Architecture Design
Insurance Premium Prediction

Revision Number – 1.1

Last Date of Revision:14– 08 - 2022

Document Version Control

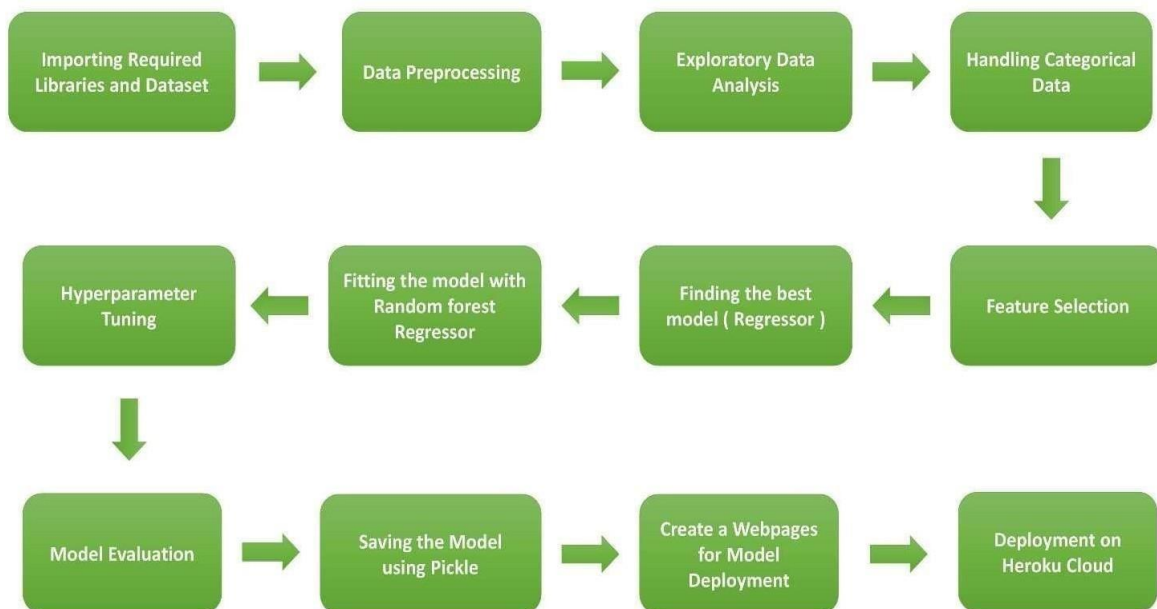| Date	Version	Description	Author
14– 08 - 2022	1.0	Abstract Introduction Architecture	Anil Naga Sai
15 – 08 - 2022	1.1	Architecture Design	Anil Naga Sai

Contents

Abstract

To give people an estimate of how much they need based on their individual health situation. After that, customers can work with any health insurance carrier and its plans and perks while keeping the projected cost from our study in mind. I am considering variables as age, sex, BMI, number of children, smoking habits and living region to predict the premium. This can assist a person in concentrating on the health side of an insurance policy rather than the ineffective part.

1. Introduction

1.1 Why this Architecture Design Document?

The main objective of the Architecture design documentation is to provide the internal logic understanding of the flight fare prediction code. The Architecture design documentation is designed in such a way that the programmer can directly code after reading each module description in the documentation.

2. Architecture

```
Importing Required      →   Data Preprocessing   →   Exploratory Data    →   Handling Categorical
Libraries and Dataset                                    Analysis                   Data
                                                                                      ↓
Hyperparameter          ←   Fitting the model    ←   Finding the best    ←   Feature Selection
    Tuning                  with Random forest       model ( Regressor )
                                Regressor
      ↓
Model Evaluation        →   Saving the Model     →   Create a Webpages   →   Deployment on
                            using Pickle              for Model                Heroku Cloud
                                                      Deployment
```

3. Architecture Design

3.1 Data Collection

The data for this project is collected from the Kaggle Dataset, the URL for the dataset is https://www.kaggle.com/datasets/noordeen/insurance-premium-prediction

3.2 Data Description

Insurance Premium dataset publicly available on Kaggle. The information in the dataset is present in one csv files named as insurance.csv. Dataset contains 1338 rows which shows the information such age, bmi, children and expenses.

3.3 Data Pre-processing

- Checked for info of the Dataset, to verify the correct datatype of the Columns.
- Checked for Null values, because the null values can affect the accuracy of the model.
- Performed One – Hot encoding on the desired columns.
- Checking the distribution of the columns to interpret its importance.

Now, the info is prepared to train a Machine Learning Model.

3.4 Modelling Process

After pre-processing the data, we visualize our data to gain insights and split into two parts, train and test data. After splitting the data, we use different-different Machine Learning Models like, Linear Regression, Random Forest Regressor, Decision Tree Regressor to predict the Insurance Premium Price.

3.5 UI Integration

HTML files are being created and are being integrated with the created machine learning model. All the required files are then integrated to the app.py file and tested locally.

3.6 Data from User

The data from the user is retrieved from the created HTML web page.

3.7 Data Validation

The data provided by the user is then being processed by app.py file and validated. The validated data is then sent to the prepared model for the prediction.

3.8 Rendering the Results

The data sent for the prediction is then rendered to the web page.

3.9 Deployment

The tested model is then deployed to Heroku. So, users can access the project from any internet device.