Insurance Premium Prediction

Written By	**Manthan Takalkar,** **Adityaraj Chaudhari**
Document Version	**0.1**
Last Revised Date	**13-10-2021**

Problem Statement

- The goal of this project is to give people an estimate of how much insurance premium 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.

- This can assist a person in concentrating on the health side of an insurance policy rather has the ineffective part.

Objective

The purpose behind this project is to build an machine learning model which will predict rough estimate of insurance premium based on some parameters like age, BMI, sex, etc. Which would help user to choose appropriate health insurance policy.

Benefits

- The uncertainty of paying for the losses incurred out of pocket has a significant impact on cash flow management. However, with an insurance policy by your side, you can tackle this uncertainty with ease. The chosen insurance provider pays in the event of happening of an insured event whenever they occur.

- Unit linked insurance plan, , invest a part of the premium into several market linked funds. This way, they enables you to invest money regularly to benft of market linked returns and fulfil yo

Dataset Description

Feature Names	Data Type	Feature Type	Description
age	Integer	Independent	Gives general information about human age.
gender	String	Independent	The sex column gives information about human gender I.e. male or female
bmi	Float	Independent	BMI column tells Body Mass index of particular person.
children	Integer	Independent	It refers whether that person has children or not.
smoker	String	Independent	This particular feature column tells whether the person is smoker or not
region	String	Independent	This feature column tells from which region the particular person belong Here 4 regions are specified : 1. southwest 2. southeast 3. northwest 4.northeast.
expenses	Float	Dependent	It shows previous medical expenses of individual person.
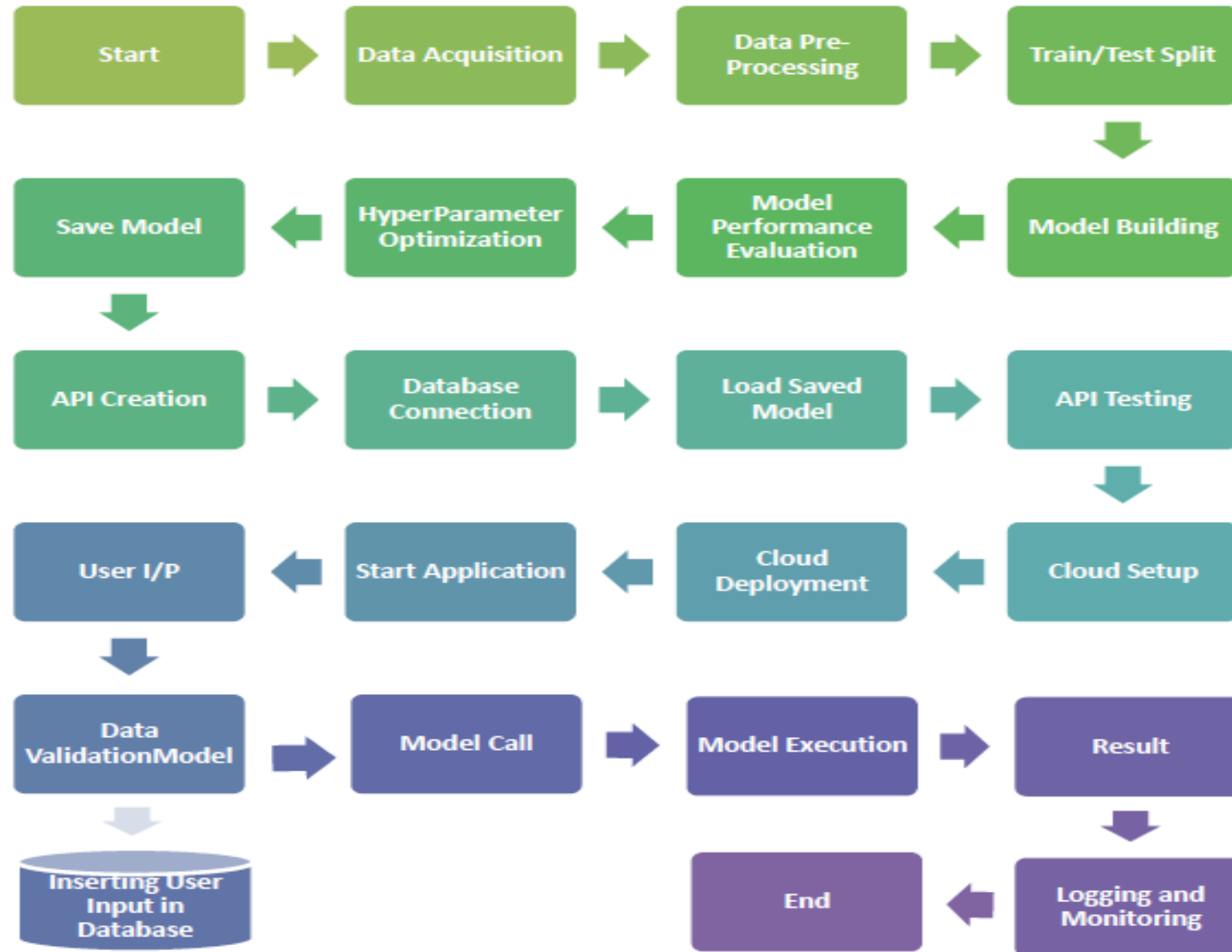
Data Acquistion

The Dataset we acquired is in CSV(Comma Separated Format) format and was accessed from the given source

Data PreProcessing and Data Splitting

Data pre-processing step can be used for handling categorical values, statistical analysis, Outlier detection etc.

Spliting the dataset into two sets I.e into Training set and Testing set. This will help us for model training and evaluation. Here we can follow hold-out cross validation approach, k-fold cross validation or stratified k-fold validation approaches which allow us to compare the performance of machine learning algorithms for our predictive modeling problem.

Model Building

After performing Train/Test split we build model by using Gradient Boosting machine learning algorithm on training set and testing set.

API Creation

API(Application Programming Interface) creation is the process of creating and exposing APIs. We can use flask frame-work to create API which act as an interface between web application. To design a web page we can use HTML, CSS, etc.

Database Connection

For data storing purpose we have used PostgreSQL database. With the help of psycopg2 we can connect PostgreSQL database to the application.For storing entered data from created web application the database connection is required.

Cloud Set-up And Deployment :

- We will be deploying the model to Heroku cloud platform. The process of deployment of Heroku starts with creating the app with any desired name and choosing the region. Then we create Heroku PostgreSQL (Database) instance. After this we connect our app with the Github and fetch required repository to successfully setup the cloud environment.

- Once the cloud setup process is successfully completed then we are ready to deploy our app to the cloud.

Q1. Which domain does the this project belongs to?

This project belongs to Insurance Domain.

Q2. From which source this dataset was accessed?

Firstly we downloaded the dataset from kaggle in csv file format then stored it in a project folder and then accessed it using pandas library from python.

Q3. What was the type of data?

The data was the combination of numerical and Categorical values.

Q4- Which steps did you followed for data preprocessing?

Step_1: Removing Outliers

To get a generalized model we removed outliers from the dataset.

2 - Feature Transformation

We converted the columns containing categorical data into numerical data using technique like One Hot Encoding

3 - Data Scaling

After transforming features we scaled down the data.

Q5.To get generalized model which ML algorithm was used?

We tried using Random Forest, XGBoost and Gradient Boosting algorithm for building the model.Out of the three models we created we found that Gradient Boosting was more generalized and gave good results so we used it.

Q6 . Which frame work was used to build API?

We used Flask Web Framework for creating the API.We used it because it is light weight python web framework for creating API's and it is easier to use and understand.

Q7. On which database the user input data was stored?

We used PostgreSQL Database for the storing the user data.One of the reasons using it is the speed of read/write operations as well as robustness of database.

Q8. Which cloud platfrom was to deploy web aplication?

We deployed our applicaton on Heroku Cloud Platform. Heroku is Paas Cloud Platform.

This cloud setup is easy and the developer does not need to take care of infrastructure related things like os,services,storage,etc.

Q9. To handle logs which approach was used?

For logging we used file logging approach that is we stored our logs into the different log files for various operations like data preprocessing , model building ,etc we created different log files.

Q10. How to preview the data stored in heroku postgreSQL database?

On Heroku cloud platform we need to login , create dataclips and then write a query to get a data from database.After executing this query we can see the data (user input) stored in Heroku PostgreSQL Database.