Low Level Document

Insurance Premium Prediction

Date issued	Version	Description	Author
5/9/2024	V1.0	LLD V1.0	Madhuri Chandanbatwe

1.0 Introduction

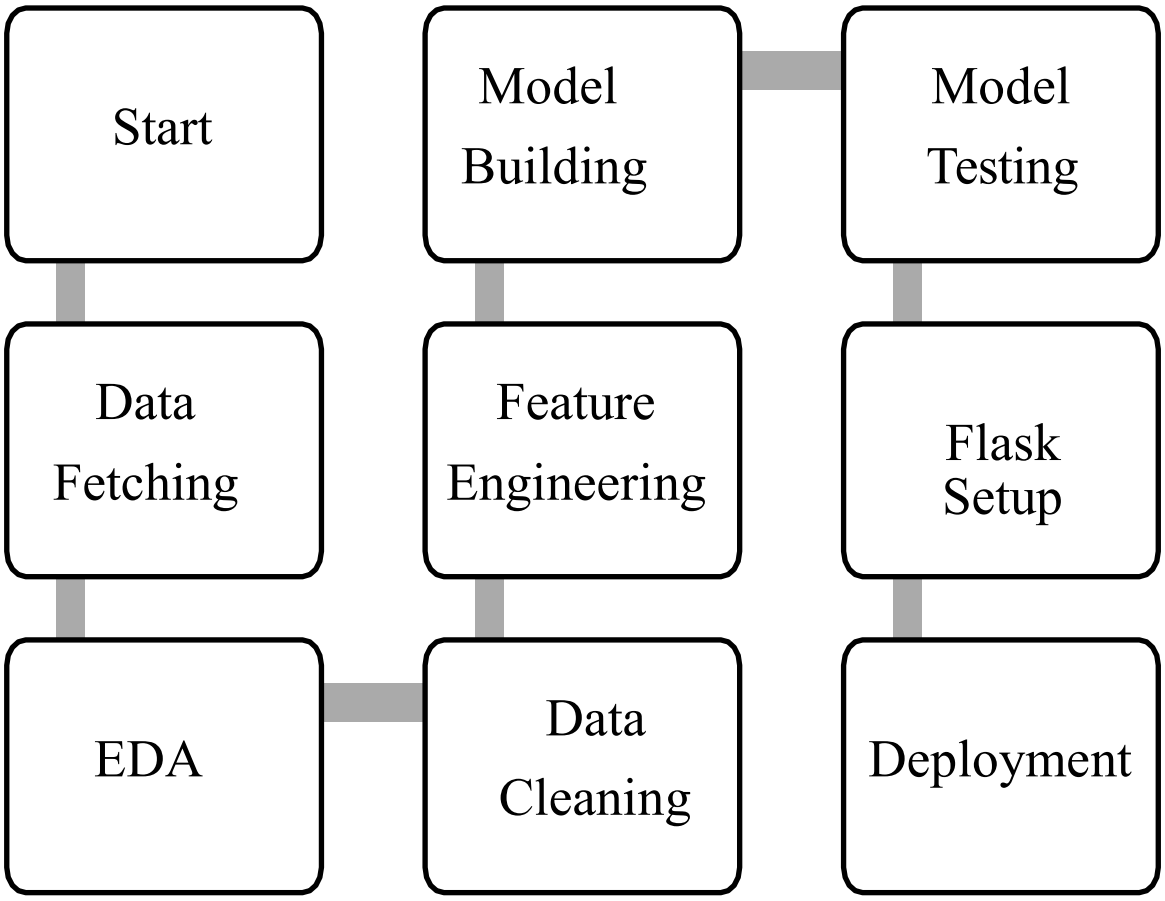
1.1 What is Low-Level Design Document?

The goal of LLD or Low-Level design document (LLDD) is to give the internal logical design of the actual program code. Low-Level design is created based on the High-Level design. 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 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.

2.0 Architecture



2.1 Data Description

The primary source of data was from Kaggle.It was a 14 gb data with a 1338 rows and 7 columns. 'age', 'sex', 'bmi', 'children', 'smoker', 'region', 'expenses' were the columns in this data.Data is in structured format and if unstructured mongodb can be used.

2.2 Exploratory Data analysis

Exploring the data by visualizing the distribution of values in some columns of the dataset, and the relationships between expenses and other columns.We checked the univariate and bivariate and multivariate analysis of the

features.Checked if the data follows the linear regression assumptions.Checked the feature importances with respect to various models.Fortunately the data was devoid of any missing values.If missing values persisted we would have imputed with various techniques like Simple Imputation, filling with mean or median or mode,KNN imputation,dropping or deleting the features, etc.

2.3 Data Preprocessing

Data preprocessing like encoding categorical features with one hot encoding,splitting the data into numerical and categorical features and scaling the data with MinMax scaling in order to avoid the impact of a particular columns was performed on the dataset.

3.4 Model Building

Model was trained on the csv data.Various machine learning models like Linear Regression,Lasso Regression,XG Boost Regressor,KNN Regressor,Ada Boost Regressor,Random Forest Regressor,Gradient Boosting regressor were applied on the data.

3.5 Model Evaluation:

Evaluation metrics for regression models like mean squared error, root mean squared error, mean absolute error,r2 score.Rsquare is been used as the final evaluation metric for getting the best model required for our premium prediction.In the end the model is been dumped to a pickle file to develop a web application framework.

3.6 Deployment

The application was deployed on AWS EC2 platform.