

# **SMARTINTERNZ**

# **INTERNSHIP REPORT**

## **HEALTH INSURANCE COST PREDICTION USING IBM WATSON AUTO-AI**

**T S VENKATA SUBRAMANIAM**  
**B.TECH ELECTRONICS AND COMPUTER(3<sup>RD</sup> YR)**  
**VELLORE INSTITUTE OF TECHNOLOGY CHENNAI**

### **I. INTRODUCTION**

**Problem statement :**

Health Insurance Cost Predictor

**Overview :**

A predictive and analysing machine learning model is built with the help of IBM watson studio, AUTOAI, and node red.

**Purpose :**

To build a machine learning model that helps the health insurance companies to provide premium offers to the customers based on certain factors and statistics.

## **II. LITERATURE SURVEY**

### **Existing Problem :**

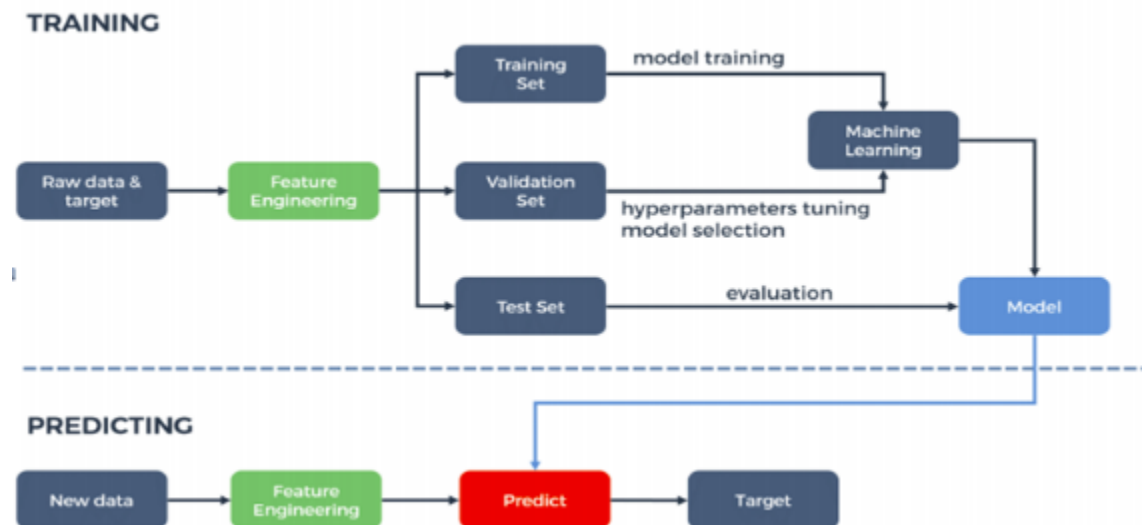
Health insurance companies face a problem in determining the premium insurances for their customers. They need a trained model which predicts what kind of service is suitable to which customer. The companies should also follow certain rules set by the health care law in that country. In order to determine premium offers the companies have to consider certain factors and also go statistically to give importance to a customer.

### **Proposed Solution:**

The main aim of this project is to provide the health insurance companies with a machine learning model using the IBM WATSON AI platform that can predict the customer's eligibility for different premium insurance services based on certain factors like age, gender, BMI, weight, previous health issues, the region he/she belongs to, if he is a smoker or not etc. The model is tested based on the accuracy and performance of the model.

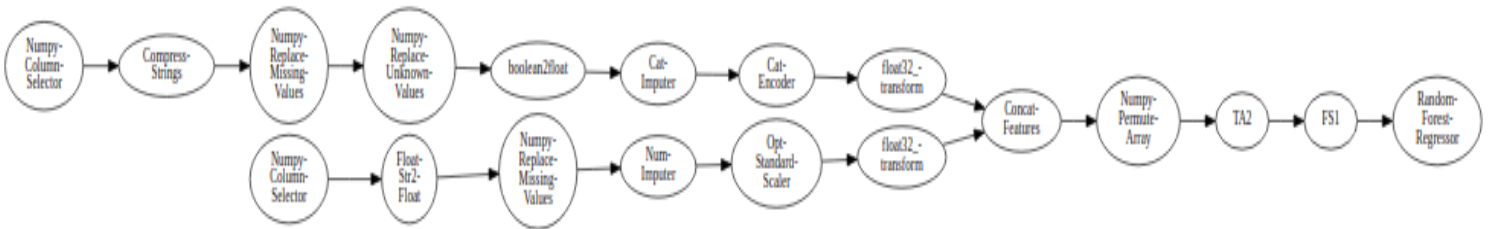
## **III. THEORETICAL ANALYSIS**

### **Block Diagram :**



## Hardware-Software Designing :

- Machine Learning algorithms using python
- IBM watson studio, AutoAI and Nodered platform



## IV. EXPERIMENTAL INVESTIGATION

1. Collection of dataset from kaggle.  
( <https://www.kaggle.com/annetxu/health-insurance-cost-prediction>)
2. Create an account for IBM Cloud platform.
3. On the IBM Watson studio platform use the Auto AI to build a model that predicts the cost of the insurance that can be delivered to a company based on the details given by their customer.
  - a. First we need to create an account on the IBM Watson studio.
  - b. Using add to project choose Auto AI
  - c. Then upload the dataset that is taken from kaggle into the data assets

IBM Cloud Pak for Data

My projects / Health Insurance Cost Predictor

Overview Assets Environments Jobs Deployments Access Control

What assets are you looking for?

Data assets

0 assets selected.

Name	Type	Created by	Last modified
insurance.csv	Data Asset	T S Venkata Subramaniam	Jul 24, 2020, 12:43 PM

AutoAI experiments

Name	Status	Model type	Last modified
Insurance cost predictor	Completed	Regression	Jul 24, 2020, 12:48 PM

Notebooks

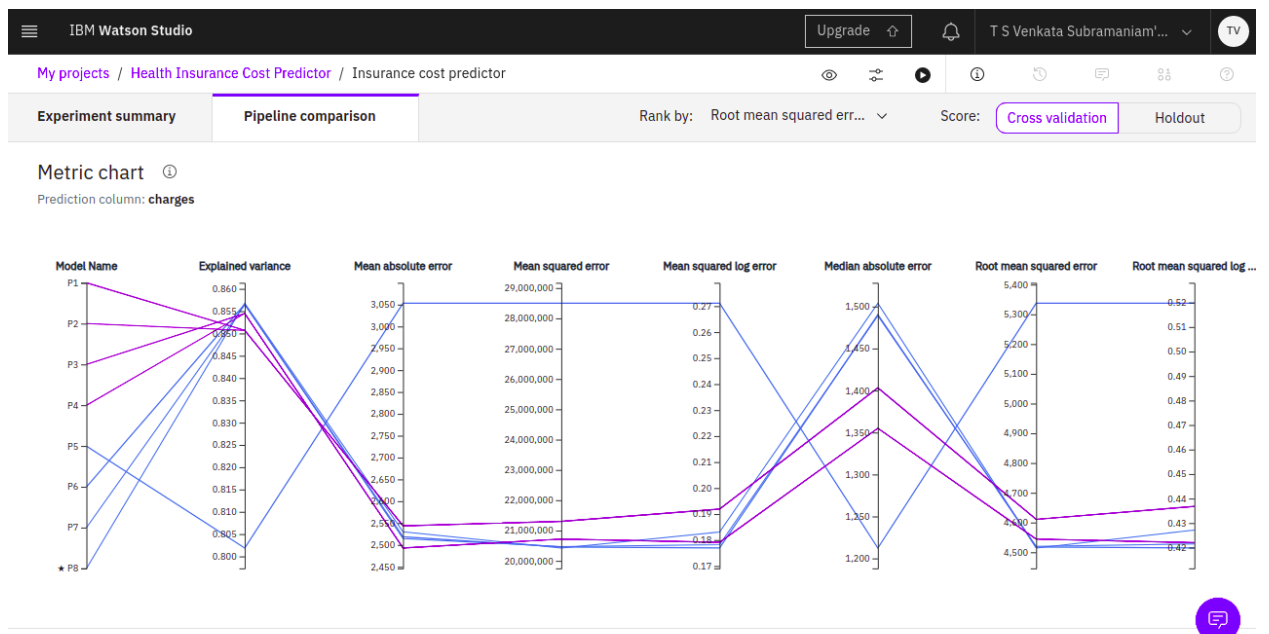
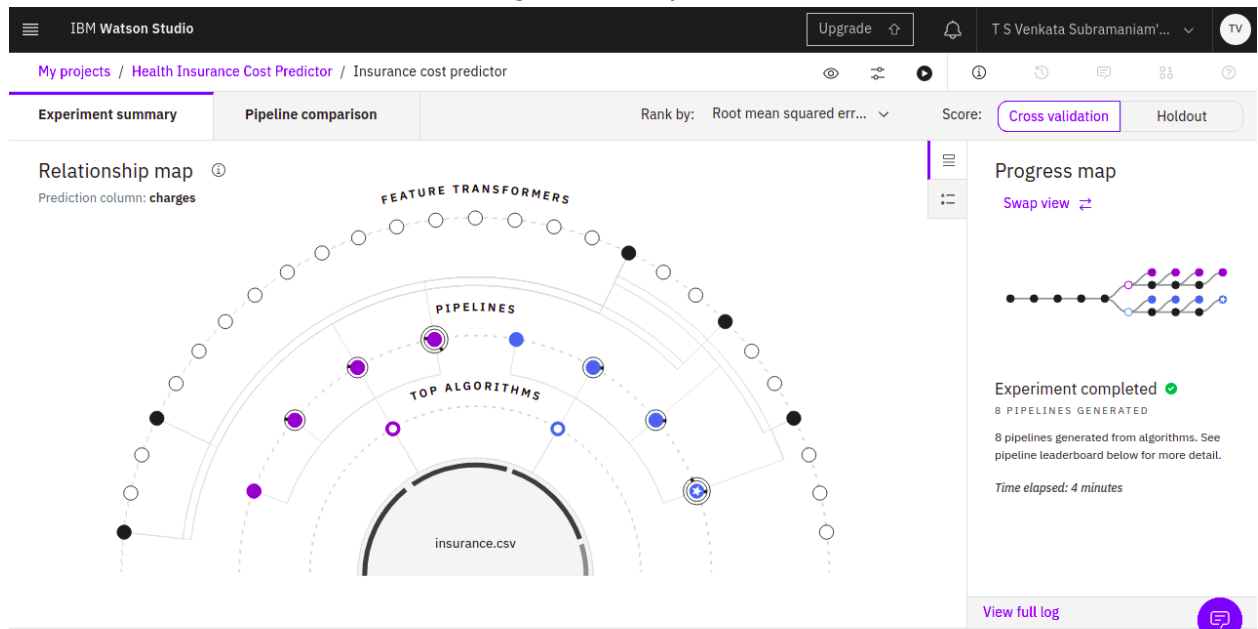
New Notebook +

Data

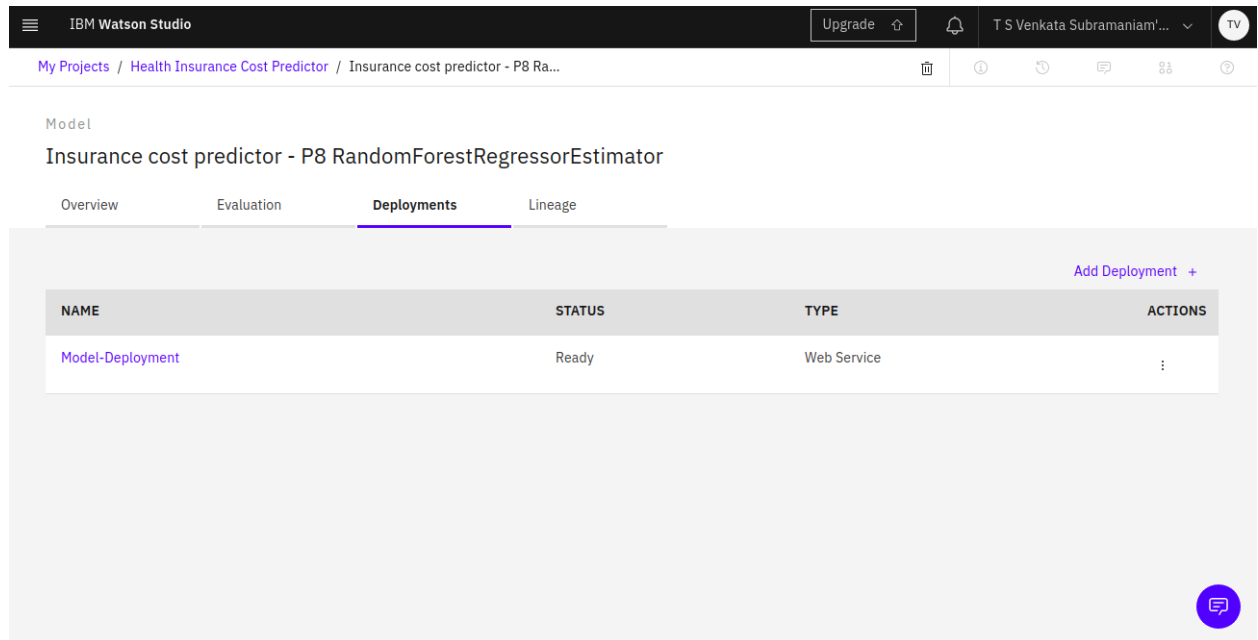
Load Files Catalog

Drop files here or browse for files to upload.

#### 4. Choose the best machine learning model to predict the cost.



## 5. Deploy the model.

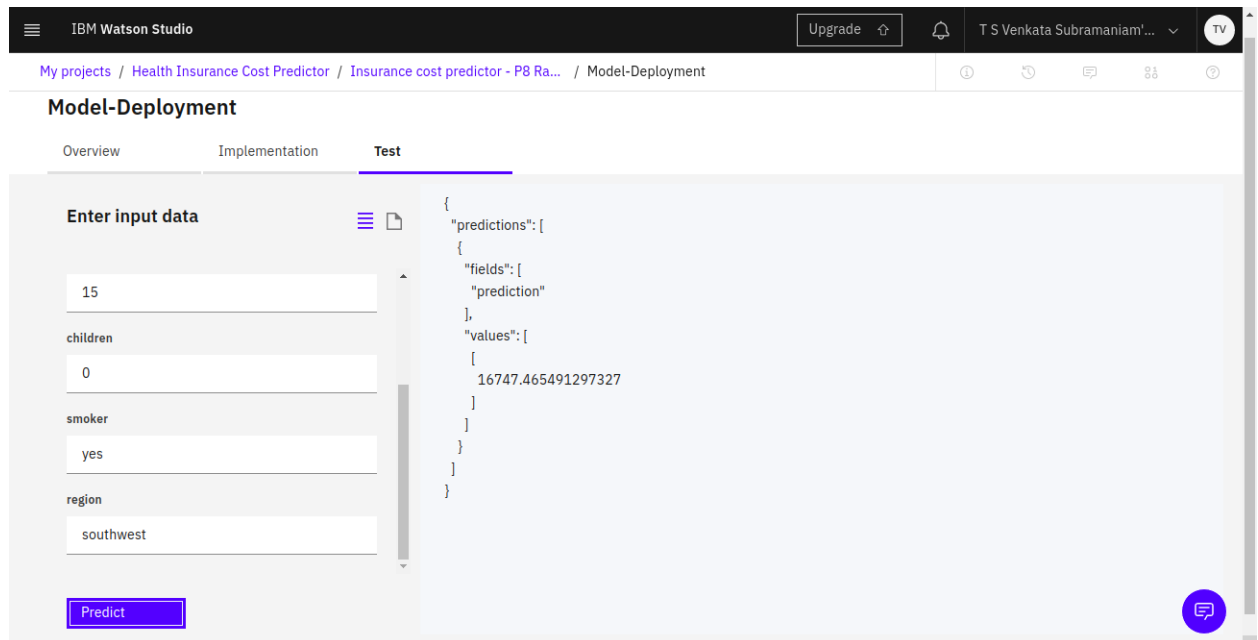


The screenshot shows the IBM Watson Studio interface. The top navigation bar includes the IBM Watson Studio logo, an 'Upgrade' button, a notification bell, and a user profile dropdown for 'T S Venkata Subramaniam'. The breadcrumb trail indicates the current location: 'My Projects / Health Insurance Cost Predictor / Insurance cost predictor - P8 Ra...'. The main content area is titled 'Model' and 'Insurance cost predictor - P8 RandomForestRegressorEstimator'. Below this, there are four tabs: 'Overview', 'Evaluation', 'Deployments' (which is active), and 'Lineage'. The 'Deployments' tab displays a table with the following data:

NAME	STATUS	TYPE	ACTIONS
Model-Deployment	Ready	Web Service	⋮

An 'Add Deployment +' button is located in the top right corner of the table area. A chat icon is visible in the bottom right corner.

## 6. Test the model against various values.

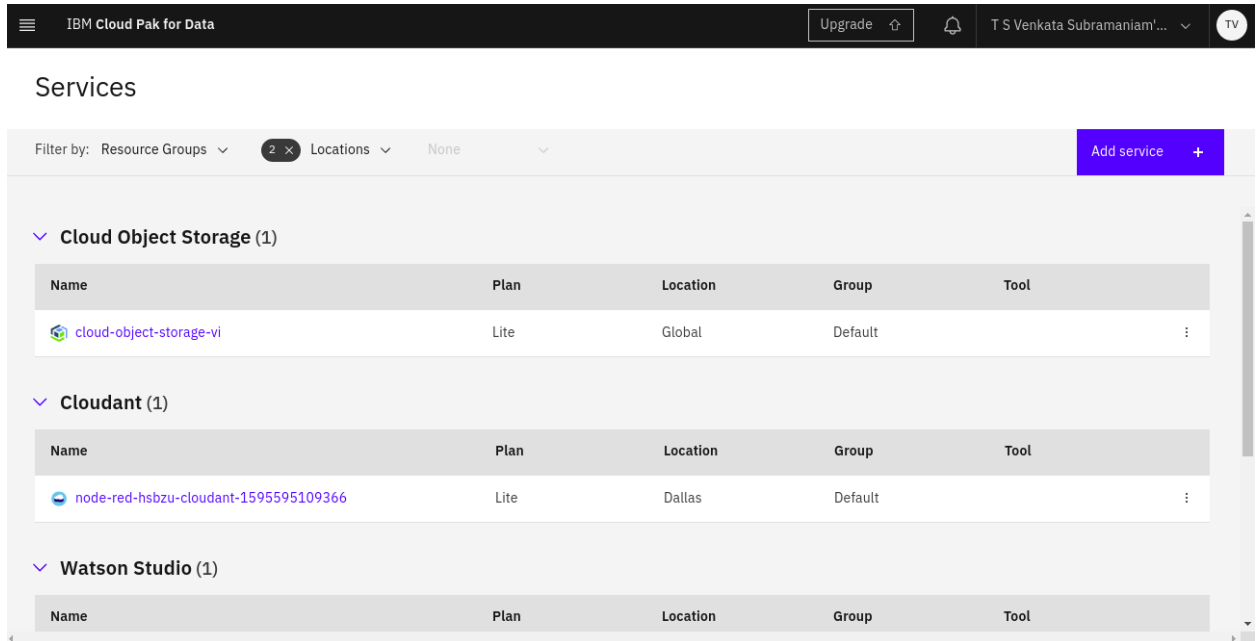


The screenshot shows the IBM Watson Studio interface with the 'Model-Deployment' tab selected. The breadcrumb trail is 'My projects / Health Insurance Cost Predictor / Insurance cost predictor - P8 Ra... / Model-Deployment'. The main content area is titled 'Model-Deployment' and has three tabs: 'Overview', 'Implementation', and 'Test' (which is active). The 'Test' tab is divided into two sections. On the left, under 'Enter input data', there are input fields for '15', 'children' (0), 'smoker' (yes), and 'region' (southwest). A 'Predict' button is at the bottom of this section. On the right, the output is displayed as a JSON object:

```
{
  "predictions": [
    {
      "fields": [
        "prediction"
      ],
      "values": [
        16747.465491297327
      ]
    }
  ]
}
```

A chat icon is visible in the bottom right corner.

7. Then create a service credential and also cloud foundry app.



The screenshot shows the 'Services' page in IBM Cloud Pak for Data. The top navigation bar includes 'IBM Cloud Pak for Data', an 'Upgrade' button, a notification bell, and a user profile 'T S Venkata Subramaniam'. The main content area is titled 'Services' and features a filter bar with 'Resource Groups', 'Locations' (2 selected), and 'None'. A table lists three service categories: Cloud Object Storage (1), Cloudant (1), and Watson Studio (1). Each category has a table with columns: Name, Plan, Location, Group, and Tool.

Name	Plan	Location	Group	Tool
cloud-object-storage-vi	Lite	Global	Default	

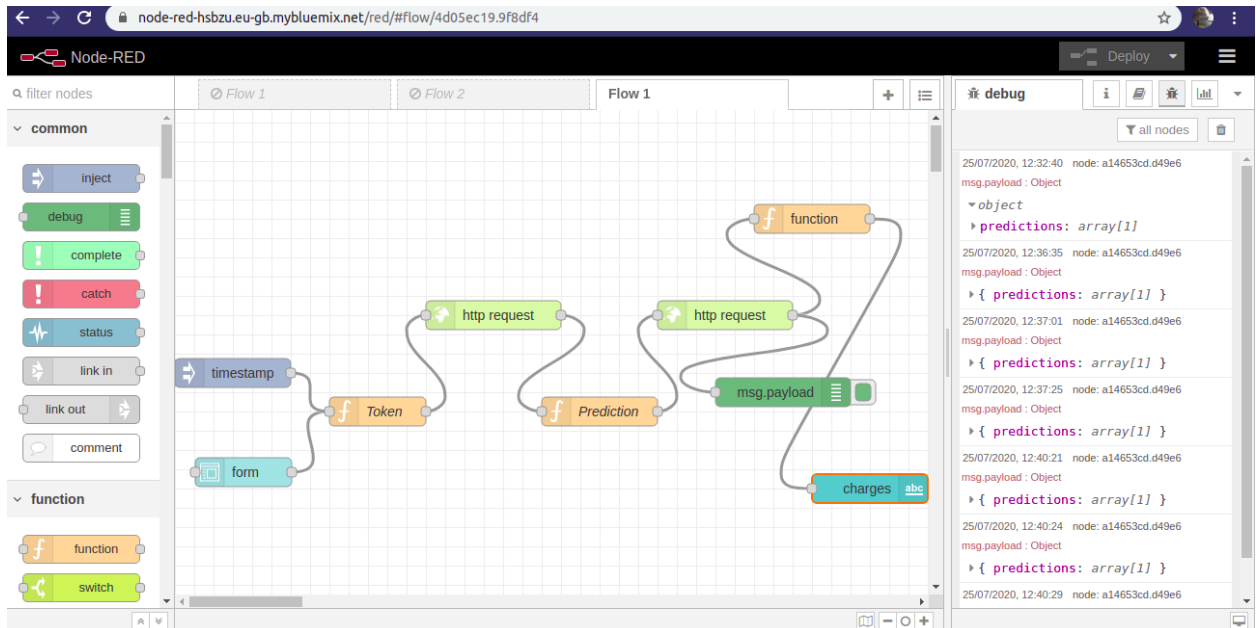
  

Name	Plan	Location	Group	Tool
node-red-hsbzu-cloudant-1595595109366	Lite	Dallas	Default	

Name	Plan	Location	Group	Tool
------	------	----------	-------	------

8. Make a node-RED flow.



The screenshot shows the Node-RED web interface in a browser. The URL is 'node-red-hsbzu.eu-gb.mybluemix.net/red/#flow/4d05ec19.9f8df4'. The interface includes a left sidebar with 'common' and 'function' node categories, a central workspace with a flow diagram, and a right sidebar with a 'debug' console. The flow diagram consists of several nodes: 'inject', 'timestamp', 'form', 'Token' (function), 'http request', 'Prediction' (function), 'msg.payload' (message), 'charges' (message), and 'function' (function). The debug console shows a series of messages with a 'predictions' array.

```
25/07/2020, 12:32:40 node: a14653cd.d49e6  
msg.payload : Object  
object  
  predictions: array[1]  
25/07/2020, 12:36:35 node: a14653cd.d49e6  
msg.payload : Object  
  predictions: array[1]  
25/07/2020, 12:37:01 node: a14653cd.d49e6  
msg.payload : Object  
  predictions: array[1]  
25/07/2020, 12:37:25 node: a14653cd.d49e6  
msg.payload : Object  
  predictions: array[1]  
25/07/2020, 12:40:21 node: a14653cd.d49e6  
msg.payload : Object  
  predictions: array[1]  
25/07/2020, 12:40:24 node: a14653cd.d49e6  
msg.payload : Object  
  predictions: array[1]  
25/07/2020, 12:40:29 node: a14653cd.d49e6
```

9. After the deployment of the model UI can be seen.

node-red-hsbzu.eu-gb.mybluemix.net/ui/#/0?socketid=NBvmIrc9E\_je1EPQAAAW

### Health Insurance Cost Predictor

Default

age \*  
21

sex \*  
male

bmi \*  
10

children \*  
2

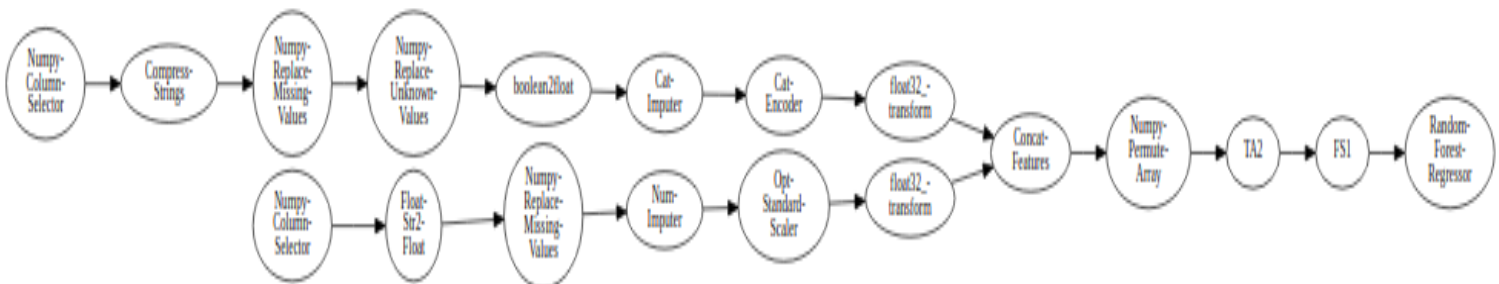
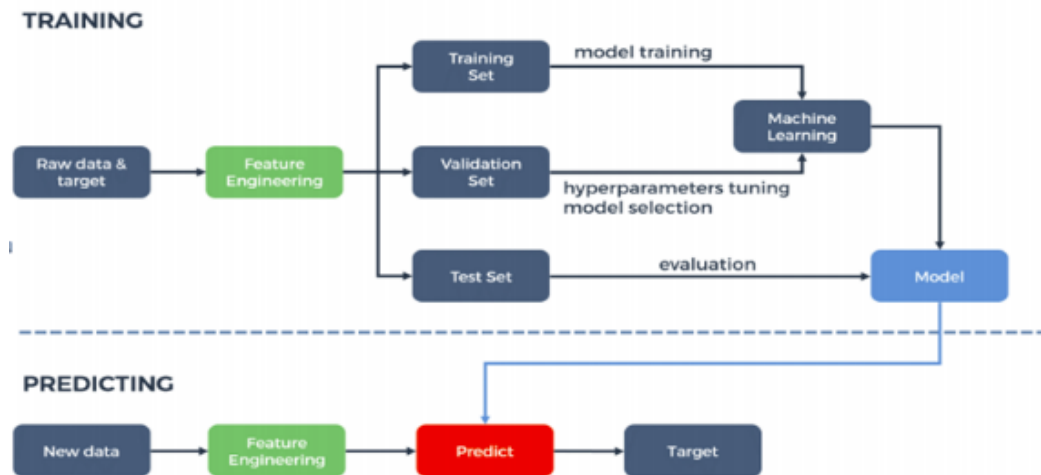
smoker \*  
no

region \*  
southeast

SUBMIT CANCEL

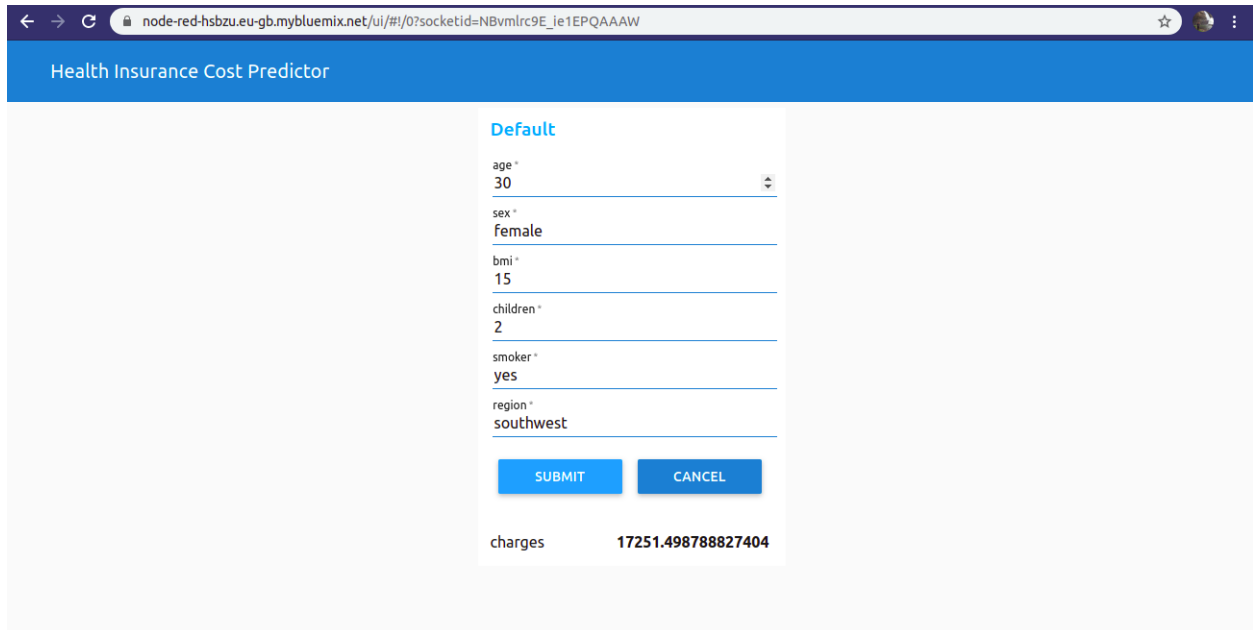
charges 6067.830590203659

## V. FLOWCHART



## **VI. RESULT**

The machine learning model that works on the random forest algorithm predicts the cost of the health insurance package that can be delivered to a customer based on the details of their customer depending upon various factors. A Nodered UI interface which helps to host the model easily on a website and interaction between the model and user in easy and effective.



The screenshot shows a web browser window with the URL `node-red-hsbzu.eu-gb.mybluemix.net/ui/#/0?socketid=NBvmIrc9E_je1EPQAAW`. The page title is "Health Insurance Cost Predictor". The interface features a "Default" form with the following fields and values:

Field	Value
age *	30
sex *	female
bmi *	15
children *	2
smoker *	yes
region *	southwest

Below the form are two buttons: "SUBMIT" and "CANCEL". At the bottom, the predicted "charges" are displayed as **17251.498788827404**.

## **VII. ADVANTAGES AND DISADVANTAGES**

### **ADVANTAGE :**

This model helps the health insurance companies to predict the cost of insurance that can be provided to a particular customer based on factors like age, gender, weight, previous health issues etc.

### **DISADVANTAGE:**

Sometimes the cost requirement may not depend on gender, region they belong to.



## **VIII. APPLICATIONS**

This model that predicts the various health insurance policies applicable to a given customer can be used by the health insurance companies in order to improve their services and can also bring new changes to their policies based on the statistics.

## **IX. CONCLUSION**

This was a great experience with Smartbridge learning new and interesting things and also applying them in innovation field. Related to my project I can say that the machine learning model that is created to predict the cost of health insurance has a wide range of application and makes the work of health insurance companies more simpler. This gives all the predictions just by giving basic details of the customer. I learnt a lot from this project and also thank all the mentors and the bootcamp that was very supportive and helpful at every point of work.

## **X. BIBLIOGRAPHY**

- Smartbridge(SmartInternz) :  
<https://www.thesmartbridge.com/summer-internship-program-2020>
- Kaggle :  
<https://www.kaggle.com/annetxu/health-insurance-cost-prediction>
- IBM Cloud services :  
<https://www.ibm.com/in-en/cloud>
- Node-Red UI :  
[https://node-red-hsbzu.eu-gb.mybluemix.net/ui/#!/0?socketid=NBvmlrc9E\\_ie1EPQAAAW](https://node-red-hsbzu.eu-gb.mybluemix.net/ui/#!/0?socketid=NBvmlrc9E_ie1EPQAAAW)