Project Report

on

PREDICTING LIFE EXPECTANCY
USING MACHINE LEARNING

By : SOHAM R. SAHARE

INTRODUCTION

Overview:

A typical Regression Machine Learning project leverages historical data to predict insights into the future. This problem statement 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. This problem statement 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.

Purpose:

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 problem statement 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.

LITERATURE SURVEY

Existing Problem:

This problem statement 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.
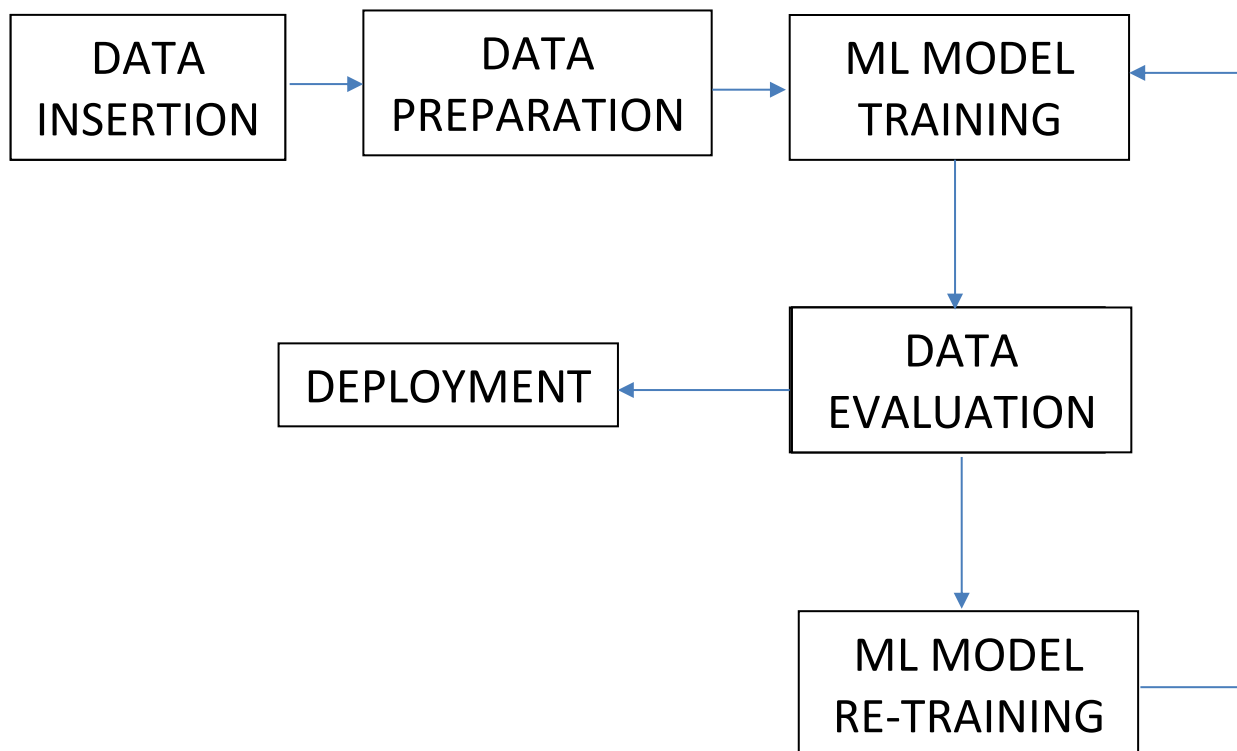
Life expectancy is a statistical measure of the average time a human being is expected to live, Life expectancy depends on various factors: Regional variations, Economic Circumstances, Sex Differences, Mental Illnesses, Physical Illnesses, Education, Year of their birth and other demographic factors.

Proposed Solution:

Designed a Regression model to predict Life Expectancy on some features such as year, GDP, education, alcohol intake, expenditure, disease related deaths in any country.

THEORITICAL ANALYSIS

Block Diagram:

```
┌──────────────┐     ┌──────────────┐     ┌──────────────┐
│     DATA     │ ──> │     DATA     │ ──> │   ML MODEL   │ <──┐
│  INSERTION   │     │ PREPARATION  │     │   TRAINING   │    │
└──────────────┘     └──────────────┘     └──────────────┘    │
                                                  │           │
                                                  v           │
                          ┌──────────────┐   ┌──────────────┐ │
                          │  DEPLOYMENT  │ <─│     DATA     │ │
                          └──────────────┘   │  EVALUATION  │ │
                                             └──────────────┘ │
                                                  │           │
                                                  v           │
                                             ┌──────────────┐ │
                                             │   ML MODEL   │ │
                                             │  RE-TRAINING │─┘
                                             └──────────────┘
```

Hardware and Software designing:

- **Hardware** : Desktop / Laptop, Internet Connectivity

- **Software** : IBM Cloud, IBM Watson Studio, Node-Red App.

EXPERIMENTAL INVESTIGATION

1) **Choose a Project :**
   Predicting Life Expectancy of a person

2) **Collection of Dataset :**
   https://www.kaggle.com/kumarajarshi/life-expectancy-who

3) **Hypothesis :**
   Based on our study and information gathered we can predict the average age of a person.

4) **Design :**
   Construct various Machine Learning Models and finally selecting the model with maximum accuracy.

5) **Conclusion :**
   Model will be able to predict the life expectancy of a person with maximum accuracy.

FLOWCHART

A flowchart is a diagram that depicts a process, system or computer algorithm. They are widely used in multiple fields to document, study, plan, improve and communicate often complex processes in clear, easy-to-understand diagrams. Flowcharts, sometimes spelled as flow charts, use rectangles, ovals, diamonds and potentially numerous other shapes to define the type of step, along with connecting arrows to define flow and sequence.
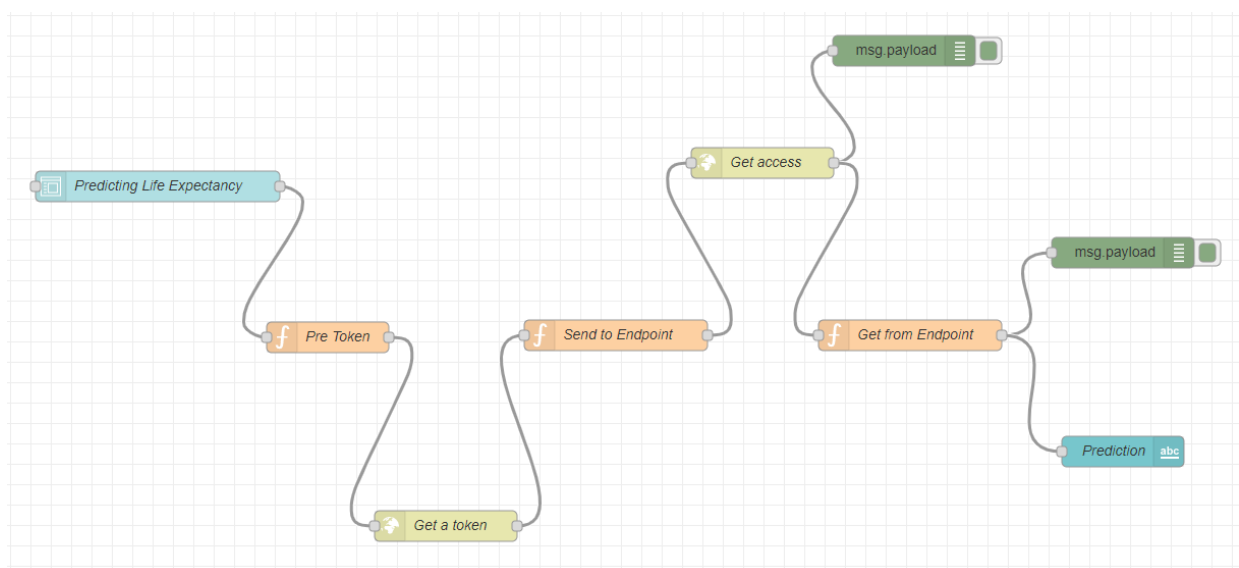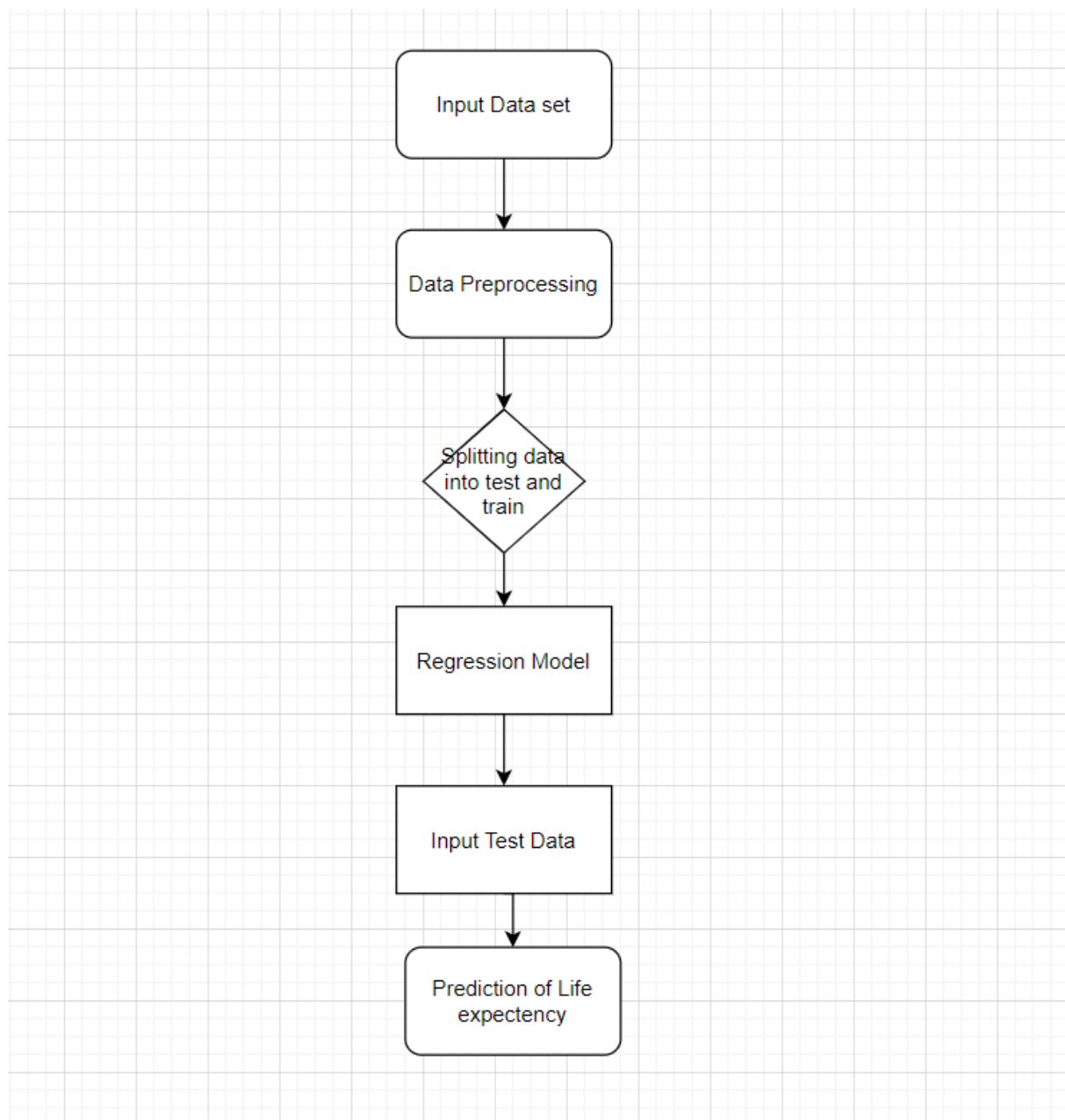

Figure 1

Figure 2

RESULT

Machine Learning Model

Prediction :
75.11941902818847

Inputs

Year
2008

Adult Mortality
1

Infant Deaths
1

Alcohol
5.61

Percentage Expenditure
36.622

Hepatitis B
99

Measles
0

BMI
52.6

Under five deaths
1

Polio
99

Total Expenditure
5.87

Diphtheria
99

HIV/AIDS
0.1

GDP
437.539

Population
2947314

thinness 1-19 years
1.6

thinness 5-9 years
1.6

Income composition of resource…
0.713

Schooling
12

Developing

SUBMIT   CANCEL

The model appears to the user in the form of an interface as shown.

The user has to fill in the inputs and click on "SUBMIT" button at the end of the form.

On clicking the "SUBMIT" button, the user will be displayed the predicted life expectancy, based on the inputs provided, at the top of the page as shown.

ADVANTAGES AND DISADVANTAGES

- ## Advantages :
  1) IBM Watson:
     - a. Process Unstructured Data.
     - b. Full fill Human Limitations.
     - c. Improves performance and abilities.
     - d. Handles large amount of data.
  2) Easy User Interface (UI).
  3) User – friendly.
  4) Easy to predict.
  5) Don't require storage space.

- ## Disadvantages :
  1) IBM Watson:
     - a. Only in English Language.
     - b. Maintenance.
  2) Requires Internet Connectivity.

APPLICATIONS

Life expectancy is the primary factor in determining an individual's risk factor and the likelihood they will make a claim. Insurance companies consider age, lifestyle choices and several other factors when determining premium rates for individual life insurance policies. It can be used by researchers to make meaningful researches out of it and thus, bring about something that will help increase the expectancy consider the impact of a specific factor on the average lifespan of people in a specific country.

CONCLUSION

Thus, we have developed a model that will predict the life expectancy of a specific demographic region based on the inputs provided.
 Various factors have a significant impact on the life span such as Adult Mortality, Population, Under 5 Deaths, Thinness 1-5 Years, Alcohol, HIV, Hepatitis B, GDP, Percentage Expenditure and many more.
User can interact with the system via a simple user interface which is in the form of a form with input spaces which the user needs to fill the inputs into.

FUTURE SCOPE

As future scope, we can connect the model to the database to have the record of predictions. This will help us analyze the trends in the life span. A model with country wise bifurcation can be made, which will help to segregate the data demographically.

BIBLOGRAPHY

Appendix : Source Code

SOURCE CODE LINK