

# 1 INTRODUCTION

## 1.1 Overview

This is the Classic Marketing bank dataset uploaded originally in the Machine Learning Repository. The dataset gives you information about a marketing campaign of a financial institution in which you will have to analyze in order to find ways to look for future strategies in order to improve future marketing campaigns for the bank.

## 1.2 Purpose

A Term Deposit is a deposit that a bank or a financial institution offers with a fixed rate (often better than just opening deposit account) in which your money will be returned back at a specific maturity time.

# 2 LITERATURE SURVEY

## 2.1 Existing Problem

There is a dataset given by that Exploring data analysis for the given there are some of variables types in numeric and nominal they are :

Features

- 1.age | int64 | age in years
- 2.job | object | type of job (categorical: ['maid','unknown','self-employed','student'])
- 3.marital | object | marital status (categorical: ['married','single','divorced'])
- 4.education | object | education background
- 5.default | object | has credit in default? (categorical: ['no','yes'])
- 6.balance | int64 | Balance of the individual
- 7.housing | object | has housing loan? (categorical: ['yes','no'])
- 8.loan | object | has personal loan? (categorical: ['no','yes'])
- 9.contact | object | contact communication type (categorical: ['unknown','cellular','telephone'])
- 10.day | int64 | last contact day of the week (categorical: ['mon','tue','wed','thu','fri'])
- 11.month | object | last contact month of year (categorical: ['may','jun','jul','aug','oct','nov','dec','jan','feb','mar','apr','sep'])
- 12.duration | int64 | last contact duration, in seconds (numeric)
- 13.campaign | int64 | number of contacts performed during this campaign and for this client
- 14.pdays | int64 | number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted)
- 15.previous | int64 | number of contacts performed before this campaign and for this client
- 16.outcome | object | outcome of the previous marketing campaign (categorical: ['unknown','other','failure','success'])

The Class that called as Target/Label it is default at last of attribute in dataset.

**Label**

deposit | object | has the client subscribed a term deposit? (binary: 'yes','no')

## 2.2 Proposed Solution

In the dataset all the predictable values are categorical values, and the class is also a categorical value so, we need Classification algorithms to classify classifiers to get the predicted values. The following Method for Classification algorithm is a Supervised Learning technique that is used to identify the category of new observations on the basis of training data. In Classification, a program learns from the given dataset or observations and then classifies new observation into a number of classes or groups. Such as, **Yes or No, 0 or 1, Spam or Not Spam, cat or dog**, etc. Classes can be called as targets/labels or categories.

The algorithm which implements the classification on a dataset is known as a classifier. we use

- **Binary Classifier:** If the classification problem has only two possible outcomes, then it is called as Binary Classifier.

**Examples:** YES or NO, MALE or FEMALE, SPAM or NOT SPAM, CAT or DOG, etc.

## 3 THEORITICAL ANALYSIS

### 3.1 Block Diagram

- Find Unwanted Columns
- Find Missing Values
- Find Features with one value
- Explore the Categorical Features
- Find Categorical Feature Distribution
- Relationship between Categorical Features and Label
- Explore the Numerical Features
- Find Discrete Numerical Features
- Relation between Discrete numerical Features and Labels
- Find Continous Numerical Features
- Distribution of Continous Numerical Features

- Relation between Continuous numerical Features and Labels
- Find Outliers in numerical features
- Explore the Correlation between numerical features
- Find Pair Plot
- Check the Data set is balanced or not based on target values in classification

## 3.2 Hardware / Software designing

The Hardware require is with good internet connection laptop/desktop

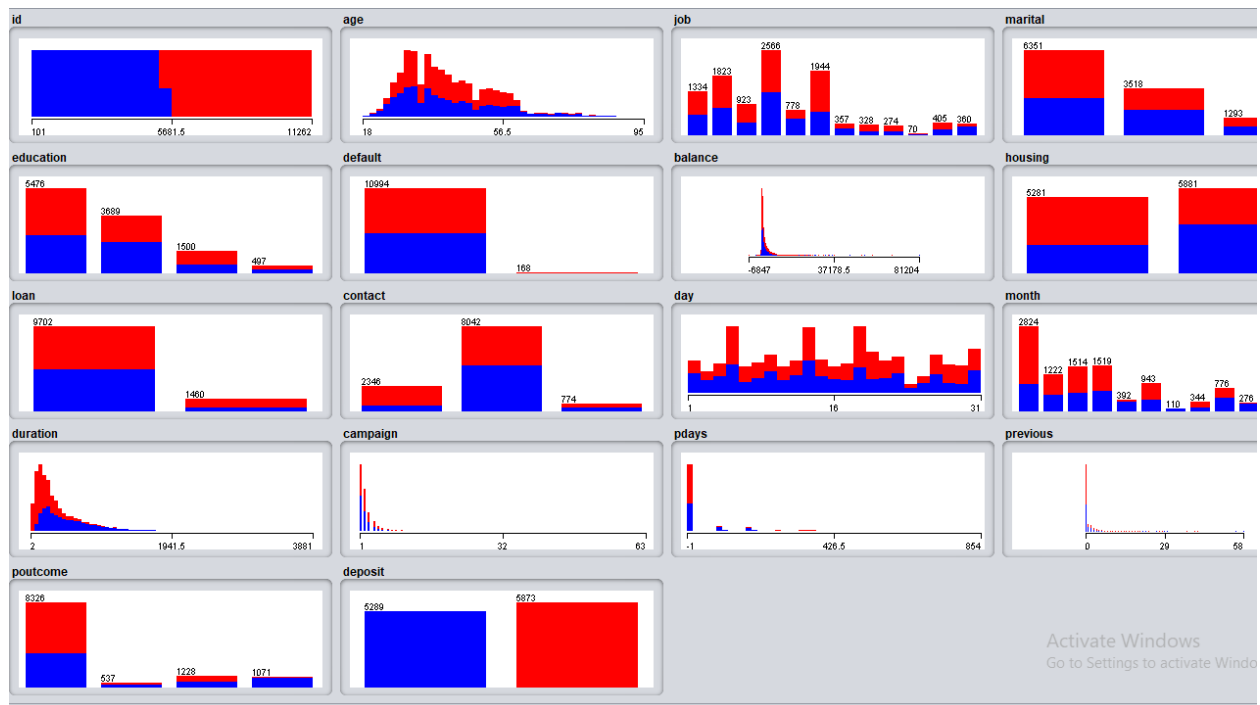
The Software is used in the project is Eclipse

## 4 EXPERIMENTAL INVESTIGATION

- there is no unwanted column present in given dataset to remove
- No missing value found
- No feature with only one value
- there are 9 categorical features
- feature job and month has highest number of categorical values
- client with job type as management records are high in given dataset and housemaid are very less
- client who married are high in records in given dataset and divorced are less
- client whose education background is secondary are in high numbers in given dataset
- default feature seems to be does not play important role as it has value of no at high ratio to value yes which can drop
- data in month of may is high and less in dec
- retired client has high interest on deposit
- client who has housing loan seems to be not interested much on deposit
- if pre campaign outcome that is outcome=succcess then, there is high chance of client to show interest on deposit
- in month of March, September, October and December, client show high interest to deposit
- in month of may, records are high but client interest ratio is very less
- there are 7 numerical features
- there is no Discrete Variables in give dataset
- there are 7 continuous numerical features
- it seems age, days distributed normally
- balance, duration, campaign, pdays and previous heavily skewed towards left and seems to be have some outliers.

- client shows interest on deposit who had discussion for longer duration
- age, balance, duration, campaign, pdays and previous has some outliers
- it seems no feature is heavily correlated with other features
- given dataset seems to be balanced.

## 5 FLOWCHART



## 6 RESULT

eclipse-workspace - org.an/src/main/java/org/an/ClassificationDemo.java - Eclipse IDE

File Edit Source Refactor Navigate Search Project Run Window Help

Problems Javadoc Declaration Console

<terminated> BankData [Java Application] C:\Users\home\p2\pool\plugins\org.eclipse.justi.openjdk.hotspot.jre.full.win32.x86\_64\_11.0.2.v20200815-0835\jre\bin\javaw.exe (May 7, 2021, 1:07:04 PM - 1:07:10 PM)

having the data to be analysingSLF4J: Failed to load class "org.slf4j.impl.StaticLoggerBinder".  
SLF4J: Defaulting to no-operation (NOP) logger implementation  
SLF4J: See <http://www.slf4j.org/codes.html#StaticLoggerBinder> for further details.

11162 rows X 18 cols

id	age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous
101	59	admin.	married	secondary	no	2343	yes	no	unknown	5	may	1042	1	-1	0
102	56	admin.	married	secondary	no	45	no	no	unknown	5	may	1467	1	-1	0
103	41	technician	married	secondary	no	1278	yes	no	unknown	5	may	1389	1	-1	0
104	55	services	married	secondary	no	2476	yes	no	unknown	5	may	579	1	-1	0
105	54	admin.	married	tertiary	no	184	no	no	unknown	5	may	673	2	-1	0
106	42	management	single	tertiary	no	0	yes	yes	unknown	5	may	562	2	-1	0

bank.csv

id	age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous
11260	32	technician	single	secondary	no	29	no	no	cellular	19	aug	156	2	-1	0
11261	43	technician	married	secondary	no	0	no	yes	cellular	8	may	9	2	172	5
11262	34	technician	married	secondary	no	0	no	no	cellular	9	jul	628	1	-1	0

Structure of bank.csv

Index	Column Name	Column Type
0	id	INTEGER
1	age	INTEGER
2	job	STRING
3	marital	STRING
4	education	STRING
5	default	STRING
6	balance	INTEGER
7	housing	STRING
8	loan	STRING
9	contact	STRING
10	day	INTEGER
11	month	STRING
12	duration	INTEGER
13	campaign	INTEGER
14	pdays	INTEGER
15	previous	INTEGER
16	outcome	STRING

Activate Windows  
Go to Settings to activate Windows

eclipse-workspace - org.an/src/main/java/org/an/ClassificationDemo.java - Eclipse IDE

File Edit Source Refactor Navigate Search Project Run Window Help

Problems Javadoc Declaration Console

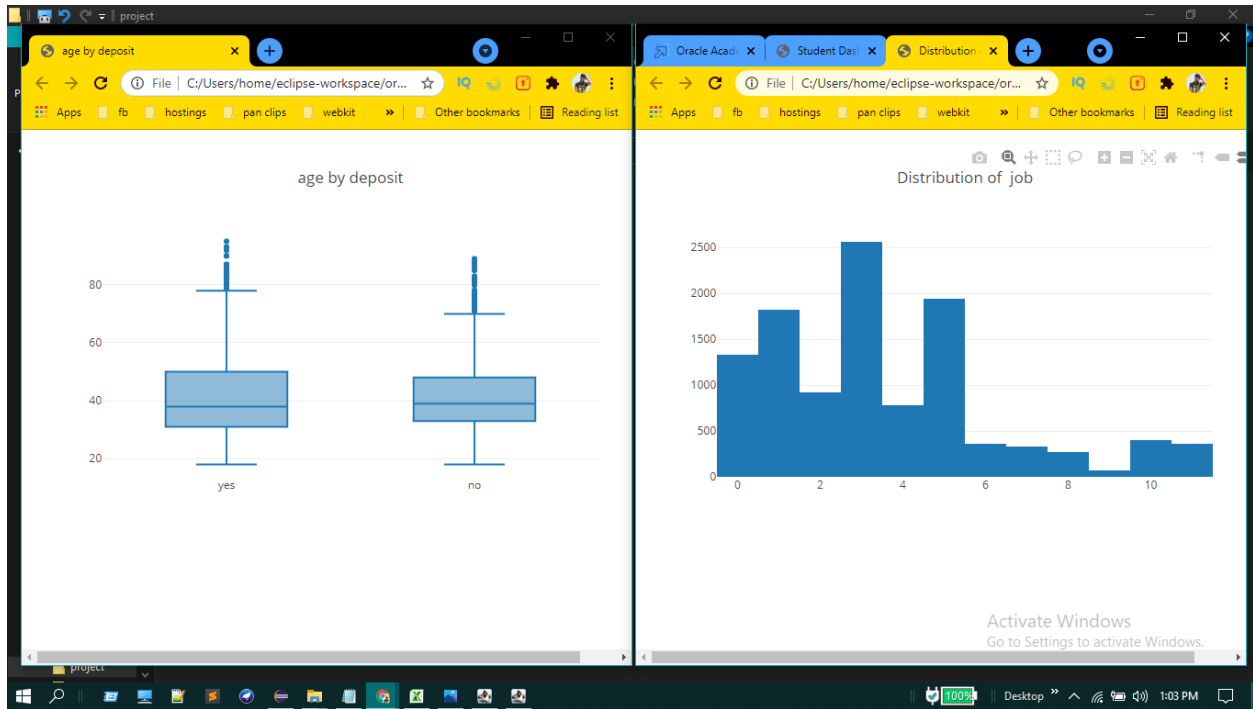
<terminated> BankData [Java Application] C:\Users\home\p2\pool\plugins\org.eclipse.justi.openjdk.hotspot.jre.full.win32.x86\_64\_11.0.2.v20200815-0835\jre\bin\javaw.exe (May 7, 2021, 1:07:04 PM - 1:07:10 PM)

	id	age	job	marital	education	default	balance	housing	loan	contact	day
Count	11162	11162	11162	11162	11162	11162	11162	11162	11162	11162	11162
Sum	63416903	460231					17861547				
Mean	5681.5	41.2319476796274					1528.5385235620836				15.6580361
Min	101	18					-6847				
Max	11262	95					81264				
Range	11161	77					88051				
Variance	10383450.5	141.92836551002986					10403291.123191				70.908854
Std. Dev	3222.3361866819546	11.91336919221552					3225.413325946149				8.4207395
Unique				12	3	4	2	2	2	3	
Top			management	married	secondary	no				cellular	
Top Freq.	775445710		2566	6351	5476	10994		5881	9702	8042	

Missing Values [loan] | Missing Values [education] | Missing Values [previous] | Missing Values [housing] | Missing Values [poutcome] | Missing Values [duration] | Missi

Summary

Activate Windows  
Go to Settings to activate Windows



```
ClassificationDemo [Java Application] C:\Users\home\p2\pool\plugins\org.eclipse.justi.openjdk.hotspot.jre.full.win32.x86_64_11.0.2.v20200815-0835\jre\bin\javaw.exe (May 7, 2021, 1:13:19 PM)

job = student: no (0.0)
marital = single
job = admin.: yes (0.0)
job = technician
age <= 33: yes (2.0)
age > 33: no (2.0)
job = services: yes (2.0)
job = management: yes (2.0/1.0)
job = retired: yes (0.0)
job = blue-collar: no (3.0/1.0)
job = unemployed: yes (0.0)
job = entrepreneur: yes (0.0)
job = housemaid: yes (0.0)
job = unknown: yes (0.0)
job = self-employed: yes (0.0)
job = student: yes (0.0)
marital = divorced: no (3.0)
campaign > 3: yes (3.0)
month = jun: yes (6.0)
month = jul: yes (1.0)
month = aug: yes (4.0)
month = oct: no (4.0/1.0)
month = nov
| marital = married: no (10.0)
| marital = single: no (5.0/1.0)
| marital = divorced: yes (1.0)
month = dec: yes (2.0)
month = jan
| marital = married: no (3.0)
| marital = single: no (2.0)
| marital = divorced: yes (1.0)
month = feb
| marital = married: no (8.0/1.0)
| marital = single: yes (3.0/1.0)
| marital = divorced: yes (1.0)
month = mar: yes (5.0/1.0)
month = apr
| loan = no
| previous <= 1: no (9.0)
```

```
eclipse-workspace - org.an/src/main/java/org/an/ClassificationDemo.java - Eclipse IDE
File Edit Source Refactor Navigate Search Project Run Window Help
ClassificationDemo [Java Application] C:\Users\home\p2\pool\plugins\org.eclipse.justj.openjdk.hotspot.jre.full.win32.x86_64_11.0.2.v20200815-0835\jre\bin\javaw.exe (May 7, 2021, 1:13:19 PM)

job = student: no (0.0)
marital = single
job = admin.: yes (0.0)
job = technician
age <= 33: yes (2.0)
age > 33: no (2.0)
job = services: yes (2.0)
job = management: yes (2.0/1.0)
job = retired: yes (0.0)
job = blue-collar: no (3.0/1.0)
job = unemployed: yes (0.0)
job = entrepreneur: yes (0.0)
job = housemaid: yes (0.0)
job = unknown: yes (0.0)
job = self-employed: yes (0.0)
job = student: yes (0.0)
marital = divorced: no (3.0)
campaign > 3: yes (3.0)
month = jun: yes (6.0)
month = jul: yes (1.0)
month = aug: yes (4.0)
month = oct: no (4.0/1.0)
month = nov
marital = married: no (10.0)
marital = single: no (5.0/1.0)
marital = divorced: yes (1.0)
month = dec: yes (2.0)
month = jan
marital = married: no (3.0)
marital = single: no (2.0)
marital = divorced: yes (1.0)
month = feb
marital = married: no (8.0/1.0)
marital = single: yes (3.0/1.0)
marital = divorced: yes (1.0)
month = mar: yes (5.0/1.0)
month = apr
loan = no
previous <= 1: no (9.0)
```

```
eclipse-workspace - org.an/src/main/java/org/an/ClassificationDemo.java - Eclipse IDE
File Edit Source Refactor Navigate Search Project Run Window Help
ClassificationDemo [Java Application] C:\Users\home\p2\pool\plugins\org.eclipse.justj.openjdk.hotspot.jre.full.win32.x86_64_11.0.2.v20200815-0835\jre\bin\javaw.exe (May 7, 2021, 1:15:04 PM)

Number of Leaves : 2352

Size of the tree : 3050

And the class deposit is... yes

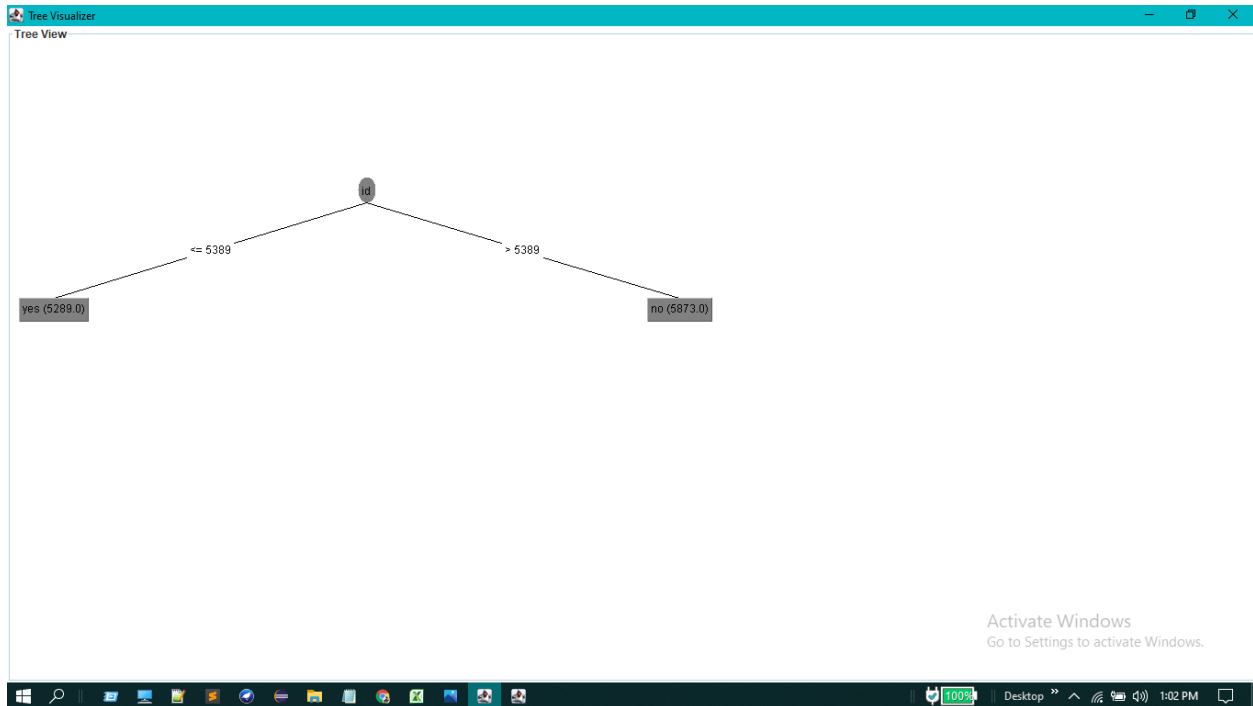
Correctly Classified Instances 9487 84.9937 %
Incorrectly Classified Instances 1675 15.0063 %
Kappa statistic 0.7003
Mean absolute error 0.2109
Root mean squared error 0.3501
Relative absolute error 42.3044 %
Root relative squared error 70.1137 %
Total Number of Instances 11162

=== Confusion Matrix ===
 a b <-- classified as
4678 619 | a = yes
1056 4817 | b = no

** Logistic Regression Evaluation with Datasets **

Correctly Classified Instances 2755 82.2634 %
Incorrectly Classified Instances 594 17.7366 %
Kappa statistic 0.6436
Mean absolute error 0.2558
Root mean squared error 0.3567
Relative absolute error 51.2943 %
Root relative squared error 71.4336 %
Total Number of Instances 3349

the expression for the input data as per algorithm is Logistic Regression with ridge parameter of 1.0E-8
Coefficients...
Variable Class
-----
age -0.0014
technician 0.147
```



eclipse-workspace - org.an/src/main/java/org/an/ClassificationDemo.java - Eclipse IDE

File Edit Source Refactor Navigate Search Project Run Window Help

ClassificationDemo [Java Application] C:\Users\home\p2\pool\plugins\org.eclipse.justj.openjdk.hotspot.jre.full.win32.x86\_64\_11.0.2.v20200815-0835\jre\bin\javaw.exe (May 7, 2021, 1:15:04 PM)

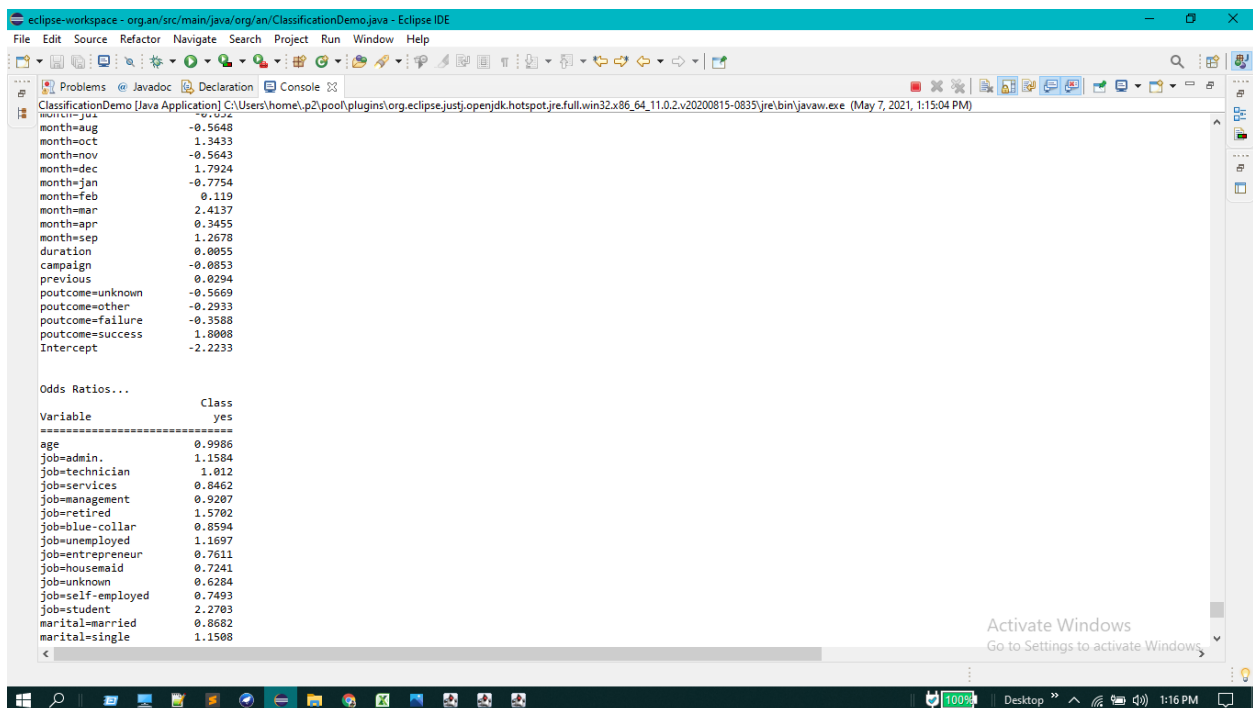
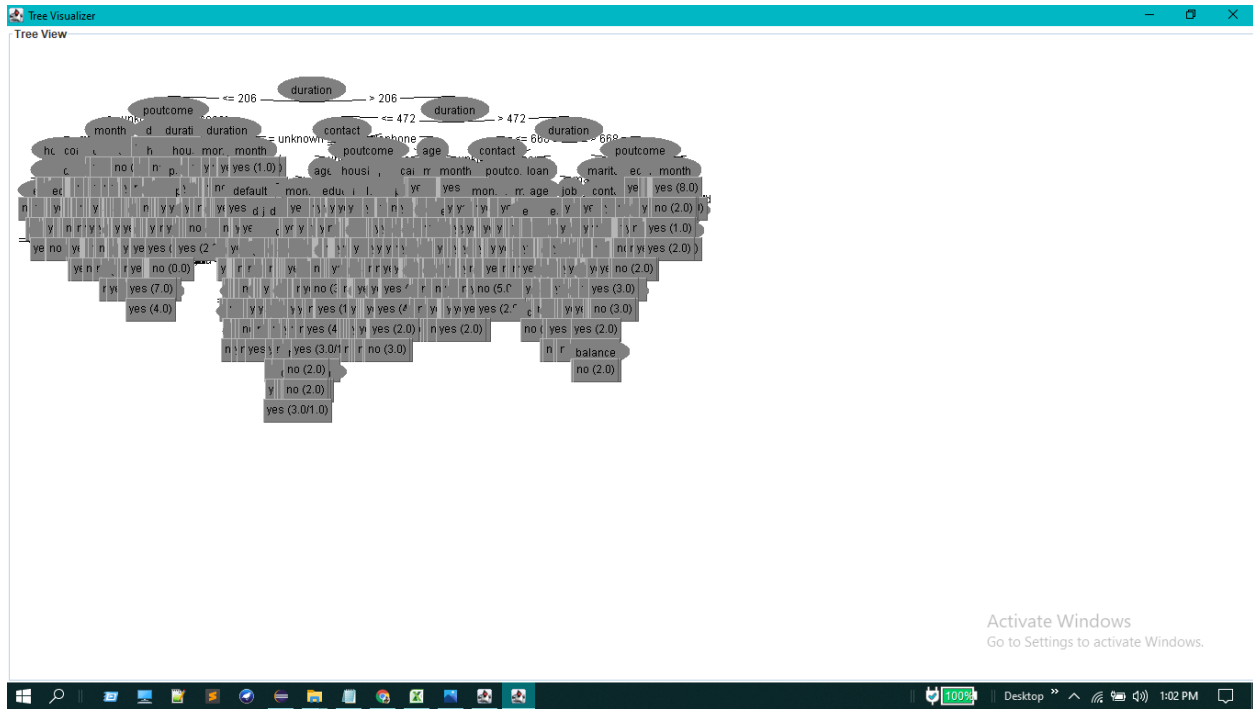
month=jan	0.0224
month=aug	-0.5648
month=oct	1.3433
month=nov	-0.5643
month=dec	1.7924
month=jan	-0.7754
month=feb	0.119
month=mar	2.4137
month=apr	0.3455
month=sep	1.2678
duration	0.0055
campaign	-0.0853
previous	0.0294
poutcome=unknown	-0.5669
poutcome=other	-0.2933
poutcome=failure	-0.3508
poutcome=success	1.8008
Intercept	-2.2233

Odds Ratios...

Variable	Class
age	0.9986
job=admin.	1.1584
job=technician	1.012
job=services	0.8462
job=management	0.9207
job=retired	1.5702
job=blue-collar	0.8594
job=unemployed	1.1697
job=entrepreneur	0.7611
job=housemaid	0.7241
job=unknown	0.6284
job=self-employed	0.7493
job=student	2.2703
marital=married	0.8682
marital=single	1.1508

Activate Windows  
Go to Settings to activate Windows.





```

eclipse-workspace - org.an/src/main/java/org/an/ClassificationDemo.java - Eclipse IDE
File Edit Source Refactor Navigate Search Project Run Window Help
ClassificationDemo [Java Application] C:\Users\home\p2\pool\plugins\org.eclipse.justj.openjdk.hotspot.jre.full.win32.x86_64_11.0.2.v20200815-0835\jre\bin\javaw.exe (May 7, 2021, 1:15:04 PM)
contact=unknown 0.3639
contact=cellular 1.8966
contact=telephone 1.7723
day 1.0057
month=may 0.7515
month=jun 2.0439
month=jul 0.5315
month=aug 0.5685
month=oct 3.8316
month=nov 0.5688
month=dec 6.0037
month=jan 0.4605
month=feb 1.1263
month=mar 11.1752
month=apr 1.4127
month=sep 3.5531
duration 1.0056
campaign 0.9182
previous 1.0298
poutcome=unknown 0.5673
poutcome=other 0.7458
poutcome=failure 0.6985
poutcome=success 6.0542

Confusion matrix:
[1259.0, 329.0]
[265.0, 1496.0]
-----
Area under the curve
0.8993762846562163
-----
[Correct, Incorrect, Kappa, Total cost, Average cost, KB relative, KB information, Correlation, Complexity 0, Complexity scheme, Complexity improvement, MAE, RMSE, RAE, RRSE, Coverag
Recall :0.85
Precision:0.82
F1 score:0.83
Accuracy:0.82
-----
Predicted label:
0.0

```

## 7 ADVANTAGES & DISADVANTAGES

High performance on non – linear **problems**, not biased by outliers, that many of the **classifications** themselves are based on subjective judgments, which may or may not be shared by everyone participating. This would lead to differences in perceived value.

	ADVANTAGE	DISADVANTAGE
Logistic Regression	Probabilistic Approach, gives information about statistical significance of features.	The assumptions of logistic regression.
Decision Tree	Interpretability, no need for feature scaling, works on both linear non – linear problems.	Poor results on Classification very small datasets, overfitting can easily occur.

Random Forest Classification	Powerful and accurate, good performance on many problems, including non – linear.	No interpretability, overfitting can easily occur, need to choose the number of trees manually.
------------------------------	---	---

## 8 APPLICATIONS

- **BANK APPLICATIONS FOR PREDICTING LABELS**
- **DETECTIONS - FRAUDS**
- **RECOGNITIONS - SPEECH**
- **CLASSIFICATIONS - DISEASES**
- **IDENTIFICATIONS - BIOMETRIC**

## 9 CONCLUSION

The dataset is split into two sets that is one is train\_data set and another is test\_data set this is getting best score by an Algorithm before building model after that model was build by tree Structure and tested/evaluated the model.

## 10 FUTURE SCOPE

In the real time that cannot be defined as predicted values. so, got the some of outliers and been removing that and trained the dataset and tested for model evaluation it gives best score.

## 11 BIBLIOGRAPHY

my references are practicing more datasets also referred many websites to learn mainly this is oracle on the boot\_camp conducted by smarinternz was ver helpful.thanks for that.