horizontal line

Project Summary

|  |  |
| --- | --- |
| ***Batch details*** | **ONLINE JANUARY 2021 BATCH ( PGPDSE-FT ONLINE JAN21 )** |
| ***Team members*** | **KADIMI CHANDRA TEJA-BJS5YIR1GX****MANOJ PRAKASH P-EY9LIV25XE****ANUSHA AINAPUR-W3BYOM5BCP****ARVIND KUMAR YADAV-P4Q62CQ7DG****ABHIJEET CHANDRAKANT GIRI -  C53IZ7EZ1D** |
| ***Domain of Project*** | **PREDICTIVE ANALYTICS** |
| ***Proposed project title*** | **VEHICLE LOAN DEFAULT PREDICTION** |
| ***Group Number*** | **GROUP 6** |
| ***Team Leader*** | **KADIMI  CHANDRA  TEJA** |
| ***Mentor Name*** | **MR. KONETI NAVEEN KUMAR YADAV** |

Date: 09-07-2021

Table of Contents

[1. OVERVIEW (50-100 words)](#_Toc76226848)

[2. Business problem statement (GOALS)](#_Toc76226849)

[3. METHODOLOGY TO BE FOLLOWED )](#_Toc76226852)

[4. REFERENCES](#_Toc76226855)

Project Details

# OVERVIEW (50-100 words)

The objective of this project is to predict whether it will be Payment default in the EMI on Vehicle Loan on due date or not using the Loan Default Prediction Dataset. The original dataset came from L&T Financial Services & Analytics Vidhya presented Data Science Fin Hack competition. There will be six phases - Business Understanding, Data Understanding, Data preparation, Modeling, Evaluation.

# Business problem statement (GOALS)

* Business Problem Understanding
* Business Objective

**Business Problem Understanding**-

* Financial institutions sustain huge number of losses due to the default of vehicle loans. This has led to the tightening up of vehicle loan policies and increased vehicle loan rejection rates. So, this requires a study to estimate the parameters of vehicle loan default.
* Analyzing this will help the financial institutions to minimize the loan default rates, which helps us gain a reliable customer base by providing the loans to correct persons.
* Following Information regarding the loan and loanee are provided in the datasets:

**Loanee Information** (Demographic data like age, Identity proof etc.)

**Loan Information** (Disbursal details, loan to value ratio etc.)

**Bureau data & history** (Bureau score, number of active accounts, the status of other loans, credit history etc.)

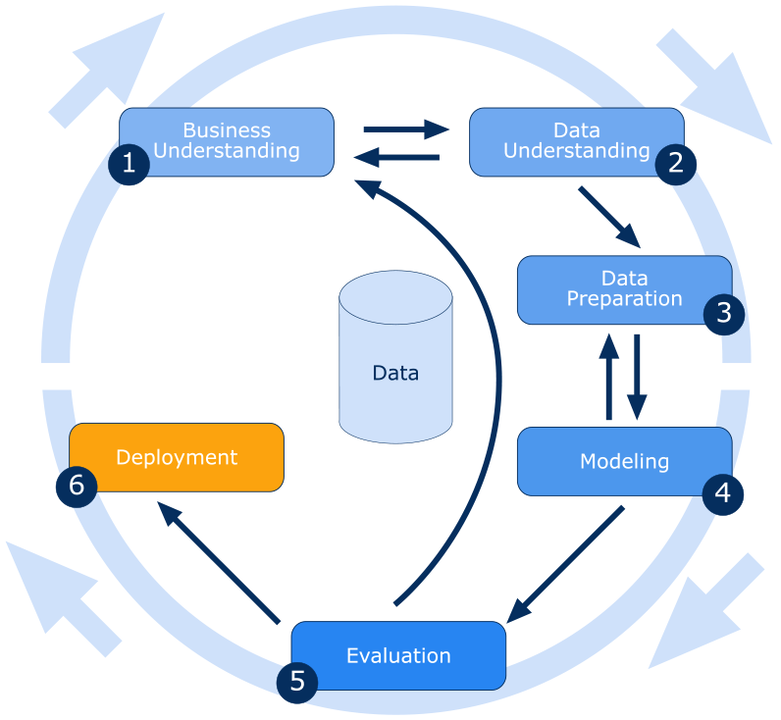
* Doing so will ensure that clients capable of repayment are not rejected and important determinants can be identified which can be further used for minimizing the default rates.

**Business Objective- (PROBLEM STATEMENT)**

The objective of this project is to predict whether it will be Payment default in the EMI on Vehicle Loan on due date or not using the Loan Default Prediction Dataset.

**Approach**

* Studying the Dataset attributes
* Doing Exploratory Data Analysis
* Doing Feature selection
* Performing Necessary Statistical methods
* Model Building using machine learning algorithms
* Model Evaluation
* Hyper-Parameter Tuning
* Best Algorithm Selection



# *\* If deployment is out of scope to the team or not advised by the mentor, please opt to leave out the step no 6.*

# 3. METHODLOGY FOR L & T LOAN DEFAULT PREDICTION

**STEP -1: BUISNESS UNDERSTANDING**

***Analyzing the parameters for vehicle loan default will help the financial institutions to minimize the loan default rates, which helps us gain a reliable customer base by providing the loans to correct persons.***

**STEP -2:** [**DATA**](https://deepnote.com/@ashish-karhade/LandT-Loan-Prediction-nNgrC4PRStOxmwveCuqx2A#00011-47ffaf9f-9082-43b9-adc1-fff9cdeef91a) **PREPARATION AND DATA CLEANING**

*Data preparation (also referred to as “data preprocessing”) is the process of transforming raw data so that data scientists and analysts can run it through machine learning algorithms to uncover insights or make predictions.*

*The main aim of****Data Cleaning****is to identify and remove errors & duplicate****data****, in order to create a reliable dataset. This improves the quality of the training****data****for analytics and enables accurate decision-making.*

***Studying required attributes in the dataset . Perform required operations on the null values and do necessary changes .Fill the missing values with appropriate values .***

**STEP-3:** [**EXPLORATORY**](https://deepnote.com/@ashish-karhade/LandT-Loan-Prediction-nNgrC4PRStOxmwveCuqx2A#00016-014ac04c-a5b6-4e9a-9a57-6353e880cc56) **DATA ANALYSIS AND VISUALIZATION**

* + 1. [Dealing with Numerical features](https://deepnote.com/@ashish-karhade/LandT-Loan-Prediction-nNgrC4PRStOxmwveCuqx2A#00028-dc1cc9a0-2020-4a68-8254-256fe0731041)
    2. [Continuous Variables](https://deepnote.com/@ashish-karhade/LandT-Loan-Prediction-nNgrC4PRStOxmwveCuqx2A#00039-215b1716-07ff-43f1-95bb-db63fd10fc9e)
    3. [Categorical Features](https://deepnote.com/@ashish-karhade/LandT-Loan-Prediction-nNgrC4PRStOxmwveCuqx2A#00052-f5299151-2f4c-4096-a74f-49d86c652b7e)
    4. [Feature Importance](https://deepnote.com/@ashish-karhade/LandT-Loan-Prediction-nNgrC4PRStOxmwveCuqx2A#00075-3c767710-41d4-4e22-a388-cae1b408e0de)
    5. Plotting necessary plots using visualization techniques
    6. Infer the points from the visualization plots

**STEP-4: MODEL BUILDING USING MACHINE LEARNING ALGORITHMS**

*A****machine learning model****is built by****learning and****generalizing from training data, then applying that acquired knowledge to new data it has never seen before to make predictions****and****fulfill its purpose. Lack of data will prevent you from****building****the****model****,****and****access to data isn't enough.*

**STEP 5: MODEL EVALUATION**

***Model evaluation****aims to estimate the generalization accuracy of a****model****on future (unseen/out-of-sample) data.*

**STEP 6: HYPER-PARAMETER TUNING**

*In****machine learning****,****hyperparameter optimization****or****tuning****is the problem of choosing a set of optimal****hyperparameters****for a****learning****algorithm. A****hyperparameter****is a****parameter****whose value is used to control the****learning****process.*

**STEP 7: BEST ALGORITHM SELECTION**

***In applied machine learning, individual algorithms should be swapped in and out depending on which performs best for the problem and the dataset.***

# REFERENCES

The references can be blogs, articles or even social media news relevant to explain the importance of the projects.

**Notes For Project Team**

*Sample Reference for Datasets (to be filled by team and mentor)*

|  |  |
| --- | --- |
| ***Original owner of data*** | **Data Science Fin Hack Competition.**  **L&T Financial Services & Analytics Vidhya.** |
| ***Data set information*** | **Train Dataset contain 2,33,154 records and 41 columns.**  **Test Dataset contain 112392 records and 40 columns. (It does not contain Target column)**  **Target Feature -The response feature is loan default which is given as: The target feature has two classes and hence it is a binary classification problem. To reiterate, the goal is to predict whether or not it will be Payment default in the first EMI on due date by Customer.**   * **Train dataset contains the training data with details on loan.** * **Test dataset contains details of all customers and loans for which the participants are to submit probability of default.** |
| ***Any past relevant articles using the dataset*** | **None** |
| ***Reference*** | **1)KAGGLE WEBSITE**  **2)https://datahack.analyticsvidhya.com/contest/ltfs-datascience-finhack-an-online-hackathon/** |
| ***Link to web page*** | **https://www.kaggle.com/sneharshinde/ltfs-av-data?select=train.csv** |

### 

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*