Low Level Design (LLD)

Insurance Premium Predictions

Revision Number: 1.1
Last date of revision: 23/11/2021

Raja Arvindan R

Document Version Control

Date Issued	Version	Description	Author
23rd Nov 2021	1.1	Initial LLD-V1.0	Raja Arvindan R

Contents

Abstract

The main foundational block of health insurance industry is to estimate the future events and measure the associated risk/value of these events, hence it is needless to say that predictive analytics is used widely to determine the risk, insurance premium enrich the overall customer experience. The health insurance industry has always been a slow-moving industry when it comes to adopting the data analytics practices into its business models. With the advent of advanced data analytics technologies, it has become important now more than ever to take advantage of such sophisticated analytics to accurately assess and predict the insurance premiums for the insured. Thus, one of the important tasks for health insurance companies is to determine the policy premiums. By using predictive modelling, the insurers can determine the policy premium for the insured based on their behaviours which are indicated by attributes such as age, BMI (Body Mass Index), smoking habits, number of children etc. This determination of premiums based on the data collected for an individual helps insurance companies in enhanced pricing, underwriting and risk selection. Additionally, it helps in making better decisions, understanding customer needs and be fair to the customers. Acquiring a comprehensive understanding of customer behaviours and habits from historical data helps insurers to anticipate future behaviours and provide the right insurance product and policy premium.

1 Introduction

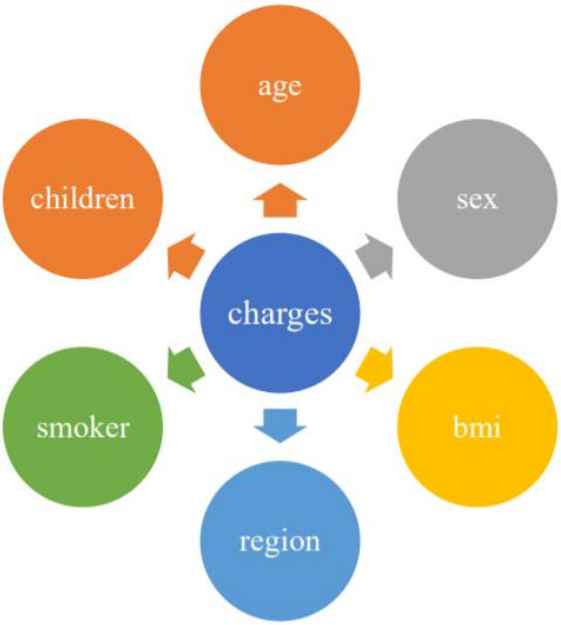
1.1 Why this Low-Level Design Document?

The purpose of this document is to present a detailed description of the Insurance Premium Prediction. It will explain the purpose and features of the system, the interfaces of the system, what the system will do, the constraints under which it must operate and how the system will react to external stimuli. This document is intended for both the stakeholders and the developers of the system and will be proposed to the higher management for its approval.

The main objective of the project is to predict the Health Insurance Prediction. The goal of this project is to allows a person to get an idea about the necessary amount required according to their own health status. Later they can comply with any health insurance company and their schemes & benefits keeping in mind the predicted amount from our project. This can help a person in focusing more on the health aspect of an insurance rather than the futile part. The main goal of the project is to predict the insurance premium charge based upon other attributes.

An Insurance Premium Prediction contains the information, such as:

- ✓ age
- ✓ sex
- ✓ children
- ✓ BMI
- ✓ smoker
- ✓ region



This project shall be delivered in two phases:

Phase 1: All the functionalities with Scikit-learn packages.

Phase 2: Integration of UI to all the functionalities.

1.2 Scope

This software system will be a Web application This system will be designed to predict health insurance premium. Premium amount prediction focuses on person's own health rather than other company's insurance terms and conditions. The models can be applied to the data collected in coming years to predict the premium. This can help not only people but also insurance companies to work in tandem for better and more health centric insurance amount. This system is designed to predict the health insurance premium based on some information like age, sex, BMI, region, smoker, children etc.

1.3 Constraints

The Insurance Premium Prediction must be user friendly, as automated as possible and users should not be required to know any of the workings.

1.4 Risks

Document specific risks that have been identified or that should be considered.

1.5 Out of Scope

Delineate specific activities, capabilities, and items that are out of scope for the project.

2 Technical specifications

2.1 Dataset

Name	Description
Age	Age of the client
BMI	Body mass index
The number of kids	Number of children the client has
Gender	Male / Female
Smoker	Weather the client is smoker or not
Region	Where the client lives southwest, southeast, northwest, northeast.

2.1.1 Insurance Premium dataset overview

To create the claim cost model predictor, we obtained the data set through the link provided by iNeuron. The data set includes seven attributes, the data set is separated into two-part the first part called training data, and the second called test data; training data makes up roughly about 80 percent of the total data used, and the rest for test data The training data set is applied to build a model as a predictor of medical insurance cost year and the test set will use to evaluate the regression model. the following table shows the Description of the Dataset.

Kids	**Have**
Gender	Male / Female 1=Male 0=Female
Smoker	whether a client is a smoker or not 1=yes 0=no
Region	where the client lives 1= southwest 2= southeast 3= northwest 4= northeast
Expenses (Target Variable)	Medical Cost the client pay

Some of the records in the dataset are following

	age	sex	bmi	children	smoker	region
0	19	female	27.9	0	yes	southwest
1	18	male	33.8	1	no	southeast
2	28	male	33.0	3	no	southeast
3	33	male	22.7	0	no	northwest
4	32	male	28.9	0	no	northwest
...	...	...	...	...	...	...
1333	50	male	31.0	3	no	northwest
1334	18	female	31.9	0	no	northeast
1335	18	female	36.9	0	no	southeast
1336	21	female	25.8	0	no	southwest
1337	61	female	29.1	0	yes	northwest

1338 rows × 6 columns

2.1.2 Input schema

Feature name	Datatype	Size	Null/Required
Age	int	2	Required

2.2 Predicting Disease

- The system displays the form where the all features are available if user can all features correctly then machine can able to predict the health insurance premium.

2.3 Logging

We should be able to log every activity done by the user.

- 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/ File logging as well.
- System should not be get stuck even after usage of many loggings. With logging we can easily debug issues so logging is a mandatory.

2.5 Deployment

1. FLASK

3 Technology stack

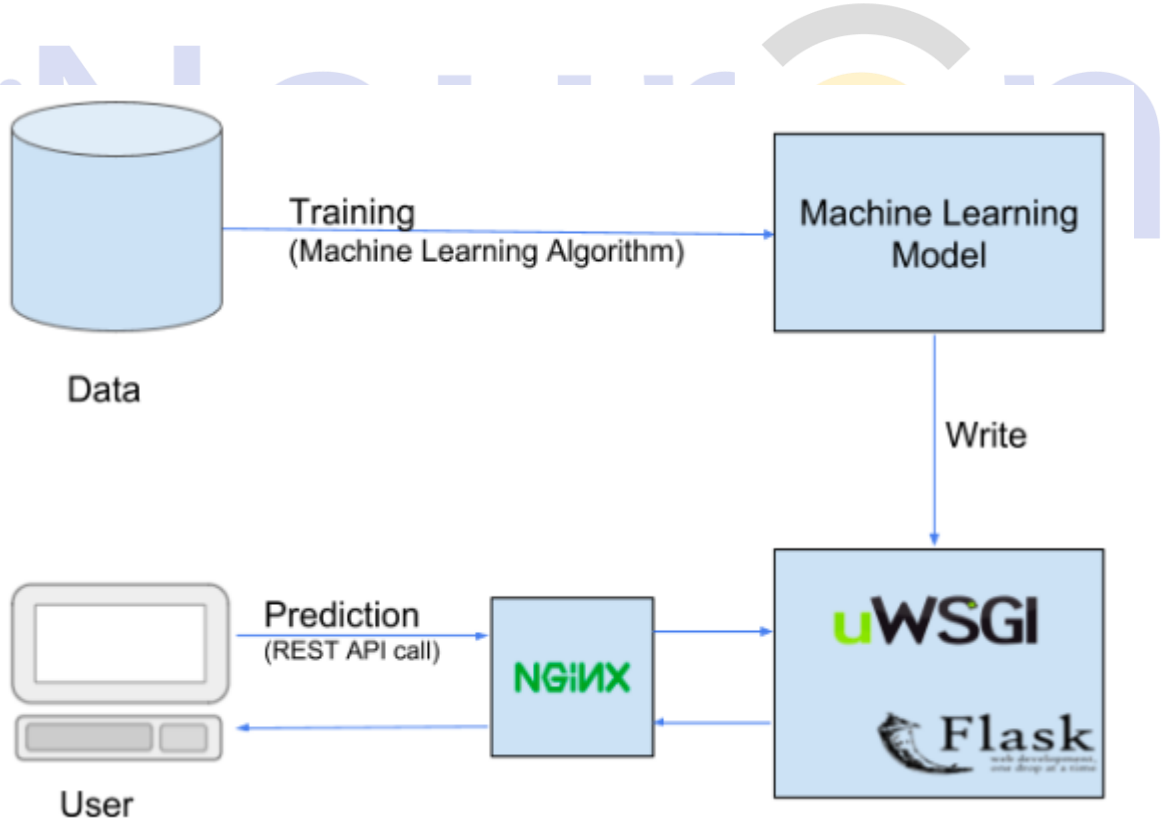
Front End	HTML/CSS/JS
Backend	Python Flask

4 Proposed Solution

Based on the actual research paper, gradient boosting gives better accuracy as compare to other so in this project we use gradient boosting regression algorithm to predict insurance. However, drawing a baseline in the form of some Machine Learning algorithm would be helpful. Why making a baseline model important? Well, to compare the performance of our actual model, let say Gradient Boosting in this case, is very important to ascertain that we are in the right direction as if performance of gradient boosting is not better than the baseline model then there is no point of using gradient boosting.

1. Actual model: Gradient Boosting.

5 Model training/validation workflow



6 User I/O workflow

7 Exceptional scenarios

Step	Exception	Mitigation	Module
23rd Nov 2021	1.1	First Draft	Raja Arvindan R

8 Test cases

Test case	Steps to perform test case	Module	Pass/Fail