**High Level Design (HLD)**

IPP (Insurance Premium Prediction) Estimated Expense Based

**Author –**

**Vikas Mishra and Sachin Gupta**

****

**Abstract**

The purpose of this project is to make people aware about the estimate for the medical expenses that how much they need given the risk profile of an individual based on his health situation which is determine keeping various factor in mind such as age, medical history, location etc. As health insurance is the necessity now a days. This exercise will equip individual with regards to estimated medical expense that he/she might incurred, thus can work with health insurance carrier to plan accordingly keeping the project cost in mind and pay for what they actually need.

The purpose of this exercise to look into different features to observe their relationship, and plot a multiple linear regression based on several features of individual such as age, physical/family condition and location against their existing medical expense to be used for predicting future medical expenses of individuals that help medical insurance to make decision on charging the premium.

**1 Introduction**

* 1. **Why this high-level design document?**

The purpose of this High-Level Design Document is to provide the necessary detail to the current project description to represent a suitable model for coding. This document also intended to help detect contradiction prior to coding, and can be used a reference manual for how the modules interact at a high level.

The HLD will:

* Present all of the design aspect and define them in detail
* Describe the user interface being implemented
* Describe the hardware and software interface
* Describe the performance requirements
* Include design features and the architecture of the project
* List and describe the non-functional attribute like:
  + Security
  + Reliability
  + Maintainability
  + Portability
  + Reusability
  + Application compatibility
  + Resource utilization
  + Serviceability
  1. **Scope**

The HLD documentation present the structure of the system, such as the database architecture, application architecture(layers), application flow (Navigation), and technology architecture. The HLD uses the non-technical to mildly-technical terms which should be understandable to the administrators of the system.

* 1. Definition

|  |  |
| --- | --- |
| **Term** | **Definitions** |
| IPP | Insurance Premium Prediction |
| Database | Data collected over for the model building |
| IDE | Integrated Development Environment |
| Microsoft Azure | Azure cloud services |

1. **General Description**
   1. **Product Perspective**

The Insurance premium prediction solution aim to help individual to forecast estimated cost of the medical expenses for the individual so that he can work with the medical insurance carrier to plan based on his study.

* 1. **Problem statement**

To predict the estimated cost of the medical expenses for the individual based on health, age, medical history and situation, this will equip him/her to work with the health insurance company to plan according to specific need and stay out of worry.

**2.3 Propose solution**

The solution propose here is to build an AI application for predicting the premium for the individual. Where based on the various factor such as age, location, medical history, number of children and if he/she is smoker or not.

Given the condition of the individual type of coverage that he/she needs to be opt for, and expected variance in premium as a result of it will enable the individual have more quality discussion with the health insurance company while looking for plan and take informed decision.

**2.4 Data Requirement**

Data for this problem has been pulled from the provided resources (Kaggle).

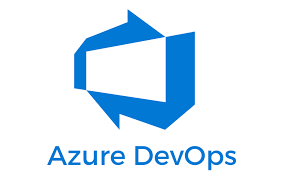
The insurance.csv dataset contains 1338 observations (rows) and 7 features (columns). The dataset contains 4 numerical features (Age, BMI, Children and Expenses) and 3 nominal features (sex, smoker and region) that were converted into factors with numerical value designated for each level.

* 1. **Tool Used**

We have leveraged the standard package to build model.

Python programming language and its framework such as numpy, pandas, and scikit-learn to build the model.

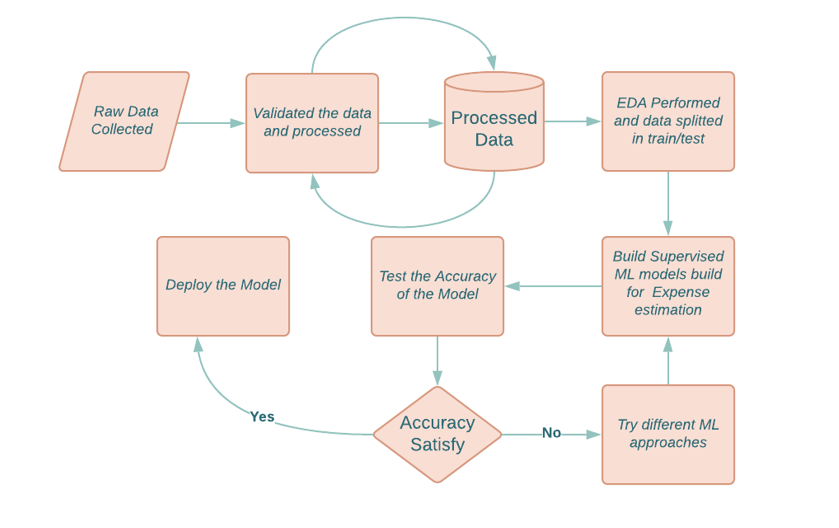
* Visual Studio code used for IDE
* For visualization of the plots, Matplotlib, Seaborn, and plotly are used.
* Azure is used for the deployment of the mode.
* PowerBI is used for the dashboard creation
* SQL server and used to retrieve, insert, update and date the database.
* We will be using the Flask for the UI/fronted
* GitHub for the version control
* Azure develops to build the CI/CD pipeline



1. **Design Details**
   1. **Process Flow**
      1. **Model Training and Evaluation**
      2. **Deployment Process**
   2. **Event Log**
2. **Design Details**
   1. **Process Flow**

In order to build the AI solution for the estimation, we will using the various technique algorithms for regression under supervised machine learning.

Below is the diagram of the flow



* 1. **Event Log**

This will enable to record the logs of the process, which will be helpful to understand what is running internally in the application

**Logging Process:**

1. The system identifies the steps where logging required
2. System should be able to log each and every step in the flow
3. Developer can choose logging method. You can choose database logging/File logging as well.
4. System should not hang even after using multiple logging.
   1. **Error Handling**

In case of error encountered, a message will be displayed to the user as to provide the information what went wrong. An error will be defined as anything that falls outside the normal and intended usage.

1. **Performance**

Insurance premium predication AI solution will be used for the expense estimation which will enable the individual to have strong foundation to conduct discussion with the health insurance company to choose the plan. Better accuracy will help to have more precise conversation, thus periodically model retraining is important for better performance and accuracy of the model.

* 1. **Reusability**

The code will be written in the modular fashion using the OOPs concept which will be easy to scale and reuse without a problem.

* 1. **Application Compatibility**

Different component of the solution will be using the Python as an interface among them. Each component will have its own task to perform, and it is the job of the Python to ensure proper transfer of information.

* 1. **Resource Utilization**

While performing the task it is likely to use the all the available resources until that function is finished.

* 1. **Deployment**



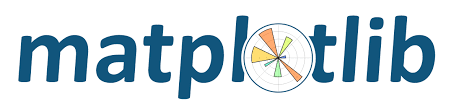
While there are various option to deploy the AI solution.

We will be using the Microsoft Azure services for model deployment.

1. **Key Studies and KPI**

We will perform and various study and understand the relationship and impact of variable and how it is ultimately contributing to the cost for the individual.

Depending the upon the nature we will either store information using the Python inbuilt library matplotlib or built dashboard out of it.

****

* 1. **KPIs (Key Performance Indicators)**

1. **Conclusion**

This application solution will enable the individual given their specific condition which is a combination of various variable provide the estimated expense of the medical. And, plan accordingly need and type of health Insurance should seek for.

1. **References**

[Insurance Premium Prediction | Kaggle](https://www.kaggle.com/noordeen/insurance-premium-prediction)

[Research Paper on Health Insurance](https://www.researchgate.net/publication/331113759_A_Study_on_Health_Insurance_Premium_Claims_Commission_and_its_Growth_of_Select_Companies_in_India)