

IDENTIFYING MONEY LAUNDERING ATTEMPTS FROM BANK TRANSACTIONS

Abstract

Money laundering is a way to conceal illegally obtained funds. It works by transferring money in elaborate and complicated financial transactions to mislead Enforcement Directorate. The objective is to make it difficult to identify the original party of the transaction. However, at the end of the convoluted scheme, the funds ultimately return back to the launderer.

The way Money laundering can occur, is the use of Shell companies, these are businesses set up as typical companies, but they do not have any assets or perform any real business activities.

My aim is to identify Shell companies using Banking Transactional Data with various attributes. We apply Ensemble Methods to classify Shell/Normal Company with high Recall and F score.

It is a Three stage Process:

- 1.**Placement:** This is the movement of cash from its source. It is carried out by Blending of Funds i.e. to hide cash with lots of other cash.
- 2.**Layering:** The purpose of this stage is to make it more difficult to detect and uncover a laundering activity.
3. **Integration:** It is at the integration stage where the money is returned to the criminal from what seem to be legitimate sources.

CCS CONCEPTS

Computing methodologies → Ensemble Methods; Modeling and simulation; Information systems → Data mining;

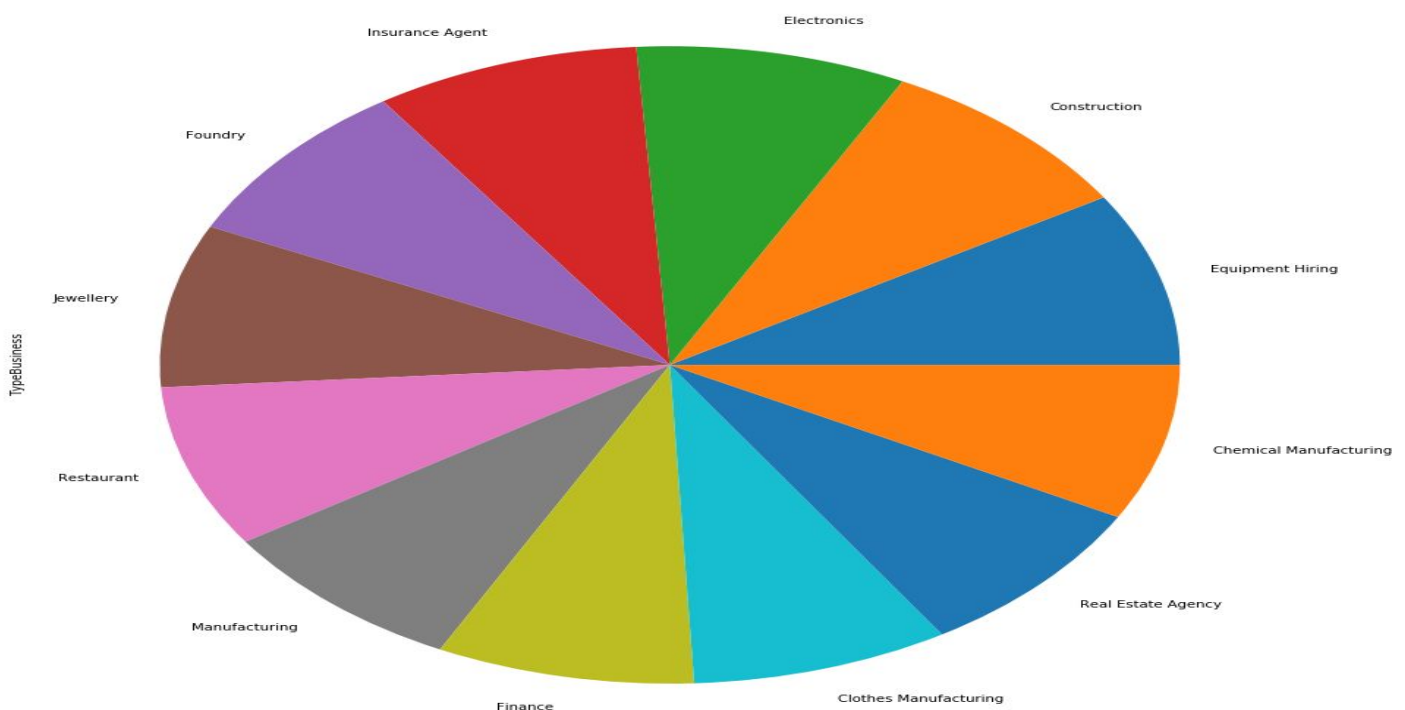
KEYWORDS

Money Laundering; Shell Companies; Bank Transaction Data

INTRODUCTION

Money Laundering can be defined as the act of making money that comes from one source to look like it comes from another source. Shell company makes illicit payments through **LAYERING**, to make it more difficult for investigating agencies, payments can be “layered”. Money is transferred from A to entity Z through multiple transactions. A pays B, B pays C, C pays D, D pays E and so on till Z gets the money.¹ On Nov’2016 **Demonetization** took place, it was a trap to catch black money holders off guard. The objectives of the note ban were checking black money and terror financing. The objectives of the note ban were checking black money and terror financing and Income Tax officials, who prepared lists of persons and taxpayers making large deposits and transaction in post-demonetisation phase. The ministry is probing deposit of over \$1 billion made by companies during the Demonetization and examining listed companies found untraceable by SEBI. Shell companies are those entities that do not contribute to any economic activity and are often set up for money laundering or tax evasion. **According to data**, more than 14,000 shell firms struck off records in January 2019 **According to Rule**, the RoC can remove a company from its records if it fails to commence business within a year of its incorporation, or has not carried out any operation for three years. According to economic activity **Industry type** of Shell Company is agriculture and allied

Types of Company in Datasets

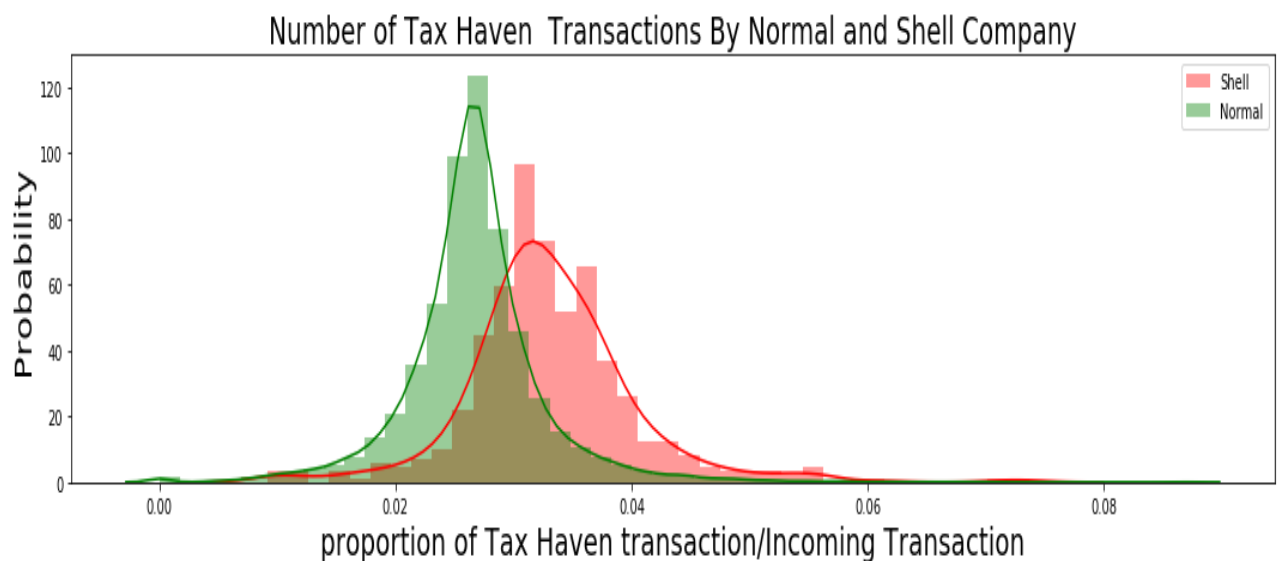


activities (32,628 companies were involved in agriculture and allied activities and 739,649 in services such as real estate, trading, finance and insurance.

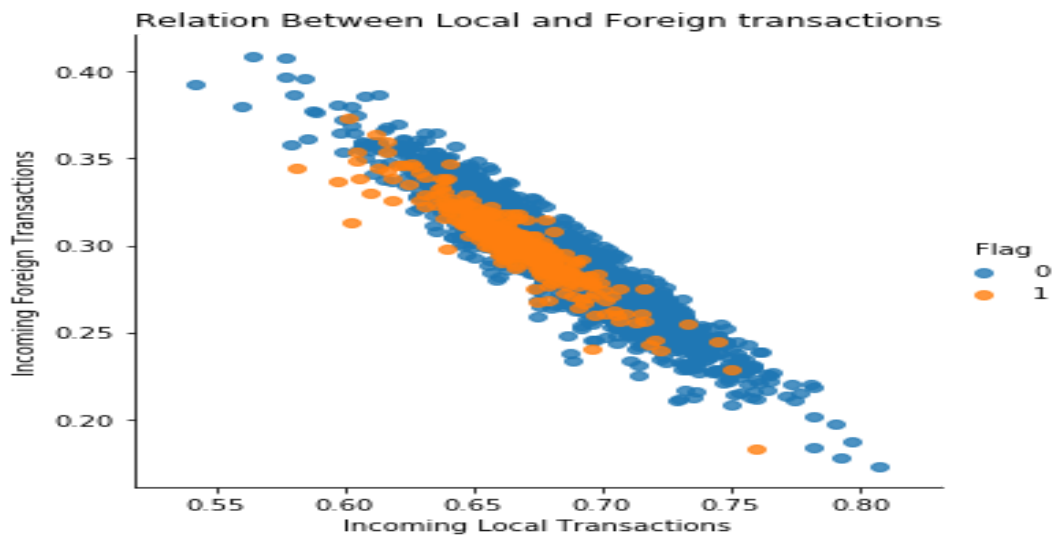
Government has issued show cause notices to 13,000 shell companies for not carrying out business activities for a prolonged time. Trusts, especially those set up in jurisdictions like the Cayman Islands, Panama and British Virgin Islands, have often been seen as routes for evading taxes. Shell Companies have transactions in this [Tax haven](#) states.

Tax haven, country or independent area where taxes are levied at a low rate.

Proportion of Tax Haven in Normal/Shell Company



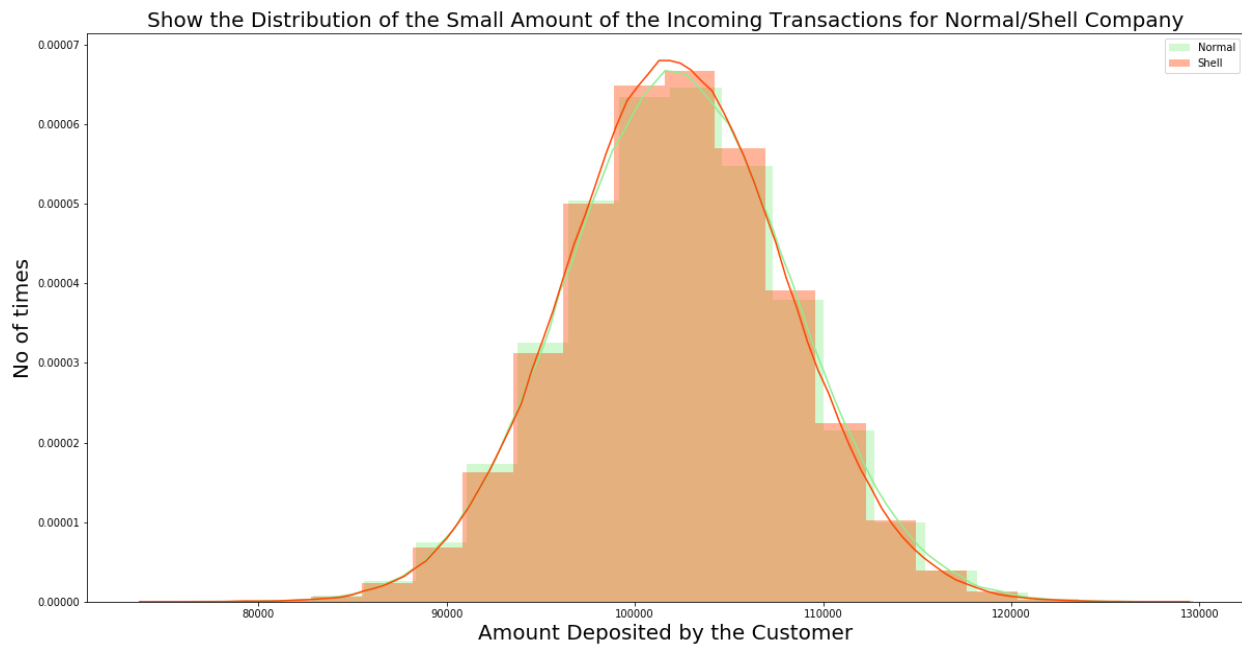
Number of Foreign Transactions performed by shell Companies are More as compared to the Local Transactions.



The most common types of criminals who need to launder money are drug traffickers, embezzlers, corrupt politicians and public officials, mobsters, terrorists and con artists. Drug traffickers are in serious need of good laundering systems because they deal almost exclusively in cash, which causes all sorts of logistics problems.

They do Structuring Deposits, known as [smurfing](#), this is a method of placement whereby cash is broken into smaller deposits of money, used to defeat suspicion of money laundering and to avoid anti-money laundering reporting requirements.

Shell Company Breaks deposit into smaller parts to avoid suspicion.

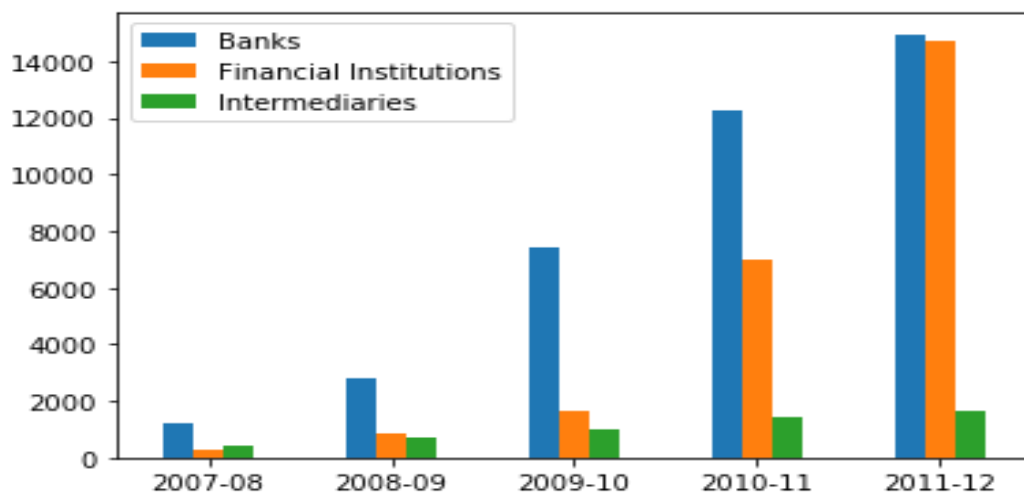


Threshold set : 1,00,000

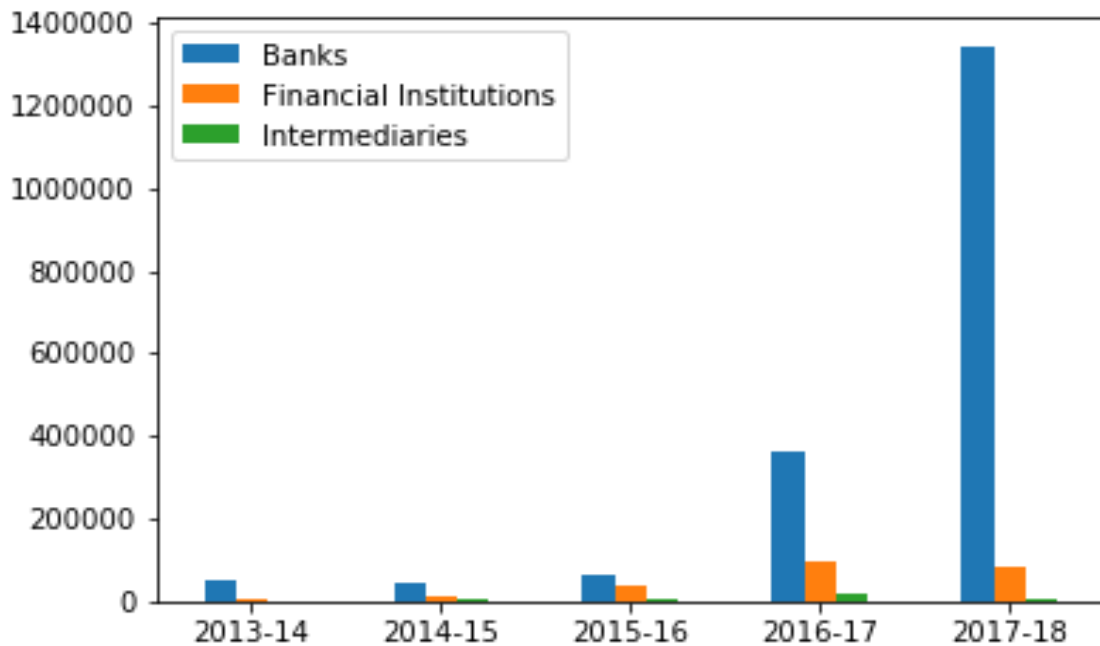
Government has taken steps to prevent Money Laundering, by Prevention of Money-Laundering Act, 2002 (PMLA).

According to the Government Suspicious Transactions Report

1 > Suspicious Transaction Reports (STRs) – [2007 – 12]



2 > Suspicious Transaction Reports (STRs) – [2013 – 18]



According to the Suspicious Transactions Alert Information, Bank Transactions have highest suspicious alert and there is an increase in the Number of alert as compared to the previous year.

Links of Forecasting Report

- <https://fiuindia.gov.in/pdfs/downloads/AnnualReport2017-18.pdf>
- https://data.gov.in/catalogsv2?format=json&offset=0&limit=9&sort%5B_score%5D=desc&query=Money+Laundering&exact_match=0
- <https://www.indiatoday.in/india/story/4-years-of-modi-govt-fight-against-black-money-explained-in-7-points-1242398-2018-05-26>
- <https://economictimes.indiatimes.com/markets/stocks/news/crackdown-on-india-shell-companies-unearths-1-billion-cash/articleshow/61198230.cms?from=mdr>

- <https://www.livemint.com/>
- <https://www.livemint.com/Politics/NN0J2ePOvGnvKJun1vj1Ql/Govt-is-sues-showcause-notices-to-nearly-13000-shell-compani.html>

SYNTHETIC DATA GENERATION

Transactions of the Normal and Shell Company

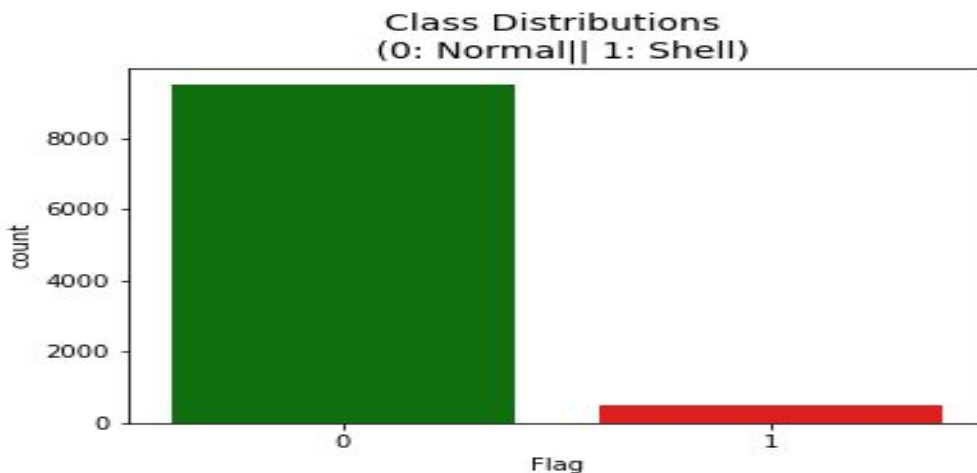
Income Tax Department and ED uses Customer Bank Transactions data to identify Fraudulent Company.

Due to lack of real-world dataset, I have generated Data using real world scenarios with Different Features.

Features are discrete and Continuous random variable with different probability distribution function and have different mean and standard deviation according to the Normal/Shell Company.

Data is Generated for the period of 1yr with the Assumption that company is located in India.

Proportion of Normal/Shell Company is 95:5



Input – N, Number of Company for which Data need to be Generated.

Output: 1. Company Data – Which is Information about the Company.

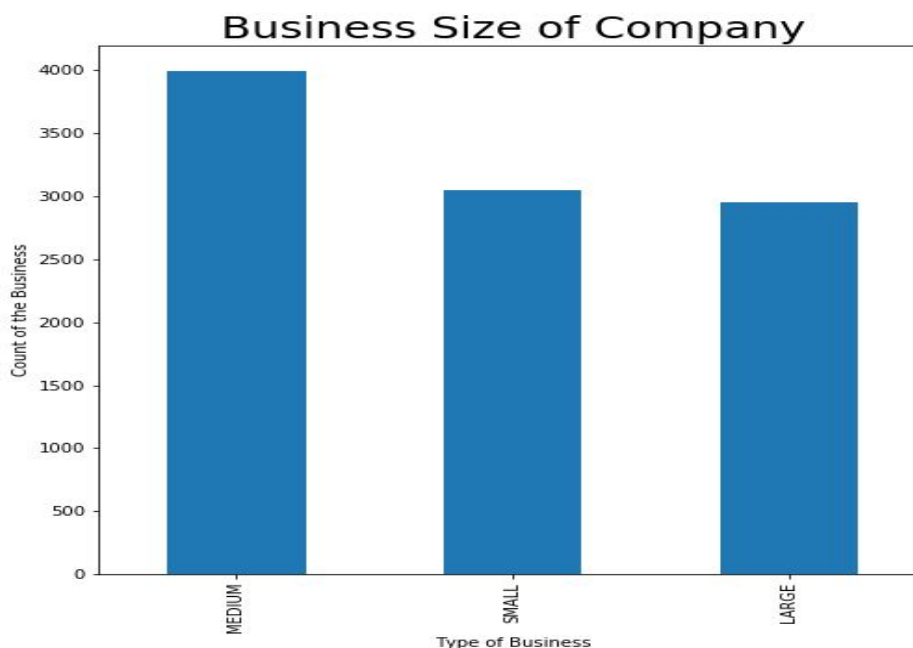
2. Transactional Data – Incoming & Outgoing Transactions of the Company for the period of 1yr.

3. Summarized Data – It is generated using Company Data and Transactional Data and create a single row vector for each Account Number.

Algorithm to Generate Data

1. Generate Company Data – Industry which are targeted as shell Companies are : Construction, Jewellery, Restaurant, Real Estate Agency, Equipment Hiring, Electronics, Manufacturing, Clothes Manufacturing, Foundry, Chemical Manufacturing, Finance, Insurance Agent.

Business Size: Small, Medium and Large.



Red Flags:

On the Basis of Red Flags

Shell Company are hard to detect due to the complex nature as it behaves similar to the other Company. It shows deviation from the behaviour of another company.

1. One of the Important Features of Shell Company is that, Number of transactions abroad and tax haven places are significantly more as compared to Normal Company.

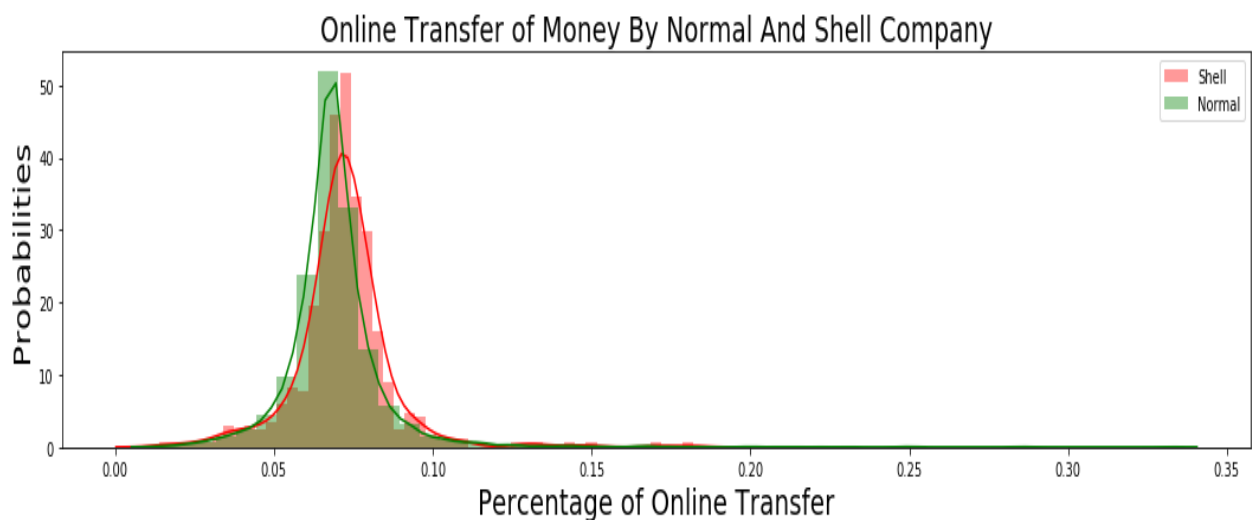
Feature used in Dataset to describe this Scenario is : Number of Foreign, Local and Tax haven Transactions.

2. Fraudulent Transactions Company prefers to use “Online Mode of Transactions” as it hides the verification from the Banks before transfer.

Feature Used: Online Transaction.

Normal Distribution Curve, Describe Number of Shell Company transactions are more as compared to the Normal.

Online Transaction of the Shell Company is more.



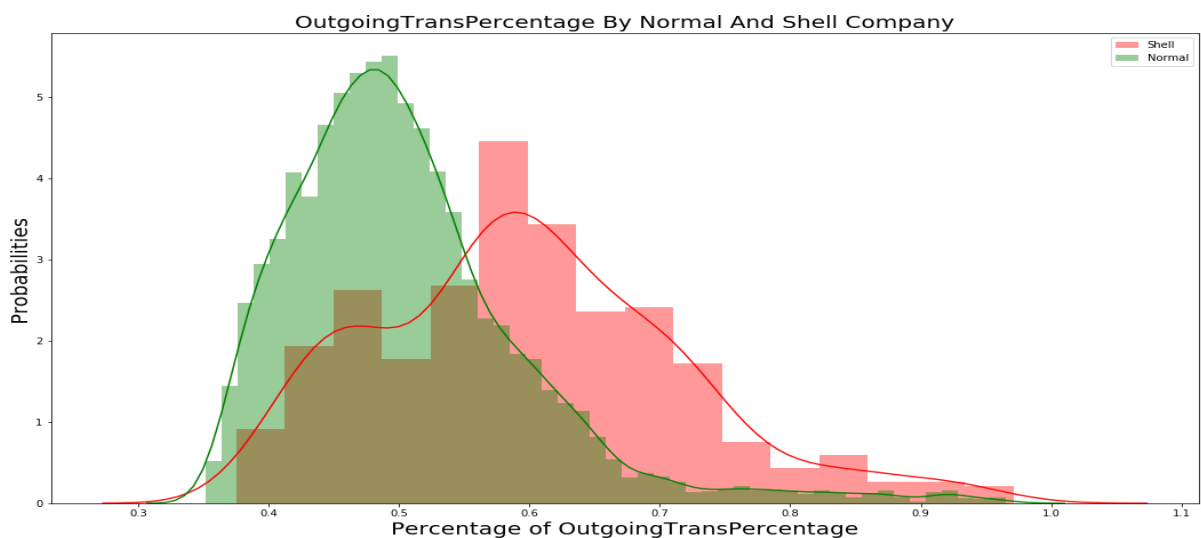
3. Important Characteristics of Shell Company is **Structuring**, it is an idea of structuring of deposit & withdrawals from bank to avoid suspicious alert.

Setting the Threshold in the Bank Transactional Data for suspicious alert (Algorithm set is to 50000).

Eg. Customer deposit Large sum of money by breaking the Amount to less than 50000 to avoid alert trigger.

4. **Another Important Feature of Shell Company is Layering**, To avoid income Tax on these earnings A shows payments are made to another shell company towards fulfillment of contracts. The money is routed to further clutch of companies – C, D, E and F, who account of it as a share capital. All these money flows are notional- No actual Money flows through this firm.

5. According to the Above mentioned Concept, There are **More Number of Outgoing Transaction Percentage** by Shell Company as compared to the Normal Account.



6. Another important Characteristics is **Integration** where laundered money is withdrawn from the legitimate account to be used for whatever purposes the criminals have in mind for it. One of the most common ways

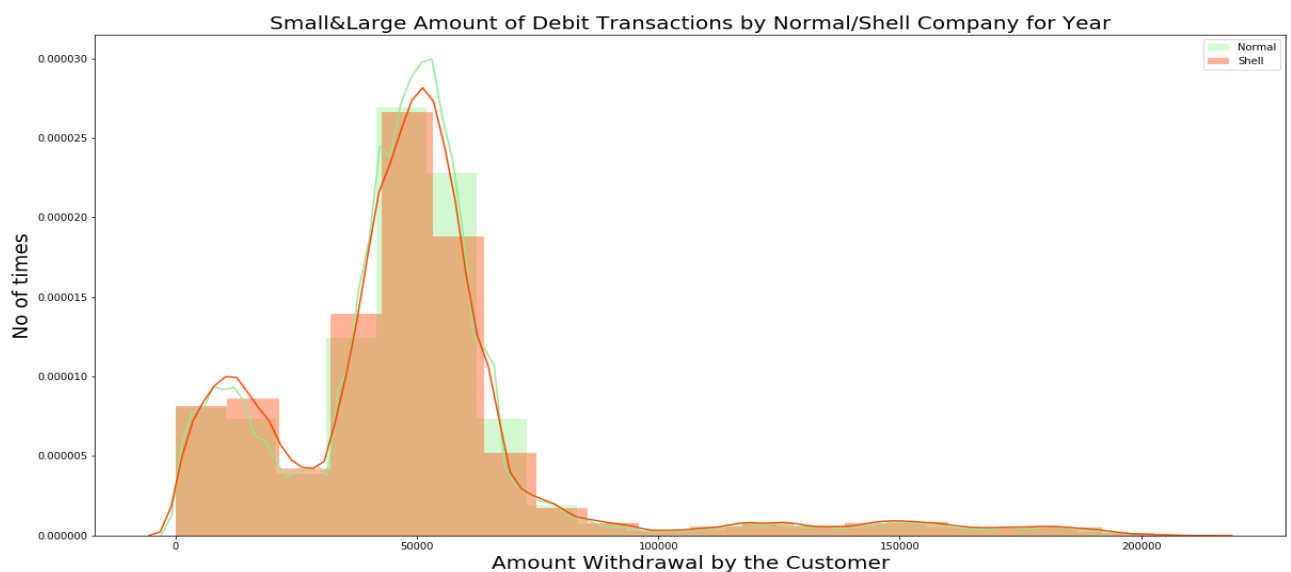
of integrating the money into the economy is through buying property, Artwork, jewellery, or high-end automobiles.

From This Scenario, Shell Company has a large amount of Outgoing Transaction as Compared to the Normal Account.

Variable Defined is : Large Amount of Outgoing Transaction.

From the graph below, Threshold set is 50,000.

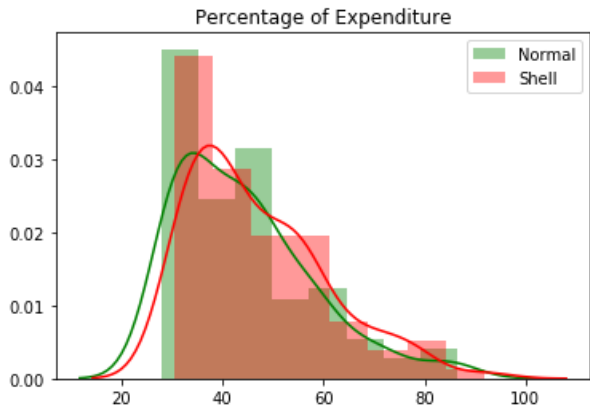
After 50000 there seems to be a layer of Shell Company on the top of the Normal Company.



Variables for Company Data

Variables	Values and variation Normal/Shell
AccountNumber	4Digit Number

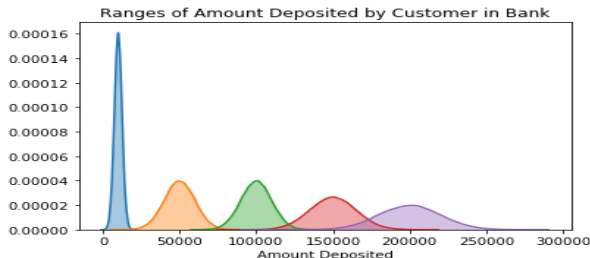
TypeBusiness	<ul style="list-style-type: none">● Construction● Jewellery● Restaurant● Real Estate Agency● EquipmentHiring● Electronics● Manufacturing● Clothes Manufacturing● Foundry● ChemicalManufacturing● Finance● Insurance Agent														
BusinessCategory	<ul style="list-style-type: none">● Retail● SemiBulk● Bulk														
RevenueSize	<ul style="list-style-type: none">● LOW● MEDIUM● HIGH														
BusinessSize	<ul style="list-style-type: none">● SMALL● MEDIUM● LARGE														
TotalEmployees	<table><tr><th>Business Size</th><th>Employees Range</th></tr><tr><td>SMALL</td><td>1 - 50</td></tr><tr><td>MEDIUM</td><td>50 - 200</td></tr><tr><td>LARGE</td><td>200 - 500</td></tr></table>			Business Size	Employees Range	SMALL	1 - 50	MEDIUM	50 - 200	LARGE	200 - 500				
Business Size	Employees Range														
SMALL	1 - 50														
MEDIUM	50 - 200														
LARGE	200 - 500														
Annual_Revenue	<table><tr><th>Revenue Size</th><th>Mean</th><th>Std</th></tr><tr><td>LOW</td><td>60000</td><td>6000</td></tr><tr><td>MEDIUM</td><td>110000</td><td>11000</td></tr><tr><td>HIGH</td><td>600000</td><td>60000</td></tr></table> <p>X = np.random.normal(mean,std,1)</p>			Revenue Size	Mean	Std	LOW	60000	6000	MEDIUM	110000	11000	HIGH	600000	60000
Revenue Size	Mean	Std													
LOW	60000	6000													
MEDIUM	110000	11000													
HIGH	600000	60000													

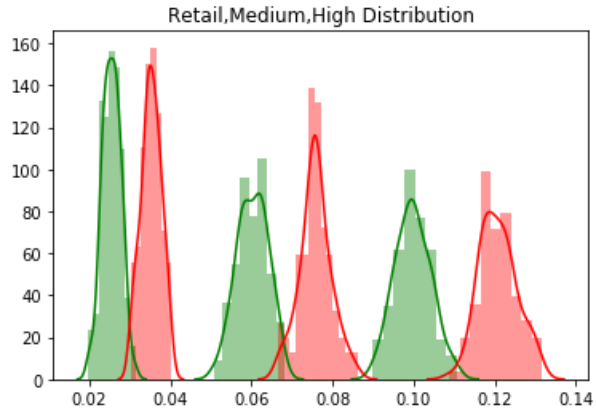
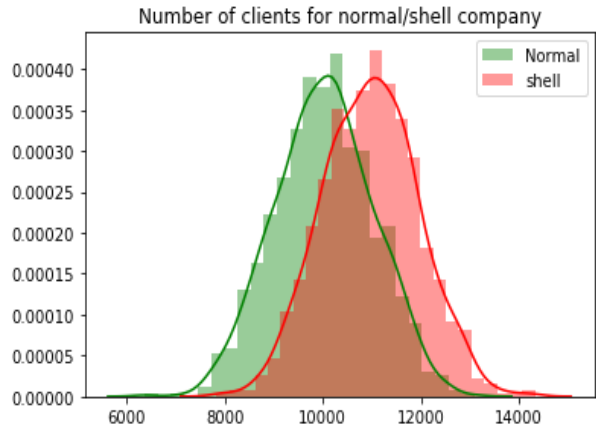
	$Y = np.random.normal(0.1 * X, 0.01 * X, 1)$ $Z = np.random.normal(X, Y, 1)$ Total_Revenue = 12 * EmployeeSize * Z
Total Owners	7,8,9
Owner History	<ul style="list-style-type: none"> • YES • NO
CompanyLocation	<ul style="list-style-type: none"> • OFFSHORE • INDIA
TaxHavenLocation	<ul style="list-style-type: none"> • YES • NO
NumberOfAccounts	1,2,3
ExpenditureAmount	PercentageExpenditure * AnnualRevenue
AnnualProfit	Profit = AnnualRevenue - Expenditure
AccountType	<ul style="list-style-type: none"> • Saving • Current
Percentage Expenditure	
Flag	N – Normal & S - Shell

Variables for Transactional Data

For Each Company, Incoming and Outgoing Transactions are Generated for period of 1yr.

Variables	Values and variation Normal/Shell
-----------	--

AccountNumber	Global Variable (Same As Company Data)
TransactionNumber	Updated for each company
TransactionDate	Any Date for year(Credit) Last day of every month(Debit)
TransactionCategory	Debit/Credit
TransactionMode	Cash/Cheque/Transfer
ExchangeType	Local/Foreign/Tax Haven
OnlineOffline	Online/Offline
Transaction Amount	<p>1.Generate Income:</p> <p>Its Generated by Combination of 5 Normal Distribution curve with different probability for Normal/Shell Company.</p>  <p>2.Generate Expenditure:</p> <p>Retail, Medium and High Business size company have different Percentage of Expenditure.</p>

	 <p>A histogram showing the distribution of values for Retail, Medium, and High categories. The x-axis ranges from 0.02 to 0.14, and the y-axis ranges from 0 to 160. There are three distinct peaks: a green one at approximately 0.03, a red one at approximately 0.04, and a green one at approximately 0.10. The red peak is the highest, reaching nearly 160.</p>
PaymentTo	<p>Credit : Revenue Debit : Salary, Utility, Supplier, ProfessionalServices, Entertainment Expenses, Personal Expenses</p>
Source Account	<p>AccountIDCompany/(Generated Through Normal Distribution Curve) Assumption: Normal Company has More Client as compared to the shell Company.</p>  <p>A histogram comparing the number of clients for Normal and shell companies. The x-axis ranges from 6000 to 14000, and the y-axis ranges from 0.00000 to 0.00040. The Normal distribution (green) is centered around 10000, and the shell distribution (red) is centered around 11000. The Normal distribution has a higher peak, reaching approximately 0.00040.</p>
Destination Account	<p>AccountIDCompany/(Generated Through Normal Distribution Curve) Assumption: Normal Company has More Client as compared to the shell Company.</p> <p>Shell company have repetition of clients after larger amount, so it will result in the less number of clients.</p>

Variables for Summarized Data

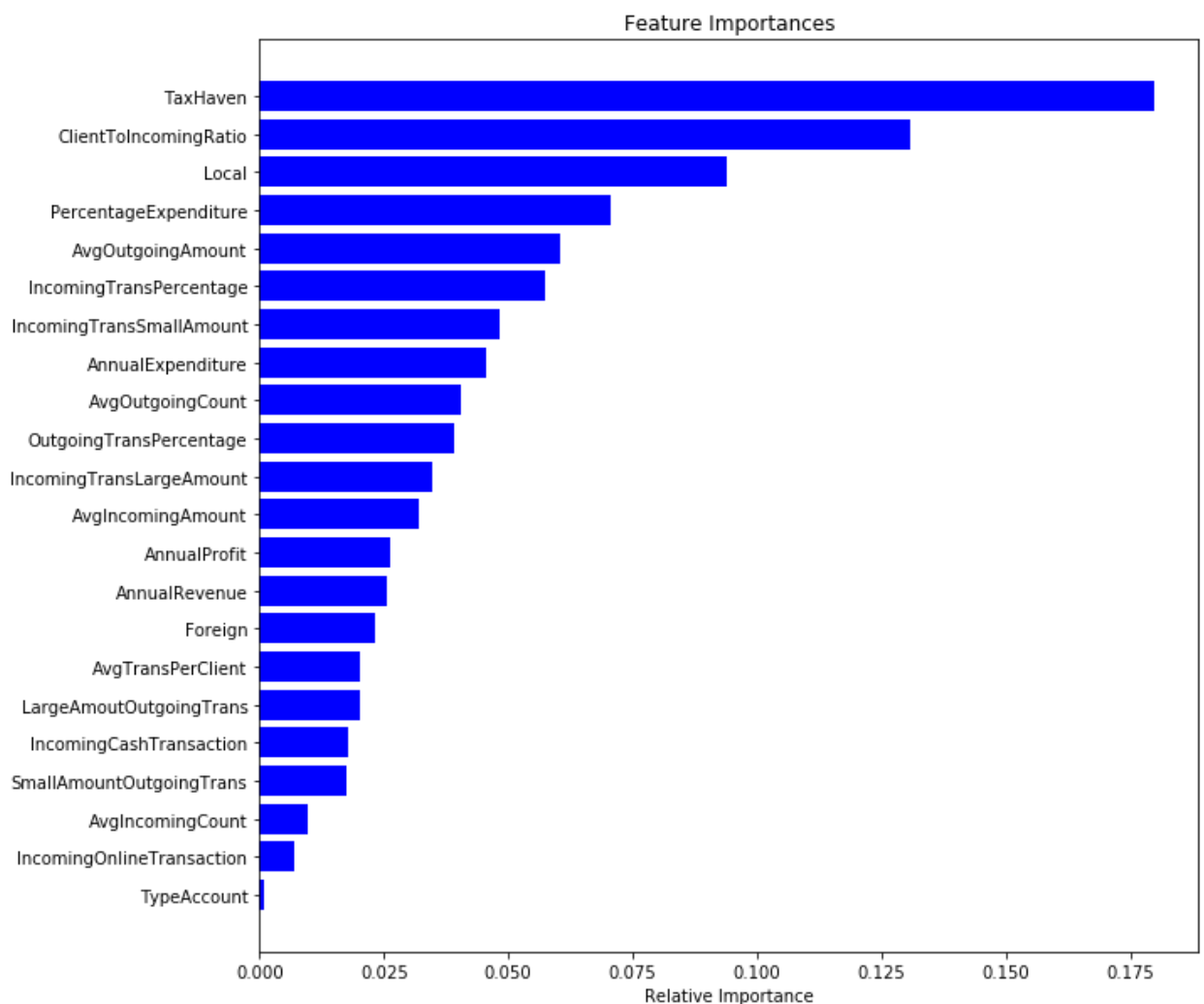
It is Generated from Company Data and Transactional Data

Variables	Variation Normal/Shell		
AccountNumber	Normal: 100001, Shell: 100951		
Foreign	Shell Company has more		
Local	Normal Company has more		
TaxHaven	Shell Company has more		
ClientToIncomingRatio	Normal Company has more		
IncomingTransSmallAmount	Shell Company has more		
IncomingTransLargeAmount	Normal Company has more		
Type Account – Saving /Current Account	Normal – More Saving Account Shell - More Current Account.		
	Company Prefers Current Account for Daily Transactions		
	Company	Saving	Current
	Honest		More
	Shell	More	
IncomingCashTransaction	Normal Company has more		
IncomingOnlineTransaction	Shell Company has more		
Annual Expenditure	Shell Company has more, because they want to pay less tax.		
Annual Revenue	Normal Company has more		
LargeAmoutOutgoingTrans	Shell Company has more		
SmallAmountOutgoingTrans	Normal Company has more		
AvgTransPerClient	Shell Company has more, Breaks large amount into small parts.		

IncomingTransPercentage	Shell Company has more
OutgoingTransPercentage	Normal Company has more
AvgIncomingAmount	Shell has small amount of outgoing transaction.
AvgOutgoingAmount	Normal Has small amount of outgoing transaction.
AvgIncomingCount	Shell Company has more, Break of large amount into small parts.
AvgOutgoingCount	Normal Company has more, Because of they have a greater number of small amounts of outgoing transaction.
Annual Profit	Normal Company has more
PercentageExpenditure	Shell Company has more percentage Expenditure.
Flag	N – Normal, S-Shell

Data Analysis

we have summarized Data using company information and transactional Data for one year in 22 features described above in Summarized Table and we need to classify whether Company is Normal/Shell (Flag Variable set to 1 – Shell and 0 – Normal). We have created here Biased data with 10000 Company where 950 were Normal Company and 50 were Shell Company. Feature Importance is calculated for Dataset.



Now we have performed Experiment using set of features,

Using 1 feature, Precision, recall and F score is = $9/142 = 0.063$

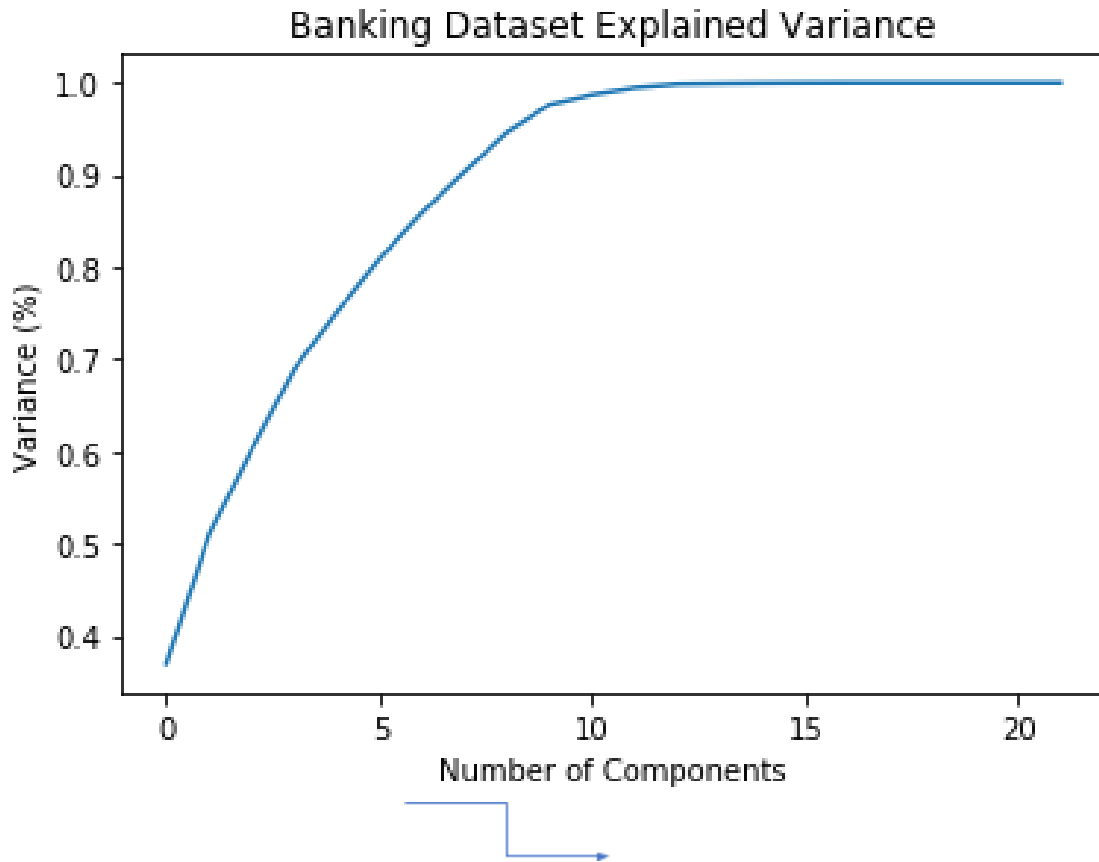
Using 2 feature, Precision, recall and F score is = $58/142 = 0.40$

Using 3 feature, Precision, recall and F score is = $70/142 = 0.49$

Using 4 feature, Precision, recall and F score is = $80/142 = 0.56$

With Increase in the Number of features there in increase in F score.

To Check Redundancy among 20 Summary Variable, we perform **PCA(Principal component analysis).**



Number of Components

This plot tells us that selecting 10 components we can preserve something around 98.8% or 99% of the total variance of the data. The PCA analysis showed that we need at least 10 principal components out of these 22 to reproduce 90% of the total variance present in the data, and at least 9 components to capture 95% of the variance.

EXPERIMENTAL RESULTS

we have adapted some of the well-known **Tree Based** Algorithm for 9500 Normal Company and 500 Shell Company.

AdaBoost is an Ensemble technique where new models are added to correct the errors made by existing models. It weights data points that are hard to detect. **Gradient boosting** use gradient descent algorithm to minimize the loss when adding new models. It is an approach where new models are created that predict the residuals or errors of prior models

and then added together to make the final prediction. **XGBoost** (Extreme Gradient Boosting) has good execution speed and Model Performance. And, In the terms of **Algorithmic Enhancements**, it has Better regularization supports (both the L1 and L2 regularization) to prevent Overfitting.

Data is Biased, we check the performance of the Model using **Precision, Recall and FScore**.

Our Assumption is, most of the shell Company should be predicted correctly. So, we want to **decrease the Number of False Negative (Because it is a Loss for the government)**. Recall is the Better Measure than precision. Here Recall refers to the “**Number of Shell Company Our model is able to detect**”.

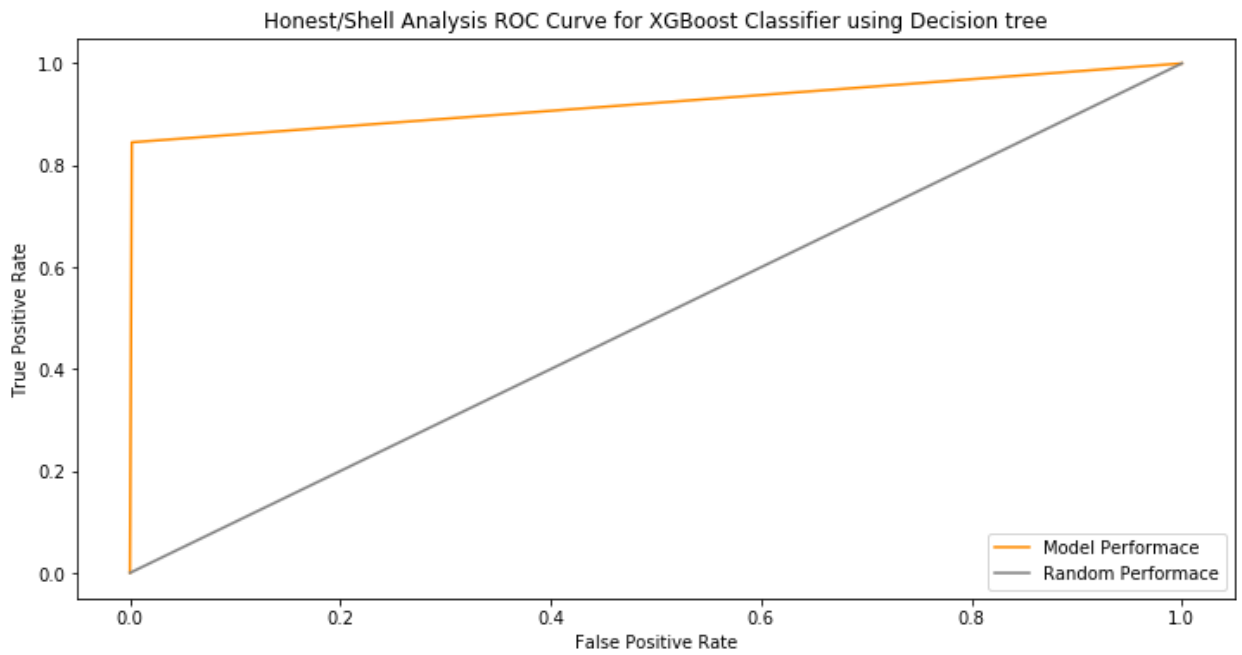
We want **increase recall** in Our Current Scenario.

Algorithm Performance:

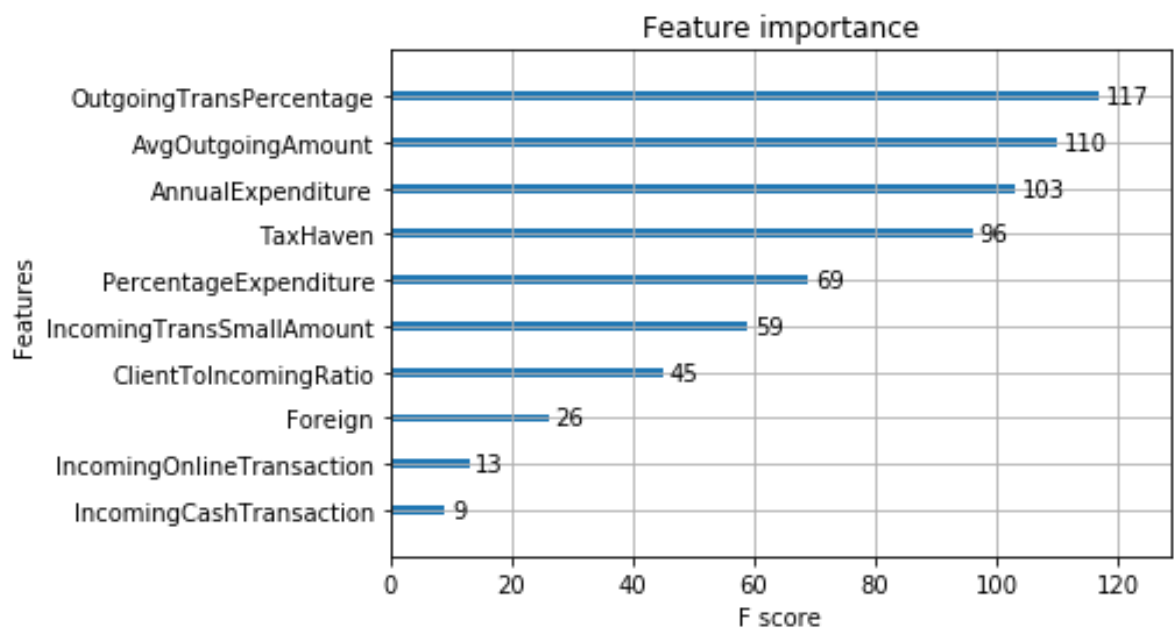
Algorithm	Precision	Recall	FScore
Gradient Boosting	0.72	0.72	0.72
XGBoost	0.96	0.84	0.88

XGBoost Performance best in the Biased Data

[ROC Curve](#) for XGBoost Classifier



Feature Importance XG- Boost Classifier



CONCLUSIONS AND FUTURE WORK

Money laundering is a way to conceal illegally obtained funds. Money laundering works by transferring money in elaborate and complicated financial transactions which mislead anyone who may seek to trace and review the transactions. The objective is to make it difficult to identify the original party to the transaction, known as the **launderer**. However, at the end of the convoluted scheme, the funds ultimately return back to the launderer.

Generated Bank Transactional Data According to the Recent News and Cases in the past. Using the Information we have created multiple variables with different Probability Distribution.

Using Summarized Variable for each company (Feature vector) to train the model to classify the Normal and Shell Company, we applied multiple tree based Models so in future we are able to classify any new company, So ED should investigate that company.

In Future, With the launch of the **Goods and Services Tax (GST)** at midnight on 1 July 2017 by the President of India, and the Government of India. Due to the Loopholes In Model GST Law there is an increase in the Number of Fraud. we would like to detect fraud and develop a strategy to fix the loopholes in GST implementation.

Submitted By:
Kriti Bajaj

