

***Project Report***

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

*By: Kanuru Thanishq*

*RSIP Career Basic ML 028*

**INDEX**

Project Details

Acknowledgment

1. INTRODUCTION

2. LITERATURE SURVEY

a. Existing problem

b. Proposed solution

3. THEORETICAL ANALYSIS

a. Block diagram

b. Hardware / Software designing

4. EXPERIMENTAL INVESTIGATION

5. FLOW CHART

6. RESULT

7. ADVANTAGES & DISADVANTAGES

8. APPLICATION

9. CONCLUSION

10. FUTURE SCOPE

11. BIBLIOGRAPHY

APPENDIX

**Project Details**

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

**Project Title** : Health Insurance Cost prediction Using Watson Auto AI

**Duration** : 4 Weeks

**Project Support** : SmartBridge Educational Services

**Project Mentor** : Mr. Rammohan Bethi

**Kickoff Date** : June 1st, 2020

**Finish Date** : June 30th, 2020

**ACKNOWLEDGEMENT**

This project has taken a considerable amount of time and resources. I would like to acknowledge the help of all of those who have made this project possible. In ﬁnical I would like to thank my supervisor Mr. Rammohan Bethi for his time, patience and guidance, and also for allowing the idea to be pursued primitively. I would also like to thank Mr. Vinay Kumar Nomula for his help. Further to these people I would like to thank the members of the Smartbridge career workshop for their technical help in setting up various codes and faults. Also, I would like to thank all of my co-interns who have worked on the Open Source projects without whose efforts this project would not have been possible.

1. **INTRODUCTION**

The project “Health Insurance Cost prediction” predicts the cost of health insurance based on some details of person which helps to improve 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 individuals. To address this important need, we conducted a systematic literature review and identified five methods for predicting healthcare costs. To enable a direct comparison of these different approaches, we empirically evaluated the predictive performance of each reported approach, as well as other state-of-the-art supervised learning methods, using data from University of Utah Health Plans for October 2013 to October 2016. The data set consisted of approximately 90,000 individuals, 6.3 million medical claims and 1.2 million pharmacy claims. In this comparative analysis, gradient boosting had the best predictive performance overall and for low to medium cost individuals. For high cost individuals, Artificial Neural Network (ANN) and the Ridge regression model, which have not been previously reported for use in healthcare cost prediction, had the highest performance.

**2. LITERATURE SURVEY**

**a. Existing problem**

The growing healthcare industry is generating a large volume of useful data on patient demographics, treatment plans, payment, and insurance coverage attracting the attention of clinicians and scientists alike. In recent years, a number of peer-reviewed articles have addressed different dimensions of data mining application in healthcare. However, the lack of a comprehensive and systematic narrative motivated us to construct a literature review on this topic.

With the wide usage of computers and internet, there has recently been a huge increase in availability of data that can be analyzed. Be it online sales information, website traffic, or user habits, data is generated every day. Such a large amount of data presents both a problem and an opportunity. The problem is that it is difficult for humans to analyze such large data. The opportunity is that this type of data is ideal and easy for computers to handle and process, because it is stored digitally in a well-formatted way and orderly manner, also computers can process data much faster than humans.

**2.b. Proposed Solution**

Health Insurance Cost prediction project consist a Machine Learning model to predict the cost using IBM Watson Auto AI Machine Learning Service. The model is deployed on IBM cloud to get the cost which can be used as API in mobile app or web app building. Here, we will be developing a web application which is built using node red service.

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

This thesis examines the application of machine learning algorithms to predict cost in effective manner. Machine learning techniques can be utilized in various areas to predict different values apart from health insurance cost prediction. Such techniques would help companies to improve their services to customers.

**3. THEORETICAL ANALYSIS**

**a. Block diagram**



**b. Hardware / Software designing**

The project has been done by using IBM Cloud in which machine learning service, Watson studio and cloud storage service (to store the data) have been created by using the options available in Catlog.

**4. 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

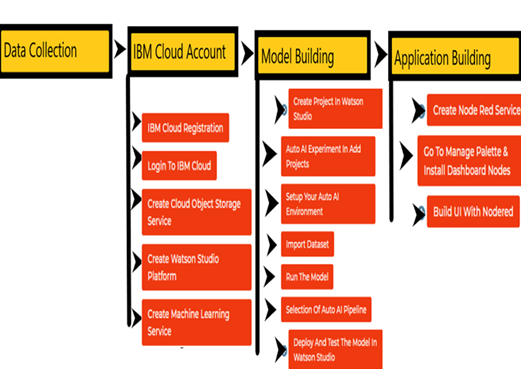
4. Design your Experiment

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.

**5. FLOW CHART**

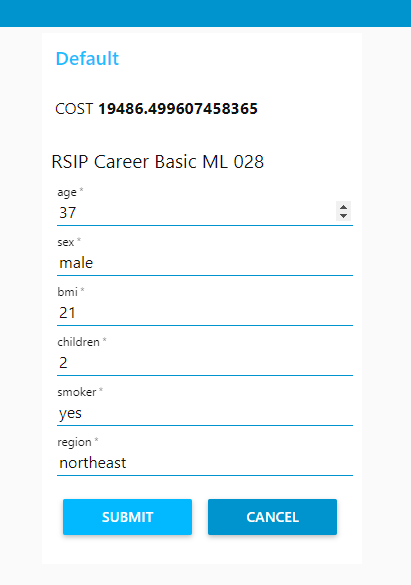


**6. RESULT**

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:



**7. 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.

**8. 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.

**9. 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.

**10. 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.

**11. BIBLIOGRAPH**

**APPENDIX**

A. Source Code

[{"id":"d7e90de0.10f78","type":"tab","label":"Resale value preditcion ","disabled":false,"info":""},{"id":"caa9fa4.22aca08","type":"ui\_form","z":"d7e90de0.10f78","name":"","label":"","group":"e9be4b44.04ea98","order":4,"width":0,"height":0,"options":[{"label":"DateCrawled","value":"dcc","type":"text","required":true,"rows":null},{"label":"Name","value":"nm","type":"text","required":true,"rows":null},{"label":"Seller","value":"se","type":"text","required":true,"rows":null},{"label":"OfferType","value":"ot","type":"text","required":true,"rows":null},{"label":"Abtest","value":"ab","type":"text","required":true,"rows":null},{"label":"VehicleType","value":"vt","type":"text","required":true,"rows":null},{"label":"YearOfRegistration","value":"yor","type":"number","required":true,"rows":null},{"label":"Gearbox","value":"gb","type":"text","required":true,"rows":null},{"label":"PowerPS","value":"pps","type":"number","required":true,"rows":null},{"label":"Model","value":"mo","type":"text","required":true,"rows":null},{"label":"Kilometer","value":"km","type":"number","required":true,"rows":null},{"label":"MonthOfRegistration","value":"mor","type":"number","required":true,"rows":null},{"label":"FuelType","value":"ft","type":"text","required":true,"rows":null},{"label":"Brand","value":"br","type":"text","required":true,"rows":null},{"label":"NotRepairedDamage","value":"nrd","type":"text","required":true,"rows":null},{"label":"DateCreated","value":"dc","type":"text","required":true,"rows":null},{"label":"NrOfPictures","value":"nop","type":"number","required":true,"rows":null},{"label":"PostalCode","value":"pc","type":"number","required":true,"rows":null},{"label":"LastSeen","value":"ls","type":"text","required":true,"rows":null}],"formValue":{"dcc":"","nm":"","se":"","ot":"","ab":"","vt":"","yor":"","gb":"","pps":"","mo":"","km":"","mor":"","ft":"","br":"","nrd":"","dc":"","nop":"","pc":"","ls":""},"payload":"","submit":"submit","cancel":"cancel","topic":"","x":70,"y":280,"wires":[["5203133f.39fe7c"]]},{"id":"c1efeac1.436418","type":"inject","z":"d7e90de0.10f78","name":"","topic":"","payload":"","payloadType":"date","repeat":"","crontab":"","once":false,"onceDelay":0.1,"x":120,"y":160,"wires":[["5203133f.39fe7c"]]},{"id":"5203133f.39fe7c","type":"function","z":"d7e90de0.10f78","name":"","func":"global.set(\"dcc\",msg.payload.dcc)\nglobal.set(\"nm\",msg.payload.nm)\nglobal.set(\"se\",msg.payload.se)\nglobal.set(\"ot\",msg.payload.ot)\nglobal.set(\"ab\",msg.payload.ab)\nglobal.set(\"vt\",msg.payload.vt)\nglobal.set(\"yor\",msg.payload.yor)\nglobal.set(\"gb\",msg.payload.gb)\nglobal.set(\"pps\",msg.payload.pps)\nglobal.set(\"mo\",msg.payload.mo)\nglobal.set(\"km\",msg.payload.km)\nglobal.set(\"mor\",msg.payload.mor)\nglobal.set(\"ft\",msg.payload.ft)\nglobal.set(\"br\",msg.payload.br)\nglobal.set(\"nrd\",msg.payload.nrd)\nglobal.set(\"dc\",msg.payload.dc)\nglobal.set(\"nop\",msg.payload.nop)\nglobal.set(\"pc\",msg.payload.pc)\nglobal.set(\"ls\",msg.payload.ls)\nvar apikey=\"V0rssd-qXs7KiARctxybiJ\_hixXPlZCZypUY84FMrtKq\";\nmsg.headers={\"content-type\":\"application/x-www-form-urlencoded\"}\nmsg.payload={\"grant\_type\":\"urn:ibm:params:oauth:grant-type:apikey\",\"apikey\":apikey}\nreturn msg;\n","outputs":1,"noerr":0,"x":230,"y":260,"wires":[["6052ffe8.1c4cf"]]},{"id":"6052ffe8.1c4cf","type":"http request","z":"d7e90de0.10f78","name":"","method":"POST","ret":"obj","paytoqs":false,"url":"https://iam.cloud.ibm.com/identity/token","tls":"","persist":false,"proxy":"","authType":"","x":230,"y":400,"wires":[["814f5f66.bc05b","ff759d15.76a83"]]},{"id":"814f5f66.bc05b","type":"function","z":"d7e90de0.10f78","name":"","func":"var dcc=global.get('dcc')\nvar nm = global.get('nm')\nvar se = global.get('se')\nvar ot = global.get('ot')\nvar ab = global.get('ab')\nvar vt = global.get('vt')\nvar yor = global.get('yor')\nvar gb = global.get('gb')\nvar pps = global.get('pps')\nvar mo = global.get('mo')\nvar km = global.get('km')\nvar mor = global.get('mor')\nvar ft = global.get('ft')\nvar br = global.get('br')\nvar nrd = global.get('nrd')\nvar dc = global.get('dc')\nvar nop = global.get('nop')\nvar pc = global.get('pc')\nvar ls = global.get('ls')\nvar token=msg.payload.access\_token\nvar instance\_id=\"53b45fb4-d437-4c6a-86d5-56f4221dc67d\"\nmsg.headers={'Content-Type': 'application/json',\"Authorization\":\"Bearer \"+token,\"ML-Instance-ID\":instance\_id}\nmsg.payload={\"input\_data\": [{\"fields\": [\"DateCrawled\", \"Name\", \"Seller\",\"OfferType\",\"Abtest\",\"VehicleType\",\"YearOfRegistration\",\"Gearbox\",\"PowerPS\",\"Model\",\"Kilometer\",\"MonthOfRegistration\",\"FuelType\",\"Brand\",\"NotRepairedDamage\",\"DateCreated\",\"NrOfPictures\",\"PostalCode\",\"LastSeen\"], \"values\": [[dcc,nm,se,ot,ab,vt,yor,gb,pps,mo,km,mor,ft,br,nrd,dc,nop,pc,ls]]}]}\nreturn msg;","outputs":1,"noerr":0,"x":410,"y":180,"wires":[["45451549.6c45ac"]]},{"id":"45451549.6c45ac","type":"http request","z":"d7e90de0.10f78","name":"","method":"POST","ret":"obj","paytoqs":false,"url":"https://eu-gb.ml.cloud.ibm.com/v4/deployments/30e4d641-9353-48ae-9591-e4fb44d4b18e/predictions","tls":"","persist":false,"proxy":"","authType":"","x":450,"y":320,"wires":[["d582f960.786778","7446c0c5.a7458"]]},{"id":"20e12bca.b526b4","type":"debug","z":"d7e90de0.10f78","name":"","active":true,"tosidebar":true,"console":false,"tostatus":false,"complete":"false","x":670,"y":240,"wires":[]},{"id":"d582f960.786778","type":"function","z":"d7e90de0.10f78","name":"","func":"msg.payload = msg.payload.predictions[0].values[0][0]\nreturn msg;\n","outputs":1,"noerr":0,"x":550,"y":140,"wires":[["b41d8a83.4827c8","20e12bca.b526b4"]]},{"id":"b41d8a83.4827c8","type":"ui\_text","z":"d7e90de0.10f78","group":"e9be4b44.04ea98","order":5,"width":0,"height":0,"name":"","label":"Price","format":"{{msg.payload}}","layout":"row-left","x":750,"y":60,"wires":[]},{"id":"ff759d15.76a83","type":"debug","z":"d7e90de0.10f78","name":"","active":true,"tosidebar":true,"console":false,"tostatus":false,"complete":"false","x":460,"y":420,"wires":[]},{"id":"7446c0c5.a7458","type":"debug","z":"d7e90de0.10f78","name":"","active":true,"tosidebar":true,"console":false,"tostatus":false,"complete":"false","x":720,"y":320,"wires":[]},{"id":"e9be4b44.04ea98","type":"ui\_group","z":"","name":"Resale value preditcion ","tab":"985891c7.4ddb4","order":1,"disp":true,"width":"8","collapse":false},{"id":"985891c7.4ddb4","type":"ui\_tab","z":"","name":"Resale Value Prediction","icon":"dashboard","disabled":false,"hidden":false}]

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.

3. Burwell SM. Setting value-based payment goals--HHS efforts to improve US health care. 2015