SUMMER INTERNSHIP REPORT

(16/05/2020 – 15/06/2020)

PROJECT TITLE:

"Predicting life expectancy using machine learning"

SUBMITTED BY:

Deepali manjunath joshi

INDEX

CHAPTER 1
INTRODUCTION

1.1   OVERVIEW

Life expectancy is an important factor which refers to the number of years a person is expected to live based on the statistical average. This factor varies by geographical area and by era. Life expectancy factor depends on various different parameters such as the year of its birth, its current age, and other demographic factors including gender. The data are then used to draw a life table. These tables can be used to predict how likely it is that a person of a given age will die before their next birthday. Life expectancy is based on an estimate of the average age that members of a particular population group will be when they die. Life expectancy varies by geographical areas. Life expectancy is predicted for every country to check the overall health of a community. The benefit of knowing life expectancy is to evaluate the performance of the government in improving the welfare of the population in general and improving health status in particular. Low life expectancy in an area must be followed by health development programs, and other social programs including environmental health, nutritional and calorie adequacy including poverty eradication programs. This project tests the potential of using machine learning techniques for predicting life expectancy from previous medical records.

1.2   PURPOSE

Life expectancy of every country is calculated to check the overall health of the country. The purpose of this project is to build a machine learning model which will estimate or predict life expectancy with the help of Life Expectancy Dataset provided by WHO (World Health Organization) which includes attributes such as country, year, alcohol intake, population, etc. and will give high accuracy output.

1. Predicting Life Expectancy in various region by providing different inputs.
2. To check accuracy of life expectancy rate using machine learning
3. It will be useful for government to take appropriate measures to control the population growth.
4. To analyze which factor affects most to the life expectancy.

CHAPTER 2
LITERATURE SURVEY

2.1 EXISTING PROBLEM

Life expectancy reflects local conditions. In less-developed countries, life expectancy at birth is relatively low, compared with more-developed countries. In some less-developed countries, life expectancy at birth may be lower than life expectancy at age 1, because of high infant mortality rates (commonly due to infectious disease or lack of access to a clean water supply).

Life expectancy is calculated by constructing a life table. A life table incorporates data on age-specific death rates for the population in question, which requires enumeration data for the number of people, and the number of deaths at each age for that population. Those numbers typically are derived from national census and vital statistics data, and from them the average life expectancy for each of the age groups within the population can be calculated.

Another life expectancy calculation is healthy life expectancy (or disability-free life expectancy), which is the average number of years a person is expected to live in good health, or without disability, given current age-specific mortality rates and disease and disability prevalence rates. Calculation of those figures requires reliable health statistics as well as mortality and census data.

2.2 PROPOSED SOLUTON:

In this project I have proposed two solutions one is with machine learning and one is using auto AI experiment. Machine learning is the process of training the computer to think and decide solutions like human. In machine learning I have used linear regression model for prediction. For this solution we have used the dataset consisting of various factors. In this system we have taken all the correlated features into consideration. Hence the output will give an average life expectancy prediction by considering all different factors. This project uses mortality factors, economic factors and many more to predict the life expectancy of a country for a given year using machine learning model.

CHAPTER 3
THEORITICAL ANALYSIS

3.1 BLOCK DIAGRAM
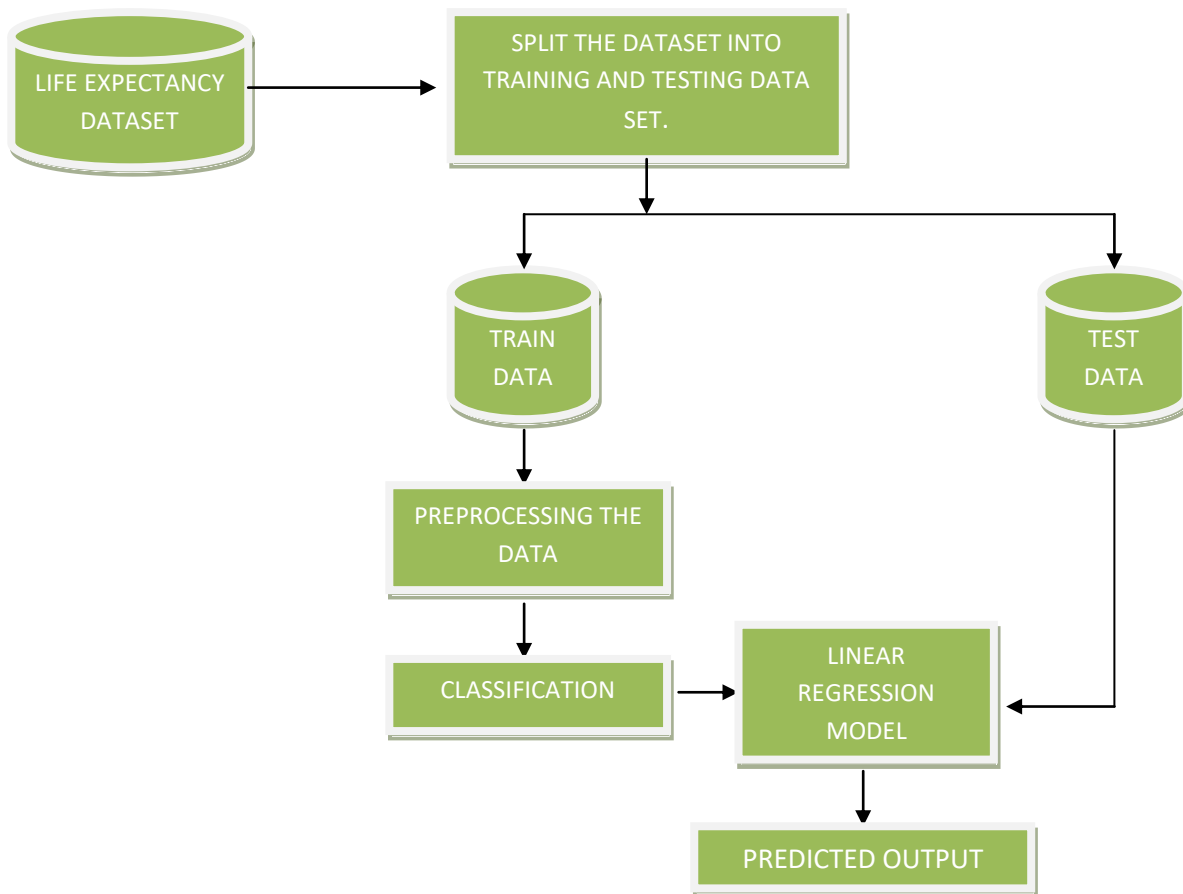
3.2 HARDWARE / SOFTWARE DESIGNING

Software Requirements:

1. IBM cloud access.
2. Watson studio
3. Node red app
4. Machine Learning service

Technical Requirements:

1. Machine Learning
2. WATSON Studio
3. Python
4. Node-Red

CHAPTER 4

EXPERIMENTAL INVESTIGATION

EXPERIMENTAL INVESTIGATION:

A] DATASET:

The Life Expectancy Data is provided by world health organization (WHO) which has different attributes:

- Country
- Year
- 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

B] EXPERIMENTAL SETTING:

1) Preprocess the Dataset:

Before applying any machine learning algorithm to a dataset, it is very important to clean the dataset and preprocess it. Data cleaning and preprocessing will remove the unwanted data and it will also deal with the null values present in it. In this project I have filled null values with 0.

2) Decide an algorithm:

In this project I have used Linear Regression algorithm we predict scores on one variable from the scores on a second variable.
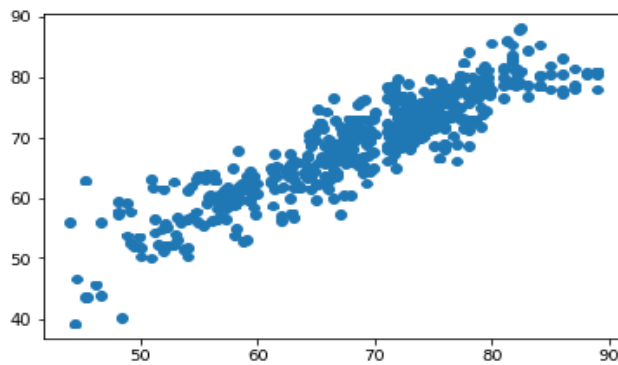

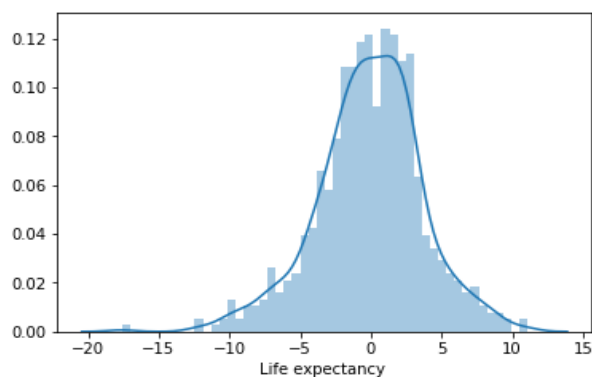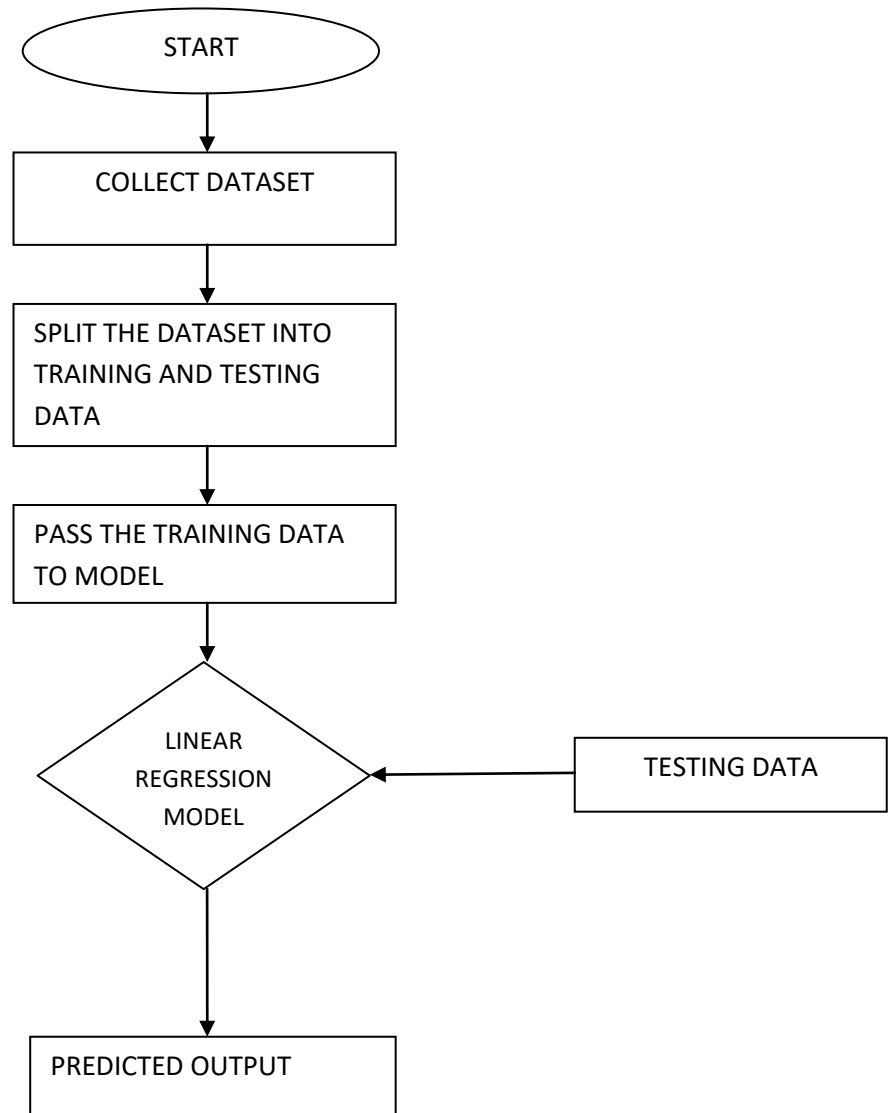Figure 1: Scatter plot of Life expectancy prediction

Figure 2: Distplot of life expectancy prediction

CHAPTER 5

FLOWCHART

```
          ┌─────────────┐
          │    START    │
          └──────┬──────┘
                 │
                 ▼
      ┌────────────────────┐
      │   COLLECT DATASET  │
      └─────────┬──────────┘
                │
                ▼
      ┌────────────────────┐
      │ SPLIT THE DATASET  │
      │ INTO TRAINING AND  │
      │   TESTING DATA     │
      └─────────┬──────────┘
                │
                ▼
      ┌────────────────────┐
      │ PASS THE TRAINING  │
      │   DATA TO MODEL    │
      └─────────┬──────────┘
                │
                ▼
         ◇ LINEAR ◇
         REGRESSION  ◄──────── TESTING DATA
           MODEL
                │
                ▼
      ┌────────────────────┐
      │  PREDICTED OUTPUT  │
      └────────────────────┘
```

START

COLLECT DATASET

SPLIT THE DATASET INTO TRAINING AND TESTING DATA

PASS THE TRAINING DATA TO MODEL

LINEAR REGRESSION MODEL

TESTING DATA

PREDICTED OUTPUT

CHAPTER 6
RESULT

RESULT:

I trained my model using linear regression and then deployed the model in a web based UI which was developed by integrating all the services using NODE-RED. Following images shows the UI of my model in which I have passed some random values as input.

CHAPTER 7
ADVANTAGES &
DISADVANTAGES

ADVANTAGES:

- Machine learning technology typically improves efficiency and accuracy due to increasing amounts of data that are processed.
- Node-red app is a very user friendly for handling.
- Machine Learning can review large volumes of data and discover specific trends and patterns that would not be apparent to humans.

DISADVANTAGES:

- ML needs enough time to let the algorithms learn and develop enough to fulfill their purpose with a considerable amount of accuracy and relevancy. It also needs massive resources to function. This can mean additional requirements of computer power for you.

CHAPTER 8
APPLICATIONS

APPLICAITIONS:

- This Project can be used as a personal health application.
- This project will also help government to take appropriate measures regarding life expectancy and to control population growth.
- Supporting timely recognition of the right moment to start Advance Care Planning.

CHAPTER 9
CONCLUSION
AND FUTURE SCOPE

CONCLUSION:

I have developed a web application for predicting life expectancy using node red app of IBM cloud. The backend of this project is linear regression model created and deployed on Watson Studio using machine learning model. Predicting exact and completely accurate rate is difficult. Additionally, human prognostication is costly, time-consuming, requires medical expertise, and is a subjective task. Without compromising prediction accuracy, the model is able to make predictions quickly, automatically and systematically, while it does not depend on human medical expertise.

CHAPTER 11
BIBILOGRAPHY &
APPENDIX

APPENDIX

- **PROJECT DETAILS:**

https://github.com/SmartPracticeschool/llSPS-INT-1681-Predicting-Life-Expectancy-using-Machine-Learning