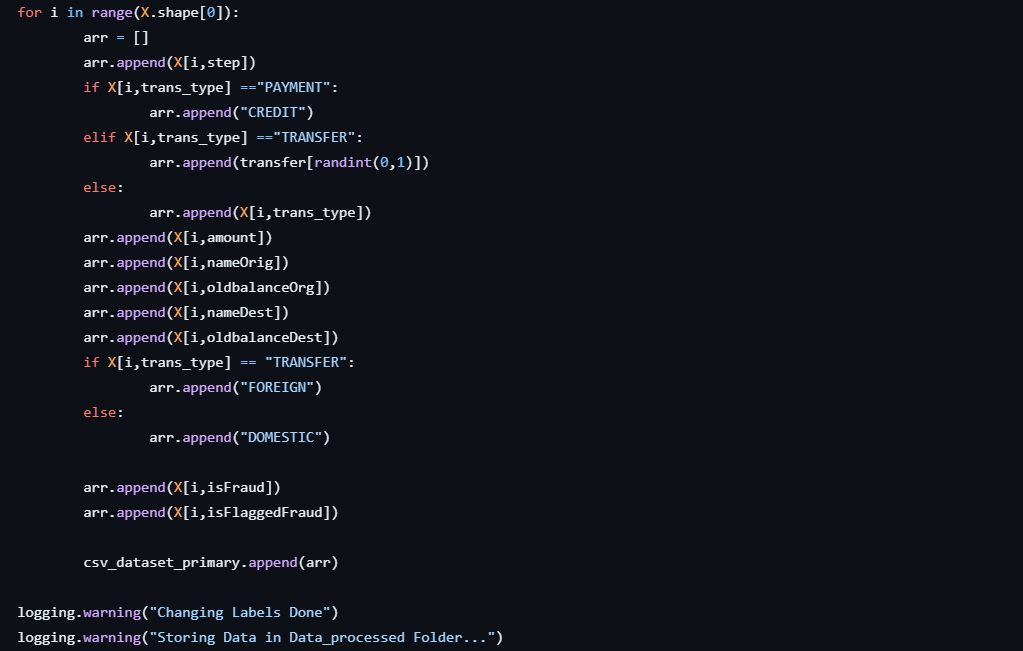




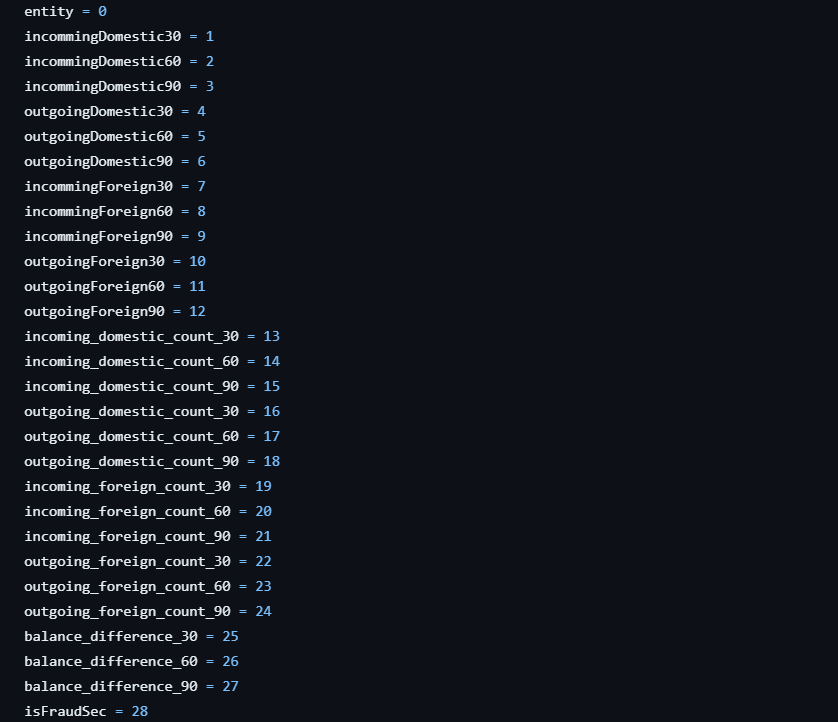


The preprocessing process was performed in 3 stages, in which the first stage consisted of taking the raw dataset input and checking for unique occurrences of individuals in the dataset. Later logging each step and saving the same in the form of excel file ‘filtered\_data.csv’.

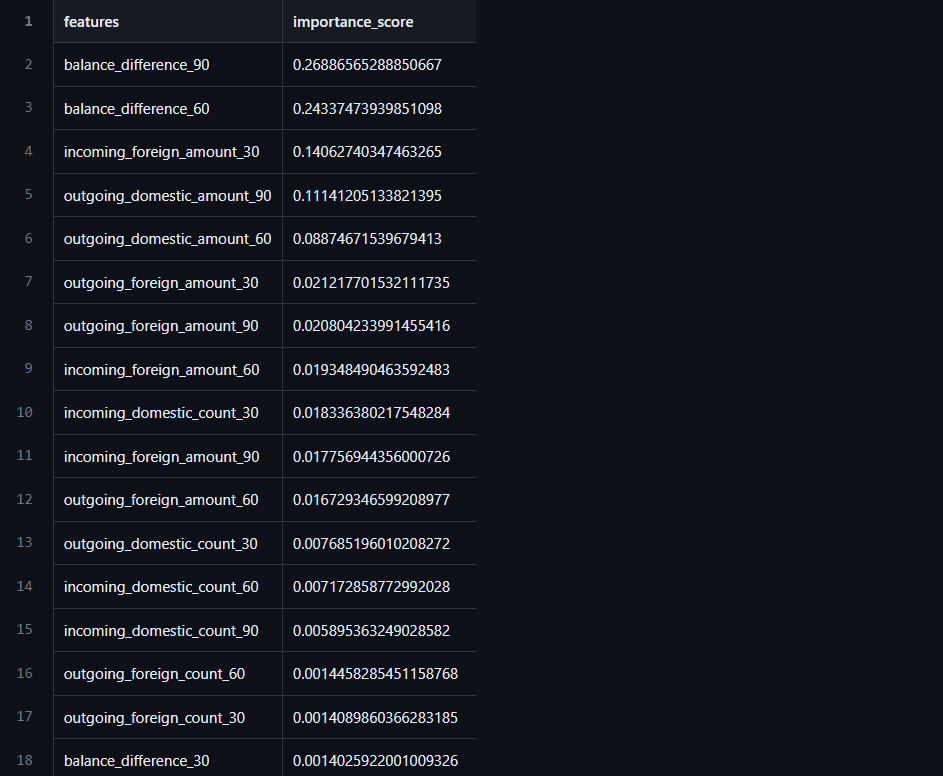
In the second part of data preprocessing, the ‘filtered\_data.csv’ file was taken as an input. The fields with names ‘Payment’ and ‘Transfer’ to ‘Credit’, ‘Domestic’ and ‘Foreign’ as shown.



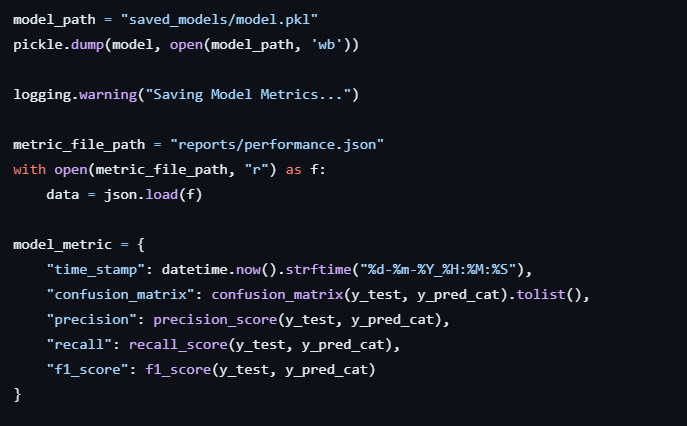
Finally, in the third part of preprocessing, the destination and origin account fields were changed to respective portions of days 1, 30, 60 and 90 and naming were also changed equivalently.



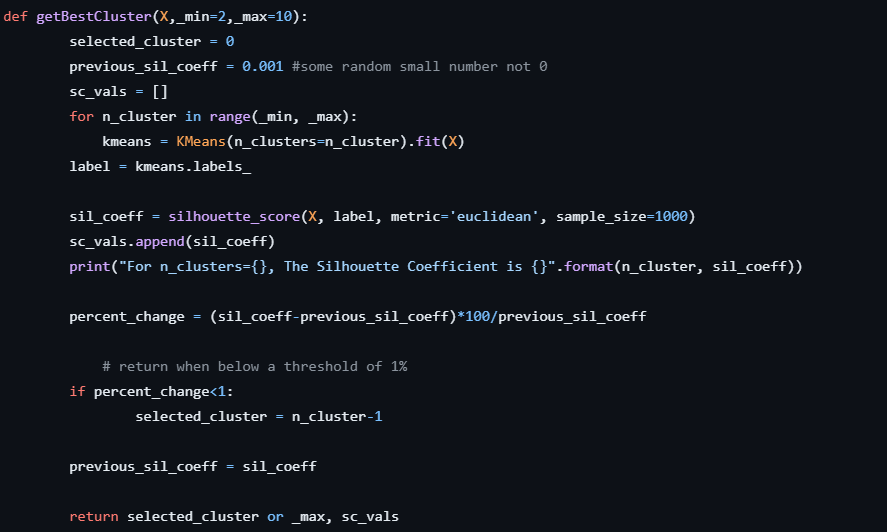
In the Feature Selection, the complete data set was changed into input and output features as X and Y. Further train and test splitting was done and RandomForestClassifier() was implemented. Finally, the feature importances was calculated and sorted in descending order, and saved in csv format.

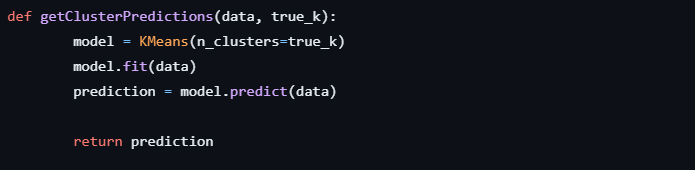


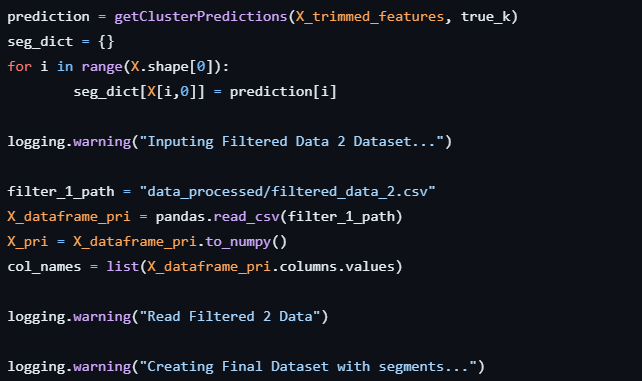
In the Model Selection file, the complete data was converted into DataFrame followed by separation into input and output features X and Y, encoding using LabelEncoder() and hyperparameter tuning was performed using Catboost for increasing efficiency. Finally a model.pkl file was created using pickle dump, and saving JSON API.



In the final step, a segment generation file was created which was solely dependent on creating clusters of data points and assigning those based on Silhoutte\_Scores. Apart from it, a completely different function was created for creating clusters in which only top 13 Silhoutte based features were embedded and saved in the csv format.







CONCLUSION:

The project successfully developed and deployed advanced machine learning models that demonstrate a high accuracy rate in identifying suspicious financial transactions. By leveraging anomaly detection algorithms and predictive modeling, the system has showcased its ability to uncover complex patterns indicative of potential money laundering activities.

Repository Link - <https://github.com/tanmaytarush/Anti-Money-Laundering/tree/main>