Low Level Design

Adult Census Income Prediction (ACIP)

Written By	Ekram Shethil
Document Version	0.3
Last Revised Date	07 Mar 2022

Document Control

Change Record:

Version	Date	Author	Comments
0.1	05 Mar 2022	Ekram Shethil	Introduction & Architecture defined
0.2	06 Mar 2022	Ekram Shethil	Architecture & Architecture Description append and updated
0.3	07 Mar 2022	Ekram Shethil	Unit Test Cases defined and appended

Reviews:

Version	Date	Reviewer	Comments
0.2	07 Mar 2022	Ekram Shethil	Document Control, Version Control and Unit Test Cases to be added

Approval Status:

Version	Review Date	Reviewed By	Approved By	Comments

Contents

1.Introduction

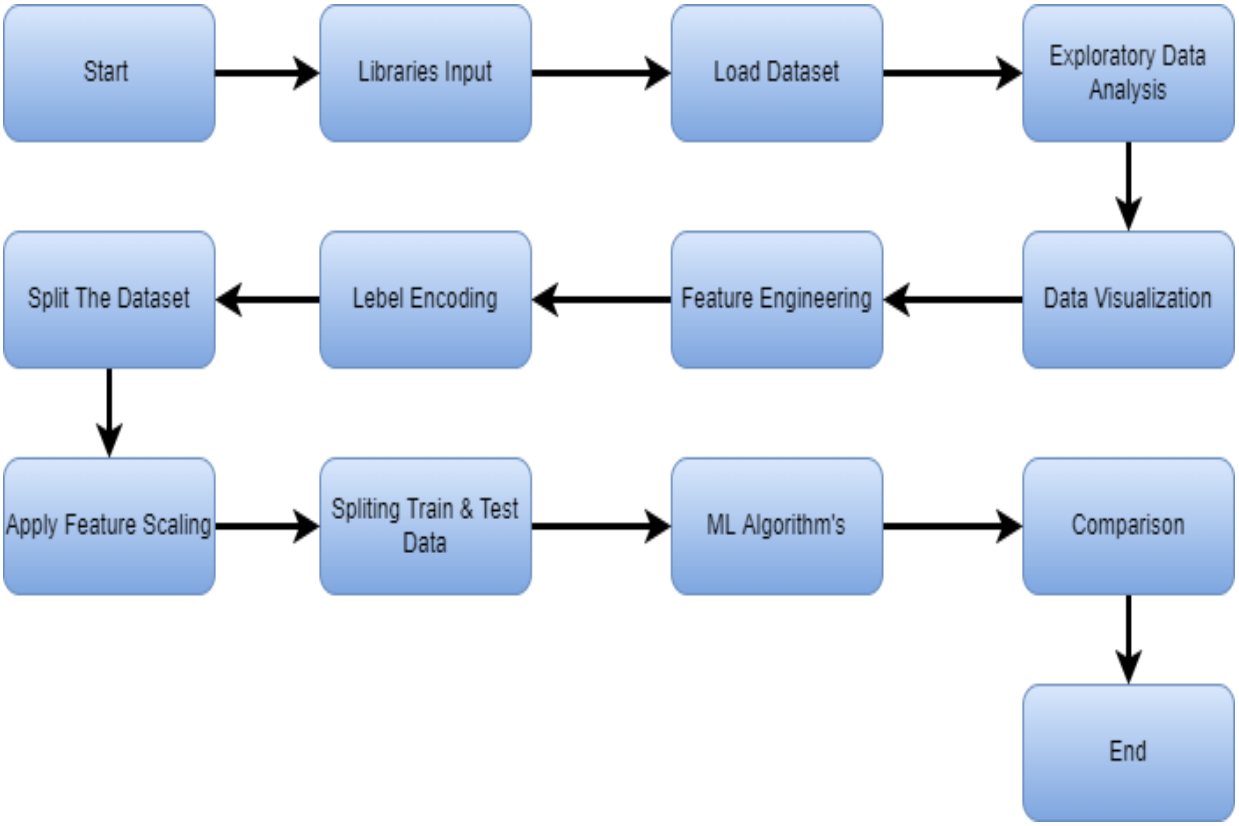
1.1 What is Low-Level design document?

The goal of LLD or a low-level design document (LLD) is to give the internal logical design of the actual program code for Adult Census Income Prediction. LLD describes the class diagrams with the methods 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

Recipe 1M+ dataset is the biggest publicly available recipe dataset. The information each recipe contains is separated in two Comma-Separated Values (CSV) files. This dataset contains 32561 recipes including 15 different ingredients.

3.2 Web Scrapping

In order to create a more complete recipe collection we will need some more datasets which will contain Adult Census Income Prediction.

3.3 Data Insertion into Database

a. Database Creation and connection – Create a database with name passed. If the database is already created, open the connection to the database.

b. Table creation in thee database.

c. Insertion of files in the table.

3.4 Export Data from Database

Data Export from Database – The data in a stored database in exported as a CSV file to be used for Data Pre-processing and Model Training.

3.5 Data Pre-processing

Data Pre-processing steps we could use are Null value handling, Imbalanced data set handling, Handling columns with standard deviation zero or below a threshold, etc.

3.6 Data Clustering

K-NN algorithm will be used to create clusters in pre-processed data. The optimum number of clusters is selected by plotting the elbow plot. The idea behind clustering is to implement different algorithms to train data in different clusters. The k-NN model is trained over preprocessed data and the model is saved for further use in prediction.

3.7 Model Building

After clusters are created, we will find the best model for each cluster. For each cluster, algorithms will be passed with the best parameters derived from Grid-Search. We will calculate the ACIP scores for models and select the model with the best score. Similarly, the models will be selected for each cluster. All the models for every cluster will be saved for use in Recommendation.

3.8 Data from User

Here we will collect financial data from user such as user age, workclass, fnlwgt, education, education-num, marital-status, occupation, relationship, race, sex, capital-gain, capital-loss, hours-per-week and country.

3.9 Data Validation

Here Data Validation will be done, given by the user.

3.10 Data Clustering

The model create during training will be loaded, and clusters for the user data will be predicted.

3.11 Model Call for Specific Cluster

Based on the cluster number, the respective model will be loaded and will be used to predict/Recommend the data for that cluster.

3.12 Recipe Recommendation & Saving Output in Database

After calling model Recipe/Output will be recommended, this output will be saved in Database and it will be reused to show the same Output if other users provide the same data.

3.13 Deployment

We will be deploying the model to ACIP.

This is a workflow diagram for the Recipe Recommendation.