PROJECT REPORT

PREDICTING LIFE EXPECTANCY USING MACHINE LEARNING

SUBMITTED BY :

ABHISHEK MOHARIR

GITHUB :

https://github.com/SmartPracticeschool/llSPS-INT-1960-Predicting-Life-Expectancy-using-Machine-Learning

VIDEO DEMONSTRATION LINK

https://drive.google.com/file/d/1zf0ca46y9GoC5LHBbpkPbqJD1ZjacxJ8/view?usp=sharing

INDEX

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, Medical Health Services in the region , country's population , GDP etc. The main objective of the project is to predict the life expectancy of a person depending on several factors based on an individual or the residing country. It is very important to predict average life expectancy of a country to analyse further requirements to increase its rate of growth or stabilise the rate of growth in that country.

Machine learning algorithms that can be used in this case are: Regression, Decision Tree, Random Forest, Clustering techniques, so that we can achieve high accuracy for our model. So this is a typical Regression Machine Learning project that leverages historical data to predict insights into the future.

1.2 Purpose

Predicting the life expectancy will give the country an idea of the factors which can be improved to increase the lifespan of the people living, like by improving the health care facilities or immunization vaccines for infants. The whole purpose is to build a machine learning model for the prediction of life expectancy . We should take full advantage of this new era advanced technology to improve the future by predicting it in the present.

2. LITERATURE SURVEY

2.1 Existing Problem

After having reviewed the existing work on the problem statement , we found that e factors used for predicting were just personal causes and not related to the surrounding, healthcare facilities, demographic, social, regional and economical factors of the country he resides . These factors can increase the accuracy of our model . We reached a conclusion that it is feasible for individuals using evolving technologies and devices wearables and mobile health monitoring devices.

2.2 Proposed Solution

The previous factors were more human based but it is important to know the economical, regional, social and demographic factors like the GDP, population, education, immunizations, history of illness, health care facility, funds allocated by the government, schemes, medical expenditures like if it is very high then people will shy away to get regular medical check-ups, and many more. So we need to focus on some additional features . Calculating a life expectancy reliably would require a sophisticated system that considers a breadth of environmental, geographic, genetic and lifestyle factors – all of which have influence. The new technologies like machine learning makes it more feasible to analyse large quantities of data . Also, tools such as deep learning and artificial intelligence can be used to consider complex variables, such as biomedical data, to predict someone's biological age . The use of deep learning and cognitive computing, such as with IBM Watson, helps to make more accurate diagnoses than using human judgement alone. This, coupled with predictive analytics and increasing computational power, means we may soon have systems, or even apps, that can calculate life expectancy.

Steps:

a) Create IBM cloud services :

Watson Studio

Machine Learning resource

Node-Red

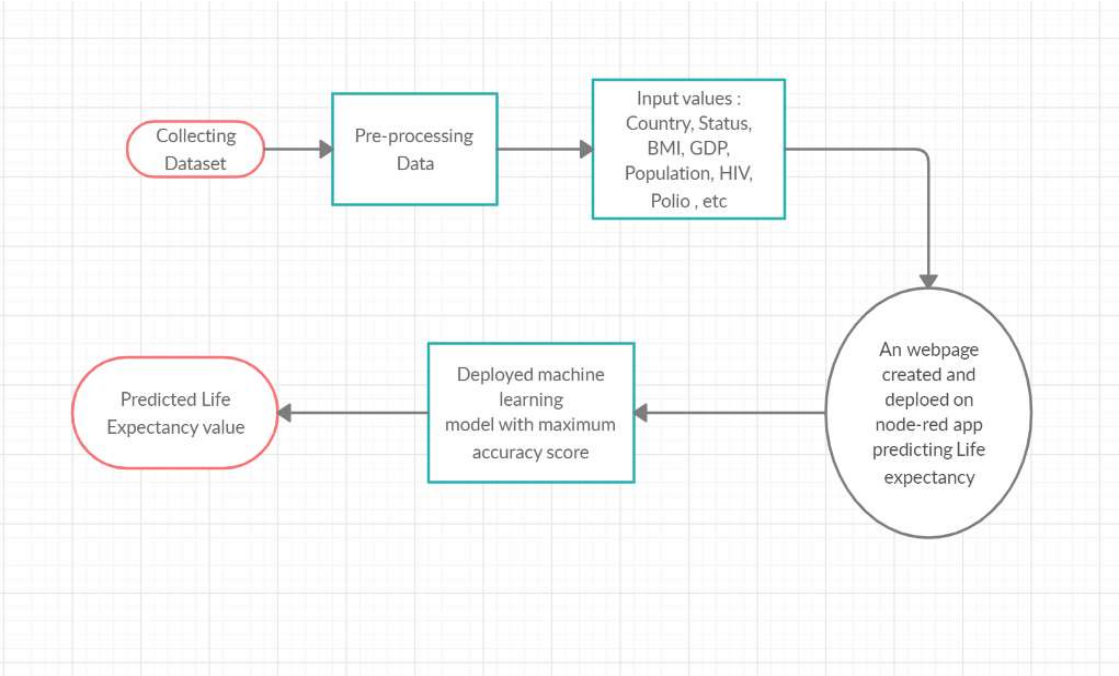
b) Configure Watson Studio

c) Create Node-Red Flow to connect all services together

d) Deploy and run Node-Red app

3. THEROTICAL ANALYSIS

3.1 Block Diagram



3.2. Hardware / Software Designing

Project Requirements: Python, IBM Cloud, IBM Watson

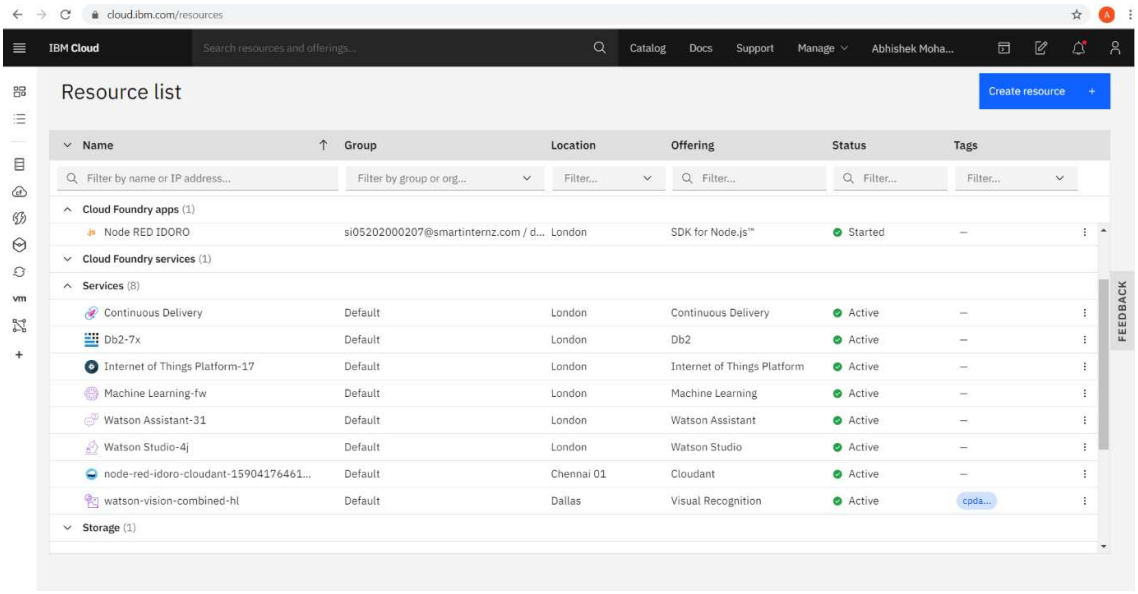
Functional Requirements: IBM cloud

Technical Requirements: ML, WATSON Studio, Python, Node-Red

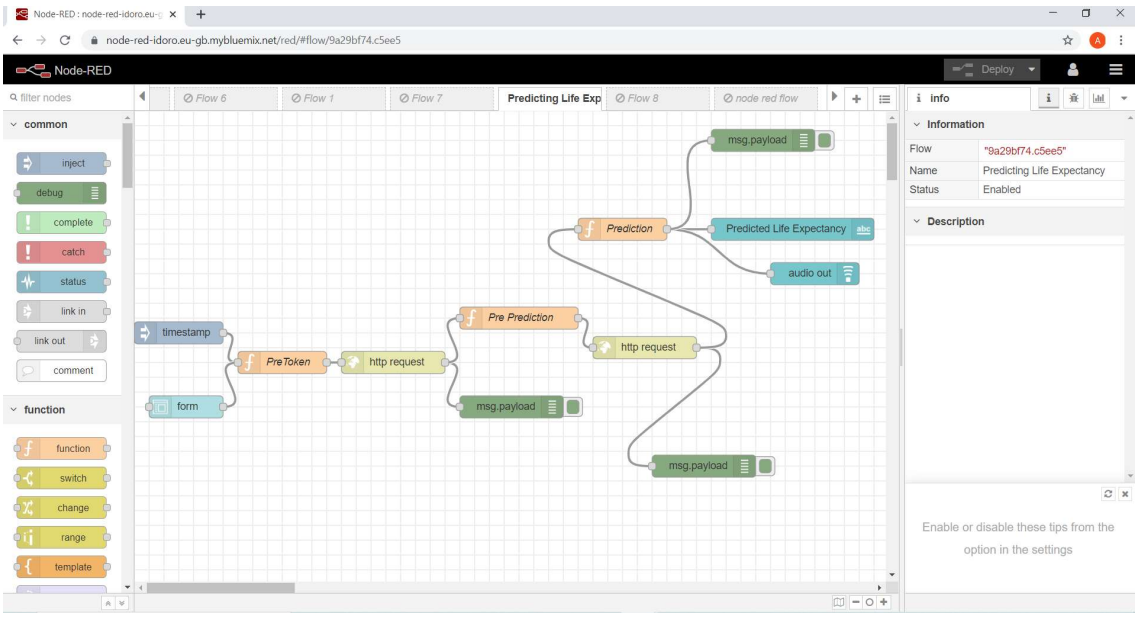
Software Requirements: Watson Studio, Node-Red

4. EXPERIMENTAL INVESTIGATIONS

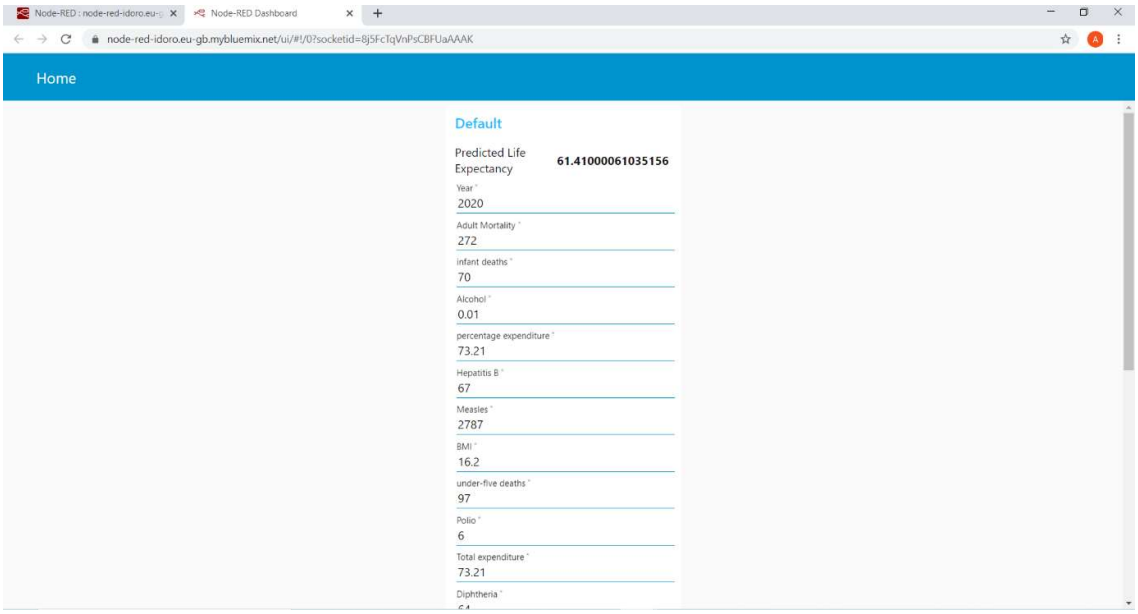
A) IBM Cloud Resource List



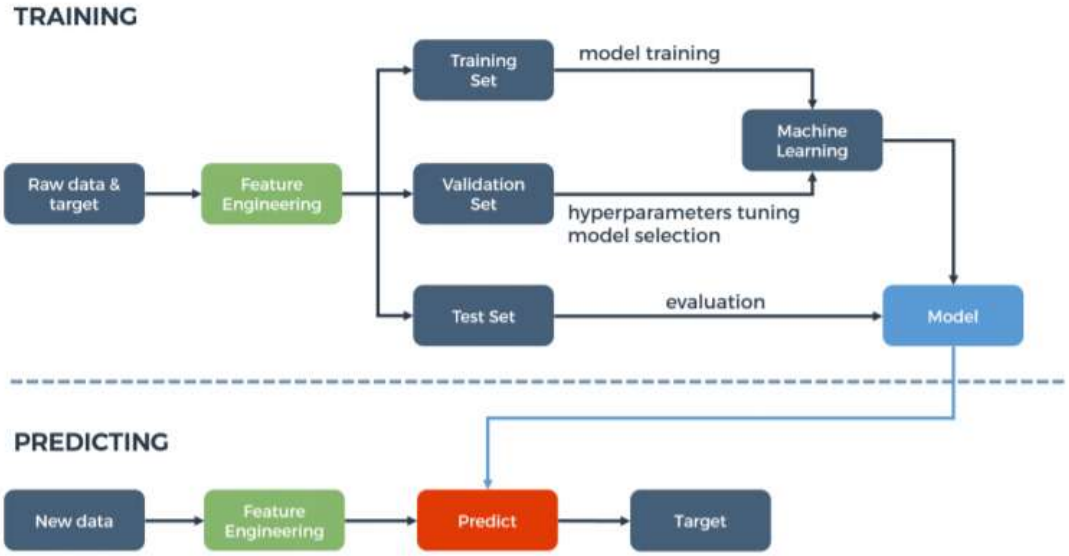
B) Node-Red Flow



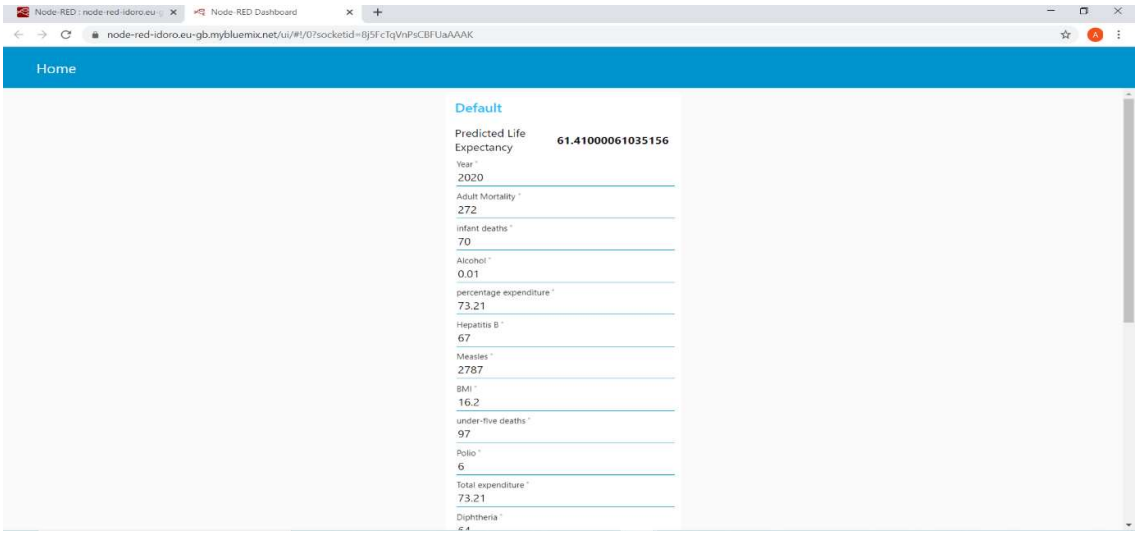
C) Life Expectancy Prediction UI



5. FLOWCHART



6 . RESULT



Some interesting correlations here:

• There is a strong positive correlation between ‘Schooling’ and ‘Life Expectancy’. This may be because education is more established and prevalent in wealthier countries. This means countries with less corruption, infrastructure, healthcare, welfare, and so forth.

• Similarly to the point above, there is a moderate positive correlation between ‘GDP’ and ‘Life Expectancy’, most likely due to the same reason.

• Surprisingly there’s a moderate positive correlation between ‘Alcohol’ and ‘Life Expectancy’. I’m guessing that this is due to the fact that only wealthier countries can afford alcohol or the consumption of alcohol is more prevalent among wealthier populations.

7. ADVANTAGES AND DISADVANTAGES

Advantages

a) User Friendly Interface: This interface requires no background knowledge of how to use it. It’s a simple interface and only ask for required values and predict the output.

b) Advantages of using IBM Cloud: Easy to use and deploy, easy to connect with UI, takes care of large storage space.

c) Reduced Costs: This is a simple webpage and can be accessed by any citizen of a country to calculate life expectancy of their country and doesnot required any kind of payment neither for designing nor for using.

d) The life expectancy predictor will give important insights and help people achieve good quality of life in future. The country can plan and improve various healthcare facilities.

Disadvantages

a) Wrong Prediction: As it depends completely on user, so if user provides some wrong values then it will predict wrong value.

b) Average Prediction: The model predicts average or approximate value with 97.07% accuracy but not accurate value.

d) User input is not saved in any database.

e) Input should be in range only to predict accurate values.

8 . APPLICATIONS

1. To analyse all the factors and plan out measures to increase the life expectancy of the country.

2. It can be used to analyse the factors for high life expectancy.

3. It is user friendly and can be used by anyone.

4. To help government prepare life insurance policies for people. This will benefit the people .

5. To analyse country's growth statistics in future years .

6. It can be used to monitor health inequalities of a country.

7. It can be used to develop statistics for country development process .

9. CONCLUSION

This user interface will be useful for the user to predict life expectancy value of their own country or any other country based on some required details such as GDP, BMI, Year, Alcohol Intake, Total expenditure etc .

10. FUTURE SCOPE

Integrating with services such as speech recognition

the website can be added with many more features to improve the user experience. The user input can be connected to the database for future purposes.

It requires much more data about 21 columns to be known prior for predicting life expectancy which can be again difficult for a normal user to gather such data so some kind of feature reduction or replacement of some features as individuals or groups can be done to make it more user friendly.

11. BIBLIOGRAPHY

https://cloud.ibm.com/docs/overview?topic=overview-whatis-platform

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

https://github.com/watson-developer-cloud/node-red-labs

https://bookdown.org/caoying4work/watsonstudio-workshop/jn.html

https://machinelearningmastery.com/columntransformer-for-numerical-and-categorical-data/

https://bookdown.org/caoying4work/watsonstudio-workshop/jn.html#deploy-model-as-webservice

https://www.youtube.com/watch?v=DBRGlAHdj48&list=PLzpeuWUENMK2PYtasCaKK4bZjaYzhW23L