

***Health Insurance Cost prediction Using Watson Auto AI***

*A*

*Internship Project Report*

*By*

**Hema Siva Rama Rayalu .Palivela**

**RSIP Career Basic ML 064**

**Project ID : SPS\_PRO\_286**

|  |  |  |
| --- | --- | --- |
|  |  |  |

**DEPARTMENT OF ELECTRONICS & COMMUNICATION**

**SCHOOL OF ENGINEERING & APPLIED SCIENCE**

**BENNETT UNIVERSITY GREATER NOIDA**

**GREATER NOIDA-201310**

JUNE 2020

**INDEX**

1. ACKNOWLEDGMENT..................................... 3

2. INTRODUCTION......................................................... 4

3. LITERATURE SURVEY.............................................. 5

4. EXPERIMENTAL INVESTIGATION........................ 7

5. FLOW CHART........................................................... 8

6. RESULT...................................................................... 8

7. ADVANTAGES & DISADVANTAGES................... 9

8. APPLICATION.......................................................... 10

9. CONCLUSION.......................................................... 10

10. FUTURE SCOPE ..................................................... 10

11. JSON FILE................................................................ 11

12. NODE RED WEB APPLICATION LINK............... 13

13. REFERENCES.......................................................... 14

**ACKNOWLEDGMENT**

This project has taken a substantial amount of your time and resources. i'd prefer to acknowledge the assistance of all of these who have made this project possible. In ﬁnical i might wish to thank my supervisor Mr. Rammohan Bethi for his time, patience and guidance, and also for allowing the concept to be pursued primitively. i might also wish to thank Mr. Vinay Kumar Nomula for his help. Further to those people i'd prefer to thank the members of the Smartbridge career workshop for his or her technical help in fitting various codes and faults. Also, I would really like to thank all of my co-interns who have worked on the Open Source projects without whose efforts this project wouldn't are possible.

**INTRODUCTION**

The project “Health Insurance Cost prediction” predicts the value of insurance supported some details of person which helps to enhance performance of Insurance companies and to provide better services to the customer.

An important informatics tool for controlling healthcare costs is accurately predicting the likely future healthcare costs of people. to deal with this important need, we conducted a scientific literature review and identified five methods for predicting healthcare costs. To enable an on the spot comparison of those different approaches, we empirically evaluated the predictive performance of every reported approach, still as other state-of-the-art supervised learning methods, using data from University of Utah Health Plans for October 2013 to October 2016. the information set consisted of roughly 90,000 individuals, 6.3 million medical claims and 1.2 million pharmacy claims. during this comparative analysis, gradient boosting had the most effective predictive performance overall and for low to medium cost individuals. for prime cost individuals, Artificial Neural Network (ANN) and also the Ridge regression model, which haven't been previously reported to be used in healthcare cost prediction, had the very best performance.

**LITERATURE SURVEY**

**Existing problem :**

The growing healthcare industry is generating an oversized volume of useful data on patient demographics, treatment plans, payment, and sum attracting the eye of clinicians and scientists alike. In recent years, variety of peer-reviewed articles have addressed different dimensions of information mining application in healthcare. However, the dearth of a comprehensive and systematic narrative motivated us to construct a literature review on this subject.  
With the wide usage of computers and internet, there has recently been an enormous increase in availability of knowledge that may be analyzed. Be it online sales information, website traffic, or user habits, data is generated daily. Such an outsized amount of information presents both an issue and a chance. the matter is that it's difficult for humans to research such large data. the chance is that this kind of information is right and simple for computers to handle and process, because it's stored digitally in an exceedingly well-formatted way and orderly manner, also computers can process data much faster than humans .

**Proposed Solution :**

Health Insurance Cost prediction project consist a Machine Learning model to predict the price using IBM Watson Auto AI Machine Learning Service. The model is deployed on IBM cloud to urge the value which might be used as API in mobile app or web app building. Here, we are going to be developing an internet application which is constructed using node red service.

We make use of the various user input values to the deployed model. The model prediction is then showcased on interface. This model is employed to predict the distribution based by using different input parameters.

This thesis examines the appliance of machine learning algorithms to predict cost in effective manner. Machine learning techniques are often utilized in various areas to predict different values except for insurance cost prediction. Such techniques would help companies to boost their services to customers.

**EXPERIMENTAL INVESTIGATION**

There are six steps in experimental investigation on of a general project:

1. Choose a Project Idea

2. Conduct Background Research

3. Compose a Hypothesis

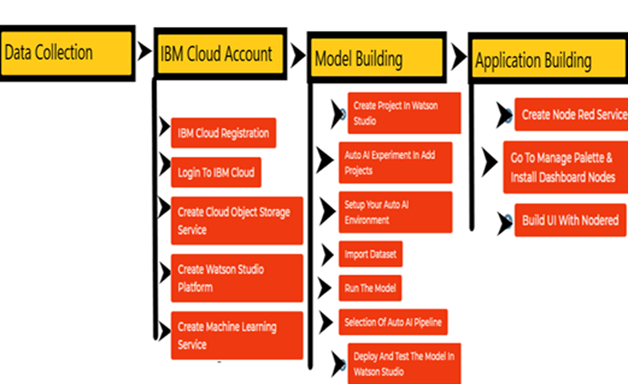
4. Design your Experiment

5. Collect Data

6. Analyse Data and Draw Conclusions

All the data has been collected considering the above factors and it has been formatted. After formatting it has been uploaded in the project and after that using Watson Studio Auto AI Experiment, it is uploaded to cloud object storage service and implemented. Based on these implementation, the value can be predicted using the data we have collected. After that application is developed using Node red Service.

**FLOW CHART**

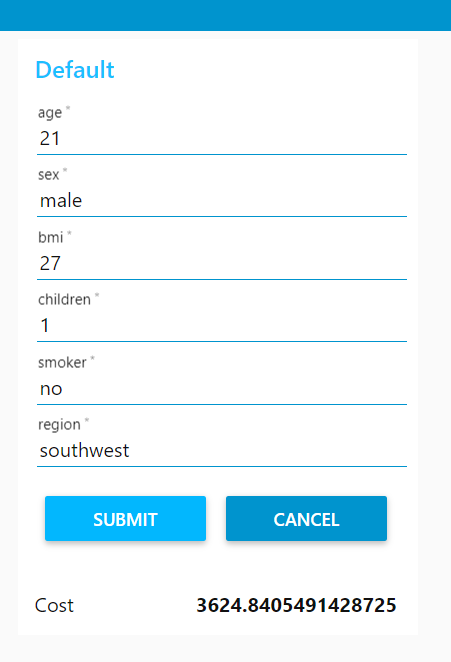


**RESULT**

After the implementation, deployment of project the result i.e. predicted health insurance cost of can be seen in Node Red UI. This value depends on different parameters. The Node Red UI provide us simple way to get the result of Auto AI Experiment.

Here is the Node Red UI predicts the cost of health insurance.

**PREDICTED VALUE:**



**ADVANTAGES & DISADVANTAGES**

The advantages are easy to implement, accessibility is fast, continuous Improvement, wide application, available 24x7, no human intervention needed. We can handle multi-dimensional and multi-variety data.

Where as the disadvantages are lack of security, loss of control on data, high error susceptibility, dependence of network/providers.

**APPLICATION**

Using The Auto AI Experiment, one can build and deploy a machine learning model with sophisticated training features. In the given project we can predict the price of the required vehicle by giving few input parameters.

**CONCLUSION**

In this project by using IBM Cloud the model processing is been done in Auto AI services in IBM cloud and then the deployment is been done in Watson studio and application is build using Node red service which has been successful as we are able to get the desired output.

**FUTURE SCOPE**

As we are developing day to day there is a continuous growth of Auto AI and Machine Learning. The web application can be used to predict the cost of the health insurance accurately, precisely and efficiently instead of n number of people being involved directly or indirectly.

**JSON FILE**

[{"id":"e4302c43.28ff3","type":"ui\_form","z":"6d5f31c.3f75ed","name":"","label":"","group":"f2bfb49.3609c48","order":1,"width":0,"height":0,"options":[{"label":"age","value":"ae","type":"number","required":true,"rows":null},{"label":"sex","value":"sx","type":"text","required":true,"rows":null},{"label":"bmi","value":"bi","type":"number","required":true,"rows":null},{"label":"children","value":"cr","type":"number","required":true,"rows":null},{"label":"smoker","value":"sk","type":"text","required":true,"rows":null},{"label":"region","value":"rg","type":"text","required":true,"rows":null}],"formValue":{"ae":"","sx":"","bi":"","cr":"","sk":"","rg":""},"payload":"","submit":"submit","cancel":"cancel","topic":"","x":171,"y":522.9999952316284,"wires":[["9902b93a.91b438"]]},{"id":"9902b93a.91b438","type":"function","z":"6d5f31c.3f75ed","name":"PreToken","func":"global.set(\"ae\",msg.payload.ae)\nglobal.set(\"sx\",msg.payload.sx)\nglobal.set(\"bi\",msg.payload.bi)\nglobal.set(\"cr\",msg.payload.cr)\nglobal.set(\"sk\",msg.payload.sk)\nglobal.set(\"rg\",msg.payload.rg)\nvar apikey=\"yg7v6CHeo\_FC-NHFOSjZN7\_VanXF-dA8vdsRiGnpyuZ4\";\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,"x":380,"y":480,"wires":[["fddf53c.b65e7b"]]},{"id":"fddf53c.b65e7b","type":"http request","z":"6d5f31c.3f75ed","name":"","method":"POST","ret":"obj","paytoqs":false,"url":"https://iam.cloud.ibm.com/identity/token","tls":"","persist":false,"proxy":"","authType":"","x":579,"y":468.9999952316284,"wires":[["2dfc2d80.67a4e2"]]},{"id":"f355d33b.e3978","type":"debug","z":"6d5f31c.3f75ed","name":"","active":true,"tosidebar":true,"console":false,"tostatus":false,"complete":"payload","targetType":"msg","x":1018.0000114440918,"y":210.99999713897705,"wires":[]},{"id":"2dfc2d80.67a4e2","type":"function","z":"6d5f31c.3f75ed","name":"Pre Prediction","func":"var ae = global.get('ae')\nvar sx = global.get('sx')\nvar bi = global.get('bi')\nvar cr = global.get('cr')\nvar sk = global.get('sk')\nvar rg = global.get('rg')\nvar token=msg.payload.access\_token\nvar instance\_id=\"62581313-25c2-48bc-981d-51395aa4935d\"\nmsg.headers={'Content-Type': 'application/json',\"Authorization\":\"Bearer \"+token,\"ML-Instance-ID\":instance\_id}\nmsg.payload={\"input\_data\": [{\"fields\": [\"age\", \"sex\", \"bmi\", \"children\", \"smoker\", \"region\"], \"values\": [[ae,sx,bi,cr,sk,rg]]}]}\nreturn msg;","outputs":1,"noerr":0,"x":776.0000076293945,"y":431.9999990463257,"wires":[["e84af46d.de7fe8"]]},{"id":"e84af46d.de7fe8","type":"http request","z":"6d5f31c.3f75ed","name":"","method":"POST","ret":"obj","paytoqs":false,"url":"https://us-south.ml.cloud.ibm.com/v4/deployments/30b4a6fb-5726-4a21-ba86-5013d6ffe4e0/predictions","tls":"","persist":false,"proxy":"","authType":"","x":869.5000076293945,"y":375.9999990463257,"wires":[["e16a7fdc.b4857","f355d33b.e3978"]]},{"id":"eb10a3ce.e9eb2","type":"ui\_text","z":"6d5f31c.3f75ed","group":"f2bfb49.3609c48","order":2,"width":0,"height":0,"name":"","label":"Cost","format":"{{msg.payload}}","layout":"row-spread","x":897.6000366210938,"y":528.2000074386597,"wires":[]},{"id":"e16a7fdc.b4857","type":"function","z":"6d5f31c.3f75ed","name":"","func":"msg.payload=msg.payload.predictions[0].values[0][0]\nreturn msg;","outputs":1,"noerr":0,"x":729.6000366210938,"y":271.60000133514404,"wires":[["eb10a3ce.e9eb2"]]},{"id":"f2bfb49.3609c48","type":"ui\_group","z":"","name":"Default","tab":"b3c1b03.83e7e5","order":1,"disp":true,"width":"6","collapse":false},{"id":"b3c1b03.83e7e5","type":"ui\_tab","z":"","name":"Home","icon":"dashboard","disabled":false,"hidden":false}]

**Node Red Web Application Link**

<https://node-red-zrqfk.eu-gb.mybluemix.net/ui>

**References**

1. The Centers for Medicare & Medicaid Services (CMS) DoHaHS, United States. National Health Expenditure Data 2016. Available from: [https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics- Trends-and-Reports/NationalHealthExpendData/index.html](https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/index.html).

2. Duncan I, Loginov M, Ludkovski M. Testing Alternative Regression Frameworks for Predictive Modeling of Health Care Costs. North American Actuarial Journal. 2016;20(1):65–87.