SMARTINTERNZ INTERNSHIP PROJECT REPORT

PROJECT TITLE:

HEALTH INSURANCE COST PREDICTION USING WATSON AUTO AI

SUBMITTED BY:

ASTHA AGNIHOTRI

INDEX

1. INTRODUCTION

1.1 Overview

Health Insurance cost prediction is basically and approximately calculated amount which any particular person is needed to pay for his/her insurance. These predictions and expectations are calculated based on several factors like:

- **Age** of the person
- **Sex** of the person
- **BMI** (Body Mass Index)
- No. of children
- If the person is a **smoker** or not
- **Region** where the person lives.

It will require a sophisticated system that will consider all the above-mentioned factors.

1.2 Purpose

The goal for the project is to create a model to predict or calculate the average cost of Health Insurance for a person that he/she must need.

2. LITRETURE SURVEY

2.1 Existing Problem

Predicting the health insurance cost may have many benefits such as a person can adjust his/her expenditure and other expenses so that it suits the best for his pocket and could have maximum savings.

Usually, health insurance companies face much difficulties to select the perfect premiums for their customers, so that the customer can get the best deal suited for him/her.

There is a need of an application which will predict the Health Insurance cost factors, as;

- Social
- Economic
- Mortality
- Immunization Factors.

The application must give prediction accurately, within fractions of second.

2.2 Proposed Solution

A regular machine learning project that processes data to predict the approximate cost of health insurance which is best suited for the person. The problem statement is targeted to predict the cost of health insurance given with the various factors.

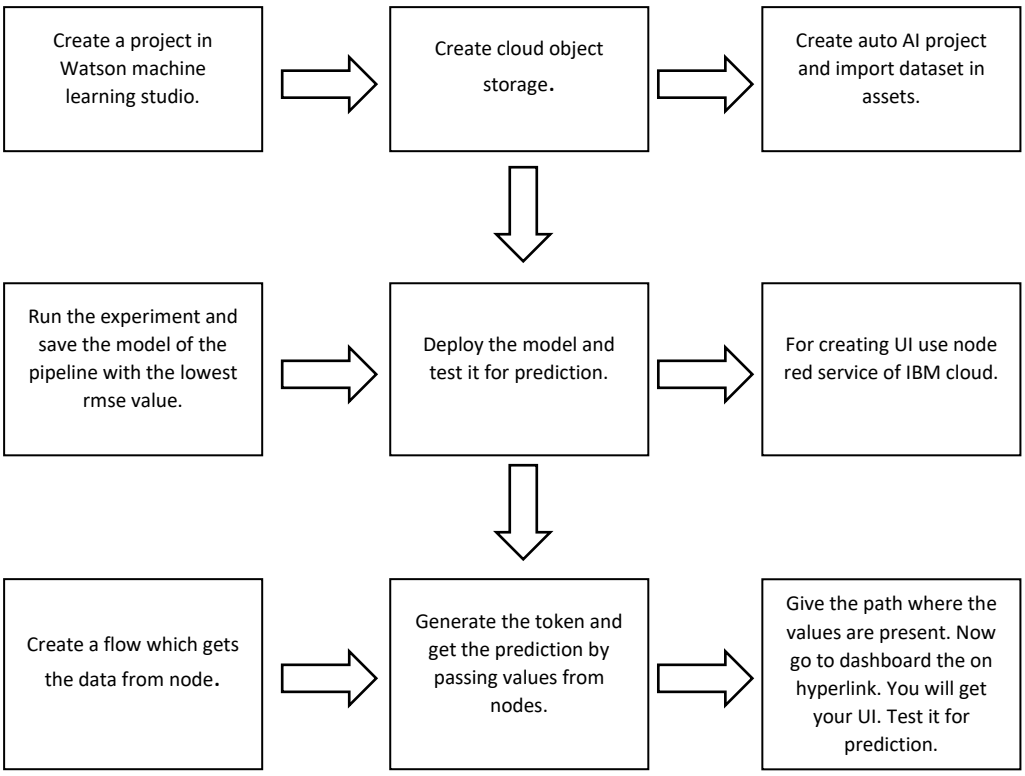
The main aim for the project is to create a model based on significant factors which will affect premiums and charges by any insurance company.

In this project we are using Multi Linear regression for the accurate prediction. An application is also built in Auto AI service in the IBM cloud which can be linked with the model so as to view the result on the UI based on the input parameters.

The cost of health insurance may vary from different factors depending upon the situation and condition of the person for whom the cost of health insurance is to be calculated.

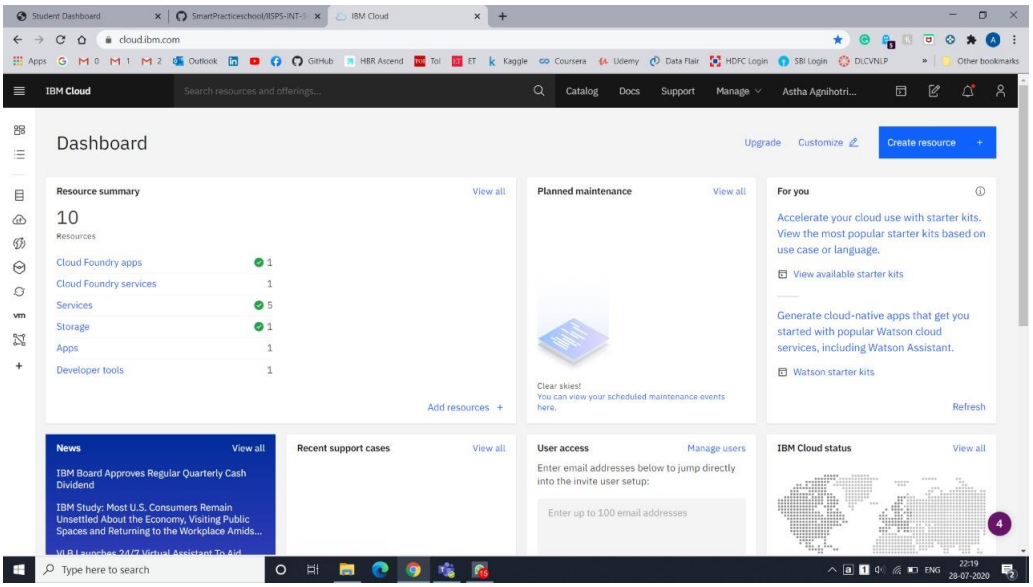
3. HARDWARE SURVEY

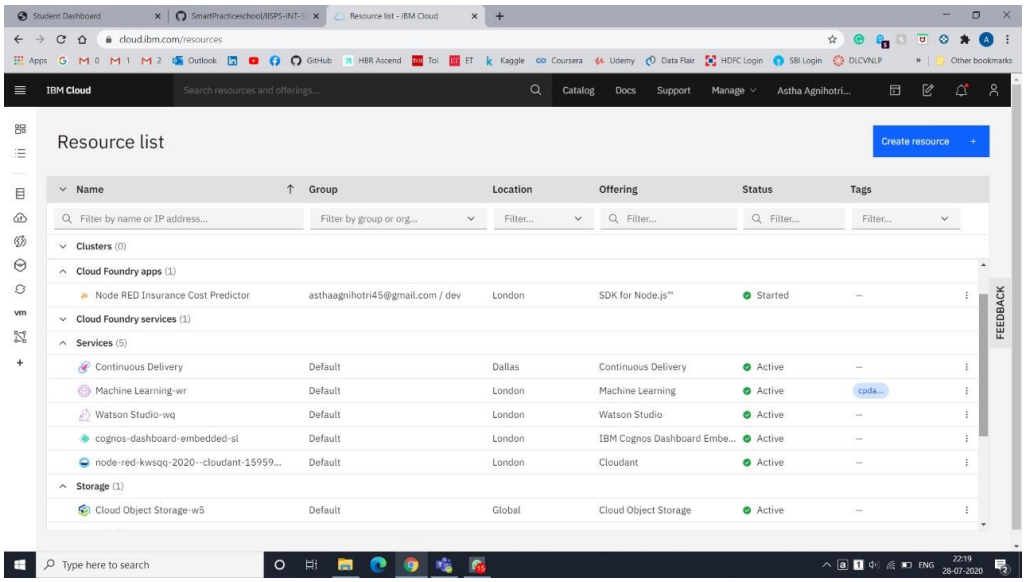
3.1 Block Diagram



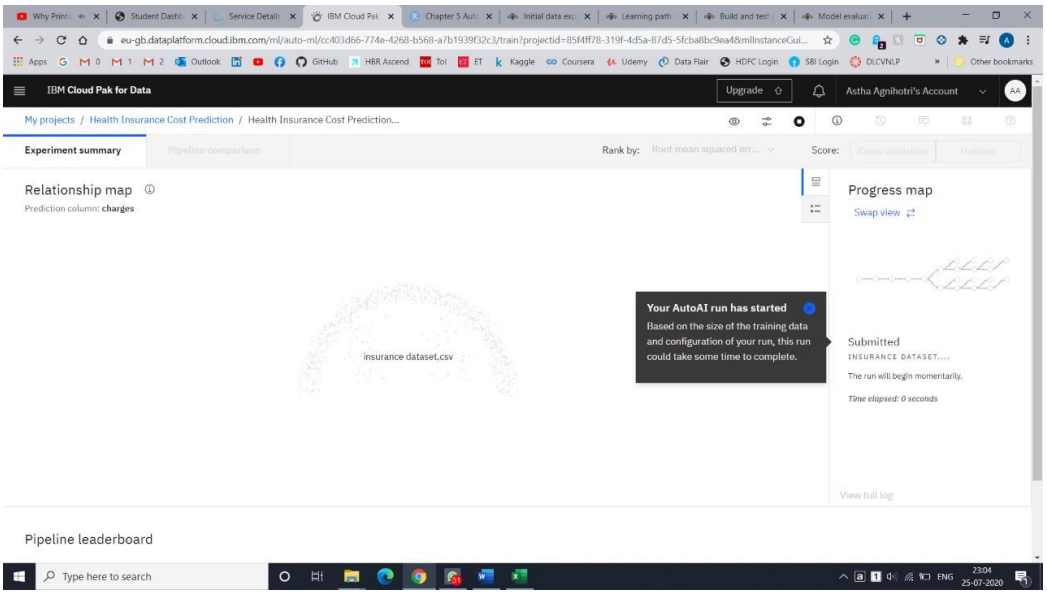
3.2 Hardware / Software Designing

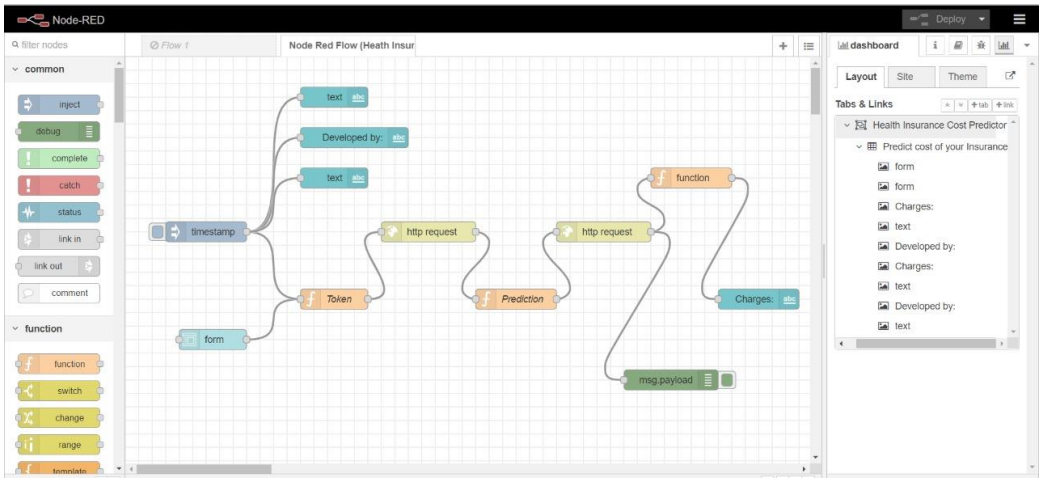
In this particular project model, the soul usage is of the software model, hence there is no special hardware involved in this project.



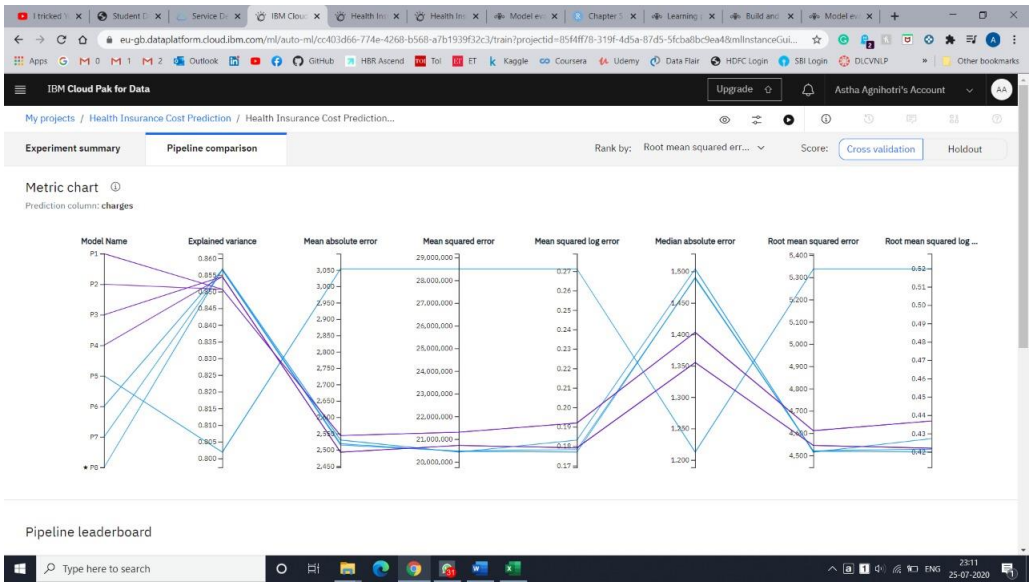


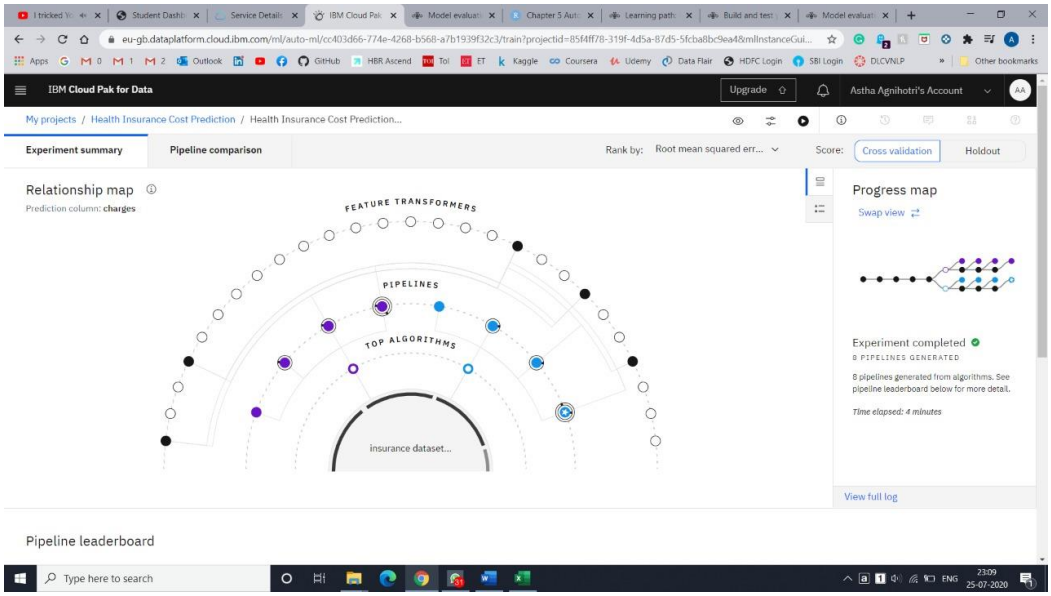
The web application created consists of the IBM Watson as the backend service and the Node-RED for the development of the user interface.





Auto AI Experiment:





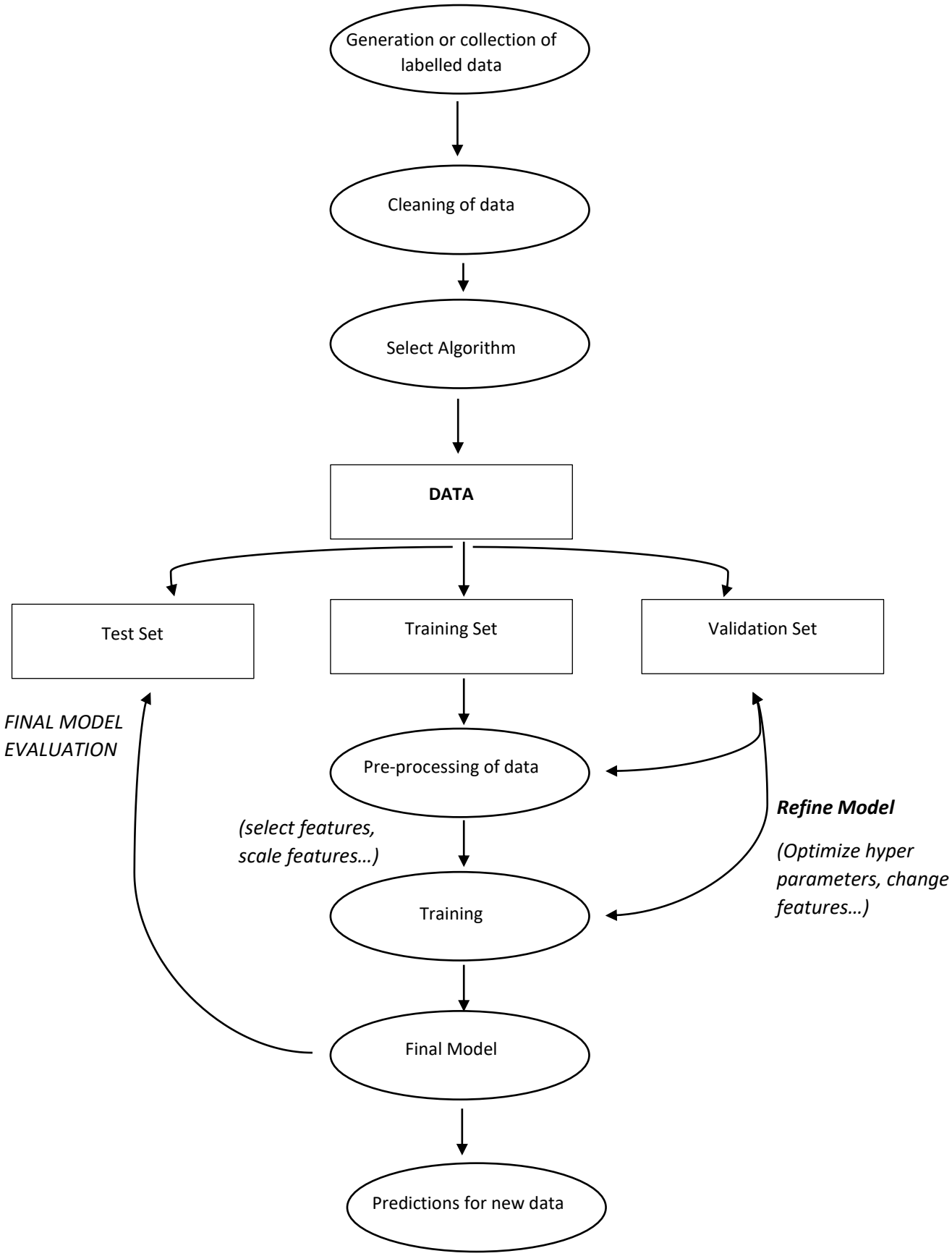
4. EXPERIMENTAL INVESTIGATIONS

Firstly, the data is collected from the given data set. Then a model is built on the IBM Watson studio machine learning to calculate the cost of health insurance. For that, an account is to be created on the IBM Watson studio then a project is added with the auto AI option. Then the data set is uploaded and the best possible way to predict the cost is chose. Finally, the model is executed and tested with the various data, factors and different circumstances.

Then a cloud foundry application is created from the Node-RED, and created a Node-RED flow. At last an API key is added with an ID and URL. And with the execution of the model the User Interface can be seen from the dashboard.

The data of the given dataset had to be updated with some the information as the available dataset had some missing information due to which the outcome was dropped from the data frame due to which the model was unable to learn from the missing values. The missing values were filled with the mean value of the respective feature for the particular country.

5. FLOWCHART


Generation or collection of labelled data
Cleaning of data
Select Algorithm
DATA
Test Set
Training Set
Validation Set
FINAL MODEL EVALUATION
Pre-processing of data
(select features, scale features…)
Refine Model
(Optimize hyper parameters, change features…)
Training
Final Model
Predictions for new data

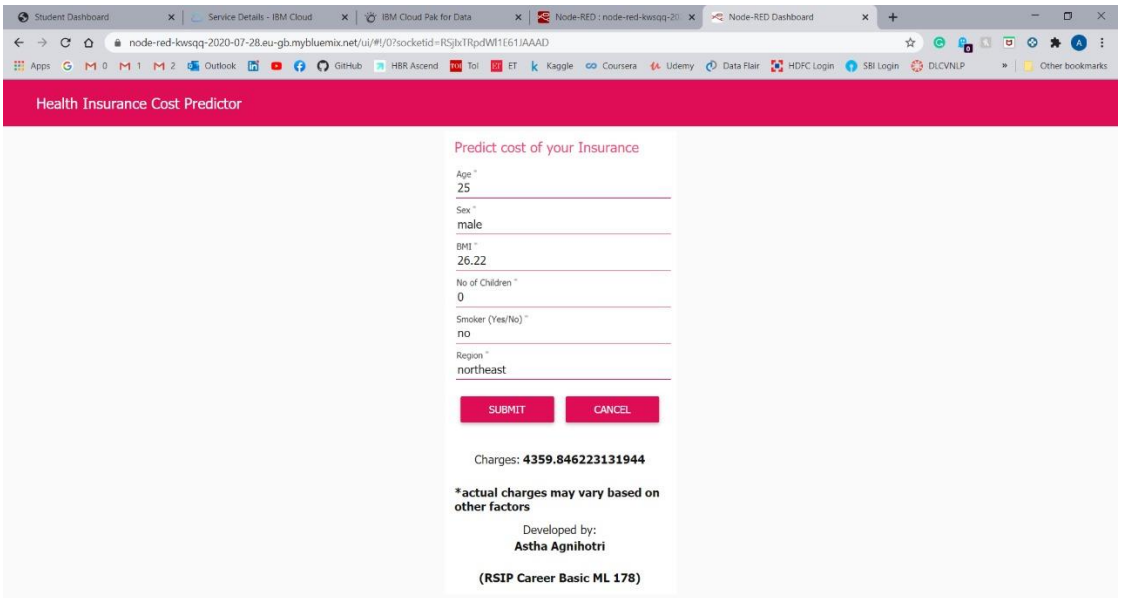
6. RESULT

The web application is created successfully with good and appropriate results.

The model is working properly as expected and fulfil all the given circumstances and it properly manages and process all the factors provided and give away the required result.
As factors the model takes the values for: Age, Sex, BMI, No. of children, Smoker or not, Region.

The web application is available in the following link:

https://node-red-kwsqq-2020-07-28.eu-gb.mybluemix.net/ui/



7. ADVANTAGES & DISADVANATAGES

Advantages:

- It easily identifies the trends and the patterns; machine learning can review a large number of data and discover specific trends and patterns that would not be apparent to humans.
- No human interaction needed. With this model, the system will automatically generate the feasible output in this case cost for health insurance without any need or help of a human presence.
- Can handle multiple variety of data in any dynamic or uncertain environments.

Disadvantages:

- This requires a massive data set to train the program model, and the data set should be unbiased and must be of good quality.
- Not every time the algorithm works properly, it tries to keep the result as precised as possible but sometimes it cannot get the perfect desired result.

8.APPLICATION

The project is very useful and can have many applications in several fields, some the best applications according to me are listed below:

- **Reduce readmissions:**
  Clinicians can receive daily guidance as to which patient are most likely to be readmitted and how they might be able to reduce that risk to be admitted again.

- **Prevent Hospital acquired Infection:**
  This might help to monitor a patient, which is more likely to catch an infection or may get other people around infected. This model will help to rack the patient and give them treatment accordingly.

- **Customer's Savings:**
  This model can also help to get the customer the exact required part of a health insurance, so that the customer can save some money and also be safe with a health insurance which is more likely to be affordable by the person himself.

9. CONCLUSION

This internship with Smartbridge turned out to be a great opportunity for me, as I get to learn much about the machine learning and GitHub, etc. I had a chance to do a professional practice in the field of Machine Learning.

Still I have a lot left to learn about the topic but this internship program was great for me. This helped me know my skills and weaknesses and also what skills do I have to work on for my personal development in the field of Machine Learning. I got to learn about the IBM Cloud and its machine learning model creation.

With the proper help and guidance from the Smartbridge team, I was able to build a successful working model for the given task. The project was well designed and established. There still are many improvements in the current model but I believe the current model fulfil all the requirement of the given project task.

10. BIBLIOGRAPHY

- https://www.youtube.com/
- https://developer.ibm.com/tutorials/how-to-create-a-node-red-starter-application/
- https://github.com/SmartPracticeschool/llSPS-INT-3458-Health-Insurance-Cost-prediction-Using-Watson-Auto-Ai
- https://node-red-kwsqq-2020-07-28.eu-gb.mybluemix.net/ui/#!/0?socketid=IbLQxinJJPu1jm7aAAAZ
- https://www.kaggle.com/kumarajarshi/life-expectancy-who