High Level Design (HLD)

Insurance Premium Prediction

Last date of revision:30-06-2023

Document Version Control

DATE ISSUED	VERSION	DESCRIPTION	AUTHORS
30-06-2023	1	Initial HLD-V1.0	Dipans Verma

Contents

Abstract

Insurance is a means of protection from financial loss in which, in exchange for a fee, a party agrees to compensate another party in the event of a certain loss, damage, or injury. It is a form of risk management, primarily used to hedge against the risk of a contingent or uncertain loss.

An entity which provides insurance is known as an **insurer**, **insurance company**, **insurance carrier**, or underwriter. A person or entity who buys insurance is known as a **policyholder**, while a person or entity covered under the policy is called an **insured**. The insurance transaction involves the policyholder assuming a guaranteed, known, and relatively small loss in the form of a payment to the insurer (a premium) in exchange for the insurer's promise to compensate the insured in the event of a covered loss. The loss may or may not be financial, but it must be reducible to financial terms. Furthermore, it usually involves something in which the insured has an insurable interest established by ownership, possession, or pre-existing relationship.

The insured receives a contract, called the insurance policy, which details the conditions and circumstances under which the insurer will compensate the insured, or their designated beneficiary or assignee. The amount of money charged by the insurer to the policyholder for the coverage set forth in the insurance policy is called the **premium**. If the insured experiences a loss which is potentially covered by the insurance policy, the insured submits a claim to the insurer for processing by a claims adjuster. A mandatory out-of-pocket expense required by an insurance policy before an insurer will pay a claim is called a **deductible** (or if required by a health insurance policy, a **copayment**). The insurer may hedge its own risk by taking out reinsurance, whereby another insurance company agrees to carry some of the risks, especially if the primary insurer deems the risk too large for it to carry..

1 Introduction

1.1 Why this High-Level Design Document?

The purpose of this High Level Design (HLD) Document is to add the necessary details 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 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 feature and the architecture of the project
- List and describe the non-functional attribute like:

- Security
- Reliability
- Maintainability
- Portability
- Reusability
- Application compatibility
- Resource utilization
- Serviceability

1.2 Scope

The HLD document 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

IPP – Insurance Premium Prediction

2 General Description

2.1 Product Perspective

The Insurance Premium Prediction solution system is a data science-based machine learning model which help us to detect the premium for people .

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 han the ineffective part.

2.3 Proposed Solution

The solution proposed here is a data science model based on machine learning can be implemented to perform above mention use cases. In first use case , we will take input from a healthy person who is not suffering from any disease and see whether proposed solution is going to check their premium. And in second use case, we will take input from an unhealthy person, already suffering from any disease and check our solution whether it is detect the premium.

2.4 Further Improvements

The premium prediction is necessary according to all the heath concerns . they can detect the money they are spending in insurance for their health.

2.5 Data Requirements

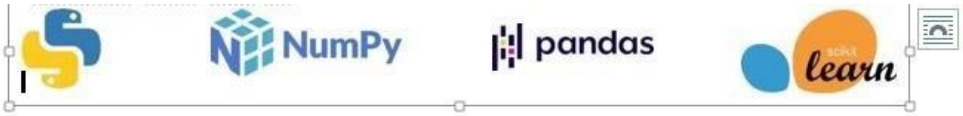
Data requirement completely depend on our problem statement.

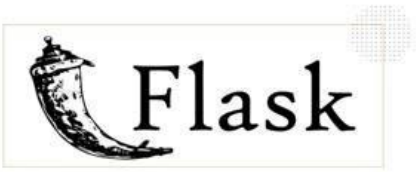
The purposes of this exercise to look into different features to observe their relationship, and plot a multiple linear regression based on several features of individual such as age, physical/family condition and location against their existing medical expense to be used for predicting future medical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

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 desginated for each level..

2.6 Tools used

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





- Pycharm is used as IDE.

- Visual Studio Code is also used as IDE.
- For visualization of the plots, Matplotlib, Seaborn and Plotly are used.
- Render is used for deployment of the model.
- MongoDB is used for DataBase operations.
- Python, Flask is used for backend development.

2.7 Constraints

The Insurance Premium Prediction solution system must be correct enough that it not mislead any report and as automated as possible and users should not be required to know any of the workings.

2.8 Assumptions

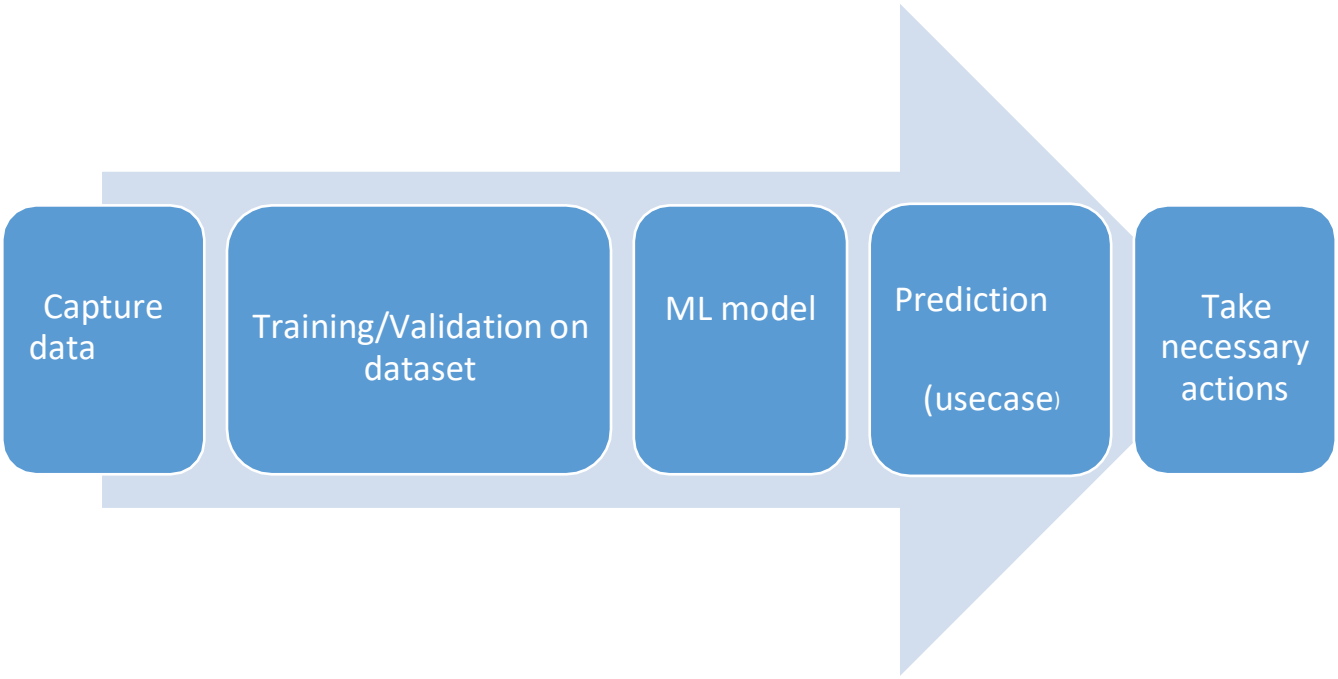
The main objective of the project is to You have to build a solution that should able to predict the premium of the personal for health insurance .

3 Design Details

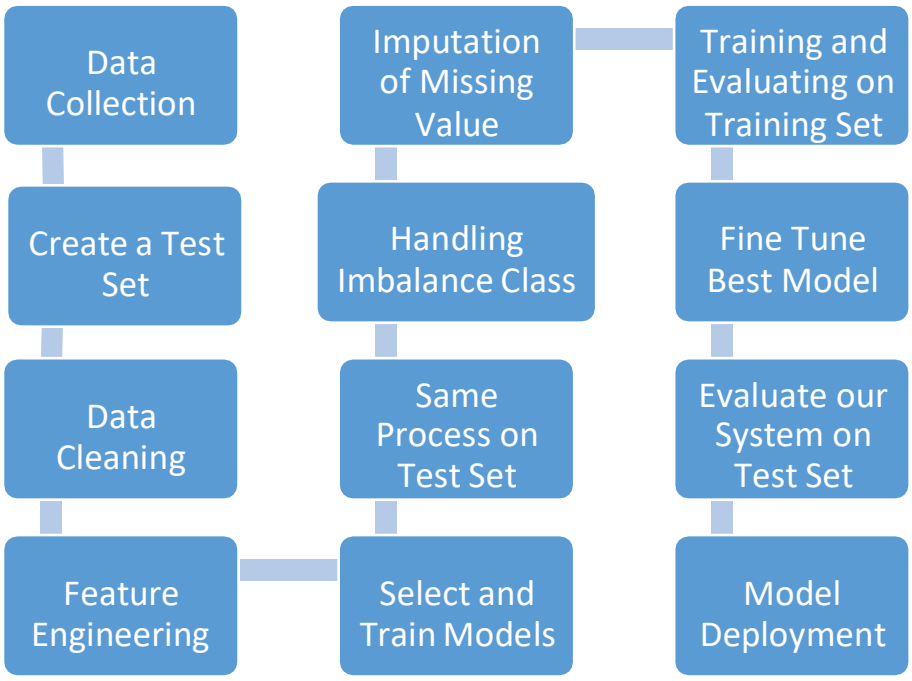
3.1 Process Flow

For detecting insurance premium, we will use machine learning base model. Below is the process flow diagram is as shown below

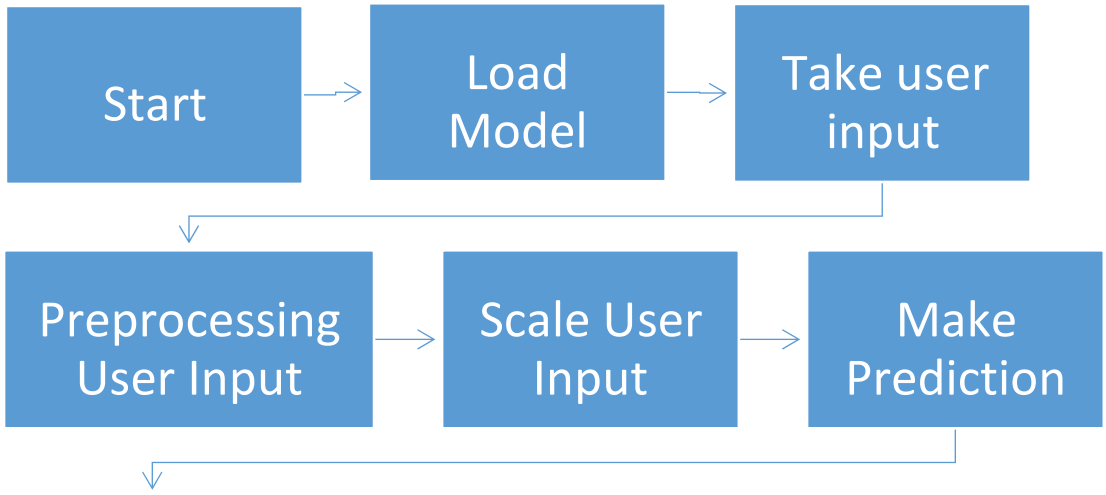
Proposed methodology



3.1.1 Model Training and Evaluation



3.1.2 Deployment Process



Result

2.6 Event 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. Developer can choose logging method. You can choose database logging/ File logging s 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.

2.7 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.

4 Performance

The machine learning based Insurance Premium Prediction solution will used for detection of insurance premium for people. So that necessary action will be taken ASP. Also model retraining is very important to improve performance.

4.1 Reusability

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

4.2 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.

4.3 Resource utilization

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

4.4 Deployment

5 Conclusion

Insurance Premium Prediction solution will take health-care domain data of those patients who have undergone diagnosis for any disease to train our machine learning model and will evaluate its performance over usecase mentioned above. And then leverage its prediction to detect disease in people having symptoms of any disease and able to alert people who is on positive side so that medical attention along with treatment will be given to that particular people as soon as possible. This solution should be as accurate as possible, so that chances of misleading reports will be taken good care of. And also taking care of premium .

6 References

KAGGLE

URL: https://www.kaggle.com/datasets/noordeen/insurance-premium-prediction