# Low Level Design

Ad-Click Prediction

| | |
|---|---|
| Written By | Vishal Gujarathi |
| Date | 23 Nov 2022 |

## Document Control

**Change Record:**

| Version | Date | Author | Comments |
|---|---|---|---|
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

**Reviews:**

| Version | Date | Reviewer | Comments |
|---|---|---|---|
| | | | |

**Approval Status:**

| Version | Review Date | Reviewed By | Approved By | Comments |
|---|---|---|---|---|
| | | | | |

# Contents

# 1. Introduction

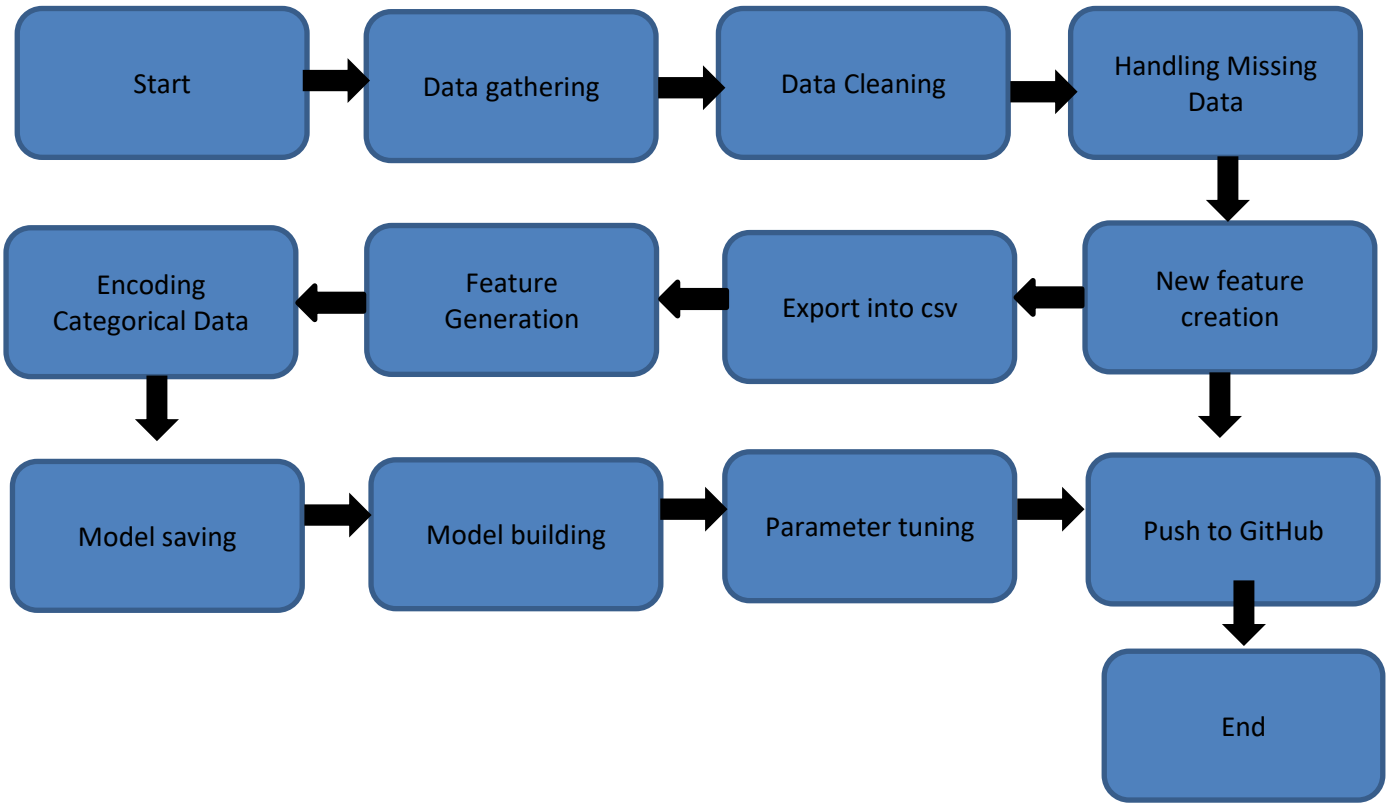
## 1.1. What is Low-Level design document?

The goal of LLD or a low-level design document (LLDD) is to give the internal logical design of the actual program code for Ad-Click Predction. LLD describes the class diagrams with themethods and relations between classes and program specs. It describes the modules so that the programmer can directly code the program from the document.

## 1.2. Scope

Low-level design (LLD) is a component-level design process that follows a step-by-step refinement process. This process can be used for designing data structures, required software architecture, source code and ultimately, performance algorithms. Overall, the data organization may be defined during requirement analysis and then refined during data design work

## 2. Architecture



# 3. Architecture Description

## 3.1. Data Description

| Variable |
|---|
| **Daily Time Spent on Site** |
| **Age** |
| **Area Income** |
| **Daily Internet Usage** |
| **City** |
| **Ad Topic Line** |
| **City** |
| **Male** |
| **Country** |
| **Timestamp** |
| **Clicked on Ad** |

## 3.2. Data Gathering

Dataset link:- Link
We got csv files train. We use these files for the data preparation purpose and EDA.

## 3.3. Data Cleaning

We observed some inconsistency in the dataset as we see there were some categorical feature that converted into numerical feature.

## 3.4. Handling Missing Data

We observed from the table above that all the values in column "Ad Topic Line" is unique, while the "City" column contains 969 unique values out of 1000. There are too many unique elements within these two categorical columns and it is generally difficult to perform a prediction without the existence of a data pattern. Because of that, they will be omitted from further analysis. The third categorical variable, i.e "Country", has a unique element (France) that repeats 9 times. Additionally, we can determine countries with the highest number of visitors

## 3.5. Feature Generation

We observed from the table above that all the values in column "Ad Topic Line" is unique, while the "City" column contains 969 unique values out of 1000. There are too many unique elements within these two categorical columns and it is generally difficult to perform a prediction without the existence of a data pattern. Because of that, they will be omitted from further analysis. The third categorical variable, i.e "Country", has a unique element (France) that repeats 9 times. Additionally, we can determine countries with the highest number of visitors

## 3.6. Feature Selection

We included all the features for model training 'Daily Time Spent on Site', 'Age', 'Area Income', 'Daily Internet Usage', 'Male', 'Month', 'Day of the month' ,'Day of the week', 'Clicked on Ad'

### 3.7. Encoding Categorical Data

Label Encoding was used for Ad Topic Line", "City", and "Country**.** But we observed that Ad Topic Line" is unique, while the "City" column contains 969 unique values out of 1000. There are too many unique elements within these two categorical columns and it is generally difficult to perform a prediction without the existence of a data pattern. Because of that, they will be omitted from further analysis. The third categorical variable, i.e "Country", has a unique element (France) that repeats 9 times.

For model selection we had used Pycaret library that will perform multiple operations and give algorithm best for data. Pycaret had given Gradient Boosting regressor as best for our data.

### 3.8 Parameter Tuning

Parameters are tuned using Grid searchCV. The parameters are tuned on Gradient Boost model.

### 3.9 Model Building

In this project, three different ML models will be developed: a Logistic Regression model, Decision Tree model and Random Forest

### 3.10 Model Saving

Model is saved using pickle library in `.pkl` format.

### 3.11 Github

The whole project directory will be pushed into the GitHub repository