HIGH LEVEL DESIGN (HLD)

**High Level Design (HLD)**

Insurance Premium Prediction

Revision Number: 1.0

Last date of revision: 16/01/2024

**JAYANTO KUMAR BHAKAT**

Page 1

**Contents**

Document Version Control.......................................................................................................3

Abstract.....................................................................................................................................4

1 Introduction..........................................................................................................................5

1.1 Why this High Level Design Document?....................................................................5

1.2 Scope............................................................................................................................5

2 Genral Description...............................................................................................................6

2.1 Problem Statement & Product Perspective..................................................................6

2.2 Proposed Solution........................................................................................................6

2.3 Technical Requirements...............................................................................................6

2.4 Data Requirements.......................................................................................................7

2.5 Tools used....................................................................................................................7

2.6 Constraints....................................................................................................................8

2.7 Assumptions.................................................................................................................8

3 Design Details......................................................................................................................9

3.1 Process flow.................................................................................................................9

3.2 Event log....................................................................................................................10

3.3 Error Handling...........................................................................................................10

4 Performance

4.1 Reusability..................................................................................................................10

4.2 Application compatibility...........................................................................................11

4.3 Deployment................................................................................................................11

5 Conclusion..........................................................................................................................11

Page 2

Document Version Control

|  |  |  |  |
| --- | --- | --- | --- |
| **Date Issued** | **Version** | **Description** | **Author** |
| 24-11-2023 | 1.0 | First version of Complete HLD | Jayanto Kumar Bhakat |
|  |  |  |  |
|  |  |  |  |

**Abstract**

Health insurance is a necessity nowadays , and almost every individual is linked with a government or private health insurance company. Factors determining the amount of insurance vary from company to company. Also people in rural areas are unaware of the fact that the government of India provide free health insurance to those below poverty line. It is very complex method and some rural people either buy some private health insurance or do not invest money in health insurance at all. Apart from this people can be fooled easily about the amount of the insurance and may unnecessarily buy some expensive health insurance,

Our project does not give the exact amount required for any health insurance company but gives enough idea about the amount associated with an individual for his/her own health insurance.

Page 4

**1 Introduction**

1.1 Why this High-Level Design Document?

The purpose of this High\_Level Design (HLD) document is to add the necessary detail to the project description to represent a suitable model and coding for application. This document is also intended to help detect contradictions before coding and can be used as a reference manual for how the modules interact at a high level.

The HLD will:

* Present all of the design aspects and define them in detail
* Describe the user interface is implemented
* Describe the hardware and software interfaces
* Describe the performance requirements
* Include design features and the architecture of the project
* List and describe the non-functional attributes like:

- Security

- Reliability

- Maintainability

- Portability

- Reusability

- Application compatibility

- Resources utilization

- Serviceability

1.2 Scope

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

Page 5

2 General Description

2.1 Problem Statement & Product Perspective

The dataset contains Age, Sex, BMI (Body Mass index), Children, Smoker, region and Expenses where I have to predict Insurance Premium Expenses with

- To detect BMI value affects the premium

- To detect smoking affects the premium of the insurance.

- To create API interface to predict the premium

2.2 Proposed Solution

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

2.3 Technical Requirements

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

* Good internet connection.
* Web Browser

For training model, the system requirements are as follows:

1. +4 GB RAM preferred

- Operating System: Windows, Linux, Mac

- Visual Studio Code / Jupyter notebook / Google colab

Page 6

2.4 Data Requirements

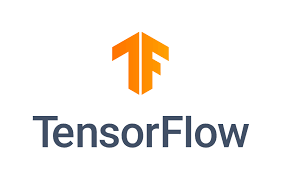
Data requirements completely depends on out problem statement.

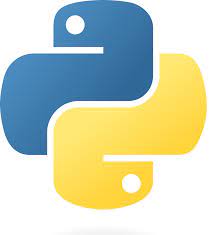
- Common separated values (CSV) file.

- Input file features/field names and its sequence should be followed as per decided.

2.5 Tools used

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





Page 7

 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 an IDE (Integrated Development Environment)

 GitHub is used as a version control system.

 Front-end development is done using HTML and CSS.

 Railway is used for the deployment of the model..

2.6 Constraints

It is useful for the user by predicting Insurance prices based on their provided details for ex: - Bmi, sex, smoker, yes/no, age, etc.

The application should be user friendly as automated as possible. Users can easily use the application and not needed to know any of the workings.

2.7 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.

Page 8

3 Design Details

3.1 Process flow

For finding the insurance premium we will be using machine learning model

Below is the process flow diagram

START-------> DATA GATHERING------>EDA----->DATA CLEANING------>FEATURE ENGINEERING----->MODEL CREATION------>MODEL TESTING------->FLASK SETUP--------->DEPLOYMENT

Page 9

3.2 Event log

The system should log every event so that the track of every detail will be known and what process is running currently could be seen.

**Initial Step-By-Step Description:**

1. The system identifies at what step logging is required.

2. The system should be able to log each and every system flow.

3. Developer can choose logging method. You can choose database logging/ File logging as well.

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

3.3 Error Handling

Should errors be encountered, an explanation will be displayed as to what went wrong? Anerror will be defined as anything that falls outside the normal and intended usage.

3.4 Optimization

Data strategy derives performance:

1. Filling missing values.
2. Replacing outliers.
3. Creating new features from cut expenses feature.
4. Hyper parameter tuning.
5. Validating score again

Page 10

4 Performance

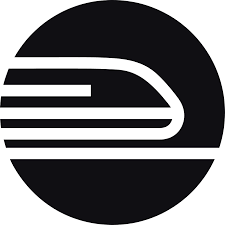
4.1 Reusability

The entire solution will be done in a modular fashion and will be API oriented. So, in the case of scaling 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 Conclusion

This system shows us the different techniques that are used to estimate the how much amount of premium required based on individual health situations. After analysis, it shows how a smoker and non-smokers affect the amount of estimate. Also, a significant difference between male and female expenses. Accuracy 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 accuracy. Our predictions help users to know how much amount premium they need based on their current health situation.

Page 11