**PROJECT REPORT**

**PREDICTING LIFE EXPECTANCY USING MACHINE LEARNING**

**BY**

**KRISHNA SHARMA S**

**1.INTRODUCTION**

**1.1 OVERVIEW**

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. This 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. The user shall be able to interact with the program using the UI built with Node-RED.

**1.2 PURPOSE**

This project uses the data of 193 and considers factor like schooling, GDP, alcohol intake and various diseases to predict the life expectancy of the people of a country.

Using the model, the user is also able to predict by giving his own input to see how life expectancy changes with respect to various factors.

**2. LITERATURE SURVEY**

Organisations around the world including WHO do maintain the data on life expectancy of various countries. There are also many models which predict life expectancy.

**2.1 EXISTING PROBLEM**

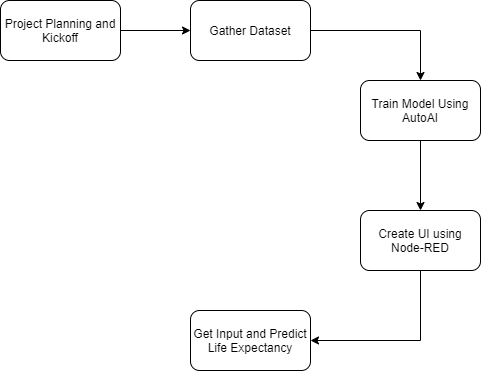
The WHO maintains annual life tables for all its member countries. It maintains a two-year cycle to update this table. Even then, not all information is available and this may cause errors.

**2.2 PROPOSED SOLUTION**

In this project, machine learning is used to predict the life expectancy. It takes in various factors from alcohol intake to schooling to provide a good prediction. The project uses linear regression to train its model and has a very user-friendly UI built using Node-RED.

**3. THEORITICAL ANALYSIS**

**3.1 BLOCK DIAGRAM**



**3.2 HARDWARE / SOFTWARE DESIGNING**

The software designing involved the following steps:

1. PROJECT PLANNING AND KICKOFF:

* Understanding the project description and analyze the data and attributes in the given dataset.
* Plan how to proceed with project.
* Learning to use Zoho writer.

2. Exploring IBM Cloud Platform

3. Exploring IBM Watson Services:

* Exploring the use cases of Watson services.
* Insight into Machine Learning with Watson.

4. Introduction to Watson Studio:

* Learn build a Machine learning Model using IBM Watson.
* Learn to use AutoAI.

5. Predicting Life Expectancy Without Python:

* Collect the Dataset for the project from Kaggle.
* Create necessary Watson Services.
* Create a Watson Studio Project.
* Create Machine Learning Service.
* Import Dataset to the project.
* Use AutoAI in Watson Studio to train the dataset and deploy the model.
* Create a UI with Node-RED and Integrate AutoAI with it.

**4. EXPERIMENTAL INVESTIGATIONS**

**The Dataset:**

The dataset used is a life expectancy dataset maintained by the World Health Organization.

The data is a csv file. The table contains data on:

• Countries

• Status

• Life Expectancy

• Adult Mortality

• 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

**AutoAI:**

Create an AutoAI experiment from your Watson Studio project. Import the dataset and train the model using linear regression. Pick the best pipeline and save the model.

**Node-RED:**

Once, we have setup the flow, we need to integrate the ML model with it. To integrate the ML Model with it we need to access the endpoint URL of our ML Model.

The form contains all the elements of the UI. All the labels are associated with a variable.

**5. FLOWCHART**

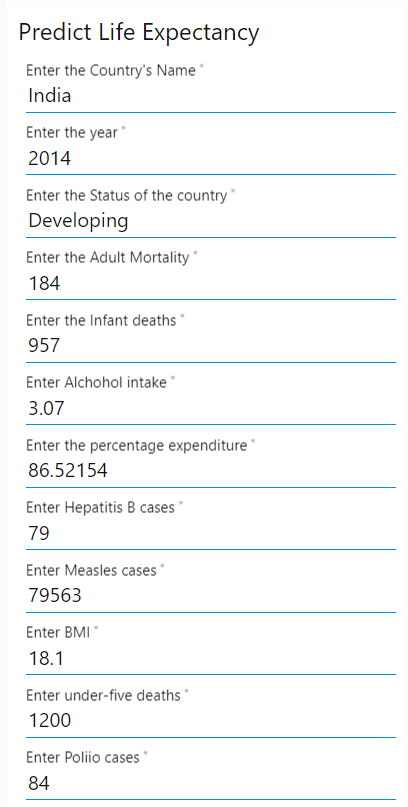


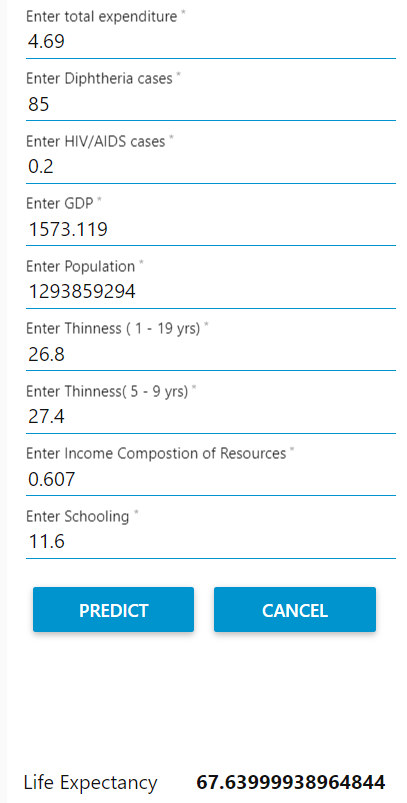
**6. RESULT**

Web based UI was developed by integrating all the services using Node-RED.

The UI can be found here:

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





**7. ADVANTAGES AND DISADVANTAGES**

**Advantages:**

Machine learning technology typically improves efficiency and accuracy thanks to the ever-increasing amounts of data that are processed. The application learns the patterns and trends hidden within the data without human intervention which makes predicting much simpler and easier. The more data is fed to the algorithm, the higher the accuracy of the algorithm.

Node-Red can provide ways for the developer to create a simple UI and integrate the ML model with it.

**Disadvantages:**

Using machine learning interface comes with its own problems. Since the whole point of it is minimize human involvement, it also makes error detection and fixing much more problematic. It takes a lot of time to identify the root cause for the problem.

Machine learning can also be very resource consuming especially while dealing with large amount of data.

Node-Red does not give many features to customize our UI when compared to UI developed with React or Angular.

**8. APPLICATIONS**

* It is useful for the government of a country. The government can change use this data to try and prioritise on certain aspects to increase the life expectancy of its population.
* Individuals may also use this to predict their life expectancy.
* Health sectors can also use this model and provide the necessary services to maintain a good life expectancy of the patients.

**9. CONCLUSION**

* This model contains various factors such as Country, Year, Status, Life Expectancy, 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.
* With the help of all these input values, the model will predict the life expectancy of people living in 193 countries.
* It has a friendly UI to get user’s input.

**10. FUTURE SCOPE**

* Results can be displayed in the form of a graph.
* The UI can be made to look more attractive.
* Other factors like mental health can also be taken into account to make the predictions more accurate.

**11. BIBLIOGRAPHY**

1. Node-RED Starter Application:

<https://developer.ibm.com/tutorials/how-to-create-a-node-red-starter-application/>

2. Watson Studio Cloud: <https://bookdown.org/caoying4work/watsonstudio-workshop/jn.html/>

3. Dataset Reference:

<https://www.kaggle.com/kumarajarshi/life-expectancy-who/>

4. IBM Cloud Services: <https://www.youtube.com/watch?v=DBRGlAHdj48&list=PLzpeuWUENMK2PYtasCaKK4bZjaYzhW23L>

**APPENDIX**

1. **Source Code**
   1. **Node-RED Flow**

|  |
| --- |
| [ |
|  | { |
|  | "id": "fdd6cf59.11f31", |
|  | "type": "tab", |
|  | "label": "Life Expectancy", |
|  | "disabled": false, |
|  | "info": "" |
|  | }, |
|  | { |
|  | "id": "f7221f5b.a9b6d", |
|  | "type": "function", |
|  | "z": "fdd6cf59.11f31", |
|  | "name": "PreToken", |
|  | "func": "global.set(\"cName\",msg.payload.cName);\nglobal.set(\"year\",msg.payload.year);\nglobal.set(\"stat\",msg.payload.stat);\nglobal.set(\"adult\",msg.payload.adult);\nglobal.set(\"infant\",msg.payload.infant);\nglobal.set(\"alchohol\",msg.payload.alchohol);\nglobal.set(\"expenditure\",msg.payload.expenditure);\nglobal.set(\"hepatitis\",msg.payload.hepatitis);\nglobal.set(\"measles\",msg.payload.measles);\nglobal.set(\"bmi\",msg.payload.bmi);\nglobal.set(\"underFive\",msg.payload.underFive);\nglobal.set(\"polio\",msg.payload.polio);\nglobal.set(\"ttlExp\",msg.payload.ttlExp);\nglobal.set(\"diphtheria\",msg.payload.diphtheria);\nglobal.set(\"hiv\",msg.payload.hiv);\nglobal.set(\"gdp\",msg.payload.gdp);\nglobal.set(\"pop\",msg.payload.pop);\nglobal.set(\"thin1to19\",msg.payload.thin1to19);\nglobal.set(\"thin5to9\",msg.payload.thin5to9);\nglobal.set(\"income\",msg.payload.income);\nglobal.set(\"school\",msg.payload.school);\nvar apikey = \"vghY2UmRDNxXLjvcbX5khaAO6AgkE2qnsdknWUKytRfA\";\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": 255, |
|  | "y": 316.0000047683716, |
|  | "wires": [ |
|  | [ |
|  | "891a7104.65704" |
|  | ] |
|  | ] |
|  | }, |
|  | { |
|  | "id": "891a7104.65704", |
|  | "type": "http request", |
|  | "z": "fdd6cf59.11f31", |
|  | "name": "", |
|  | "method": "POST", |
|  | "ret": "obj", |
|  | "paytoqs": false, |
|  | "url": "https://iam.cloud.ibm.com/identity/token", |
|  | "tls": "", |
|  | "persist": false, |
|  | "proxy": "", |
|  | "authType": "", |
|  | "x": 454, |
|  | "y": 305, |
|  | "wires": [ |
|  | [ |
|  | "40e332ed.fe8b9c" |
|  | ] |
|  | ] |
|  | }, |
|  | { |
|  | "id": "8a88afe9.d7746", |
|  | "type": "inject", |
|  | "z": "fdd6cf59.11f31", |
|  | "name": "", |
|  | "topic": "", |
|  | "payload": "", |
|  | "payloadType": "date", |
|  | "repeat": "", |
|  | "crontab": "", |
|  | "once": false, |
|  | "onceDelay": 0.1, |
|  | "x": 100.5, |
|  | "y": 230.00000286102295, |
|  | "wires": [ |
|  | [ |
|  | "f7221f5b.a9b6d" |
|  | ] |
|  | ] |
|  | }, |
|  | { |
|  | "id": "bcf4317f.10e8c", |
|  | "type": "debug", |
|  | "z": "fdd6cf59.11f31", |
|  | "name": "", |
|  | "active": true, |
|  | "tosidebar": true, |
|  | "console": false, |
|  | "tostatus": false, |
|  | "complete": "payload", |
|  | "targetType": "msg", |
|  | "x": 893.0000114440918, |
|  | "y": 47.00000190734863, |
|  | "wires": [] |
|  | }, |
|  | { |
|  | "id": "40e332ed.fe8b9c", |
|  | "type": "function", |
|  | "z": "fdd6cf59.11f31", |
|  | "name": "Pre Prediction", |
|  | "func": "var cName = global.get(\"cName\",msg.payload.cName);\nvar year = global.get(\"year\",msg.payload.year);\nvar stat = global.get(\"stat\",msg.payload.stat);\nvar adult = global.get(\"adult\",msg.payload.adult);\nvar infant = global.get(\"infant\",msg.payload.infant);\nvar alchohol = global.get(\"alchohol\",msg.payload.alchohol);\nvar expenditure = global.get(\"expenditure\",msg.payload.expenditure);\nvar hepatitis = global.get(\"hepatitis\",msg.payload.hepatitis);\nvar measles = global.get(\"measles\",msg.payload.measles);\nvar bmi = global.get(\"bmi\",msg.payload.bmi);\nvar underFive = global.get(\"underFive\",msg.payload.underFive);\nvar polio = global.get(\"polio\",msg.payload.polio);\nvar ttlExp = global.get(\"ttlExp\",msg.payload.ttlExp);\nvar diphtheria = global.get(\"diphtheria\",msg.payload.diphtheria);\nvar hiv = global.get(\"hiv\",msg.payload.hiv);\nvar gdp = global.get(\"gdp\",msg.payload.gdp);\nvar pop = global.get(\"pop\",msg.payload.pop);\nvar thin1to19 = global.get(\"thin1to19\",msg.payload.thin1to19);\nvar thin5to9 = global.get(\"thin5to9\",msg.payload.thin5to9);\nvar income = global.get(\"income\",msg.payload.income);\nvar school = global.get(\"school\",msg.payload.school);\nvar token=msg.payload.access\_token\nvar instance\_id=\"48bbedc7-82d9-42e5-b264-bb3233048f5d\"\nmsg.headers={'Content-Type': '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\"], \"values\": [[cName,year,stat,adult,infant,alchohol,expenditure,hepatitis,measles,bmi,underFive,polio,ttlExp,diphtheria,hiv,gdp,pop,thin1to19,thin5to9,income,school]]}]}\nreturn msg;", |
|  | "outputs": 1, |
|  | "noerr": 0, |
|  | "x": 560, |
|  | "y": 220, |
|  | "wires": [ |
|  | [ |
|  | "925860dd.0be24" |
|  | ] |
|  | ] |
|  | }, |
|  | { |
|  | "id": "925860dd.0be24", |
|  | "type": "http request", |
|  | "z": "fdd6cf59.11f31", |
|  | "name": "", |
|  | "method": "POST", |
|  | "ret": "obj", |
|  | "paytoqs": false, |
|  | "url": "https://eu-gb.ml.cloud.ibm.com/v4/deployments/df9e6bce-a1d8-40f8-b91c-2891a2bb1eeb/predictions", |
|  | "tls": "", |
|  | "persist": false, |
|  | "proxy": "", |
|  | "authType": "", |
|  | "x": 610, |
|  | "y": 140, |
|  | "wires": [ |
|  | [ |
|  | "ad1c3331.044ee" |
|  | ] |
|  | ] |
|  | }, |
|  | { |
|  | "id": "44bd134e.f45d1c", |
|  | "type": "ui\_text", |
|  | "z": "fdd6cf59.11f31", |
|  | "group": "ea41acd0.838fa", |
|  | "order": 2, |
|  | "width": 0, |
|  | "height": 0, |
|  | "name": "", |
|  | "label": "Life Expectancy", |
|  | "format": "{{msg.payload}}", |
|  | "layout": "row-spread", |
|  | "x": 920, |
|  | "y": 240, |
|  | "wires": [] |
|  | }, |
|  | { |
|  | "id": "ad1c3331.044ee", |
|  | "type": "function", |
|  | "z": "fdd6cf59.11f31", |
|  | "name": "", |
|  | "func": "msg.payload=msg.payload.predictions[0].values[0][0]\nreturn msg;", |
|  | "outputs": 1, |
|  | "noerr": 0, |
|  | "x": 790, |
|  | "y": 140, |
|  | "wires": [ |
|  | [ |
|  | "bcf4317f.10e8c", |
|  | "44bd134e.f45d1c" |
|  | ] |
|  | ] |
|  | }, |
|  | { |
|  | "id": "392ed0d.05d883", |
|  | "type": "ui\_form", |
|  | "z": "fdd6cf59.11f31", |
|  | "name": "Life Expectancy Form", |
|  | "label": "Predict Life Expectancy", |
|  | "group": "ea41acd0.838fa", |
|  | "order": 1, |
|  | "width": 0, |
|  | "height": 0, |
|  | "options": [ |
|  | { |
|  | "label": "Enter the Country's Name", |
|  | "value": "cName", |
|  | "type": "text", |
|  | "required": true, |
|  | "rows": null |
|  | }, |
|  | { |
|  | "label": "Enter the year", |
|  | "value": "year", |
|  | "type": "number", |
|  | "required": true, |
|  | "rows": null |
|  | }, |
|  | { |
|  | "label": "Enter the Status of the country", |
|  | "value": "stat", |
|  | "type": "text", |
|  | "required": true, |
|  | "rows": null |
|  | }, |
|  | { |
|  | "label": "Enter the Adult Mortality", |
|  | "value": "adult", |
|  | "type": "number", |
|  | "required": true, |
|  | "rows": null |
|  | }, |
|  | { |
|  | "label": "Enter the Infant deaths", |
|  | "value": "infant", |
|  | "type": "number", |
|  | "required": true, |
|  | "rows": null |
|  | }, |
|  | { |
|  | "label": "Enter Alchohol intake", |
|  | "value": "alchohol", |
|  | "type": "number", |
|  | "required": true, |
|  | "rows": null |
|  | }, |
|  | { |
|  | "label": "Enter the percentage expenditure", |
|  | "value": "expenditure", |
|  | "type": "number", |
|  | "required": true, |
|  | "rows": null |
|  | }, |
|  | { |
|  | "label": "Enter Hepatitis B cases", |
|  | "value": "hepatitis", |
|  | "type": "number", |
|  | "required": true, |
|  | "rows": null |
|  | }, |
|  | { |
|  | "label": "Enter Measles cases", |
|  | "value": "measles", |
|  | "type": "number", |
|  | "required": true, |
|  | "rows": null |
|  | }, |
|  | { |
|  | "label": "Enter BMI", |
|  | "value": "bmi", |
|  | "type": "number", |
|  | "required": true, |
|  | "rows": null |
|  | }, |
|  | { |
|  | "label": "Enter under-five deaths", |
|  | "value": "underFive", |
|  | "type": "number", |
|  | "required": true, |
|  | "rows": null |
|  | }, |
|  | { |
|  | "label": "Enter Poliio cases", |
|  | "value": "polio", |
|  | "type": "number", |
|  | "required": true, |
|  | "rows": null |
|  | }, |
|  | { |
|  | "label": "Enter total expenditure", |
|  | "value": "ttlExp", |
|  | "type": "number", |
|  | "required": true, |
|  | "rows": null |
|  | }, |
|  | { |
|  | "label": "Enter Diphtheria cases", |
|  | "value": "diphtheria", |
|  | "type": "number", |
|  | "required": true, |
|  | "rows": null |
|  | }, |
|  | { |
|  | "label": "Enter HIV/AIDS cases", |
|  | "value": "hiv", |
|  | "type": "number", |
|  | "required": true, |
|  | "rows": null |
|  | }, |
|  | { |
|  | "label": "Enter GDP", |
|  | "value": "gdp", |
|  | "type": "number", |
|  | "required": true, |
|  | "rows": null |
|  | }, |
|  | { |
|  | "label": "Enter Population", |
|  | "value": "pop", |
|  | "type": "number", |
|  | "required": true, |
|  | "rows": null |
|  | }, |
|  | { |
|  | "label": "Enter Thinness ( 1 - 19 yrs)", |
|  | "value": "thin1to19", |
|  | "type": "number", |
|  | "required": true, |
|  | "rows": null |
|  | }, |
|  | { |
|  | "label": "Enter Thinness( 5 - 9 yrs)", |
|  | "value": "thin5to9", |
|  | "type": "number", |
|  | "required": true, |
|  | "rows": null |
|  | }, |
|  | { |
|  | "label": "Enter Income Compostion of Resources", |
|  | "value": "income", |
|  | "type": "number", |
|  | "required": true, |
|  | "rows": null |
|  | }, |
|  | { |
|  | "label": "Enter Schooling ", |
|  | "value": "school", |
|  | "type": "number", |
|  | "required": true, |
|  | "rows": null |
|  | } |
|  | ], |
|  | "formValue": { |
|  | "cName": "", |
|  | "year": "", |
|  | "stat": "", |
|  | "adult": "", |
|  | "infant": "", |
|  | "alchohol": "", |
|  | "expenditure": "", |
|  | "hepatitis": "", |
|  | "measles": "", |
|  | "bmi": "", |
|  | "underFive": "", |
|  | "polio": "", |
|  | "ttlExp": "", |
|  | "diphtheria": "", |
|  | "hiv": "", |
|  | "gdp": "", |
|  | "pop": "", |
|  | "thin1to19": "", |
|  | "thin5to9": "", |
|  | "income": "", |
|  | "school": "" |
|  | }, |
|  | "payload": "", |
|  | "submit": "Predict", |
|  | "cancel": "Cancel", |
|  | "topic": "", |
|  | "x": 100, |
|  | "y": 360, |
|  | "wires": [ |
|  | [ |
|  | "f7221f5b.a9b6d" |
|  | ] |
|  | ] |
|  | }, |
|  | { |
|  | "id": "ea41acd0.838fa", |
|  | "type": "ui\_group", |
|  | "z": "", |
|  | "name": "Default", |
|  | "tab": "6ced1270.359c2c", |
|  | "order": 1, |
|  | "disp": false, |
|  | "width": "6", |
|  | "collapse": false |
|  | }, |
|  | { |
|  | "id": "6ced1270.359c2c", |
|  | "type": "ui\_tab", |
|  | "z": "", |
|  | "name": "Home", |
|  | "icon": "home", |
|  | "disabled": false, |
|  | "hidden": false |
|  | } |
|  | ] |

**1.2AutoAI Notebook:**







