HIGH LEVEL DESIGN (HLD)

Insurance Premium Prediction

Prajwal Krishna

Document Version Control

Date Issued	Version	Description	Author
0.1.10.2021	V1.0	Initial HLD- V1.0	Nishant Banjade

Contents

Abstract

Looking at the present context of the world, we often see a lot of uncertainties everyday. Some lose their lives early, some may get into accidents. Insurance is the one of the solutions for helping people get some relief in their lives because by paying a certain premium, we get some amount of money just in case we meet some devastating accidents or uncertainties.

The main objective of the project is to help predict the premium amount of the insurance using features such as age, sex as well as regions of the

person. Advancement in machine learning and artificial intelligence made it easier to predict such things. For this project, we utilized various regression models such as Linear regression, DecisionTreeRegressor, RandomForest as well as ElasticNet to predict the premium. After the model is evaluated, ElasticNet is chosen as it was giving better accuracy and report as compared to others.

1.0 Introduction

1.1 Why this High-Level Design Document?

The purpose of this High-Level document is to add necessary details to current project description to represent a suitable model for coding. This document is used as a reference manual for how the model interacts at a high-level.

The HLD will

- Presents all design aspects and defines them in detail.
- Describe the user interface being implemented.
- Describe the hardware and software interfaces.
- Describe the performance requirements.
- Include design features and the architecture of the project.

1.2 Scope

The HLD document presents the structure of the system, such as the database architecture, application architecture, and technology architecture. The HLD uses non-technical to middle-technical terms which should be understandable to the administrators of the system.

1.3 Definitions

Term	Description
Database	Collection of all the information
IDE	Integrated Development Environment
API	Application Programming Interface
KPI	Key Performance Indicator
VSCode	Visual Studio Code
EDA	Exploratory Data Analysis

2.0 General Description

2.1 Product Perspective

The Insurance premium estimation is a machine learning based predictive model which will help us to predict the premium of the personal for health insurance.

2.2 Problem Statement

To develop an API interface to predict the premium of insurance using people individual health data and analyzing the following:

- To detect BMI value affects the premium.
- To detect smoking affects the premium of the insurance.
- To create API interface to predict the premium

2.3 Proposed Solution

The solution proposed here is an estimating premium of insurance based on people's health data and this can be implemented to perform above mentioned use cases. In the first case, analyzing how BMI values affect the people's health as well as the premium of the insurance. In the second case, if the model detects the smoking affecting the premium, we will inform people. And in the last use case, we will be making an interface to predict the premium.

2.4 Further Improvements

2.5 Technical Requirements

The solution can be a cloud-based or application hosted on an internal server or even be hosted on a local machine. For accessing this application below are the minimum requirements:

- Good internet connection.
- Browsers such as Brave, chrome, Mozilla firefox

For training model, the system requirements are as follows:

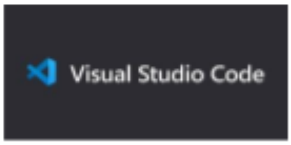
- +4 GB RAM preferred
- Operation System: Windows, Linux, Mac
- vim / Visual Studio Code / Jupyter notebook

2.6 Data Requirements

- Comma separated values (CSV) file.
- Input file feature/field names and its sequence should be followed as per decided.

2.7 Tools Used

Python programming language and frameworks such as NumPy, Pandas, Scikit-learn, Plotly, Flask are used to build the whole model.

- Pandas is an open-source Python package that is widely used for data analysis and machine learning tasks.

- NumPy is the most commonly used package for scientific computing in Python.
- Plotly is an open-source data visualization library used to create interactive and quality charts/graphs.
- Scikit-learn is used for machine learning.
- Flask is used to build API.
- VS Code is used as IDE (Integrated Development Environment)
- GitHub is used as a version control system.
- Front end development is done using HTML/CSS.
- Heroku is used for deployment of the model.

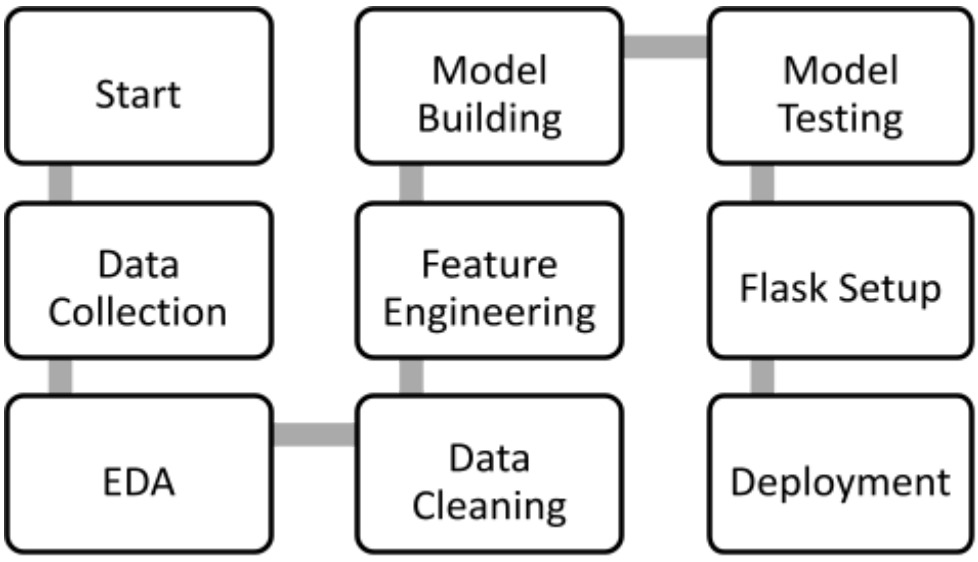
2.8 Constraints

2.9 Assumptions

The main objective of the project is to develop an API to predict the premium for people on the basis of their health information. Machine learning based regression model is used for predicting above mentioned cases on the input data.

3.0 Design Details

3.1 Process Flow



3.2 Event Log

The system should log every event so that the user will know what process is running internally.

Initial Step-By-Step Description:

- The system identifies at what step logging is required.
- The system should be able to log each and every system flow.
- Developers can choose logging methods. You can choose database logging.

System should not hang out even after using so many loggings.

4.0 Performance

4.1 Reusability

The entire solution will be done in modular fashion and will be API oriented. So, in the case of the scaling of the application, the components are completely reusable.

4.2 Application Compatibility

The interaction with the application is done through the designed user interface, which the end user can access through any web browser.

4.3 Deployment

5.0 Conclusion

This system shows us the different techniques that are used in order to estimate the amount of premium required on the basis of individual health situation. After analyzing it shows how a smoker and non-smokers affect the amount of estimate. Also, there is a significant difference between male and female expenses. Accuracy, which plays a key role in prediction-based systems. From the results we could see that Gradient Boosting turned out to be the best working model for this problem in terms of the accuracy. Our predictions help users to know how much amount premium they need on the basis of their current health situation.

High Level Design (HLD)