PROJECT REPORT

Topic: Predicting Life Expectancy using Machine Learning

Submitted by:

Toyada Pattanaik Silicon Institute of Technology

CONTENT

1	INTRODUCTION
	1.1 Overview
	1.2 Purpose
2	LITERATURE SURVEY
	2.1 Existing Problem
	2.2 Proposed Solution
3	THEORETICAL ANALYSIS
	3.1 Block Diagram
	3.2 Hardware/Software Designing
4	EXPERIMENTAL INVESTIGATIONS
5	FLOWCHART
6	RESULT
7	ADVANTAGES AND DISADVANTAGES
8	APPLICATIONS
9	CONCLUSION
10	FUTURE SCOPE
11	BIBLIOGRAPHY
	A Source Code

1 INTRODUCTION

1.1 Overview

This is a typical Regression Machine Learning project that leverages historical data to predict insights into the future. This project is aimed at predicting Life Expectancy rate of a country given various features.

Life expectancy is a statistical measure of the average time a human being is expected to live, Life expectancy depends on various factors: Regional variations, Economic Circumstances, Sex Differences, Mental Illnesses, Physical Illnesses, Education, Year of their birth and other demographic factors. The project provides a way to predict average life expectancy of people living in a country when various factors such as year, GDP, education, alcohol intake of people in the country, expenditure on healthcare system and some specific disease related deaths that happened in the country are given.

1.2 Purpose

The project aims to train a machine learning model on a set of historical data in order to predict the life expectancy of a country when the various parameters affecting it are changed or randomized. It enables the user to input values for different variables and visualize how it affects the life expectancy.

2 LITERATURE SURVEY

2.1 Existing Problem

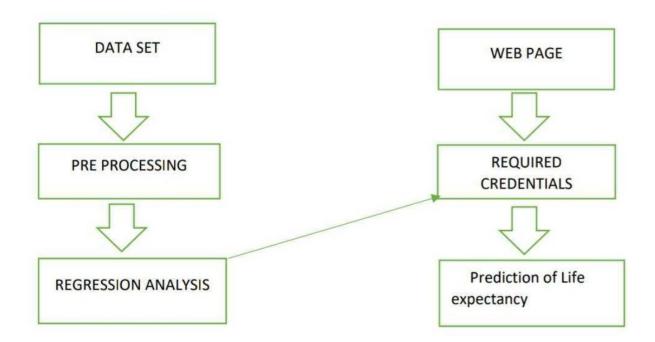
Life expectancy is a statistical measure of the average time a human being is expected to live, Life expectancy depends on various factors: Regional variations, Economic Circumstances, Sex Differences, Mental Illnesses, Physical Illnesses, Education, Year of their birth and other demographic factors. We need to know how one specific factor among these affects life expectancy.

2.2 Proposed Solution

Our project aims to solve this problem by performing Regression Analysis on the historical dataset provided by WHO. It will enable us to understand how life expectancy depends on the various factors specified in the dataset. It will be beneficial so as to direct our focus on the factors which greatly affect life expectancy.

3 THEORETICAL ANALYSIS

3.1 Block Diagram



3.2 Hardware/Software Designing

To complete my project, I have used,

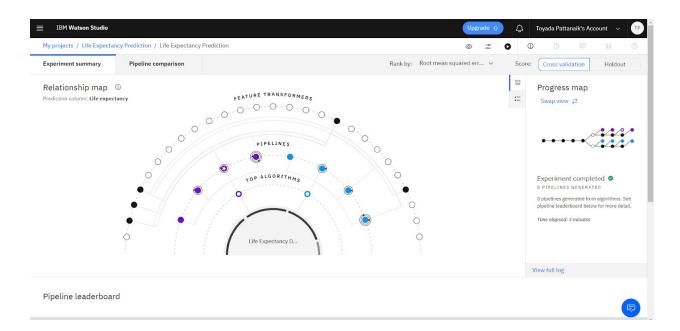
- IBM Cloud
- IBM Watson
- Node-Red

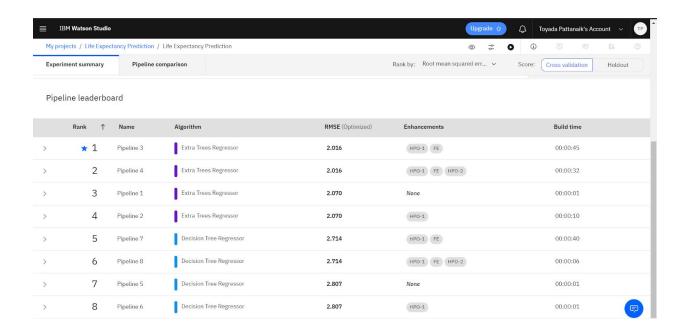
IBM Cloud was used to store the dataset and the machine learning services used in the project. Regression Analysis was carried out using the Auto AI feature of IBM Watson Studio. Node-Red, based on JSON, was used to create the web page which integrated the machine learning model to show results for the user inputted values.

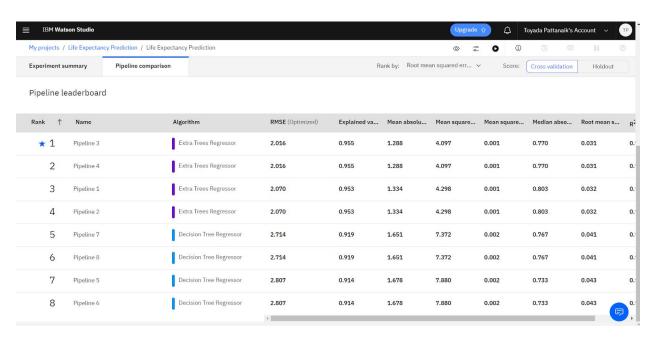
4 EXPERIMENTAL INVESTIGATIONS

We carried our Regression Analysis on the dataset in order to determine the dependency of Life Expectancy on various factors. In statistical modeling, regression analysis is a set of statistical processes for estimating the relationships between a dependent variable and one or more independent variables.

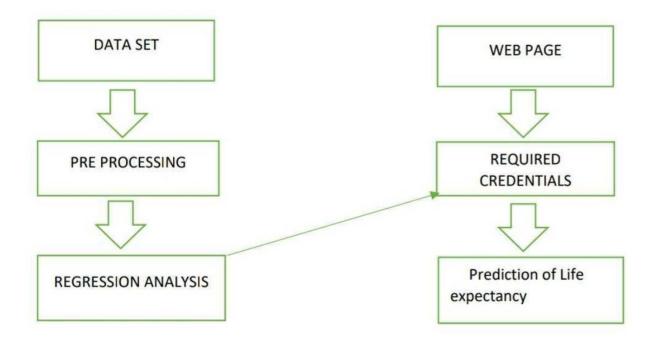
Regression analysis was performed using the Auto AI feature in the Watson Studio which helps automate the coding process.





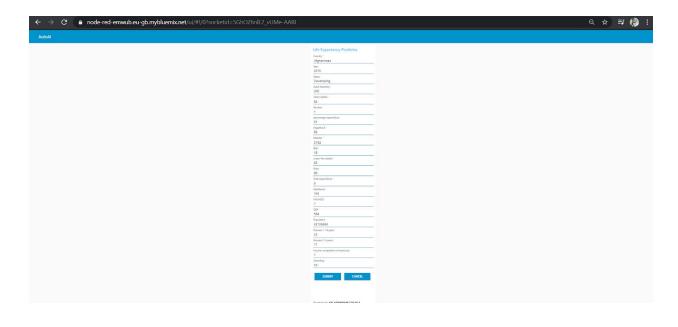


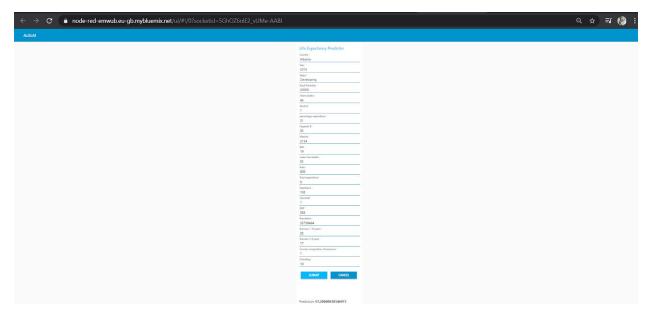
5 FLOWCHART



6 RESULT

At the end of the project, we had a simple web page with a form where users can enter random values for different variables affecting life expectancy and can visualize how all of them affect it differently. The Node-Red flow takes the inputted values and sends it to the machine learning model to process it. After processing, the result is sent back which is displayed below the form.





7 ADVANTAGES AND DISADVANTAGES

Advantages:

• It helps to provide some idea of where future data points will be. It enables a way to see into the future.

• It helps us to understand life expectancy's degree of dependency on various factors specified on the dataset.

Disadvantages:

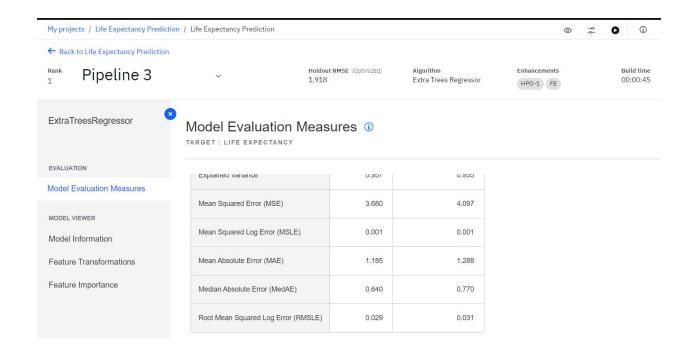
- It is limited to the linear relationship.
- Also, the regression solution is dense (because no regularization is applied).
- It is subject to overfitting.
- It is easily affected by outliers.

8 APPLICATIONS

- It can be used to predict life expectancy in future scenarios when factors will vary and dependency will change.
- It will help to direct focus towards resolving diseases which have a greater impact on life expectancy.
- Planning and management of data can be done more efficiently to produce better results.

9 CONCLUSION

After performing regression analysis using Auto AI, we conclude that Pipeline 3 is the most optimised method to predict life expectancy. The evaluation measures of our model is shown below,



10 FUTURE SCOPE

Machine Learning has huge potential to predict life expectancy in future. As we go forward, we can train our model on a massive scale of data which will in turn generate more accurate predictions. We can develop softwares which will evolve with time along with the ever changing factors.

In the end, we will have an advanced flexible system which will be able to predict the life expectancy of an individual or a country when specific values are provided.

11 BIBLIOGRAPHY

A. Source Code

Python Notebook of ML Model:

JSON Script of Node-Red flow:

```
[{"id":"7f4dbbe6.6c04f4","type":"tab","label":"Life
                                                                                    Expectancy
                                                                                                                Prediction
                                                                                                                                          using
                                                                                                                                                              Auto
AI", "disabled": false, "info": ""}, {"id": "10c3b9f3.5185a6", "type": "ui tab", "z": "", "name": "Home", "icon": "dashboard",
"disabled":false, "hidden":false \ , \"id":"50805e17.a9c9d", "type":"ui group", "z":"", "name":"Life
Predictor", "tab": "10c3b9f3.5185a6", "order": 2, "disp": true, "width": "6", "collapse": false }, {"id": "b05e5433.ba3db8", "t
ype":"ui base","theme":{"name":"theme-light","lightTheme":{"default":"#0094CE","baseColor":"#0094CE","base
Font":"-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-titleb".
ar-backgroundColor":{"value":"#0094CE","edited":false},"page-backgroundColor":{"value":"#fafafa","edited":false
e}, "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-backgroundColor": {"value": false}, "widget-backgroundColor: {"value": false}, "widget-backgroundColor: {"value": false}, "widget-backgroundColor: {"value": false}, "widget-backgroundColor: {"value": false}, "wid
d":false}, "widget-borderColor": {"value": "#ffffff", "edited":false}, "base-font": {"value": "-apple-system, BlinkMacSys
temFont,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":"d6500cd.4fb42f",
"type":"ui group","z":"","name":"Life
Predictor", "tab": "5475484d.671f98", "order": 2, "disp": true, "width": "6", "collapse": false }, { "id": "5475484d.671f98", "t
ype":"ui tab","z":"","name":"AutoAI","icon":"dashboard","order":2,"disabled":false,"hidden":false},{"id":"1d648c
89.cc6c93", "type": "function", "z": "7f4dbbe6.6c04f4", "name": "Pre
token","func":"global.set(\"co\",msg.payload.co)\nglobal.set(\"ve\",msg.payload.ye)\nglobal.set(\"st\",msg.payload.s
t)\nglobal.set(\"adult\\",msg.payload.adult)\nglobal.set(\"infant\",msg.payload.infant)\nglobal.set(\"alc\\",msg.payload
.alc)\nglobal.set(\"per exp\\",msg.payload.per exp)\nglobal.set(\"hepB\",msg.payload.hepB)\nglobal.set(\"meas\",ms
g.payload.meas)\nglobal.set(\"bmi\",msg.payload.bmi)\nglobal.set(\"under\",msg.payload.under)\nglobal.set(\"polio\
",msg.payload.polio)\nglobal.set(\"tot exp\",msg.payload.tot exp)\nglobal.set(\"diph\",msg.payload.diph)\nglobal.se
t(\"hiv\",msg,payload.hiv)\nglobal.set(\"gdp\",msg,payload.gdp)\nglobal.set(\"popu\",msg,payload.popu)\nglobal.set
("thin1\",msg.payload.thin1)\nglobal.set(\"thin5\",msg.payload.thin5)\nglobal.set(\"income\",msg.payload.income)\
nglobal.set(\"school\",msg.payload.school)\nvar
apikey=\"KjaZGdmAHJl5ECDEa3br-Zw81bTBq1o1vZx3MBeKhu8I\";\nmsg.headers={\"content-type\":\"applicati
on/x-www-form-urlencoded"\nmsg.payload={\"grant_type\":\"urn.ibm:params:oauth:grant-type:apikey\",\"apikey\
":apikey}\nreturn
msg;","outputs":1,"noerr":0,"x":260,"y":280,"wires":[["a89041c3.5ab2d"]]}, {"id":"a89041c3.5ab2d","type":"http
request","z":"7f4dbbe6.6c04f4","name":"","method":"POST","ret":"obj","paytoqs":false,"url":"https://iam.cloud.ib
m.com/identity/token","tls":"","persist":false,"proxy":"","authType":"","x":430,"y":280,"wires":[["1d9b245f.737dec
","b599d5fc.31f438"]]},{"id":"9a43f05f.d48b9","type":"debug","z":"7f4dbbe6.6c04f4","name":"","active":true,"tosi
debar":true, "console":false, "tostatus":false, "complete": "payload", "targetType": "msg", "x":890, "y":60, "wires": []}, {"i
d":"1d9b245f.737dec","type":"function","z":"7f4dbbe6.6c04f4","name":"sendto
                                                                                                                   Endpoint", "func": "var
global.get('co')\nvar ye = global.get('ye')\nvar st = global.get('st')\nvar adult = global.get('adult')\nvar infant =
global.get('infant')\nvar alc = global.get('alc')\nvar per exp = global.get('per exp')\nvar hepB
```

global.get('hepB')\nvar meas = global.get('meas')\nvar bmi = global.get('bmi')\nvar under = global.get('under')\nvar polio = global.get('polio')\nvar tot exp = global.get('tot exp')\nvar diph = global.get('diph')\nvar hiv = global.get('hiv')\nvar gdp = global.get('gdp')\nvar popu = global.get('popu')\nvar thin1 = global.get('thin1')\nvar thin5 = global.get('thin5')\nvar income = global.get('income')\nvar school = global.get('school')\nvar token=msg.payload.access token\nvar instance $id=\"919d0abc-349b-4648-a6fb-ee98423bc04e\"\nmsg.headers=\{'Content-Type': a content-Type': a cont$ 'application/json',\"Authorization\":\"Bearer \"+token,\"ML-Instance-ID\":instance id\\nmsg.payload=\\"input data\\": [\\"fields\\": [\\"Country\\\", \\"Year\\\\, \"Status\", \"Adult Mortality\", \"infant deaths\", \"Alcohol\", \"percentage expenditure\", \"Hepatitis B\", \"Measles \", \" BMI \", \"under-five deaths \", \"Polio\", \"Total expenditure\", \"Diphtheria \", \" HIV/AIDS\", \"GDP\", \"Population\", \" thinness 1-19 years\", \" thinness 5-9 years\", \"Income composition of resources\", \"Schooling\"], [[co,ye,st,adult,infant,alc,per exp,hepB,meas,bmi,under,polio,tot exp,diph,hiv,gdp,popu,thin1,thin5,income,school]]}]}\nreturn msg;","outputs":1,"noerr":0,"x":540,"y":200,"wires":[["c9ac629d.203e8"]]},{"id":"c9ac629d.203e8","type":"http request","z":"7f4dbbe6.6c04f4","name":"","method":"POST","ret":"obj","paytoqs":false,"url":"https://eu-gb.ml.clou d.ibm.com/v4/deployments/ee8233e7-b357-4733-9e77-fcc2394a3492/predictions", "tls":"", "persist": false, "proxy":"", "authType":"","x":710,"y":300,"wires":[["68b090c1.1a8b5","bd68dc9a.8971c"]]},{"id":"54923a3.f6d56c4","type":" ui form", "z": "7f4dbbe6.6c04f4", "name": "", "label": "", "group": "d6500cd.4fb42f", "order": 1, "width": 0, "height": 0, "opt ions":[{"label":"Country", "value":"co", "type":"text", "required":true, "rows":null}, {"label":"Year", "value":"ye", "type ":"number", "required":true, "rows":null \, \{"label": "Status", "value": "st", "type": "text", "required":true, "rows":null \, \{"l abel":"Adult Mortality", "value": "adult", "type": "number", "required": true, "rows": null}, { "label": "infant ber", "required": true, "rows": null}, {"label": "percentage expenditure", "value": "per exp", "type": "number", "required": true, "rows": null }, { "label": "Hepatitis B", "value": "hepB", "type": "number", "required": true, "rows": null}, { "label": "Measles", "value": "meas", "type": "number ","required":true,"rows":null},{"label":"BMI","value":"bmi","type":"number","required":true,"rows":null},{"label": "under-five deaths", "value": "under", "type": "number", "required": true, "rows": null}, { "label": "Polio", "value": "polio", "type": "num ber", "required": true, "rows": null \}, \{ "label": "Total

expenditure", "value": "tot_exp", "type": "number", "required": true, "rows": null}, { "label": "Diphtheria ", "value": "diph", "type": "number", "required": true, "rows": null}, { "label": "

msg;\n","outputs":1,"noerr":0,"x":670,"y":80,"wires":[["9a43f05f.d48b9","f502ea73.da9788"]]},{"id":"b599d5fc.31 f438","type":"debug","z":"7f4dbbe6.6c04f4","name":"","active":true,"tosidebar":true,"console":false,"tostatus":false, "complete":"payload","targetType":"msg","x":550,"y":380,"wires":[]},{"id":"68b090c1.1a8b5","type":"debug","z": "7f4dbbe6.6c04f4","name":"","active":true,"tosidebar":true,"console":false,"tostatus":false,"complete":"payload","ta

 $\label{total control of the contro$