**IMPLEMENTATION AND ANALYSIS OF PROJECT**

**About Dataset:**

Term deposits are major source of income for a bank.  
In term deposit your money is invested for an agreed rate of interest over a fixed amount of time, or fixed term.

The best way to sell term deposits by bank is telephonic marketing.

For telephonic marketing we don’t call every customer but only targeted customers.

**Data Understanding**

The data is related to the direct marketing campaigns of a Portuguese banking institution.

We need to predict whether a customer may take term plan or not.

Business Idea: - Predict Customers which can take term plan.

Supervised Binary Classification Project

Domain – Banking Domain

**Implementation**

Data Shape – (45211, 17)

**Data Preprocessing**

**Data Cleaning**

**1. Age**

* No Missing Values or null values
* Outlier Handling **–** Above whiskers only 1 % of values so applied outlier treatment

**2. Balance**

* No Missing Valuesor null values.
* The data was positive skewed so transformation was required to reduce skewness in data
* The column has negative and zero values so log transformation is not possible, so applied cube root transformation
* Outlier Handlingafter cube root transformation

**3. Duration**

* No Missing Valuesor null values
* The data was positive skewed so transformation was required to reduce skewness in data
* The column has negative and zero values so log transformation is not possible, so applied cube root transformation
* Outlier Handlingafter cube root transformation

**4. Campaign**

* No Missing Valuesor null values
* The data was positive skewed so transformation was required to reduce skewness in data
* Applied cube root transformation to maintain uniformity with other numerical columns
* Outlier Handling

job – club

* self-employed, entrepreneur, housemaid into 1 as self-employed
* unemployed, unknown, student into 1 as unemployed

marital – keep as it is

education – keep as it is

default – drop as most of the values are same

housing – keep as it is

loan – drop as most of the values are same

contact – keep as it is

day – keep as it is

month – keep as it is for transformation phase

pdays – drop

previous – drop

poutcome – drop

New Data Shape – (45211,12)

**Transformation**

1. Age – Raw, Binning
2. Job – No Transformation
3. Marital – No Transformation
4. Education - No Transformation
5. Housing - No Transformation
6. Contact - No Transformation
7. Day – Raw, Week of the month,
8. Month – Financial Quarters , Drop raw variables
9. Balance – No Transformation
10. Duration – Raw, Binning, ILATE Transformation
11. Campaign – Keep it as it is

**Data Encoding**

Nominal Variables – One Hot Encoding and drop the originals

Ordinal Variables – Label Encoding As they have natural order

Numerical Variables – Do Nothing

**MODELLING**

Logistic Regression

Decision Tree

Random Forest Classifier

Naïve Bayes Classifier

**MODEL ANALYSIS**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **MODEL** | **ACCURACY** | **PRECISION** | **RECALL** | **F1-SCORE** |
| **DECISION TREE**  **CLASSIFIER** | 0.88 | FOR NO-0.92 FOR YES-0.41 | FOR NO-0.92 FOR YES-0.42 | FOR NO-0.92 FOR YES-0.41 |
| **NAÏVE BAYES CLASSIFIER** | 0.83 | FOR NO-0.91 FOR YES-0.32 | FOR NO-0.92 FOR YES-0.41 | FOR NO-0.90 FOR YES-0.34 |
| **RANDOMFOREST CLASSIFIER** | 0.88 | FOR NO-0.91 FOR YES-0.59 | FOR NO-0.97 FOR YES-0.27 | FOR NO-0.94 FOR YES-0.37 |

**INFERENCE**

According to me best model Random forest classifier.

Improvement for prediction in case of yes scenario will be better .