Project Proposal

**Project Title:**

# Bank Marketing

* The data is related with direct marketing campaigns (phone calls) of a Portuguese banking institution. The classification goal is to predict if the client will subscribe a term deposit (variable y).

1. **Problem statement and business use -case**

* There has been a revenue decline for the Portuguese bank and they would like to know what actions to take. After investigation, they found out that the root cause is that their clients are not depositing as frequently as before. Knowing that term deposits allow banks to hold onto a deposit for a specific amount of time, so banks can invest in higher gain financial products to make a profit. In addition, banks also hold better chance to persuade term deposit clients into buying other products such as funds or insurance to further increase their revenues. As a result, the Portuguese bank would like to identify existing clients that have higher chance to subscribe for a term deposit and focus marketing effort on such clients.
* To resolve the problem, we suggest a classification approach to predict which clients are more likely to subscribe for term deposits.

1. **Data science work flow**

* We are going to predict the success of bank telemarketing based on data collected from bank marketing based on data collected from bank marketing campaigns in Portugal

Data provided us has factors like

Input variables:

# bank client data:

1 . age (numeric)

2 . job : type of job (categorical: "admin.","blue-collar","entrepreneur","housemaid","management","retired","self-employed","services","student","technician","unemployed","unknown")

3 - marital : marital status (categorical: "divorced","married","single","unknown"; note: "divorced" means divorced or widowed)

4 - education (categorical: "basic.4y","basic.6y","basic.9y","high.school","illiterate","professional.course","university.degree","unknown")

5 - default: has credit in default? (categorical: "no","yes","unknown")

6 - housing: has housing loan? (categorical: "no","yes","unknown")

7 - loan: has personal loan? (categorical: "no","yes","unknown")

# related with the last contact of the current campaign:

8 - contact: contact communication type (categorical: "cellular","telephone")

9 - month: last contact month of year (categorical: "jan", "feb", "mar", ..., "nov", "dec")

10 - day\_of\_week: last contact day of the week (categorical: "mon","tue","wed","thu","fri")

11 - duration: last contact duration, in seconds (numeric).

12 - campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)

13 - pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted)

14 - previous: number of contacts performed before this campaign and for this client (numeric)

15 - poutcome: outcome of the previous marketing campaign (categorical: "failure","nonexistent","success")

# social and economic context attributes

16 - emp.var.rate: employment variation rate - quarterly indicator (numeric)

17 - cons.price.idx: consumer price index - monthly indicator (numeric)

18 - cons.conf.idx: consumer confidence index - monthly indicator (numeric)

19 - euribor3m: euribor 3 month rate - daily indicator (numeric)

20 - nr.employed: number of employees - quarterly indicator (numeric)

Output variable (desired target):

21 - y - has the client subscribed a term deposit? (binary: "yes","no")

* The dataset is consisted of 45,211 customer data
* We want to analyse different factors that would be impacting the revenue of Portuguese bank.

1. **EDA questions**

* Is job has effect on subscription of a term deposit.
* Is Marital status has effect on subscription of a term deposit.
* Is education has effect on subscription of a term deposit.
* Is housing loan has effect on subscription of a term deposit.
* Is personal loan has effect on subscription of a term deposit.
* Is type of contact has effect on subscription of a term deposit
* Is month of the year has effect on subscription of a term deposit
* Is day of week has effect on subscription of a term deposit
* Is outcome of the previous marketing campaign has effect on subscription of a term deposit

1. **Data Analytics**

* **Descriptive** analysis

Duration has effect on subscription of term deposit

Blue collar job people ‘s subscription is very less

* **Diagnostic** analysis

Blue-collar job may not be interested in subscription of term deposit as the bank have rate of interest per annum.

* **Predictive** analysis

Duration of the calls should increase.

Count of Blue collar job subscribers should increase

* **Prescriptive** analysis

Rate per annum and some attractive offers should provide to attract blue collar job people.

Duration of the calls to customers and clients can increase the subscribes.so hire a person who can speak well to customers.

1. **How our problem can be solved?**

* Provide some attractive offers to customers
* For Students some scholarships to be provided
* Formalities for loan should be easy .and loan provided in average rate of interest
* Be more friendly with Entrepreneur type of customers .

1. **Who will be impacted?**

Blue collars and entrepreneur type of Customers will be highly impacted by these system, If it function in proper way. So that revenue will increase.`

1. **Data collection**

Dataset-is in the form of excel sheet

Dataset in the form of excel file.

Dataset has 41188 rows × 21columns

Dataset file is in the form of.csv

It has size of 5.1mb

Datatypes

|  |  |  |
| --- | --- | --- |
| Field | Datatype | Description |
| Age | integer | Age of the customer |
| Job | Object(categorical) | Type of job , (admin., blue-collar, entrepreneur, housemaid, management retired, self-employed, services, student, technician, unemployed, unknown |
| Marital | Object(categorical) | divorced, married, single, unknown |
| Education | Object(categorical) | basic.4y, basic.6y, basic.9y, high.school, illiterate, professional.course, university.degree, unknown |
| Default: | Object(categorical) | has credit in default  : no, yes, unknown. |
| Housing | Object(categorical) | has housing loan? no, yes, unknown |
| Loan | categorical | has personal loan? no, yes, unknown) |
| Contact | Categorical | Communication type cellular,telephone |
| Month | Categorical | last contact month of year jan feb, mar,nov, dec |
| Dayofweek | Categorical | last contact day of the week  mon,tue,wed,thu,fri |
| Duration | Numeric | If duration=0 then y='no' |
| Campaign | Numeric | number of contacts performed during this campaign and for this client |
| Pdays | Numeric | number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted) |
| Previous | Numeric | number of contacts performed before this campaign and for this client |
| Poutcome | Categorical | outcome of the previous marketing campaign  failure,nonexistent, success. |
| Emp.var.rate | Numeric | employment variation rate - quarterly indicator |
| Cons.price.idx: | Numeric | consumer price index - monthly indicator |
| Cons.conf.idx | Numeric | consumer confidence index - monthly indicator |
| Euribor3m | Numeric | euribor 3 month rate - daily indicator |
| Nr.employed | Numeric | Number of employees - quarterly indicator |

**9. Data Pre-processing ,preparation and feature engineering**

* Identifying data quality-
* Check for data inconsistency, all datatypes are correct and no encoding issues.
* Handling missing values
* Treating categorical values
* Treating outliers
* Do scaling
* Feature engineering

**10. Machine Learning and Model Selection**

We are using logistic regression as we need to predict whether the client will subscribe (yes/no)to a term deposit