**Project Summary**

| Batch Details | PGP-DSE October’21 |
| --- | --- |
| Team Members | Nikhil Mishra,Piyush Dalal,Rijul Choudhary,Anupriya Bhandari,Prakhar Raj Gupta,Ayush Dagar,Deepansh Saxena,Shivam Sharma |
| Domain of Project | Finance and Risk Analytics |
| Proposed Project Title | Predict the probability of customer’s loan approval |
| Group Number | 2 |
| Team Leader | Nikhil Mishra |
| Mentor Name | Mr. Ankush Bansal |

**Dataset Name-** Banking Loan Prediction

**Introduction to the problem/domain/background details:**

The dataset chosen belongs to the domain of financial services. It contains customer loan history. The bank also analyzes whether the customer can be a qualified lead or not. Leads are customers who are interested in the bank’s products and services. To get as many leads as possible is one of the main goals of any financial institution.

**Problem Statement:**

The problem to be solved here is to identify leads’ segments having a high conversion ratio i.e, lead to buying a financial product. The focus of the bank here is to increase the number of leads getting into the conversion funnel. Customer acquisition is being done through digital channels like search, display, email campaigns and via affiliate partners. As expected, the bank sees differential conversion depending on the sources and the quality of these leads. We aim to segment these leads and understand the customer potential.

The customers’ loan amount requested and the source of the lead will be used to analyze the probability of customers' conversion.

-------------------------------------------------------------------------------------------------------------------------------

*A bank has recently started generating leads through digital channels to cope with the marketing strategy of competitor banks.They source leads through various channels like search, display, email campaigns and via affiliate partners.The bank staff is not able to prioritize the leads that are more likely to get converted into their customers.The marketing and sales manager of the bank is wondering if he/she could just predict the lead conversion using the data captured via digital channels on the basis of the data of their existing customers.*

*The Bank wants to identify leads’ segments having a high conversion ratio i.e, lead to buying a financial product. The focus of the bank here is to increase the number of leads getting into the conversion funnel. As expected, the bank sees differential conversion depending on the sources and the quality of these leads. We aim to segment these leads and understand the customer potential.*

**Business problem/ Impact in business of your problem/Need for this study/Abstract (Executive summary):**

Banks play a major role in the creation of credit, which leads to an increase in production, employment, and consumer spending, thereby boosting the economy. The creation of credit is done by providing loans to individuals and businesses. However, loans can only be disbursed after ascertaining the quality of customers as risks are involved. Hence, the quality of leads needs to be thoroughly analyzed. For the profitability of the financial institution, it is also necessary for them to understand whether the existing customers can be leads for other financial products as well.

---------------------------------------------------------------------------------------------------------------------------------

*By optimizing marketing campaigns with predictive analytics, the Bank can generate new customer responses or purchases more accurately. Predictive models can help the Bank to attract their future most valued customers.*

**Variable identification:**

Independent Variables:

* ID
* Gender
* DOB
* Lead*Creation*Date
* City\_Code
* City\_Category
* Employer\_Code
* Employer\_Category1
* Employer\_Category2
* Monthly\_Income
* Customer*Existing*Primary*Bank*Code
* Primary*Bank*Type
* Contacted
* Source
* Source\_Category
* Existing\_EMI
* Loan\_Amount
* Loan\_Period
* Interest\_Rate
* EMI
* Var1

Target Variable: **Approved**

**Variable information/Data description:**

| **Columns** | **Column Type** | **Column Description** |
| --- | --- | --- |
| ID | Alpha-Numeric | Unique Customer ID |
| Gender | Categorical | Gender of the applicant |
| DOB | Dates | Date of Birth of the applicant |
| Lead\_Creation\_Date | Dates | Date on which lead was created |
| City\_Code | Categorical | Code representing the city where the lead puts up |
| City\_Category | Categorical | Type of city - Metropolitan,Urban and Rural |
| Employer\_Code | Alpha-Numeric | Unique Employer ID of customer's workplace assigned by the bank |
| Employer\_Category1 | Categorical | 3 Broader Employer Categories assigned by Bank Management |
| Employer\_Category2 | Categorical | 4 Anonymised Sub-Categories assigned by Bank Management |
| Monthly\_Income | Numerical | Customer's Monthly Income in Dollars |
| Customer\_Existing\_Primary\_Bank\_Code | Alpha-Numeric | Codes representing customer's utilized bank names |
| Primary\_Bank\_Type | Categorical | Customer's Primary Bank type- Public and Private |
| Contacted | Categorical | Whether Customer contact details are verified - Yes/No |
| Source | Alpha-Numeric | Categorical Variable representing source of lead |
| Source\_Category | Categorical | 7 broader categories of source captured by the bank management |
| Existing\_EMI | Numerical | EMI of Existing Loans in Dollars |
| Loan\_Amount | Numerical | Loan Amount Requested in dollars |
| Loan\_Period | Numerical | Period of loan requested in years |
| Interest\_Rate | Numerical | Interest Rate of Submitted Loan Amount |
| EMI | Numerical | EMI of Requested Loan Amount in dollars |
| Var1 | Categorical | Anonymized Categorical Variable on Lead Rating captured by Bank Management |
| Approved | Categorical | Whether a loan is Approved or not (1-0) . Customer is Qualified Lead or not (1-0) |

**Timeline Chart (Weekly plan)**: Tentative weekly plan that you will be following.

| Understanding the Domain/Dataset/Problem Statement/Synopsis | Week 1 |
| --- | --- |
| Studying the variables, Data Cleaning | Week 2 |
| Pre Processing Data, Encoding, Feature Engineering | Week 3 |
| Basic Model Preparation | Week 4 |
| Interim Report and presentation prep | Week 5 |
| Data Modeling | Week 6 |
| Data Modeling and Interpretation of Model, Visualization | Week 7, 8 |
| Final Report and Presentation | Week 9 |

**Future Work and Methodology :**

The entire project will be adapted with the data mining life cycle as under the methodology of Cross-Industry Standard Process for Data Mining (CRISP-DM) process model.It’s a flexible and easily customizable model used for both modeling as well as data exploration and visualization in identifying particular patterns.

The six phases of CRISP-DM data life cycle are as follows -

* Business Understanding
* Data Understanding
* Data Preparation
* Modeling
* Evaluation
* Deployment

Here, since the problem is based on a typical classification model,we could use classification techniques or logistic regression while dealing with the model. The learning method for the classification might be Random Forest or in the form of software, Xgboost might also be applicable.

**References (Data set source/Journals/articles):**

https://www.kaggle.com/arashnic/banking-loan-prediction

**Declaration**: This is to declare that the dataset that we are using for our capstone project does not have any relevant legality associated to it and can be used to showcase the work we do on it as a presentation in Great Learning