**NITTE MEENAKSHI INSTITUTE OF TECHNOLOGY**

(AN AUTONOMOUS INSTITUTION, AFFILIATED TO VISVESVARAYA TECHNOLOGICAL UNIVERSITY, BELGAUM,

APPROVED BY AICTE & GOVT.OF KARNATAKA

Logo

Description automatically generated

**Learning Activity Project Proposal**

on

**Title: BANKING DATASET**

*Submitted in partial fulfilment of the requirement for the award of Degree of*

*Bachelor of Engineering*

*in*

*Computer Science and Engineering*

**Faculty In-charge: Dr. VANI VASUDEVAN**

**Submitted by:**

N.NAGA JAYANTH 1NT19CS118

C.VINEETH 1NT19CS061

T.EKESWAR REDDY 1NT19CS198

V.LAXMI DEEPAK 1NT19CS214

**INTRODUCTION:**

Data mining is defined as extracting information from huge sets of data. We can say that data mining is the procedure of mining knowledge from data.

Term deposits are a major source of income for a bank. A term deposit is a cash investment held at a financial institution. Your money is invested for an agreed rate of interest over a fixed amount of time, or term. The bank has various outreach plans to sell term deposits to their customers such as email marketing, advertisements, telephonic marketing, and digital marketing.

Telephonic marketing campaigns remain one of the most effective ways to reach out to people. However, they require huge investment as large call centers are hired to execute these campaigns. Hence, it is crucial to identify the customers most likely to convert beforehand so that they can be specifically targeted via call.

**DATA MINING TASK:**

* Classification – To create data structures of predefined classes
* Prediction – It uses regression analysis and detect the missing values of data
* Outlier analysis–To discard noise
* Correlation analysis -To reduce dataset, to get more accuracy
* Association analysis -Uncovering the relationship between data and deciding the rules of the association.
* Data discrimination compares common features of class which is under study. Output can be represented in many forms like graphs, pie charts & curves.

**DATA SET:**

We took the dataset from Kaggle website, and it has the attributes

Age,job,maritalstatus,education,default,balance,housing,loan,contact,day,month,duration,campaign,pday,previous,poutcome,Y.

1 - age (numeric)  
2 - job : type of job (categorical: "admin.","unknown","unemployed","management","housemaid","entrepreneur","student",  
"blue-collar”, “self-employed","retired","technician","services")  
3 - marital : marital status (categorical: "married","divorced","single"; note: "divorced" means divorced or widowed)  
4 - education (categorical: "unknown","secondary","primary","tertiary")  
5 - default: has credit in default? (Binary: "yes”, “no")  
6 - balance: average yearly balance, in euros (numeric)  
7 - housing: has a housing loan? (Binary: "yes”, “no")  
8 - loan: has a personal loan? (binary: "yes”, “no")  
# related with the last contact of the current campaign:  
9 - contact: contact communication type (categorical: "unknown","telephone","cellular")  
10 - day: last contact day of the month (numeric)  
11 - month: last contact month of year (categorical: "Jan", "Feb", "Mar", …, "Nov", "Dec")  
12 - duration: last contact duration, in seconds (numeric)  
# other attributes:  
13 - campaign: number of contacts performed during this campaign and for this client (numeric, includes the last contact)  
14 - pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric, -1 means client was not previously contacted)  
15 - previous: number of contacts performed before this campaign and for this client (numeric)  
16 - poutcome: outcome of the previous marketing campaign (categorical: "unknown","other","failure","success")

Output variable (desired target):  
17 - y - has the client subscribed to a term deposit? (Binary: "yes”, “no")

Missing Attribute Values: None

The dataset contains 45211 rows and 17 columns.

**METHODS AND MODELS:**

To come out with a good output we use the following methods:

* Decision Tree algorithms
* Outliers for missing values
* Graphs for visualization
* KNN algorithm
* Min-max normalization

**ASSESSMENT:**

We use python3 to do classifications and tasks. By applying a data mining algorithm, we find accuracy. Comparing training and testing sets class attributes we conclude on reducing dataset using correlation or not. If we are satisfied with the accuracy, we will be continuing with the original dataset.

**PRESENTATION AND VISUALIZATION:**

The data is related to direct marketing campaigns (phone calls) of a Portuguese banking institution. The classification goal is to predict if the client will subscribe to a term deposit or not.

For visualization, the graphs would be a good fit. Using matplotlib in python the graphs are created.

**ROLES:**

We worked as a team in finding dataset and preparing proposal.

As our team has a good understanding and common ideas we did everything with unity.

**SCHEDULE:**

|  |  |
| --- | --- |
| **Date** | **Task to be completed** |
| 03/01/2022 | Proposal uploaded in Github |
| 16/01/2022 | Completion of implementation |
| 17/01/2022 | Final submission of the report |

**BIBLIOGRAPHY**

1. [www.kaggle.com](http://www.kaggle.com)
2. [**https://www.kaggle.com/prakharrathi25/banking-dataset-marketing-targets**](https://www.kaggle.com/prakharrathi25/banking-dataset-marketing-targets)
3. [**https://www.tutorialspoint.com/data\_mining/dm\_quick\_guide.htm**](https://www.tutorialspoint.com/data_mining/dm_quick_guide.htm)
4. [**https://www.javatpoint.com/data-mining**](https://www.javatpoint.com/data-mining)