**High Level Design (HLD)**

**Insurance Premium Prediction**

Revision 1.0.1

Date: 17/11/2023

Description: Initial revision HLD

|  |  |  |  |
| --- | --- | --- | --- |
| **Date** | **Revision** | **Description** | **Author** |
| 17/11/2023 | 1.0.1 | Initial revision (1.0.1) | Robin Mitra |
|  |  |  |  |
|  |  |  |  |

**Document Version Control**

**Contents**

Document Version Control…………………………………………………………………………………………………………… 2

Abstract……………………………………………………………………………………………………………………………………….. 4

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

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

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

1.3 Definitions………………………………………………………………………………………………………….. 5

2 General Description…………………………………………………………………………………………………………. 6

2.1 Product Perspective……………………………………………………………………………………………. 6

2.2 Problem statement…………………………………………………………………………………………….. 6

2.3 PROPOSED SOLUTION………………………………………………………………………………………… 6

2.4 FURTHER IMPROVEMENTS………………………………………………………………………………… 6

2.5 Technical Requirements……………………………………………………………………………………… 6

2.6 Data Requirements…………………………………………………………………………………………….. 6

2.7 Tools used………………………………………………………………………………………………………….. 7

2.7.1 Hardware Requirements……………………………………………………………………….. 7

2.8 Constraints………………………………………………………………………………………………………….. 7

2.9 Assumptions……………………………………………………………………………………………………….. 7

3 Design Details………………………………………………………………………………………………………………….. 8

3.1 Process Flow……………………………………………………………………………………………………….. 8

3.1.1 Model Training and Evaluation……………………………………………………………….. 8

3.1.2 Deployment Process………………………………………………………………………………. 9

3.2 Event log…………………………………………………………………………………………………………….. 9

3.3 Performance……………………………………………………………………………………………………….. 9

3.5 Reusability………………………………………………………………………………………………………….. 9

3.6 Application Compatibility……………………………………………………………………………………. 9

3.7 Resource Utilization……………………………………………………………………………………………. 9

3.8 Deployment……………………………………………………………………………………………………….. 9

4 Dashboards…………………………………………………………………………………………………………………….. 10

4.1 KPls (Key Performance Indicators)………………………………………………………………………. 10

5 Conclusion………………………………………………………………………………………………………………………. 10

**Abstract**

An insurance premium is the regular payment made by an individual or entity to an insurance company in exchange for coverage. The cost of the premium is calculated based on a variety of factors, including the type of insurance (such as auto, health, life, or property), the level of coverage, and the risk profile of the policyholder (such as age, health status, or driving history). These payments allow the insurer to maintain financial reserves to fulfill its contractual obligations when policyholders make claims. Understanding how premiums are determined is crucial for individuals and businesses to select appropriate coverage and manage their financial protection efficiently.

**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 current project description to represent a suitable model for coding. This document is also intended to help detect contradictions prior to 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 being 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:

o Security

o Reliability

o Maintainability

o Portability

o Reusability

o Application compatibility

o Resource utilization

o 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 architecture. The HLD uses non-technical to mildly-technical terms which should be understandable to the administrators of the system.

**1.3 Definitions**

|  |  |
| --- | --- |
| Term | Description |
| Database | Collection of all the information monitored by this system |
| IDE | Integrated Development Environment |
| EDA | Exploratory Data Analysis |

**2 General Description**

**2.1 Product Perspective**

The Insurance Premium Calculator system is a machine learning-based premium calculation model which will help us to get the premium amount for the customers.

**2.2 Problem statement**

The goal of this project is to give people an estimate of how much they need based on

their individual health situation. After that, customers can work with any healthinsurance carrier and its plans and perks while keeping the projected cost from ourstudy in mind. This can assist a person in concentrating on the health side of aninsurance policy rather than the ineffective part.

**2.3 Proposed Solutions**

The solution proposed here is that first we get the basic details from the consumer like “Age”, ”Sex”, “BMI”, etc. Then in second case we put these details in the Machine Learning Model. Then in last stage that is in third case it calculates the approx. Premium amount, which help them to buy the premium or not.

**2.4 Further Improvements**

In this model we use few variable i.e. “Age”, ”Sex”, ”BMI”, ”Children”, ”Smoking”, ”Region”. For further improvement we add more variables like “Income” or “Occupation” etc.

**2.5 Technical Requirements**

I used python version 3.8 with some important libraries to develop a machine learning model, which accurately predicts the Insurance premium based on its details.

**2.6 Data Requirements**

Data requirement completely depend on our problem statement.

• We need past data that is balanced and must have at least 1000 customers details.

• We require past premium expenses for each customers with annotation

Data dictionary as follows:

|  |  |  |
| --- | --- | --- |
| **Name** | **Data Type** | **Description** |
| Age | Integer | Input variable |
| Sex | String | Input variable |
| BMI | Decimal | Input variable |
| Children | Integer | Input variable |
| Smoker | String | Input variable |
| Region | String | Input variable |
| Expenses | Decimal | Output variable |

**2.7 Tools and Technologies used**

**2.7.1 Hardware Requirements**

• Jupyter notebook is used for EDA and experimentation with various ML algorithms with the help of pandas, numpy, matplotlib, seaborn, scikit learn, statsmodels and xgboost

libraries.

• PyCharm is an IDE used for development and deployment of the solution with logging. Used python version 3.7 and libraries include logging, pandas, numpy, scikit learn, statsmodels, Xgboost, pickle, flask and html 5.

• GitHub is used as a version control system.

**2.8 Constraints**

Prediction system must be user friendly, errors free and users should not be required to know about any of the workings.

**2.9 Assumptions**

Assumptions for insurance premium prediction in machine learning include accurate and representative data, relevance of selected risk factors, independence of events, stationarity, homoscedasticity, linear relationships, normality of residuals, and minimal collinearity. Regular monitoring and model adaptation are essential for dynamic real-world scenarios.

**3 Design Details**

**3.1 Process Flow**

**3.1.1 Model Training and Evaluation**

**3.1.2 Deployment Process**

**3.2 Event log**

In this project, I used the “logging” library in both the development and deployment stages, which keeps logging the events at every step into the “.log” files. One of the advantages of event logging is, it makes debugging much easier, like we can directly go to that specific line of code, having errors.

**3.3 Performance**

The ML based Insurance Premium prediction application is used for predicting the Insurance premium based on its age and other factors. So, it should be as accurate as possible, so that it will not mislead the user. Also, the model retraining is very important to keep it relevant if the new factors are added in future or to improve the performance.

**3.4 Reusability**

The code written and the components used have an ability to be reused without any problem.

**3.5 Application compatibility**

The different components or modules of this project use python version 3.8 as their interface between them. Each component has its own task to perform, and it is the job of the python version to ensure proper transfer of the information.

**3.6 Resource utilization**

In this project, any task may likely to use all the processing power available in the system, until it is accomplished.

**3.7 Deployment**

**3.8 User Interface**

Designed user interface using HTML with CSS styling. It looks as per the below image**.**

**4 Dashboards**

**4.1 KPls (Key Performance Indicators)**

Key performance indicators (KPIs) for insurance premium prediction models include Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Percentage Error (MPE), R-squared, accuracy, precision, recall, F1 score, AUC-ROC, and feature importance. These metrics collectively assess model accuracy, efficiency, and ability to predict premiums accurately.

**5 Conclusion**

**Insurance premium prediction is used to predict the premium of the personal for health insurance based on the given input entities, which enables the user to determine cost and save money on effective plans.**