

PROJECT REPORT

Title: Predicting Life Expectancy using ML

Name: Sumit Singh

Email: sumit01du@gmail.com

Application ID- SPS_APL_20200003832

1. Introduction

1.1 Overview

Life expectancy means the estimation of the numbers of years the individual will survive. It further depends on age, sex and location where the person is living. It also helps in the treatment of patients which are underlying with certain diseases or illness. For eg. by predicting life expectancy we can analyze the aggressiveness of any disease.

A large amount of data that is generated today is unstructured, which requires processing to generate insights. After pre-processing the data set we will remove the noise from it. After removing the noise we will clean the data set.

IBM Watson, machine learning and node red are an integral part of an analysis. The end product will be a web page where you need to give all the required inputs and then submit it. Afterwards it will predict the life expectancy value based on regression technique.

i) Project Requirements: Python, IBMCloud, IBMWatson.

ii) Functional Requirements: IBMcloud, Dataset.

iii) Technical Requirements: ML, WATSONStudio, Python, Node-Red.

iv) Software Requirements: WatsonStudio, Node-Red.

v) Project Deliverable:

- Python notebook containing all the code.
- A node red application which can input data and outputs a prediction for life expectancy.
- A JSON file containing the architecture of node red project.
- URL of the node red application.

1.2 Purpose

The purpose of the project is to design a model for predicting Life Expectancy rate of a Country given various features such as year, GDP, education, alcohol intake of people in the country, Expenditure on health care system and some specific disease related deaths that happened in the country are given. This project analyses the provided data set and creates a model to predict life expectancy and machine learning can benefit public health researchers with analyzing thousands of variables to obtain data regarding life expectancy.

We can use demographics of selected regional areas and multiple behavioral health disorders across regions to find correlation between individual behavior indicators and behavioral health outcomes. IBM Watson machine learning and node red are an integral part of analysis. The end product will be a web page where you need to give all the required inputs and then submit it. Afterwards it will predict the life expectancy value based on regression technique.

2. Literature Survey

2.1 Problem Statement

The typical regression model that can predict average life expectancy of the country based on some user inputted values such as GDP, BMI, HIV/AIDS, Year, Alcohol intake and etc.

2.2 Proposed Solution

Our proposed system makes this whole process of calculating Life Expectancy much easier so any one can calculate the Life Expectancy without any domain knowledge. Our proposed system makes calculation automated and this system has a predicting tool which can predict the Life Expectancy from various attributes value.

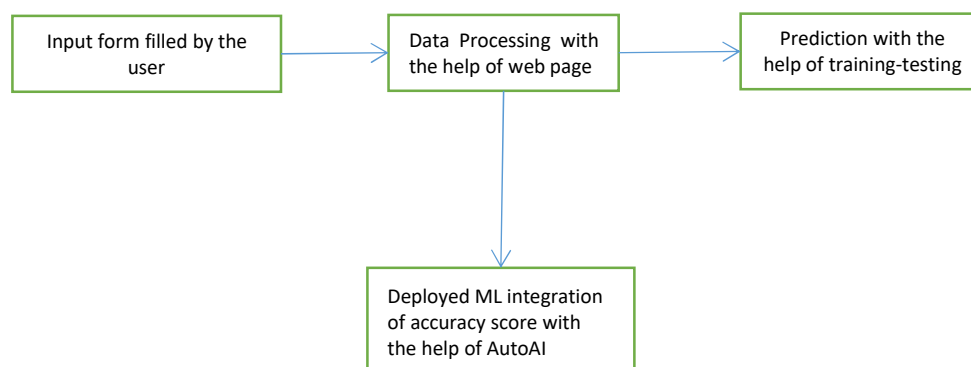
So, using machine learning technique we suppose to predict the value of Life Expectancy based on some common attributes like year, GDP, education, alcohol intake of people in the country, expenditure on health care system .We can find this data and get the Life Expectancy value based on their Country and Year.

We will follow the following steps:

- a) Create IBM cloud services
- b) Configure Watson Studio
- c) Create Node-Red Flow to connect all services together
- d) Deploy and run Node-Red app

3. Theoretical Analysis

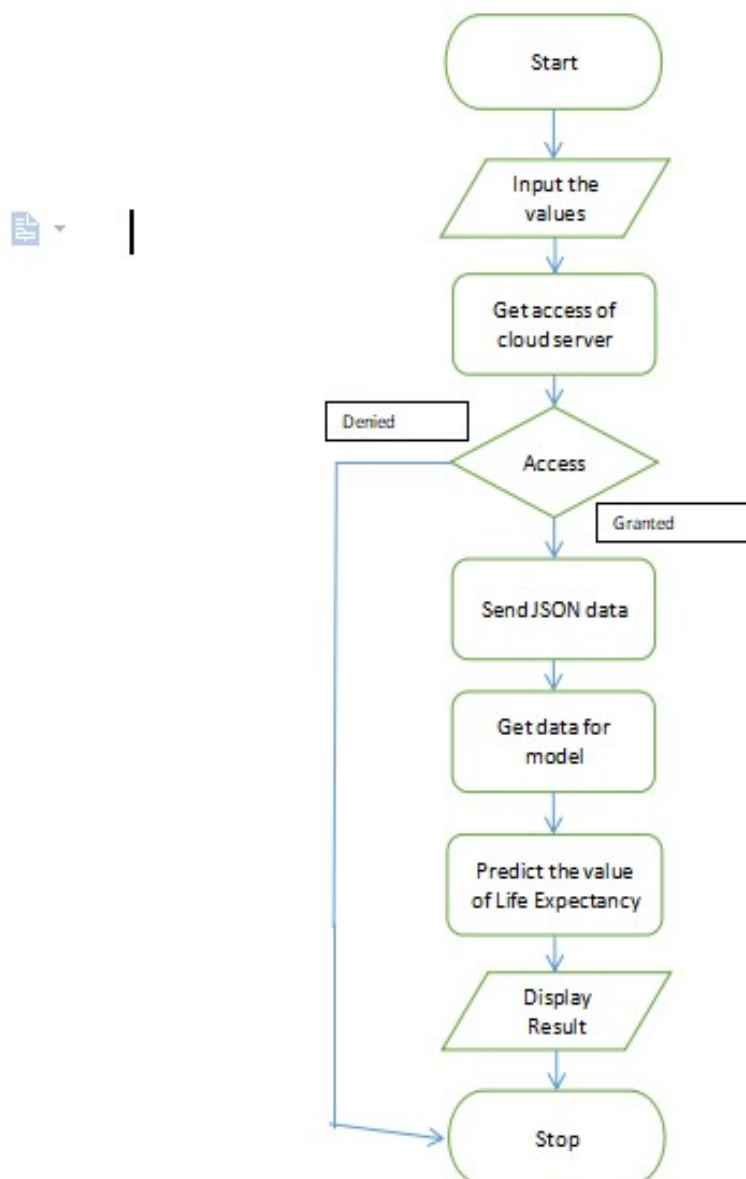
3.1 Block Diagram



3.2 Project Deliverable

- Python notebook containing all the code.
- A node red application which can input data and outputs a prediction for life expectancy.
- A json file containing the architecture of node red project.
- URL of the node red application.

4. Flowchart



5. Results

The notebook attached consists of the detailed steps involved in the pipeline. First the data was imported from cloud storage object of IBM Cloud and stored in a data frame. Later it was checked for null values and necessary steps was taken (filling the null values with the mean of the column). Data Visualization was performed on different columns to find the relation between life-expectancy and other variables. For training phase 20% data was separated for testing purpose. Different pipelines were created for numerical and categorical columns.

Machine Learning Model

Prediction **71.847**

BMI *	324
HIV/AIDS *	424
thinness 1-19 years *	4
thinness 5-9 years *	4
Adult Mortality *	4
Alcohol *	4
Country *	yemen
Diphtheria *	333
GDB *	3
Measles *	3
Polio *	3
Population *	33123
Schooling *	3
Status *	developed
Total expenditure *	344
Year *	2015
infant deaths *	3
percentage expenditure *	442
under-five deaths *	3

PREDICT

CANCEL

6. Applications

- a) It can be used to monitor health inequalities of a country.
- b) It can be used to develop statistics for country development process.
- c) It can be used to analyze the factors for high life expectancy.
- d) It is user friendly and can be used by anyone.

7. Future Scope

Integrating a data science dashboard which shows different visualizations of Life Expectancy as per the Country and Year. A system to update our model parameters when there is a change in consistency of data like when a sudden epidemic occurs or during a recession, then all the attributes used in our model need to be updated to provide most precise life expectancy.

8. 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. The advantages of longer life span outweigh its disadvantages. The benefits people and the world can get from a higher life expectancy are irreplaceable and undeniable. It is a truth that life expectancy is a symbol of civilization and better life. Knowing an estimate of how much life we have left pushes us to achieve different things. Higher life expectation is also perceived as greater quality of life and greater income of society.

Our project has automated the entire task of rigorous calculation and removed errors in the existing system and gives the life expectancy to the user. This information can be useful to the society as stated above and this method is also much cheaper than hiring people to do the calculations.

9. Bibliography:

- <https://developer.ibm.com/tutorials/how-to-create-a-node-red-starter-application>
- <https://nodered.org/>
- <https://cloud.ibm.com/docs/overview?topic=overview-what-is-platform>
- <https://github.com/watson-developer-cloud/node-red-labs>
- <https://www.youtube.com/embed/r7E1TJ1HtM0>
- <https://bookdown.org/caoying4work/watsonstudio-workshop/jn.html>
- <https://www.kaggle.com/kumarajarshi/life-expectancy-who>

- <https://www.youtube.com/watch?v=DBRGIAHdj48&list=PLzpeuWUENMK2PYtasC>
- [aKK4b](https://www.youtube.com/watch?v=Jtej3Y6uUng)
- <https://www.youtube.com/watch?v=Jtej3Y6uUng>
- <https://bookdown.org/caoying4work/watsonstudio>