

CHAPTER – 1

INTRODUCTION

1.1 Problem Definition

Health insurance is a necessity nowadays, and almost every individual is linked with a government or private health insurance company. Factors determining the amount of insurance vary from company to company.

The objective of the proposed work is to predict the insurance premium that an individual pays at the end of the year which helps them to estimate amount that they had to pay for the insurance company.

1.2 Project Overview

The goal of this project is to allow 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.

This project aims at building a web App that automatically estimates premium cost by taking the input values. Using IBM Auto AI, we automate all of the tasks involved in building predictive models for different requirements. You create a model from a data set that includes the age, gender, BMI, number of children, smoking preferences, region, and charges to predict the health insurance premium cost that an individual pays.

CHAPTER – 2

LITERATURE SURVEY

2.1 Existing System

People in rural areas are unaware of the fact that the government of India provide free health insurance to those below poverty line. It is very complex method and some rural people either buy some private health insurance or do not invest money in health insurance at all. Apart from this people can be fooled easily about the amount of the insurance and may unnecessarily buy some expensive health insurance. Also there is lack of accuracy in estimating the premium costs.

2.1 Proposed System

Judicious use of predictive analysis has empowered health insurers to improve their premium pricing accuracy, create customized health insurance plans and services, and build stronger customer relationships.

Thus, the main goal of this project is to predict the insurance premiums based on the behavioural data collected from the individuals so that insurance companies can make useful and accurate predictions.

Based on these predictions, they can then evaluate the following decisions and make better judgement calls:

- Which individuals deserve which kind of insurance plan?
- Based upon an individual's behaviour, predicting their premium helps in better risk management.

Using IBM Auto AI, we automate all of the tasks involved in building predictive models for different requirements. You create a model from a data set that includes the age, gender, BMI, number of children, smoking preferences, region, and charges to predict the health insurance premium cost that an individual pays .

- The user creates an IBM Watson Studio Service, IBM Cloud Object Storage Service on IBM Cloud.
- The user uploads the insurance premium data file into Watson Studio.
- The user creates an Auto AI Experiment to predict an insurance premium on Watson Studio.
- Auto AI uses Machine Learning Service to create several models, and the user deploys the best performing model.
- We use the Node-RED web application to connect to the deployed model and predict the insurance.

2.3 DATA DESCRIPTION

This dataset contains the information on individual attributes such as sex, age, smoking habits etcetera.

It has:

- 1338 rows
- 7 columns

Description of columns:

- age – age of primary beneficiary
- sex – gender of the beneficiary.

It has two categories:

o Male

o Female

- bmi – Body Mass Index, providing an understanding of body weights that are relatively high or low relative to height, objective index of body weight (kg/m^2) using the ratio of height to weight, ideally 18.5 to 24.9

- children – Number of children covered by the health insurance / Number of dependents.
- smoker – describing whether a person is a smoker or a non-smoker.

It has 2 values:

- o Yes
- o No

- region – the beneficiary's residential area in the US.

It has 4 region values:

- o Northeast
- o Southeast
- o Southwest
- o Northwest

- Expenses– Individual insurance premiums billed by health insurance.

Column	Description	Data Type
Age	Primary beneficiary's age	integer
Sex	Beneficiary's gender (male, female)	character
BMI	Body mass index, providing an understanding of body weights that are relatively high or low relative to height, ratio of height to weight	numeric

Children	No of children covered by health insurance/ no of dependents	integer
Smoker	Whether a person is a smoker(yes) or not(no)?	character
Region	Beneficiary's residential area (northeast, northwest, southeast, southwest)	character
Expenses	Individual insurance premiums billed by health insurance	numeric

2.4 Project Methodology

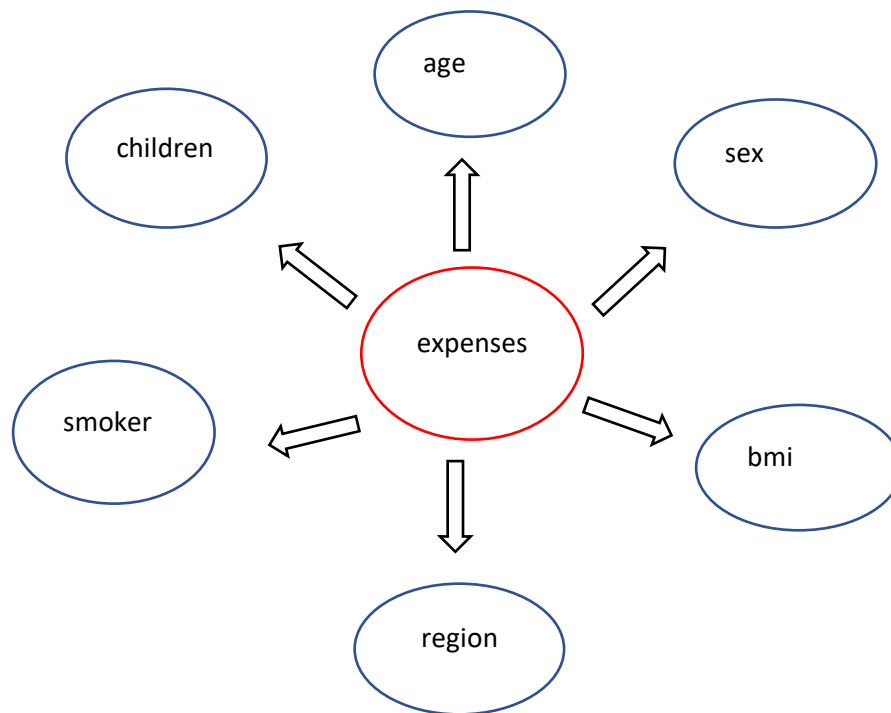
The main goal of the project is to predict the insurance premium charge based upon other attributes.

The dependent variable is Expenses.

There are 6 independent variables:

- age
- sex
- children
- bmi
- smoker
- region

The following figure shows the independent and dependent variables:



CHAPTER-3

SYSTEM ANALYSIS & DESIGN

3.1 Requirement Specification

Software Requirements

- Operating System
- Watson Studio
- Node-Red

Hardware Requirements

RAM : 16GB and Higher

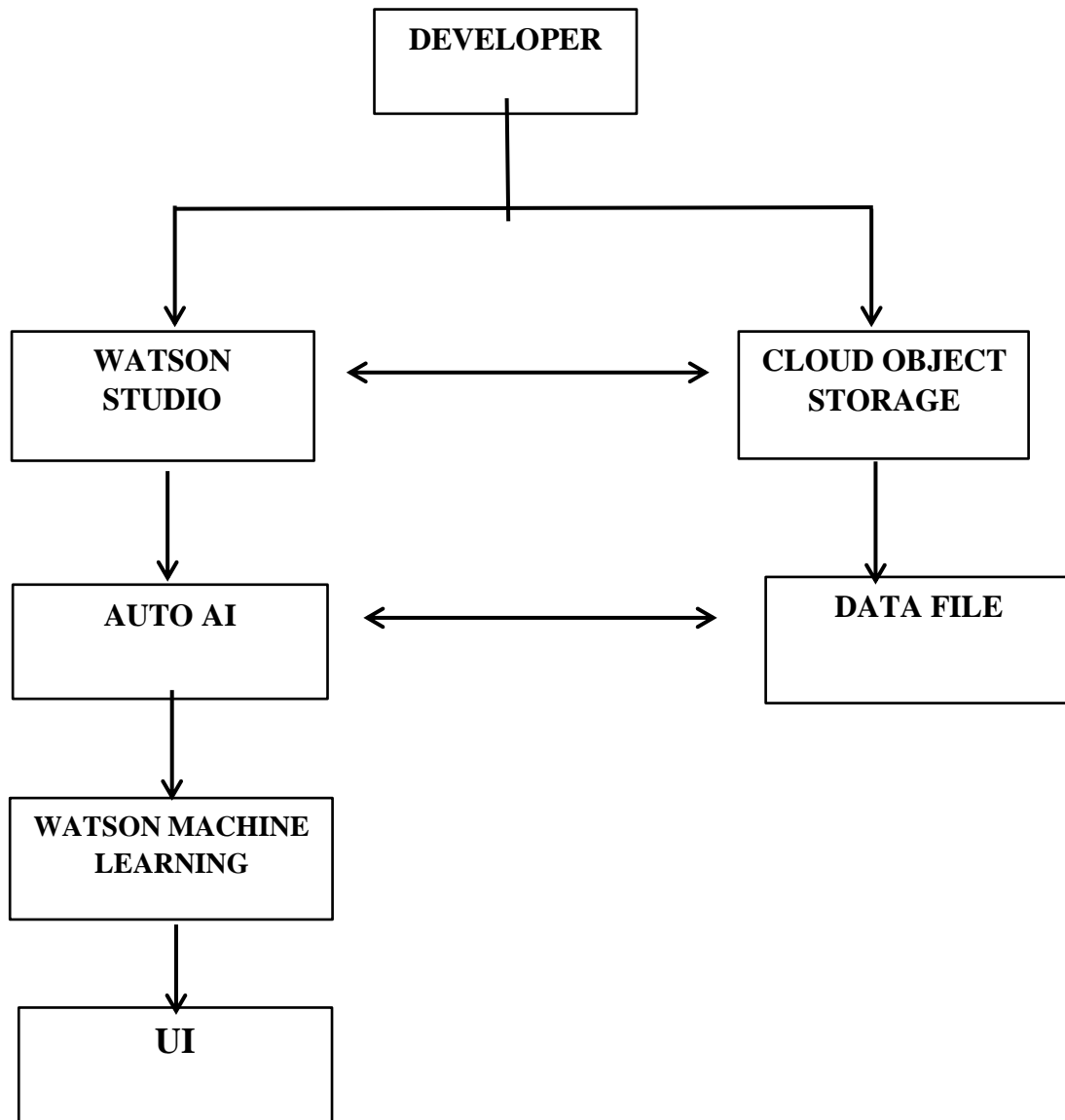
Processor : Intel i5 and above

Hard Disk : 1000GB: Minimum

Machine Learning

Machine learning (ML) is a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning algorithms use historical data as input to predict new output values.

3.2 Design and Architecture



3.3 Algorithms and Sample Code

Algorithms

Gradient Boosting

Gradient Boosting is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees. It builds the model in a stage-wise fashion like other boosting methods do, and it generalizes them by allowing optimization of an arbitrary differentiable loss function.

Random Forests

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean/average prediction (regression) of the individual trees. Random decision forests correct for decision trees habit of overfitting to their training set. Random forests generally outperform decision trees, but their accuracy is lower than gradient boosted trees. However, data characteristics can affect their performance.

Source Code:

```
[  
  {  
    "id": "99ef54bf.e4e7b8",  
    "type": "tab",  
    "label": "Flow 1",  
    "disabled": false,  
    "info": ""  
  },  
  {  
    "id": "d5a31cbe.4ba7e",  
    "type": "tab",  
    "label": "Flow 2",  
    "disabled": false,  
    "info": ""  
  },  
]
```

```
{
  "id": "d8484d88.60c0b",
  "type": "tab",
  "label": "Flow 3",
  "disabled": false,
  "info": ""
},
{
  "id": "7ac7e0a3.b5b36",
  "type": "ui_tab",
  "z": "",
  "name": "Prediction",
  "icon": "dashboard",
  "disabled": false,
  "hidden": false
},
{
  "id": "d4c57ed6.2cf01",
  "type": "ui_base",
  "theme": {
    "name": "theme-light",
    "lightTheme": {
      "default": "#0094CE",
      "baseColor": "#0094CE",
      "baseFont": "-apple-system,BlinkMacSystemFont,Segoe UI,Roboto,Oxygen-Sans,Ubuntu,Cantarell,Helvetica Neue,sans-serif",
      "edited": true,
      "reset": false
    },
    "darkTheme": {
      "default": "#097479",
      "baseColor": "#097479",
```

```
"baseFont": "-apple-system,BlinkMacSystemFont,Segoe UI,Roboto,Oxygen-Sans,Ubuntu,Cantarell,Helvetica Neue,sans-serif",
```

```
"edited": false
```

```
},
```

```
"customTheme": {
```

```
"name": "Untitled Theme 1",
```

```
"default": "#4B7930",
```

```
"baseColor": "#4B7930",
```

```
"baseFont": "-apple-system,BlinkMacSystemFont,Segoe UI,Roboto,Oxygen-Sans,Ubuntu,Cantarell,Helvetica Neue,sans-serif"
```

```
},
```

```
"themeState": {
```

```
"base-color": {
```

```
"default": "#0094CE",
```

```
"value": "#0094CE",
```

```
"edited": false
```

```
},
```

```
"page-titlebar-backgroundColor": {
```

```
"value": "#0094CE",
```

```
"edited": false
```

```
},
```

```
"page-backgroundColor": {
```

```
"value": "#fafafa",
```

```
"edited": false
```

```
},
```

```
"page-sidebar-backgroundColor": {
```

```
"value": "#ffffff",
```

```
"edited": false
```

```
},
```

```
"group-textColor": {
```

```
"value": "#1bbfff",
```

```
"edited": false
```

```
},
```

```
"group-borderColor": {
  "value": "#ffffff",
  "edited": false
},
"group-backgroundColor": {
  "value": "#ffffff",
  "edited": false
},
"widget-textColor": {
  "value": "#111111",
  "edited": false
},
"widget-backgroundColor": {
  "value": "#0094ce",
  "edited": false
},
"widget-borderColor": {
  "value": "#ffffff",
  "edited": false
},
"base-font": {
  "value": "-apple-system,BlinkMacSystemFont,Segoe UI,Roboto,Oxygen-
Sans,Ubuntu,Cantarell,Helvetica Neue,sans-serif"
}
},
"angularTheme": {
  "primary": "indigo",
  "accents": "blue",
  "warn": "red",
  "background": "grey"
}
},
"site": {
```

```
"name": "Node-RED Dashboard",
"hideToolbar": "false",
"allowSwipe": "false",
"lockMenu": "false",
"allowTempTheme": "true",
"dateFormat": "DD/MM/YYYY",
"sizes": {
  "sx": 48,
  "sy": 48,
  "gx": 6,
  "gy": 6,
  "cx": 6,
  "cy": 6,
  "px": 0,
  "py": 0
}
},
{
  "id": "c904f915.dff018",
  "type": "ui_group",
  "z": "",
  "name": "Enter the values",
  "tab": "7ac7e0a3.b5b36",
  "order": 1,
  "disp": true,
  "width": 11,
  "collapse": false
},
{
  "id": "b1b11140.4e4ef",
  "type": "inject",
```

```
"z": "99ef54bf.e4e7b8",
"name": "",
"repeat": "",
"crontab": "",
"once": false,
"topic": "",
"payload": "Hello Node-RED!",
"payloadType": "string",
"x": 141,
"y": 61,
"wires": [
  [
    "f2f2649a.0d0d98"
  ]
],
},
{
  "id": "f2f2649a.0d0d98",
  "type": "debug",
  "z": "99ef54bf.e4e7b8",
  "name": "",
  "active": true,
  "console": "false",
  "complete": "false",
  "x": 401,
  "y": 87,
  "wires": []
},
{
  "id": "2e80c577.137bea",
  "type": "ui_form",
  "z": "d5a31cbe.4ba7e",
```

```
"name": "",
"label": "",
"group": "c904f915.dff018",
"order": 4,
"width": 0,
"height": 0,
"options": [
  {
    "label": "Age",
    "value": "a",
    "type": "number",
    "required": true,
    "rows": null
  },
  {
    "label": "Children",
    "value": "c",
    "type": "number",
    "required": true,
    "rows": null
  },
  {
    "label": "BMI",
    "value": "b",
    "type": "number",
    "required": true,
    "rows": null
  }
],
"formValue": {
  "a": "",
  "c": "",
```

```
    "b": ""
  },
  "payload": "",
  "submit": "submit",
  "cancel": "cancel",
  "topic": "",
  "x": 50,
  "y": 120,
  "wires": [
    [
      "ae0c0868.03e568"
    ]
  ]
},
{
  "id": "f3a24d6c.85a0d",
  "type": "ui_dropdown",
  "z": "d5a31cbe.4ba7e",
  "name": "",
  "label": "Gender",
  "tooltip": "",
  "place": "Select option",
  "group": "c904f915.dff018",
  "order": 1,
  "width": 0,
  "height": 0,
  "passthru": true,
  "multiple": false,
  "options": [
    {
      "label": "Female",
      "value": "female",
```



```
      "type": "str"
    },
    {
      "label": "Male",
      "value": "male",
      "type": "str"
    }
  ],
  "payload": "",
  "topic": "",
  "x": 80,
  "y": 200,
  "wires": [
    [
      "25d635b7.60f72a"
    ]
  ]
},
{
  "id": "5b770323.2d0c0c",
  "type": "ui_dropdown",
  "z": "d5a31cbe.4ba7e",
  "name": "",
  "label": " smoker",
  "tooltip": "",
  "place": "Select option",
  "group": "c904f915.dff018",
  "order": 2,
  "width": 0,
  "height": 0,
  "passthru": true,
  "multiple": false,
```

```
"options": [  
  {  
    "label": "YES",  
    "value": "Yes",  
    "type": "str"  
  },  
  {  
    "label": "NO",  
    "value": "No",  
    "type": "str"  
  }  
],  
"payload": "",  
"topic": "",  
"x": 80,  
"y": 260,  
"wires": [  
  [  
    "c8a30def.5a2ab"  
  ]  
]  
},  
{  
  "id": "cfcad839.1d69d8",  
  "type": "ui_dropdown",  
  "z": "d5a31cbe.4ba7e",  
  "name": "",  
  "label": " Region",  
  "tooltip": "",  
  "place": "Select option",  
  "group": "c904f915.dff018",  
  "order": 3,
```

```
"width": 0,  
"height": 0,  
"passthru": true,  
"multiple": false,  
"options": [  
  {  
    "label": "southwest",  
    "value": "southwest",  
    "type": "str"  
  },  
  {  
    "label": "southeast",  
    "value": "southeast",  
    "type": "str"  
  },  
  {  
    "label": "northwest",  
    "value": "northwest",  
    "type": "str"  
  },  
  {  
    "label": "northeast",  
    "value": "northeast",  
    "type": "str"  
  }  
],  
"payload": "",  
"topic": "",  
"x": 80,  
"y": 320,  
"wires": [  
  [  

```

```

        "363621ae.2bd44e"
    ]
]
},
{
    "id": "ae0c0868.03e568",
    "type": "function",
    "z": "d5a31cbe.4ba7e",
    "name": "global variables for form node",
    "func":
"global.set(\"age\",msg.payload.a)\nglobal.set(\"children\",msg.payload.a)\nglobal.set(\"BMI\",msg.p
ayload.a)\nvar apikey= \"_RBR-
B6V1z17YEvxnny5RloXmZ5dDWmGSVWtYtMj2tvV\" \nmsg.headers={\"content-
type\":\"application/x-www-form-
urlencoded\" }\nmsg.payload={\"grant_type\":\"urn:ibm:params:oauth:grant-
type:apikey\", \"apikey\":apikey}\nreturn msg;",
    "outputs": 1,
    "noerr": 0,
    "initialize": "",
    "finalize": "",
    "x": 310,
    "y": 120,
    "wires": [
        [
            "ed746ca6.148b3",
            "abad1f6b.ac031"
        ]
    ]
},
{
    "id": "c8a30def.5a2ab",
    "type": "function",
    "z": "d5a31cbe.4ba7e",
    "name": "smoker",

```

```
"func": "global.set(\"smoker\",msg.payload)\nreturn msg;",
"outputs": 1,
"noerr": 0,
"initialize": "",
"finalize": "",
"x": 300,
"y": 260,
"wires": [
  [
    "d39d628c.d6384",
    "82ad0635.c17f48"
  ]
]
},
{
  "id": "363621ae.2bd44e",
  "type": "function",
  "z": "d5a31cbe.4ba7e",
  "name": "Region",
  "func": "global.set(\"region\",msg.payload)\nreturn msg;",
  "outputs": 1,
  "noerr": 0,
  "initialize": "",
  "finalize": "",
  "x": 280,
  "y": 320,
  "wires": [
    [
      "629a5067.f19e7"
    ]
  ]
},
```

```
{
  "id": "25d635b7.60f72a",
  "type": "function",
  "z": "d5a31cbe.4ba7e",
  "name": "sex",
  "func": "global.set(\"sex\",msg.payload)\nreturn msg;",
  "outputs": 1,
  "noerr": 0,
  "initialize": "",
  "finalize": "",
  "x": 290,
  "y": 200,
  "wires": [
    [
      "d39d628c.d6384"
    ]
  ]
},
{
  "id": "ed746ca6.148b3",
  "type": "debug",
  "z": "d5a31cbe.4ba7e",
  "name": "",
  "active": false,
  "tosidebar": true,
  "console": false,
  "tostatus": false,
  "complete": "false",
  "statusVal": "",
  "statusType": "auto",
  "x": 510,
  "y": 180,
```

```
"wires": []  
},  
{  
  "id": "d39d628c.d6384",  
  "type": "debug",  
  "z": "d5a31cbe.4ba7e",  
  "name": "",  
  "active": false,  
  "tosidebar": true,  
  "console": false,  
  "tostatus": false,  
  "complete": "false",  
  "statusVal": "",  
  "statusType": "auto",  
  "x": 450,  
  "y": 260,  
  "wires": []  
},  
{  
  "id": "82ad0635.c17f48",  
  "type": "debug",  
  "z": "d5a31cbe.4ba7e",  
  "name": "",  
  "active": false,  
  "tosidebar": true,  
  "console": false,  
  "tostatus": false,  
  "complete": "false",  
  "statusVal": "",  
  "statusType": "auto",  
  "x": 470,  
  "y": 320,
```

```
"wires": []
},
{
  "id": "629a5067.f19e7",
  "type": "debug",
  "z": "d5a31cbe.4ba7e",
  "name": "",
  "active": false,
  "tosidebar": true,
  "console": false,
  "tostatus": false,
  "complete": "false",
  "statusVal": "",
  "statusType": "auto",
  "x": 470,
  "y": 400,
  "wires": []
},
{
  "id": "abad1f6b.ac031",
  "type": "http request",
  "z": "d5a31cbe.4ba7e",
  "name": "",
  "method": "POST",
  "ret": "obj",
  "paytoqs": "ignore",
  "url": "https://iam.cloud.ibm.com/identity/token",
  "tls": "",
  "persist": false,
  "proxy": "",
  "authType": "",
  "x": 610,
```



```
"y": 100,
"wires": [
  [
    "3cc9e9bb.ee4ce6",
    "78e2338f.c1f7ec"
  ]
],
{
  "id": "3cc9e9bb.ee4ce6",
  "type": "debug",
  "z": "d5a31cbe.4ba7e",
  "name": "",
  "active": false,
  "tosidebar": true,
  "console": false,
  "tostatus": false,
  "complete": "false",
  "statusVal": "",
  "statusType": "auto",
  "x": 730,
  "y": 180,
  "wires": []
},
{
  "id": "78e2338f.c1f7ec",
  "type": "function",
  "z": "d5a31cbe.4ba7e",
  "name": "",
  "func": "var age = global.get(\"age\")\nvar sex = global.get(\"sex\")\nvar bmi =\nglobal.get(\"bmi\")\nvar children = global.get(\"children\")\nvar smoker = global.get(\"smoker\")\nvar\nregion = global.get(\"region\")\nvar token=msg.payload.access_token\nmsg.headers={'Content-Type':\n'application/json','Authorization':'Bearer\n\"+token,\"Accept\":\"application/json\"}\nmsg.payload={\"input_data\":[{\"fields\":[\"age\",\"sex\",\""
```

```
bmi\","\children\","\smoker\","\region\","\values\":[[age,sex,bmi,children,smoker,region]]}}\nreturn  
msg;"
```

```
    "outputs": 1,
```

```
    "noerr": 0,
```

```
    "initialize": "",
```

```
    "finalize": "",
```

```
    "x": 850,
```

```
    "y": 120,
```

```
    "wires": [
```

```
        [
```

```
            "c83f37bd.ddfe48"
```

```
        ]
```

```
    ]
```

```
},
```

```
{
```

```
    "id": "c83f37bd.ddfe48",
```

```
    "type": "http request",
```

```
    "z": "d5a31cbe.4ba7e",
```

```
    "name": "",
```

```
    "method": "POST",
```

```
    "ret": "obj",
```

```
    "paytoqs": "ignore",
```

```
    "url": "https://us-south.ml.cloud.ibm.com/ml/v4/deployments/62d266c1-56e5-4b9c-913d-  
b21db7f6c321/predictions?version=2020-12-24",
```

```
    "tls": "",
```

```
    "persist": false,
```

```
    "proxy": "",
```

```
    "authType": "",
```

```
    "x": 990,
```

```
    "y": 120,
```

```
    "wires": [
```

```
        [
```

```
            "d8d0502b.e580c",
```

```
        "ed43eb74.83e908"
    ]
]
},
{
    "id": "d8d0502b.e580c",
    "type": "debug",
    "z": "d5a31cbe.4ba7e",
    "name": "",
    "active": true,
    "tosidebar": true,
    "console": false,
    "tostatus": false,
    "complete": "false",
    "statusVal": "",
    "statusType": "auto",
    "x": 1030,
    "y": 180,
    "wires": []
},
{
    "id": "ed43eb74.83e908",
    "type": "function",
    "z": "d5a31cbe.4ba7e",
    "name": "",
    "func": "msg.payload= msg.payload.predictions[0].values[0][0]\nreturn msg;",
    "outputs": 1,
    "noerr": 0,
    "initialize": "",
    "finalize": "",
    "x": 1200,
    "y": 120,
```

```
"wires": [  
  [  
    "5826d8c0.59da18",  
    "ca7d0094.916ed"  
  ]  
],  
{  
  "id": "5826d8c0.59da18",  
  "type": "debug",  
  "z": "d5a31cbe.4ba7e",  
  "name": "",  
  "active": true,  
  "tosidebar": true,  
  "console": false,  
  "tostatus": false,  
  "complete": "false",  
  "statusVal": "",  
  "statusType": "auto",  
  "x": 1290,  
  "y": 200,  
  "wires": []  
},  
{  
  "id": "ca7d0094.916ed",  
  "type": "ui_text",  
  "z": "d5a31cbe.4ba7e",  
  "group": "c904f915.dff018",  
  "order": 5,  
  "width": 0,  
  "height": 0,  
  "name": ""
```

```
"label": "The premium will be ",  
"format": "{{ msg.payload }}",  
"layout": "row-spread",  
"x": 1460,  
"y": 120,  
"wires": []  
}  
]
```

CHAPTER – 4

OUTPUTS

node-red-qavft-2020-12-24.eu-gb.mybluemix.net/ui/#/0?socketid=1IZN6SxmRtH_5_ySAAAM

Health Insurance Premium Prediction

Enter the values

Gender: Male

smoker: NO

Region: southeast

Age: 12

Children: 0

BMI: 12

The premium will be **2603.4071892768993**

Student Dashboard | Node-RED : node-red-qavft-2020-12-24.eu-gb.mybluemix.net/ui/#/0?socketid=1IZN6SxmRtH_5_ySAAAM

Health Insurance Premium Prediction

Enter the values

Gender: Female

smoker: NO

Region: northwest

Age: 12

Children: 0

BMI: 12

The premium will be **2480.3940685387875**

node-red-qavft-2020-12-24.eu-gb.mybluemix.net/ui/#/0?socketid=1IZN6SxmRTH_5_ySAAAM

Health Insurance Premium Prediction

Enter the values

Gender: Male

smoker: YES

Region: southeast

Age: 50

Children: 2

BMI: 26

SUBMIT **CANCEL**

The premium will be **10578.711713932384**

Type here to search

13:40 04-01-2021

node-red-qavft-2020-12-24.eu-gb.mybluemix.net/ui/#/0?socketid=1IZN6SxmRTH_5_ySAAAM

Health Insurance Premium Prediction

Enter the values

Gender: Female

smoker: NO

Region: southeast

Age: 50

Children: 2

BMI: 26

SUBMIT **CANCEL**

The premium will be **10653.337136000151**

Type here to search

13:44 04-01-2021

CHAPTER 5

CONCLUSION

Based on the findings of the project, it can be said that predictive modeling has tremendous benefits for the health insurance industry in determining how much the premium should be charged to the insured based upon his/her behaviors and health habits. Health insurance companies can then accurately charge the premium based upon a specific individual's attributes. This will not only help the individuals in getting charged the right amount of premium for their health insurance but will also help in forging better relationships and a level of trust between the insurance company and the insured. Based on these predictions, the health insurance providers can then evaluate the following decisions :

- Which individuals deserve which kind of insurance plan?
- How much the premium should be charged based on an individual's behaviours?
- Based upon an individual's behaviour, predicting their premium helps in better risk management.

It helps forge trust between the customer and the insurance company. Thus, it is important for a health insurance company to collect and analyse the data such as a person's age, BMI, health data to accurately predict the risk and charge accurate premiums to cover that risk.

REFERENCES

1. Nyce, Charles (2007), Predictive Analytics White Paper(PDF), American Institute for Chartered Property Casualty Underwriters/Insurance Institute of America
2. Conz, Nathan (September 2, 2008), "Insurers Shift to Customer-focused Predictive Analytics Technologies", Insurance & Technology
3. Rencher, Alvin C.; Christensen, William F. (2012), "Chapter 10, Multivariate regression – Section 10.1, Introduction", Methods of Multivariate Analysis, Wiley Series in Probability and Statistics, 709 (3rd ed.), John Wiley & Sons, p. 19, ISBN 9781118391679.
4. "Linear Regression (Machine Learning)" (PDF). University of Pittsburgh.
5. Ho, Tin Kam (1995). Random Decision Forests (PDF). Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 14–16.