PROJECT REPORT DOCUMENT

PREDICTING LIFE EXPECTANCY USING MACHINE LEARNING

Submitted in fulfillment for the completion of

the project at Smartbridge

Submitted by

VALAVALA VENKATA LAKSHMI GAYATRI

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 including: financial status, regional variations, economic circumstances, gender differences, mental illnesses, physical illnesses, education, year of their birth, demographic factors and many more.

This project is aimed at predicting Life Expectancy rate of a country, given various factors, using Machine Learning algorithms.

In this project, we have taken the dataset related to life expectancy.Health factors for 193 countries has been collected from the WHO data repository website and its corresponding economic data was collected from the United Nations website.

1.2.Purpose

Since the observations are based on different countries, it will be easier for a country to determine the predicting factor which is contributing to lower value of life expectancy. This will help in suggesting a country which area should be given importance in order to efficiently improve the life expectancy of its population.

2.LITERATURE SURVEY

2.1.Existing Problem:

Life expectancy is a statistical measure of the average time a human being is expected to live. This is based on various health, education and economic factors.
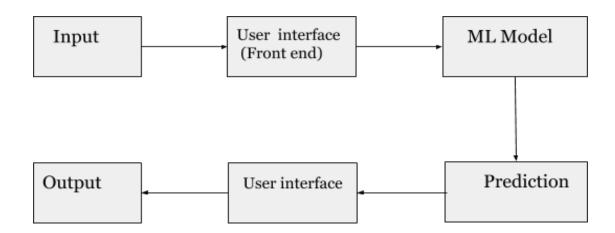
Although there have been a lot of studies undertaken in the past on factors affecting life expectancy considering demographic variables, income composition and mortality rates. It was found that the effect of immunization and human development index was not taken into account in the past.

2.2.Proposed Solution:

Design a regression model that takes into consideration all the available factors to predict accurate average life expectancy

3.THEORETICAL ANALYSIS

3.1.Block Diagram

3.2 Software /Hardware designing:

SOFTWARE REQUIREMENTS:-

- IBM Cloud

- IBM Watson Studio

- Node-red App

This problem statement is aimed at predicting average life expectancy of people living in a country given various features.

Firstly,the required data set is collected from https://www.kaggle.com/kumarajarshi/life-expectancy-who . Secondly, an IBM Watson studio project is created in the Watson Studio service, provided by IBM Cloud. Then, the data set is imported into this project and a machine learning model is formulated.This model can be opened in the Watson Studio project and tested giving the inputs.

A User Interface (UI) is designed for the project using Node-RED. In this User Interface the users can type in data regarding their country's gdp, alcohol intake, diseased count, deaths count, income and expenditure, education level, etc. and get the output which is the predicted life expectancy of the country.

4.EXPERIMENTAL INVESTIGATIONS:

1. Choose a Project Idea:

Predicting Life Expectancy of a person.

2. Conduct Background Research:

https://www.kaggle.com/kumarajarshi/life-expectancy-who

https://ourworldindata.org/life-expectancy-how-is-it-calculated-and-how-should-it-be-interpreted

3. Compose a Hypothesis:

Based on our study and information gathered we can predict the average age of a person.
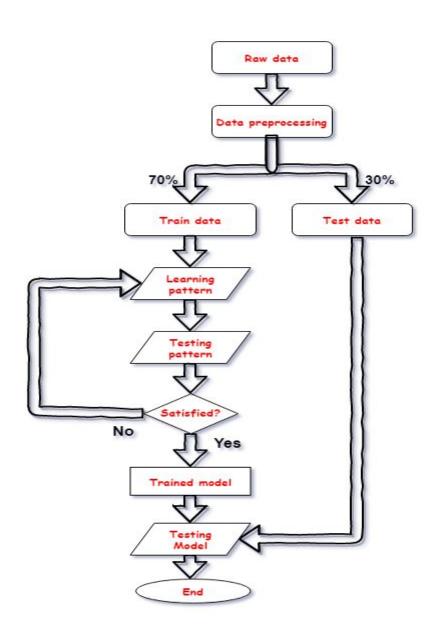
4. Design your Experiment:

First we need to collect the suitable data for our problem statement. Next we need to construct the regression model for this problem.

5. Draw Conclusions:

After construction of our model, we can predict the average life expectancy of a person.

5.FLOW CHART:

6.Ouput:

Default

Prediction	64.95333333333333

Status
1

Adult Mortality
89

Alcohol
98

percentage expenditure
98

Hepatitis B
241

BMI
36.7

under-five deaths
102

Polio
80

Total expenditure
67

Diphtheria
89

HIV/AIDS
80

GDP
900

thinness 1-19 years
36

thinness 5-9 years
36

Income composition of resources
73

Schooling
9.2

SUBMIT   CANCEL

Advantages:

1. It is used for life expectancy prediction.
2. It can be used to improve the factors that affect life expectancy.
3. It is helpful in medical fields.

Disadvantages:

1. Data can be wrong which leads to inaccurate predictions.
2. Accuracy is less.

APPLICATIONS:

1. The project can help governments to keep track of their countries' health status so they can plan for the future accordingly
2. The government can plan and develop their health infrastructures by keeping the most correlated factors in mind.
3. The project can be used as a basis to develop personalized health applications.

FUTURE SCOPE:

Some future improvements can be made on current project. They are as follows:

1. As more data comes, that can be fed to the model for more accurate predictions.

2. Currently, the project is just a web application. It can be developed to support other platforms like Android, IOS and Windows Mobile.

3. Other regression models can also be used for prediction and later the best among them should be chosen.

4. User interface can be modified for various countries according to their regional languages.

CONCLUSION:

The end product is an web page created and deployed on node-red app of IBM cloud. The backend of web page is a Decision Tree Regressor Model with 93 R2 score created and deployed on watson studio using machine learning service.

The web-page has input fields similar to dataset columns such as Country, BMI, percentage expenditure, Alcohol etc and an output field named as prediction i.e similar to dataset column Life expectancy which gives the life expectancy prediction based on the given values.

BIBLIOGRAPHY:

● https://www.youtube.com/

● Dataset: https://www.kaggle.com/kumarajarshi/life-expectancy-who

● IBM Tutorials: https://developer.ibm.com/tutorials/

● Node-RED Tutorials: https://nodered.org/docs/tutorials/

APPENDIX:

● Code:https://github.com/SmartPracticeschool/llSPS-INT-1621-Predicting-Life-Expectancy-using-Machine-Learning

● Node-RED Flow:https://node-red-pcmzv.eu-gb.mybluemix.net/ui/