Project Report

By: AMIT SINGH

PREDICTING LIFE EXPECTANCY

USING MACHINE LEARNING

INDEX

S.NO	TITLE	PAGE NO.
1	Introduction	3
2	Literature Survey	4
3	Theoretical Analysis	5
4	Experimental Investigations	6
5	Flowchart	8
6	Result	9
7	Advantages and Disadvantages	10
8	Applications	11
9	Conclusion	11
10	Future Scope	12
11	Bibliography	12
12	Source Code	13

1. Introduction

1.1 Overview

Life expectancy is a statistical measure of the average time a person is expected to live, based on the year of its birth, its current age, and other demographic factors including gender.

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. However, as life expectancy is calculated based on averages, a person may live for many years more or less than expected.

1.2. Purpose

The aim of the project is to develop a model that will predict the life expectancy of the 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 in the dataset. After Successful Creation of the model, model will be deployed on IBM Watson and Node-red flow will be created.

2. Literature Survey

2.1. Existing Problem

As a result of the evolution of biotechnologies and related technologies humans can enjoy longer life expectancies than previously before. Predicting a human's life expectancy has been a long-term question to humankind.

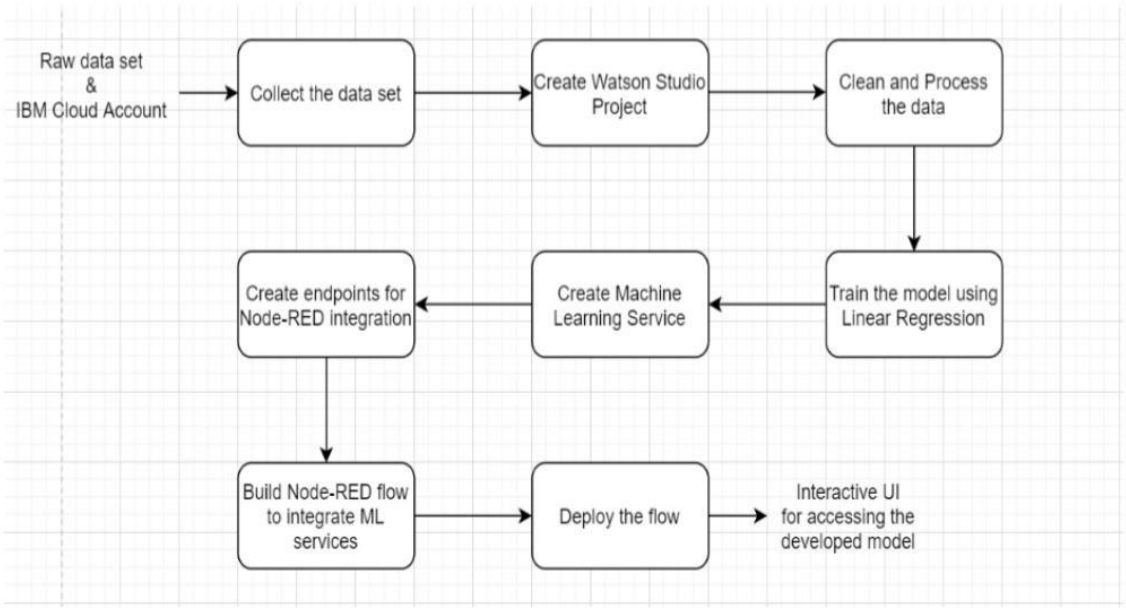
Currently, there are various smart devices and applications that provide wellness and fitness tracking. Some apps offer health-related data such as sleep monitoring, heart rate measuring, and calorie expenditure collected and processed by the devices and servers in the cloud. However, no actual works provide the Personalized Life expectancy.

2.2. Proposed Solution

This Project will try to aim at accounting all the critical factors possible for efficient prediction. The project relies on the accuracy of data and for that the dataset has been taken from kaggle. The data-set related to life expectancy, health factors for 193 countries have been collected from the WHO data repository website and its corresponding economic data was collected from the United Nations website. Among all categories of health-related factors, only those critical factors were chosen which are more representative. The Dataset contain data of the years 2000 to 2015 of all the countries. Dataset covers all the factors possible and data will be preprocessed and various regression algorithms will be applied to train the model .Best model will be chosen and deployed. IBM Cloud Platform is used to develop this project which provides a great environment and services.

3. Theoretical Analysis

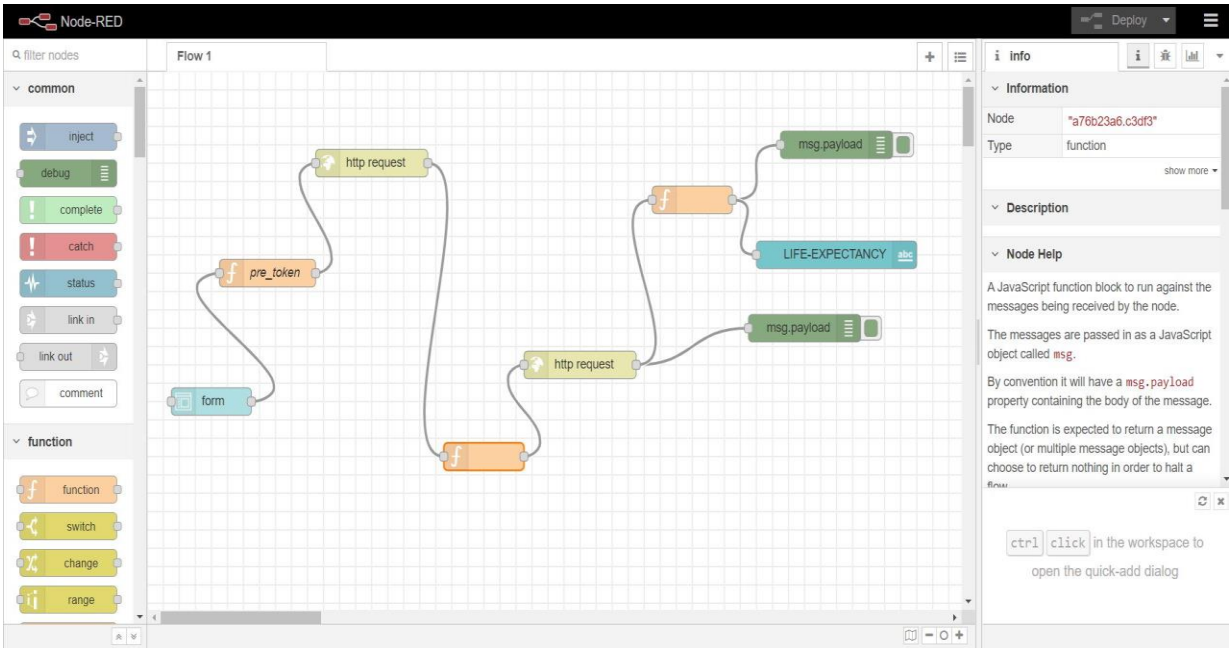
3.1. Block Diagram



3.2. Hardware/ Software Designing

Software Requirements are:-

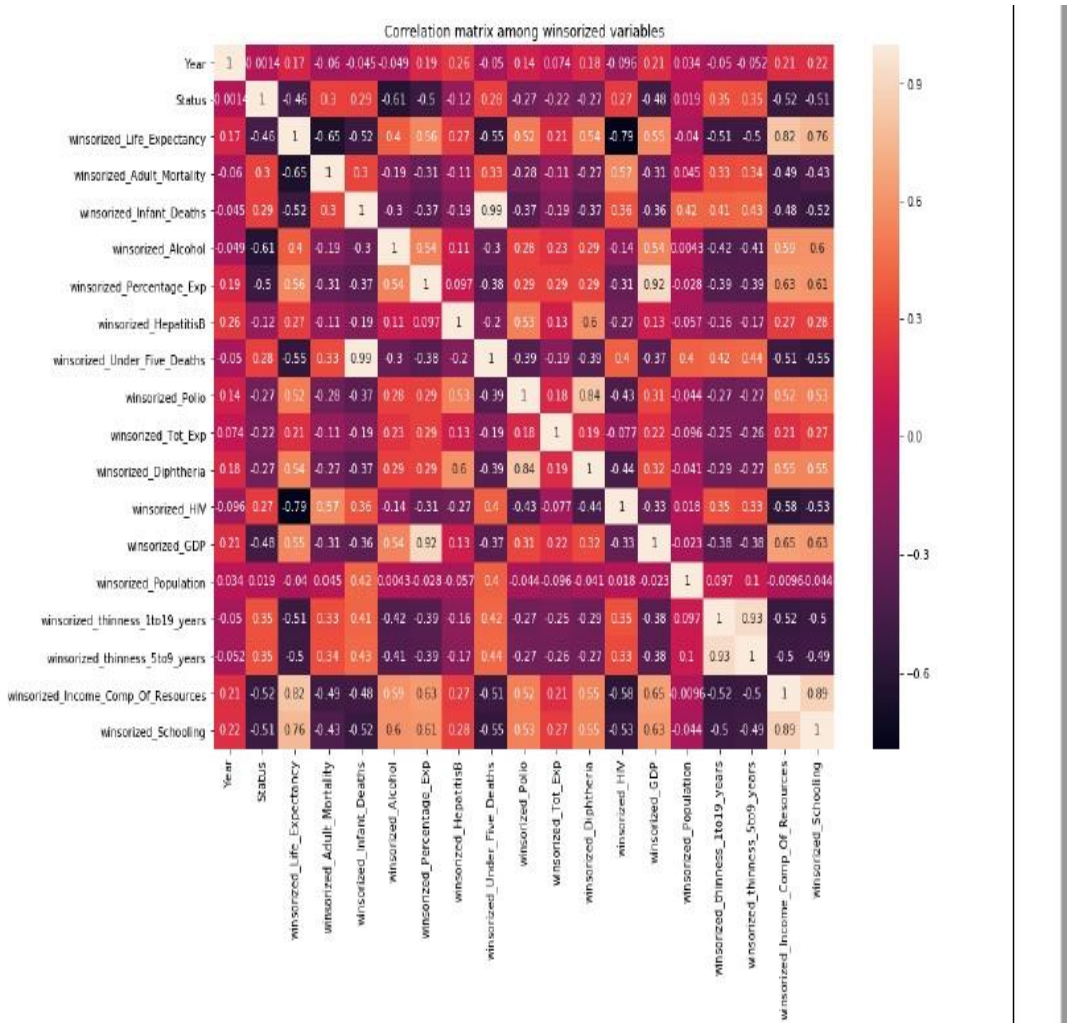
- Programming Language: Python
- Machine Learning libraries (NumPy, SciPy, matplotlib, scikit-learn, pandas)
- Cloud Service: IBM Cloud
- IDE like Jupyter
- IBM Watson
- Node-red App to create WEB UI


Node-RED
Deploy
filter nodes
Flow 1
info
common
inject
debug
complete
catch
status
link in
link out
comment
function
function
switch
change
range
http request
pre_token
form
http request
msg.payload
LIFE-EXPECTANCY
msg.payload
Information
Node
"a76b23a6.c3df3"
Type
function
show more
Description
Node Help
A JavaScript function block to run against the messages being received by the node.
The messages are passed in as a JavaScript object called msg.
By convention it will have a msg.payload property containing the body of the message.
The function is expected to return a message object (or multiple message objects), but can choose to return nothing in order to halt a
ctrl click in the workspace to open the quick-add dialog

4. Experimental Investigations

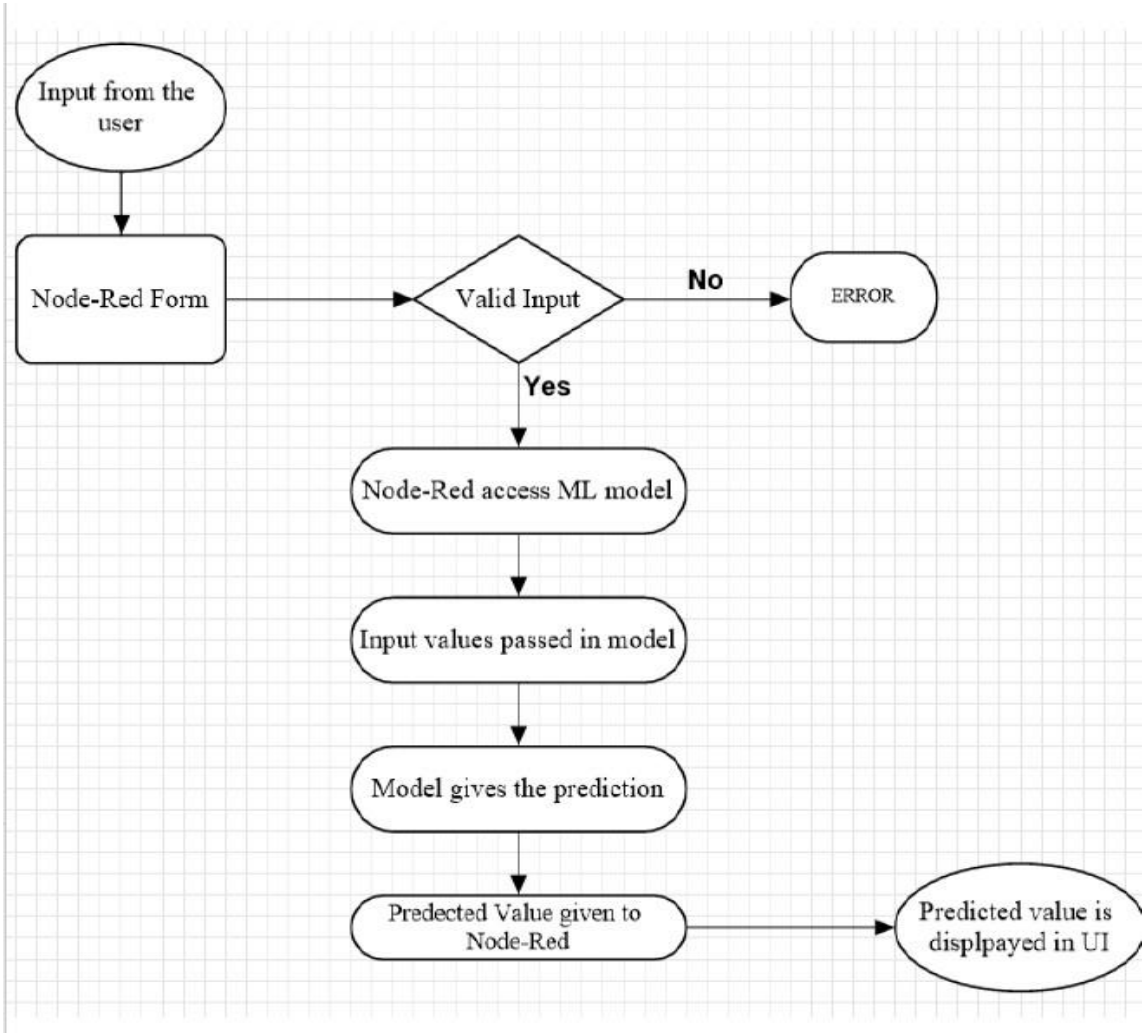
The WHO data set for life expectancy contains 22 columns in total. The data set has historical data of life expectancy for the period between 2000 to 2016. There are numerous rows with null values for some of the columns and outliers. These null values are replaced with the mean for the respective columns. All the columns except the "Country" and "Status" are of integer type. The country column is not included in the model training because of its less relation to the "Life Expectancy" column. The "Status" column is changed to integer type such that "Developing" is mapped to 0 and "Developed" is mapped to 1. After these steps, the data is ready for the model to be trained.

Analyzing the relations between various features can help us improve the performance of the model as well as decide which model would be more suitable. We decided using many visualizations.

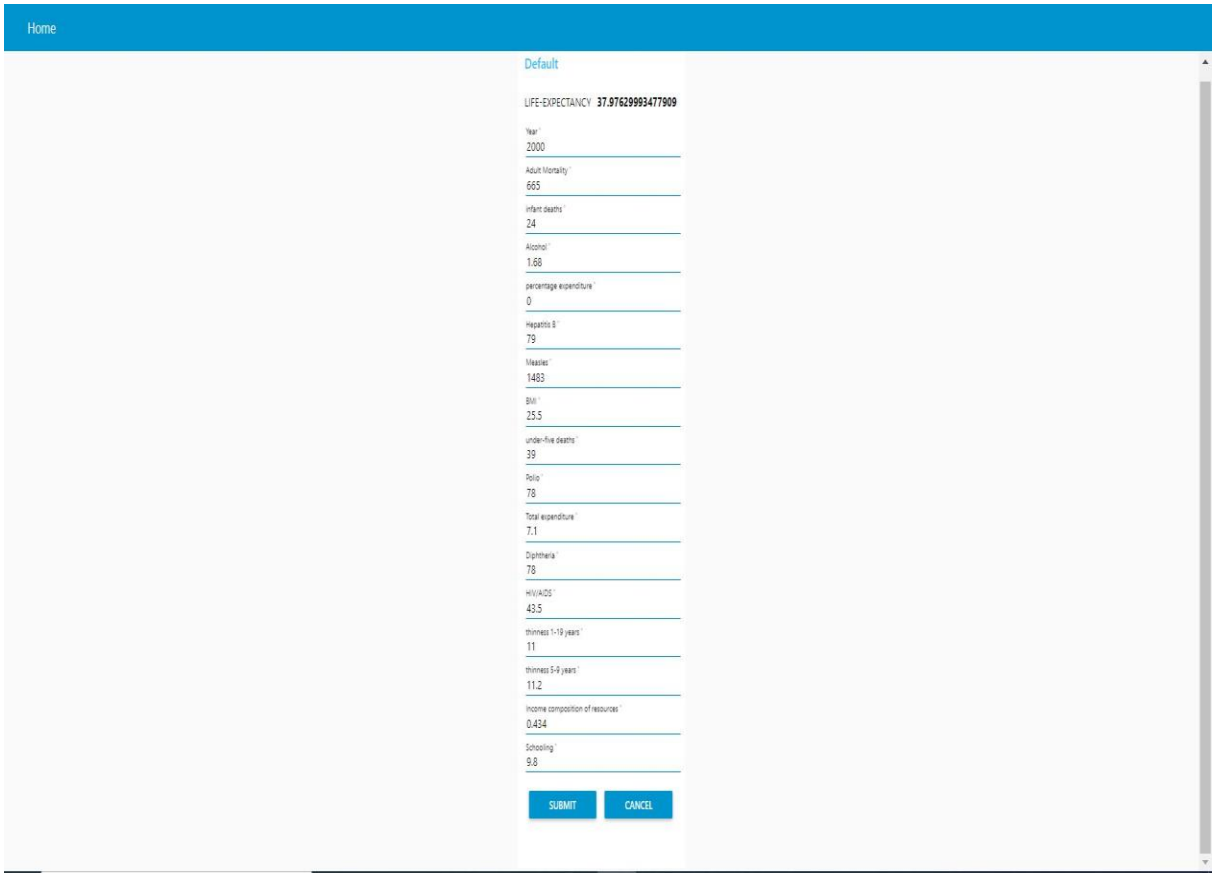


Heat map of dataset

5. Flowchart


Input from the user
Node-Red Form
Valid Input
No
ERROR
Yes
Node-Red access ML model
Input values passed in model
Model gives the prediction
Predected Value given to Node-Red
Predicted value is displpayed in UI

6. Result


Home
Default
LIFE-EXPECTANCY 37.97629993477909
Year
2000
Adult Mortality
665
infant deaths
24
Alcohol
1.68
percentage expenditure
0
Hepatitis B
79
Measles
1483
BMI
25.5
under-five deaths
39
Polio
78
Total expenditure
7.1
Diphtheria
78
HIV/AIDS
43.5
thinness 1-19 years
11
thinness 5-9 years
11.2
Income composition of resources
0.434
Schooling
9.8
SUBMIT
CANCEL

7. Advantages and Disadvantages

Advantages:

One of the most significant advantages of embedding machine learning algorithms is their ability to improve over time. Machine learning technology typically enhances 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 more manageable. As the amount of data used increases, accuracy also enhances with it. It is also the critical component in technologies for automation.

Using Node-Red also simplifies the effort put into a creating the front-end. The programmer doesn't need extensive knowledge of HTML and JavaScript. It also makes the integration between Machine learning model and the UI much more comfortable.

Disadvantages:

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

Machine learning can also be very time-consuming. When the size of the data fed to the machine learning is enormous, the computational cost and the time taken to train the model on the data increases drastically. It can increase the value of resources required to implement the application on a large scale.

At the same time, Node-Red does not give many features to customize our UI.

8. Applications

1) Personalized Life Expectancy: Individuals can predict their life expectancy by inputting values in the corresponding fields. Which could help make people more aware of their general health, and its improvement or deterioration over time. It may motivate them to make healthier lifestyle choices.

2) Government: It could help the government bodies take appropriate measures to control the population growth and also direct the utilization of the increase in human resources and skill set acquired by people over many years. Increase in life expectancy also leads to a rise in the "manpower" of a nation.

3) Health Sector: The outcome will help the healthcare sector to fund and provide better services to those in need.

4) Insurance Companies: Insurance sector will be able to provide individualized services to people based on life expectancy outcomes and factors.

9. Conclusion

Predicting the lifespan of human beings can significantly alter our lives. Human behavior and activities are unpredictable, and it may almost be impossible to predict lifespan correctly. However, with the help of Machine learning algorithms such as Regression models, we can get close to predicting a roundabout value.

This breakthrough can widely impact health sectors and economic sectors by improving the resources, funds and services provided to the familiar people. It can also increase the ease of access to individuals.

With the help of Machine Learning algorithms, one can ease the process of automating the application and predicting the expectancy with reasonable accuracy. It also reduces the effort and time put into deploying the application and making it more accessible to the users.

10. Future Scope

For future use, one can integrate the life expectancy prediction by providing suggestions and medications to the individual using the application. It will help predict as well as increase the individual's life expectancy.

The scalability and flexibility of the application will also improve with the advancement in technology.

Also, with the growth in Artificial Neural networks and deep learning, one can integrate that with our existing application. With the help of Convolutional Neural networks and Computer vision, we can also try to take into account the physical health and appearance of a person. Mental health can also be taken into account while predicting life expectancy with the help of sentiment analysis systems as well.

11. Bibliography

- https://developer.ibm.com/tutorials/how-to-create-a-node-red-starter-application/
- https://bookdown.org/caoying4work/watsonstudio-workshop/jn.html
- https://bookdown.org/caoying4work/watsonstudio-workshop/jn.html#deploy-model-as-webservice
- https://www.ibm.com/watson/products-services
- https://medium.com/swlh/predicting-life-expectancy-w-regression-b794ca457cd4

12. Appendix

12.1. Source Code

```
'https://eu-gb.ml.cloud.ibm.com/v3/wml_instances/f4bd37cf-5dd1-45e4-8018-
d98f70719514/deployments/a1f8be0a-40fb-497f-868f-d46279a374ad/online'
```

Services Used:

- **Watson Studio**
- **IBM Cloud Function**
- **Node-Red**