

iNeuron.ai

Insurance Premium Prediction

High Level Design (HLD) Documentation

Himanshu Banodha

June 26, 2024

Document Version Control

Date Issued	Version	Description	Author
26/06/2024	1	Initial HLD – V1.0	Himanshu Banodha

Table of Contents

Document Version Control	1
1. Introduction.....	3
1.1. Why this High-Level Design Document?	3
1.2. Scope	3
1.3. Definitions	3
2. General Description	3
2.1. Product Perspective	3
2.2. Problem Statement	3
2.3. PROPOSED SOLUTION	3
2.4. Data Requirements	4
2.5. Tools used	4
2.6. Assumptions	4
3. Design Details	5
3.1. Process Flow	5
3.2. Event log.....	5
3.3. Error Handling.....	5
3.4. Performance	5
3.5. Reusability	5
3.6. Application Compatibility	6
3.7. Resource Utilization	6
3.8. Deployment.....	6
4. Conclusion	6

1. Introduction

1.1. Why this High-Level Design Document?

This 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 aspect 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

1.2. Scope

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

1.3. Definitions

Term	Description
Database	Collection of all the information monitored by this system
IDE	Integrated-Development Environment

2. 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

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 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 than the ineffective part.

2.3. PROPOSED SOLUTION

Using all the standard techniques used in the life cycle of a Data Science project starting from Data Exploration, Data Cleaning, Feature Engineering, Model Selection, Model

Building and Model Testing and also building a frontend where a user can fill their information in the form input and get the output instantly.

2.4. Data Requirements

Data requirement completely depends on our problem statement.

We need the dataset from Insurance Company. Required dataset should contain the following features:

- Age of the client.
- Whether the client is smoker or no.
- The region of the client.
- Number of children the client has.
- The body mass index of the client.
- Gender of the client.

2.5. Tools used

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

- VS Code is used as an IDE.
- For visualization of the plots Matplotlib & Seaborn are used.
- Front end development is done HTML/CSS.
- Flask is used for backend development.
- GitHub is used as a version control system.



2.6. Assumptions

The main objective of this project is to implement the use case as previously mentioned (2.2 problem statement) for new dataset that comes through the form.

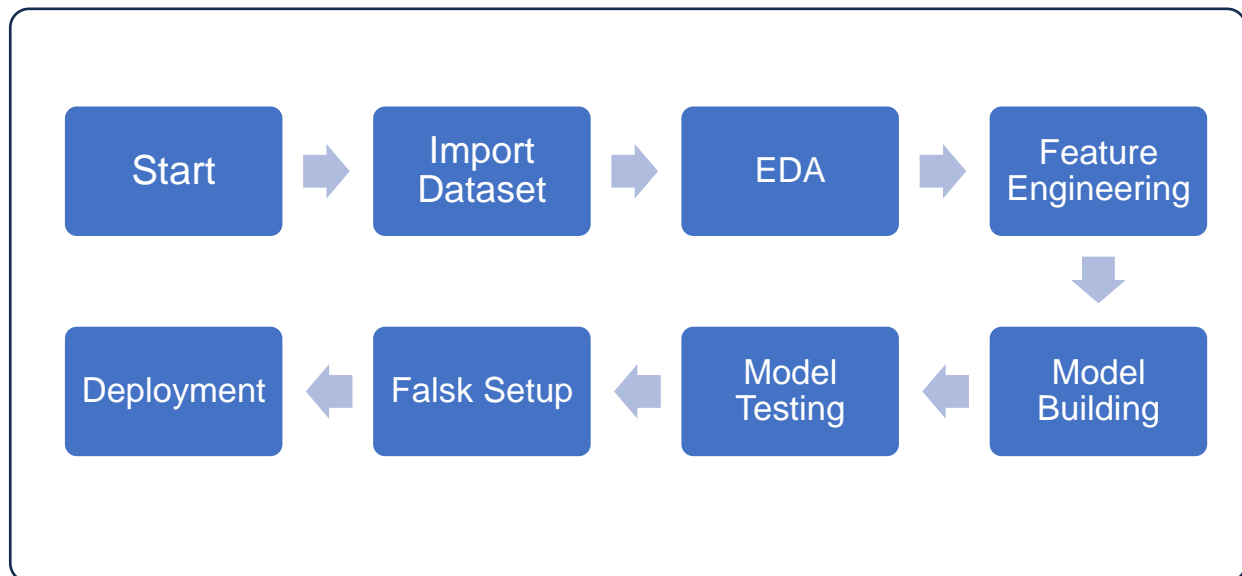
It is assumed that all aspects of this project have the ability to work together as the designer is expecting and also the data on which our model is trained is as correct as possible.

3. Design Details

3.1. Process Flow

For identifying the defaulters, we will use a machine learning base model. Below is the process flow diagram as shown below.

Proposed Methodology:



3.2. Evant log

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

Initial Step-by-Step Description:

1. The system identifies at what step logging required
2. The system should be able to log each and every system flow
3. Developers can choose logging methods. You can choose database logging/ File logging as well.
4. System should not hang even after using so many loggings. Logging just because we can easily debug issues so logging is mandatory to do.

3.3. Error Handling

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

3.4. Performance

The Insurance Premium Predictor tool is used to predict the expenses of the person on the insurance premium based on the health condition and other parameters.

Also, model retraining is very important to improve the performance.

3.5. Reusability

The code written and the components used should have the ability to be reused with no problems.

3.6. Application Compatibility

The different components for this project will be using Python as an interface between them. Each component will have its own task to perform, and it is the job of the Python to ensure proper transfer of information.

3.7. Resource Utilization

When any task is performed, it will likely use all the processing power available until that function is finished.

3.8. Deployment

*Localhost

4. Conclusion

The system shows that the different techniques that are used in order to estimate how much amount of premium is required on the basis of individual health situation. After analysing it shows how a smoker & non-smoker affecting the amount of estimate. Also, significant difference between male & female expenses.

Accuracy, which plays a key role in prediction-based system. From the results we could see that Gradient Boosting Algorithm turned out to be the best working model for this problem in terms of the accuracy.

Our predictions help user to know how much amount premium they need on the basis of their current health situation.