COMP1831 Coursework

Submitted by:- **Sohel Arsad Salam**
Student ID:- **001166291**

Subject:- **COMP1831 – Technologies for Anti-Money Laundering and Financial Crime**

SCHOOL OF COMPUTING AND MATHEMATICAL SCIENCE

Table of Content:

1. A Background Section:

1.1. What is Money Laundering:

Money laundering is an illegal act of bringing in a large amount of money earned through a crime. Subsidies for drug trafficking or fear-based oppressors appear to have been created from genuine sources. Money from crime is considered dirty and interactions make it "wash out" and look pristine. This is a serious financial fraud that is similarly abused by top criminals and street criminals. Most monetary organizations are hostile to the Money Laundering Prevention (AML) clause introduced to identify and prevent this behavior. This is an illegal way to make "dirty" money look real and not make bad money. Hoodlums uses a variety of tax evasion strategies to clean up the unearned profits. Internet banking and digital currencies have made it easier for criminals to move around unnoticed and withdraw cash. Avoiding tax evasion has become a global effort, and one of its goals now includes subsidizing fear-based oppressors.

There are three phases in Money Laundering. The principal stage, Placement, is when illicitly acquired reserves are brought into the genuine economy. The subsequent stage, Layering, is when hoodlums conceal the starting points of the unlawful assets by reallocating them in more than one way which darkens the source. The last stage, Integration, is the point at which the illegal assets are separated and utilized, presently masked as legitimate money.

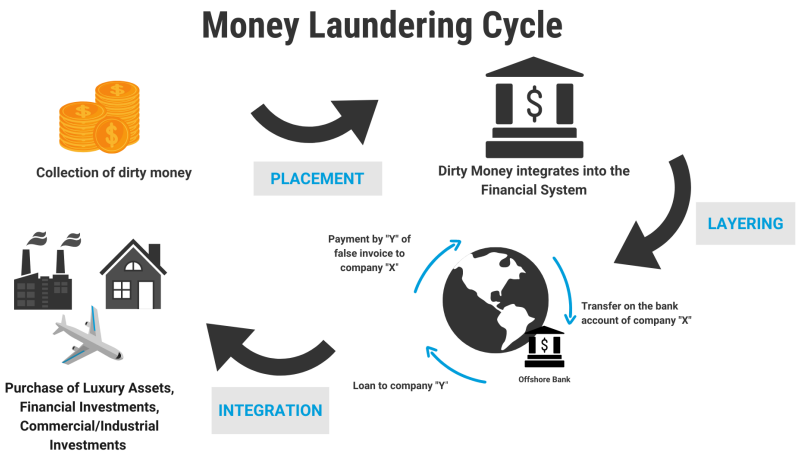


Figure 1:Money Laundering Cycle

1.2. What is Financial Technology(FinTech):

FinTech is used to represent new technologies aimed at improving and computerizing the use of money transfers and money management. At its core, Fintech uses specific programs and calculations used in personal computers and increasingly mobile phones to help organizations, entrepreneurs and buyers improve their financial activities, cycles and lives. Used for. FinTech also includes the order of events and the use of encrypted forms of money such as Bitcoin. This fragment of FinTech may see

most stocks, but the real big bucks are in the traditional global financial industry and its trillions of dollars in market capitalization.

1.3. Money Laundering in the UK and International Market:

Money Laundering can possibly compromise the UK's National Security, Prosperity and worldwide standing. The UK benefits from a functioning and dynamic business climate, upheld by a set number of limitations on laying out a business. The simplicity with which an organization can be laid out is oftentimes taken advantage of by lawbreakers who set clearly authentic organizations both inside UK and overseas, however which are basically an instrument for laundering illegal assets. The UK has one of the world's biggest and most open economies with London being especially alluring for abroad financial backers. Research proposes, the UK is the world's driving net exporter of monetary administrations, close to being a significant place for proficient administrations that help monetary administrations. Similarly, the World Bank positions the UK, 8th on the planet for its simplicity of doing business. These variables make the UK appealing for real business, however, additionally open the UK to money laundering risks from abroad.

Figure 2:Global AML Regulators.

In this globalized world, the monetary economy relies heavily on the good functioning of financial institutions. Financial institutions should operate on legal, competent and moral standards. The infamy of financial fundamentals integrity is very important. In such a difficult situation, if the assets of criminal activity and money laundering activity are allowed to flow into financial institutions, it will adversely affect the reputation and will be published in the newspaper the next day. However, when workers and chiefs are rewarded, the reserves created in the money laundering exercises can move and provide the basis for funding to criminal organizations. Evidence of such accomplices undermines the integrity of financial institutions. By the

time various financial institutions perform such exercises, the economy may be at stake and changes in cash and transaction rates will occur. Finally, humiliation and cheating are compensated. Effective tax evasion undermines the honesty of society as a whole and undermines the electoral system and legal standards.

1.4. Reduced Effectiveness of existing AML solutions:

Money Laundering is an alarming issue for the worldwide economy, with the aggregates included differently assessed at somewhere in the range of 2 and 5 percent of worldwide GDP. Monetary foundations are expected by controllers to help battle illegal tax avoidance and have contributed billions of dollars to agree. By the by, the punishments these foundations cause for consistent disappointment keep on ascending: in 2017, fines were broadly announced as having added up to $321 billion beginning around 2008 and $42 billion out of 2016 alone.2 This recommends that not entirely settled to break down yet additionally that lawbreakers are turning out to be progressively modern. Client risk-rating models are one of three essential devices utilized by monetary organizations to recognize money laundering. The models conveyed by most organizations today depend on an appraisal of chance elements like the client's occupation, compensation, and the financial items utilized. The data is gathered whenever a record is opened, yet it is rarely refreshed. These contributions, alongside the weighting each is given, are utilized to compute a gamble rating score.

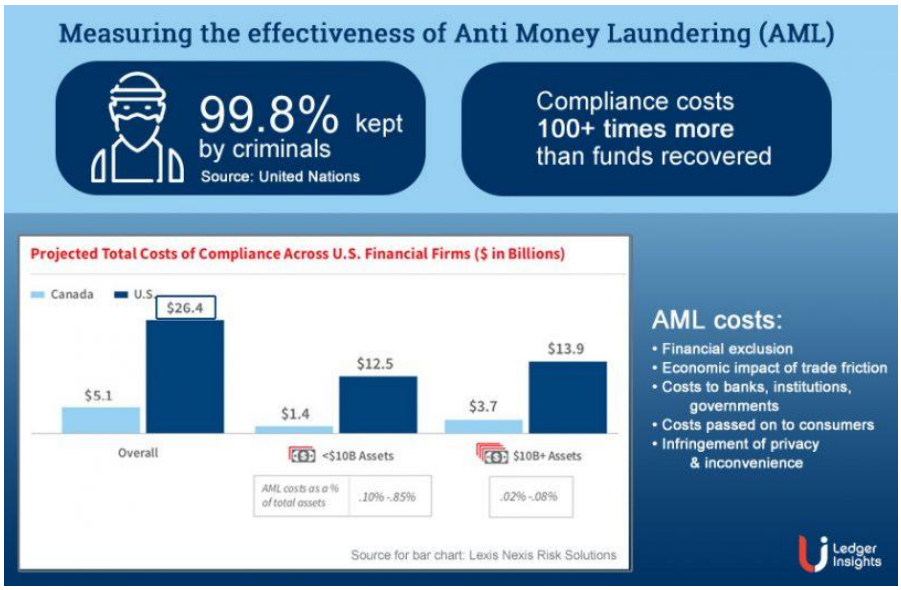


Figure 3:Global AML Regulators.

2. High-level Explanation of the service for AML:

The risk-based approach ought to be the foundation of a successful AML as it plays a crucial role in managing risks. Machine Learning is a subnet of AI that trains computer frameworks to learn from data. Distinct examples are settled on choices with negligible human intercession. ML includes planning a succession of activities to take care of an issue naturally through experience and

advancing example acknowledgement algorithms with restricted or no human intercession i.e. it is a strategy for information investigation that uses computerized scientific model structure. Respondents refer to ML and natural language processing as an extraordinary advantage to AML for regulators and supervisors. ML allegedly offers the best benefit through its capacity to gain from existing frameworks, diminishing the requirement for manual contribution to observing, lessening bogus up-sides and distinguishing complex cases, as well as working with risk the board. ML applications are helpful for distinguishing irregularities and anomalies, distinguishing and disposing of copy data to further develop information quality and examination. For instance, Deep Learning (DL) is a high level sort of AI in which neural networks (algorithms motivated by the human mind) with various layers gain from a lot of information in exceptionally independent ways. DL algorithms play out an assignment more than once, each time tweaking it a little to work on the result, empowering machines to take care of intricate issues without human intercession.

3. Data Access Needs:

A variety of data is needed to make the AML solution productive. Like having to do a background check, we need to find out the investment point of funds and also the assets that are being purchased with the funds. An immense part of money laundering is purchasing resources that can be flipped for multiple times the original price, for instance, Auctions, Art Exhibitions, Casinos and so on. People often spend millions in these areas and what looks like legit funds could be a transaction from an illegal arms deal, Human trafficking, Drug Deals and whatnot. Keeping in check these details will give us enough intel to figure out any illicit funds or unethical businesses going on.

Personal data is also required as it gives more insights towards AML. Collection of the data like bank details, business transaction and personal transaction summed up to a nexus and makes the flow of money hard to comprehend. We need to analyze these data to make sure the money is legitimate and that it is excluded from any illegal business. Once we successfully trace back all the transactions to its source, we can then find out all the illegal activities.

4. Data Privacy:

Data ingested from either source is encrypted using the most advanced encryption available. Only those who are working on an individual's background are granted access to that individual's data. To gain access, employees must fill out a form detailing the reason for the access, the required data, and the duration of the access. The form is forwarded to a higher authority, where it is analyzed to determine if the application is legal. Some forms allow employees restricted access. All order-permitted orders are accessed weekly by another department, and if an unjustified order-permitted order is found, it will be investigated and all staff involved will be thoroughly questioned. The hybrid hosting approach is an ideal solution for this type of project. Hybrid hosting is a combination of public and private hosting. Organizations can host web applications on public servers and connect to private, more secure data storage servers to ensure data protection while ensuring all benefits such as

availability and scalability. Hybrid hosting plans combine the benefits of public and private clouds to meet organizational concerns and goals related to cloud migration. The most popular cloud platforms are AWS, Azure, Oracle, Google and IBM. A detailed privacy policy outlining how data is collected, stored and used. It also details how to share and request data from other institutions and the conditions for doing so. It also describes how long you want to keep your data and how to keep it private and confidential. It also mentions the regulations that apply to all employees to keep their data private. We could also has a Terms of Service document that describes how customers use or request our services. It also mentions what they arrange by becoming our customers.

5. Architecture Plan:

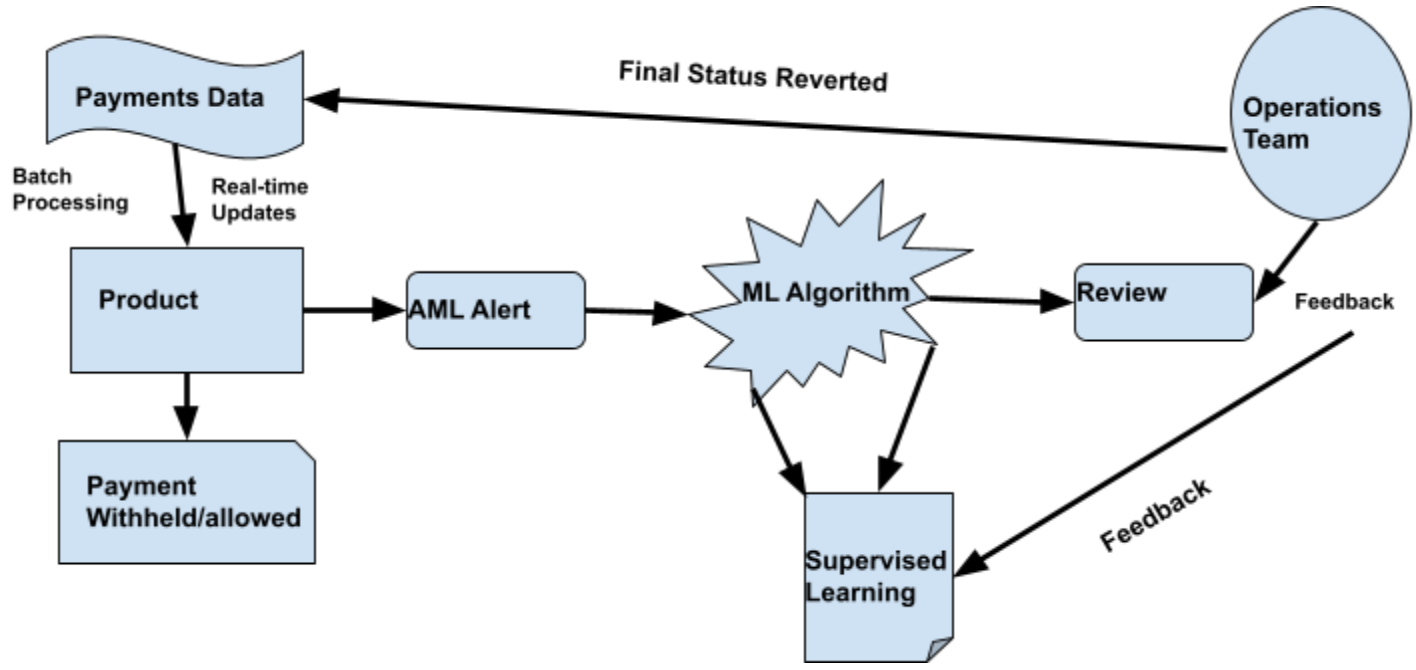


Figure 4:Architecture Diagram.

6. Tools and Methodologies:

The solution is focused on how new advancements might help locales furthermore, regulators become more viable in the execution of AML. Specifically, computerized arrangements which empower a superior comprehension, evaluation and moderation of

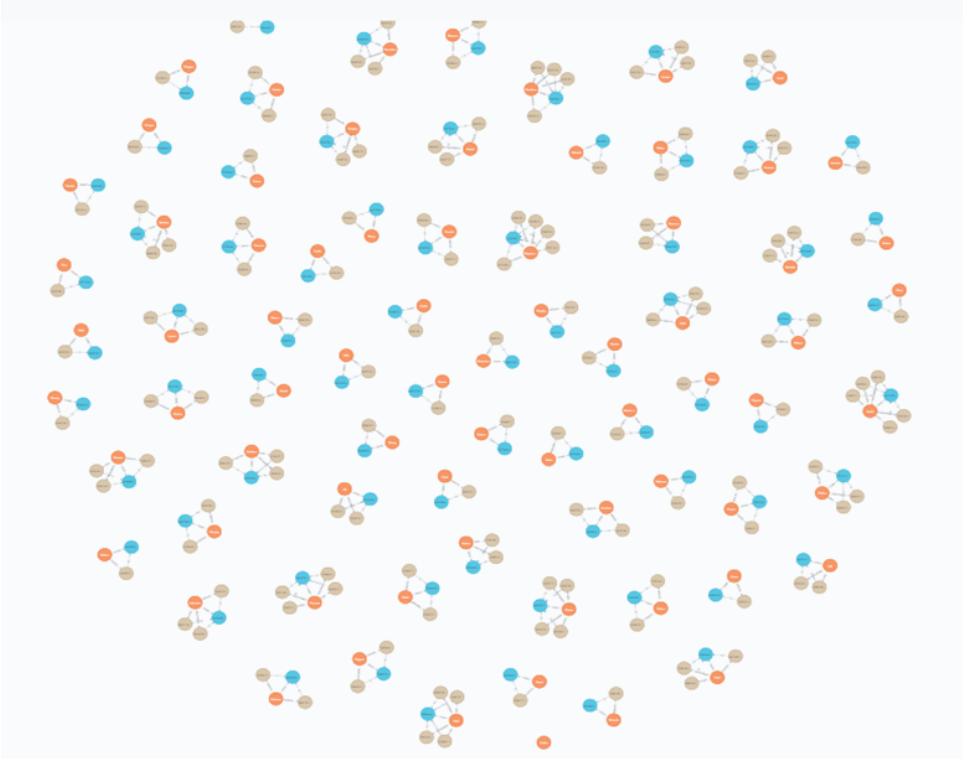
dangers, client a reasonable level of effort and observing, and correspondence with managers might help accomplishing adequacy in the execution of AML principles. Advanced technologies like ML algorithms are being implemented to target the requirement for verification and identification of customers. Additionally, considering the situations where promising innovations have not been effectively sent and recognizes difficulties or deterrents to their viable use. It also investigates whether composed worldwide activity is expected to empower more prominent utilization of creative innovation based answers for AML goals. This incorporates examining primary difficulties. For instance issues of information quality, changing inheritance frameworks, cost limitations and the absence of administrative motivators.

Exercise 2: Placement Analysis withNeo4j

Performing Exploratory Analysis: The following code of Cypher language is used for exploratory analysis.

```
MATCH (p:Person)-[:PERFORMED]-(d:Deposit)-[:TO]→(a:Account) RETURN p,a,d
```

Determining any sorts of outliers in the data and looking for patterns by visualizing it:



Performing Temporary Analysis: Investigating the deposits for a specific time duration i.e. one week's of deposits and looking for any suspicious activities.
Time stamping the dataset for the oldest and latest deposits:

min_date	min_date_epoch	max_date	max_date_epoch
"1997-11-21T06:10:32Z"	880092632000	"2022-02-12T04:31:20Z"	1644640280000

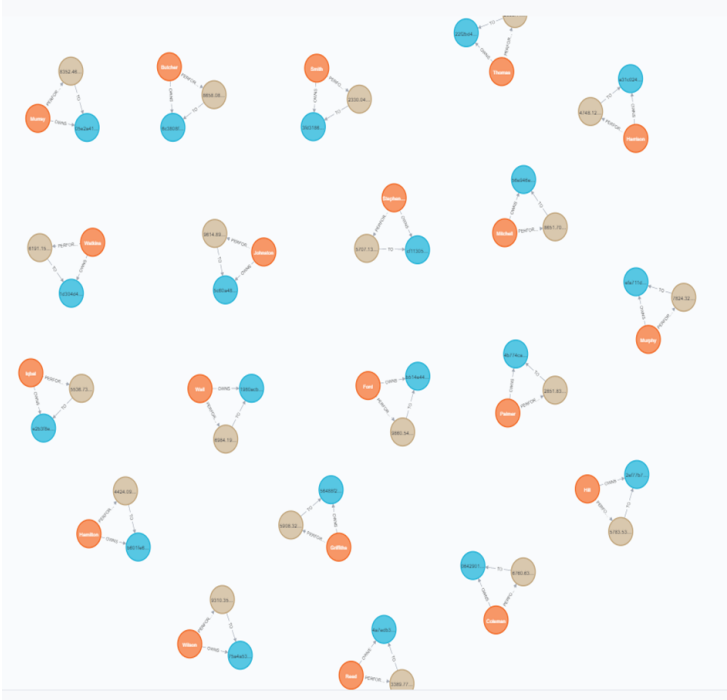
Following Cypher code is to get the large sums deposited:

```
UNWIND range(880092632000,1644640280000, 86400000) AS t
MATCH (p:Person)-[:PERFORMED]-(d:Deposit)-[:TO]→(a:Account)
WHERE d.timestamp.epochMillis ⩾ t AND d.timestamp.epochMillis < t +
86400000
RETURN datetime({epochMillis:t}), COUNT(*) as deposit_count ORDER BY
deposit_count desc
```

datetime({epochMillis:t})	deposit_count
"2011-08-27T06:10:32Z"	22
"2017-08-09T06:10:32Z"	15
"2011-08-28T06:10:32Z"	14
"2017-08-10T06:10:32Z"	9
"2020-07-14T06:10:32Z"	2

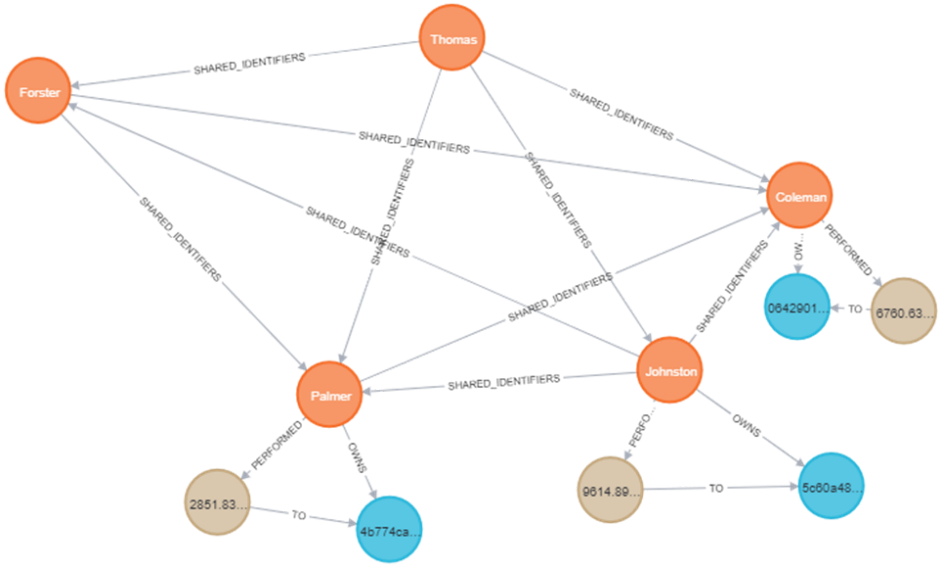
Taking the highest deposit count into consideration we are going to investigate more in that specific window.

The following diagram clearly suggest that the duration for most of the transaction is 1 hour:



Following code is to visualize virtual relationships:

```
MATCH (p:Person)-[:PERFORMED]-(d:Deposit)-[:TO]→(a:Account)
MATCH (p)-[:SHARED_IDENTIFIERS]-(o:Person)
WITH *,datetime("2011-08-27T06:10:32Z") as window_start
WHERE d.timestamp.epochMillis ⩾ window_start.epochMillis AND
d.timestamp.epochMillis < window_start.epochMillis + 86400000
RETURN p,d,a,o
```



Getting the Suspicious deposits:

"d.obj_id"
"8294135a-0df4-4310-bf75-3e741a9f00cb"
"c3cefc39-6b93-4831-be31-e34972b1eda2"
"8294135a-0df4-4310-bf75-3e741a9f00cb"
"8294135a-0df4-4310-bf75-3e741a9f00cb"
"49561b2f-d271-49bf-8a56-0f30f03a69a5"
"8294135a-0df4-4310-bf75-3e741a9f00cb"
"49561b2f-d271-49bf-8a56-0f30f03a69a5"
"c3cefc39-6b93-4831-be31-e34972b1eda2"
"49561b2f-d271-49bf-8a56-0f30f03a69a5"
"49561b2f-d271-49bf-8a56-0f30f03a69a5"
"c3cefc39-6b93-4831-be31-e34972b1eda2"
"c3cefc39-6b93-4831-be31-e34972b1eda2"

Exercise 4: ML Implementation

Reading the Dataset:

The dataset in CSV format is uploaded to colab notebook and was assigned to the dataframe with the help of PANDAS:

```
[4]  # Uploading the CSV file in to Google Colab

     from google.colab import files
     uploaded = files.upload()
```

```
Choose Files transactions.csv
• transactions.csv(text/csv) - 16107659 bytes, last modified: 5/6/2022 - 100% done
Saving transactions.csv to transactions.csv
```

```
#  Reading the data in to pandas dataframe
import io
df = pd.read_csv(io.BytesIO(uploaded['transactions.csv']))
```

Printing the properties of the dataset:

Getting the mean, SD, min-max, quartiles:

```
# Basic Statistics properties of the amount column
df['amount'].describe().apply(lambda x: format(x, 'f'))
```

```
count      200000.000000
mean       180811.166550
std        329179.963404
min             0.320000
25%         12016.120000
50%         68721.045000
75%        229079.070000
max      10000000.000000
Name: amount, dtype: object
```

Suspicious vs Non-Suspicious transactions:

The ratio is being calculated using the following line of code:

```
[11] #  The ratio of fraudlent vs non fraudelent
     print('Proportion of Fraudlent vs non fraudlent: {:.2f}%'.format(df['isFraud'].value_counts()[1]/df['isFraud'].size * 100))
```

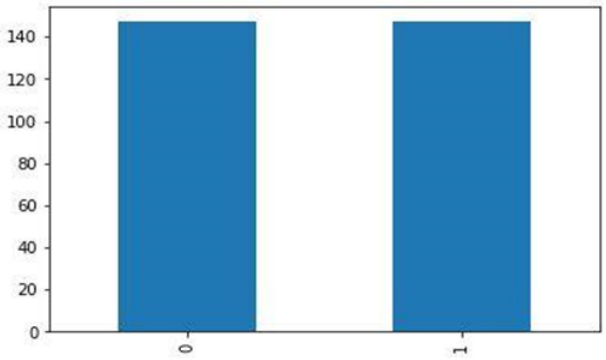
```
Proportion of Fraudlent vs non fraudlent: 0.07%
```

Preparing the Dataset:

Dropping out the major values to make it complementing with minor values using the following lines of codes:

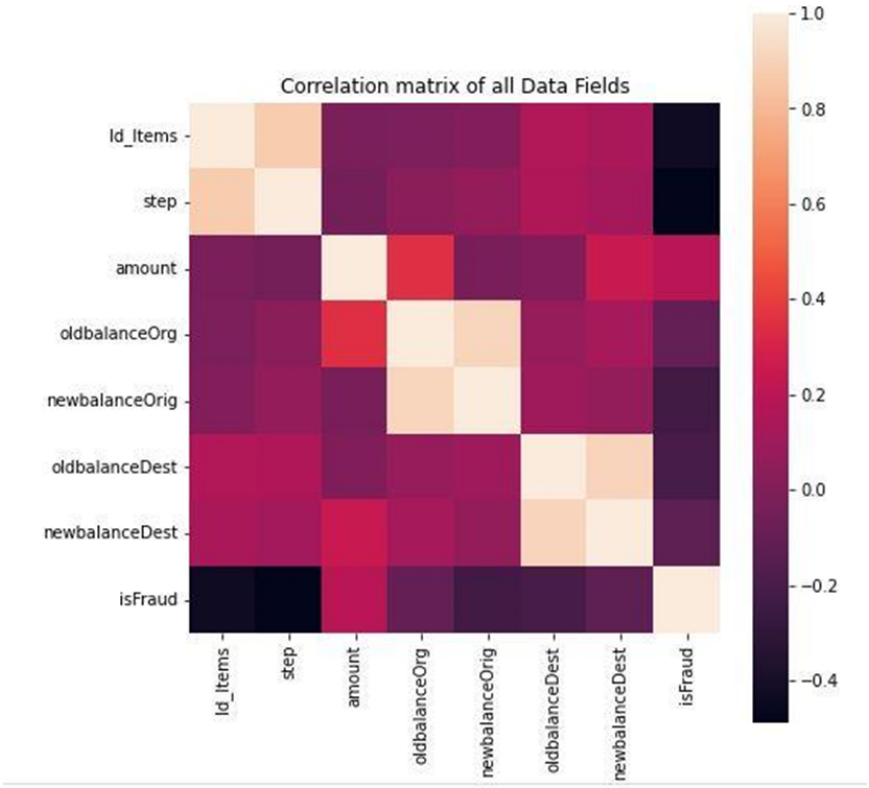
```
# Balancing the Classes with User Sampling
suspicious_count = len(df[df['isFraud'] == 1])
non_suspicious_count = len(df[df['isFraud'] == 0])

sampled_non_suspicious = df[df['isFraud'] == 0].sample(n=suspicious_count)
suspicious = df[df['isFraud'] == 1]

sampled_non_suspicious.reset_index(drop=True, inplace=True)
suspicious.reset_index(drop=True, inplace=True)

balanced_df = pd.concat([sampled_non_suspicious, suspicious]).reset_index(drop=True);
balanced_df['isFraud'].value_counts().plot(kind='bar')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f3b9b95b190>
```



Visualizing the Correlation HeatMap:

corr() function is used to get the correlation against the columns and are shown in the following heatmap:



Creating pair-plots for all the variables:

Using the pair plot function we calculated the plot as shown in the following figure:

```
# creating the pair plots for all the fields of the dataset
sns.pairplot(balanced_df, hue="isFraud")
```

Adding new columns for Classification:

New columns being created for classification with the help of existing columns:

```
features = balanced_df.copy(deep=True)
features['nameOriginal'] = features['nameOrig'].str.contains('C')
features['nameDestination_C'] = features['nameDest'].str.contains('C')
features['nameDestination_M'] = features['nameDest'].str.contains('M')
```

Adding dummies columns for categorical values:

Using the dummies function, we created dummies for categorical values and also added to the existing dataframe as shown below:

```
categorical_dict = {
    "type": "category",
    "nameOriginal": "category",
    "nameDestination_M": "category",
    "nameDestination_C": "category"
}

X = X.astype(categorical_dict)

for col in categorical_dict.keys():
  X = X.join(pd.get_dummies(X[[col]], prefix=col)).drop([col], axis=1)
```

Creating decision tree classifier:

To classifies suspicious and non-suspicious transaction we will be using decision tree classifier and splitting the dataframe into training data and test data:

```
# Splitting the data to train and test
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=0)
```

```
# Building the Decision tree classifier
from sklearn.tree import DecisionTreeClassifier
dt_clf = DecisionTreeClassifier(random_state=1)
dt_clf = dt_clf.fit(X_train,y_train)
y_pred = dt_clf.predict(X_test)
```

Evaluating the ML Model:

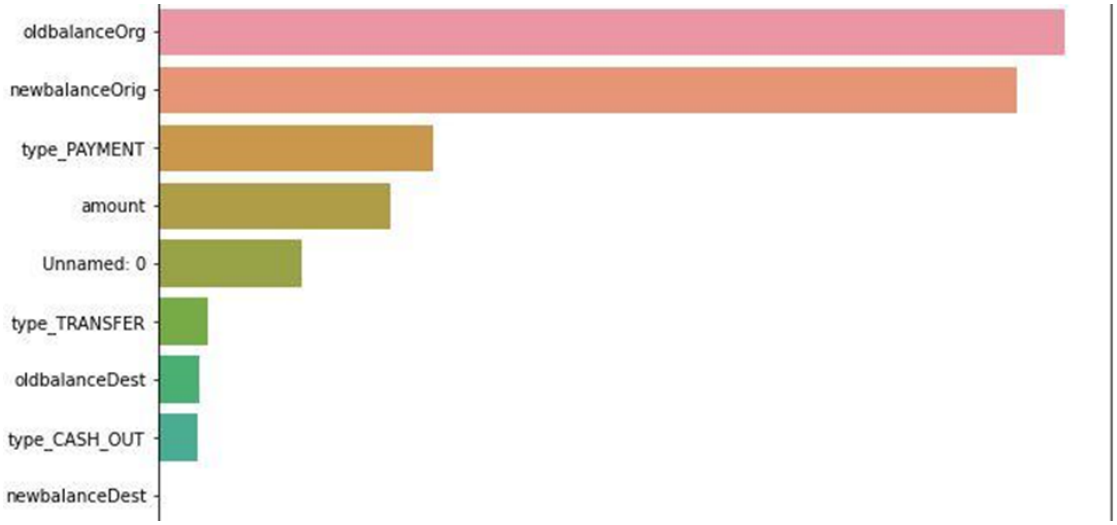
Calculating the accuracy of the model using different parameters and the result is shown below:

```
from sklearn import metrics
print(metrics.classification_report(y_test, y_pred))
```

```
              precision    recall  f1-score   support

           0       0.95      0.90      0.92        39
           1       0.92      0.96      0.94        50

    accuracy                           0.93        89
   macro avg       0.93      0.93      0.93        89
weighted avg       0.93      0.93      0.93        89
```

About the Model:

The plot we got from decision tree helped us to deduce the conclusion that oldbalanceOrg has the highest percentage in the following the visualized chart:



Bibliography:

FATF (2012), International Standards on Combating Money Laundering and the Financing of Terrorism & Proliferation ('The FATF Forty Recommendations'), FATF, Paris, www.fatf-gafi.org/recommendations.html.

FATF (2013b), Guidance on National Money Laundering and Terrorist Financing Risk Assessment, FATF, Paris https://www.fatf-gafi.org/publications/methodsandtrends/documents/nationalmoneylaunderingandterroristfinancingriskassessment.html

Chase, I. (2020), Doing What is Right: Financial Inclusion Needs Better Incentives, RUSI, https://rusi.org/commentary/doing-what-right-financial-inclusion-needs-better-incentives.

EBA (2021), Opinion of the European Banking Authority on the risks of money laundering and terrorist financing affecting the European Union's financial sector, https://www.eba.europa.eu/eba-highlights-key-money-laundering-and-terrorist-financing-risksacross-eu.

Broeders D. and Prenio J. (2018), Innovative technology in financial supervision (SupTech) – the experience of early users, https://www.bis.org/fsi/publ/insights9.pdf.