PROJECT REPORT OF PREDICTING LIFE EXPECTANCY USING MACHINE LEARNING

Project Manager:SAKSHI JAIN
Date: 15/06/2020

CONTENTS:

1.INTRODUCTION:

1.1 Overview:

Life expectancy is the average number of years a person in a population could expect to live after age x. It is the life table parameter most commonly used to compare the survival experience of populations. The age most often selected to make comparisons is 0.0 (i.e., birth), although, for many substantive and policy analyses, other ages such as 65+ and 85+ are more relevant and may be used (e.g., for determining person-years of Medicare and Social Security benefit entitlement).

In order to predict life expectancy rate of a given country, we will be using Machine Learning algorithms to draw inferences from the given dataset and give an output. For better usability by the customer, we are also going to be creating a UI for the user to interact with using Node-Red.

1.2 Purpose

Life expectancy is perhaps the most important measure of health. Life expectancy increases due to healthcare improvements like the introduction of vaccines, the development of drugs or positive behavior changes like the reduction in smoking or drinking rates.

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.

2. LITERATURE SURVEY

2.1 Existing Solution:

As a result of the evolution of biotechnologies and related technologies such as the development of sophisticated medical equipment, humans are able to enjoy longer life expectancies than previously before. Predicting a human's life expectancy has been a long-term question to humankind. Many calculations and research have been done to create an equation despite it being impractical to simplify these variables into one equation.

Currently there are various smart devices and applications such as smartphone apps and wearable devices that provide wellness and fitness tracking. Some apps provide 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 existing works provide the Personalized Life expectancy.

2.2 Proposed Solution:

he project tries to build a model based on the given dataset. The first step was to clean the data, this included detecting and dealing with both missing values and outliers. The variables and dataset were given a general description so that a better understanding of what the variables mean could be gathered. Then both explicit and inexplicit missing values were detected. Inexplicit missing values were values that didn't make sense for a variable given the nature of the data.

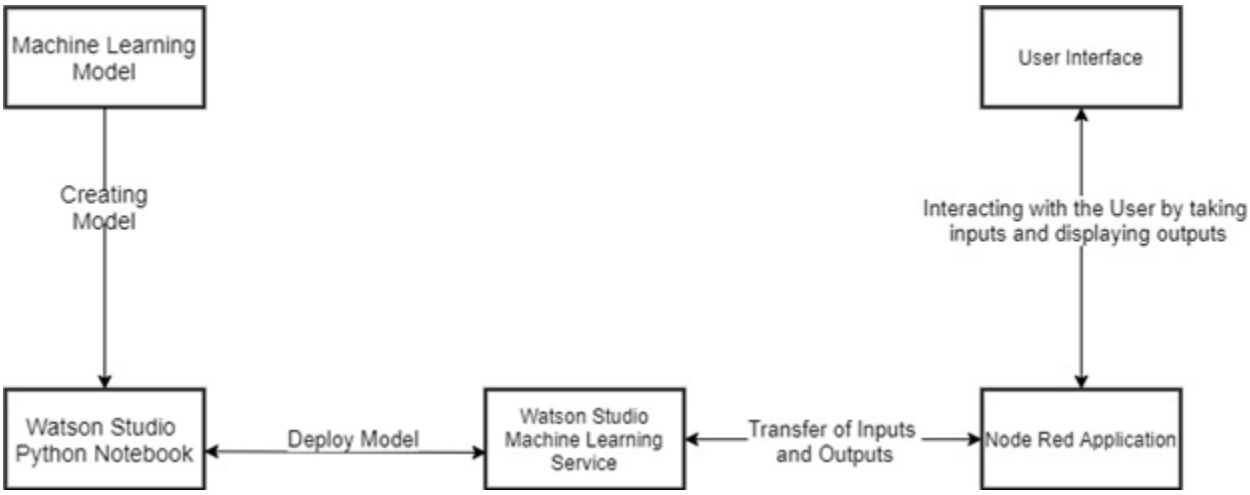
Our first step in this project is DATA PRE PROCESSING , is a crucial step that helps enhance the quality of data to promote the extraction of meaningful insights from the data. It refers to the technique of preparing the raw data to make it suitable for a building and training. We will drop required columns which will not be used in Regression. Analyzing data sets to summarize their main characteristics, often with visual methods is done. We build coefficient matrix and we obtain boxplots to analyze the outliers.

We train our regression models we will need to first split up our data into an A

array that contains the features to train on, and a B array with the target variable, here it is "Life Expectancy column". We split the data into a training set and a testing set. We will train out model on the training set and then use the test set to evaluate the model and the best model is chosen to evaluate the predictions.

3. THEORITICAL ANALYSIS

3.1 Block Diagram:



3.2 Hardware/Software Designing:

Project Requirements:

i) Functional Requirements:

To be able to predict the life expectancy accurately using Machine Learning models.

ii)Technical Requirements:

Any working laptop/PC with minimum 2.2Ghz processor and at least 8GB of memory with an Internet connection.

iii) Software Requirements:

a) Python

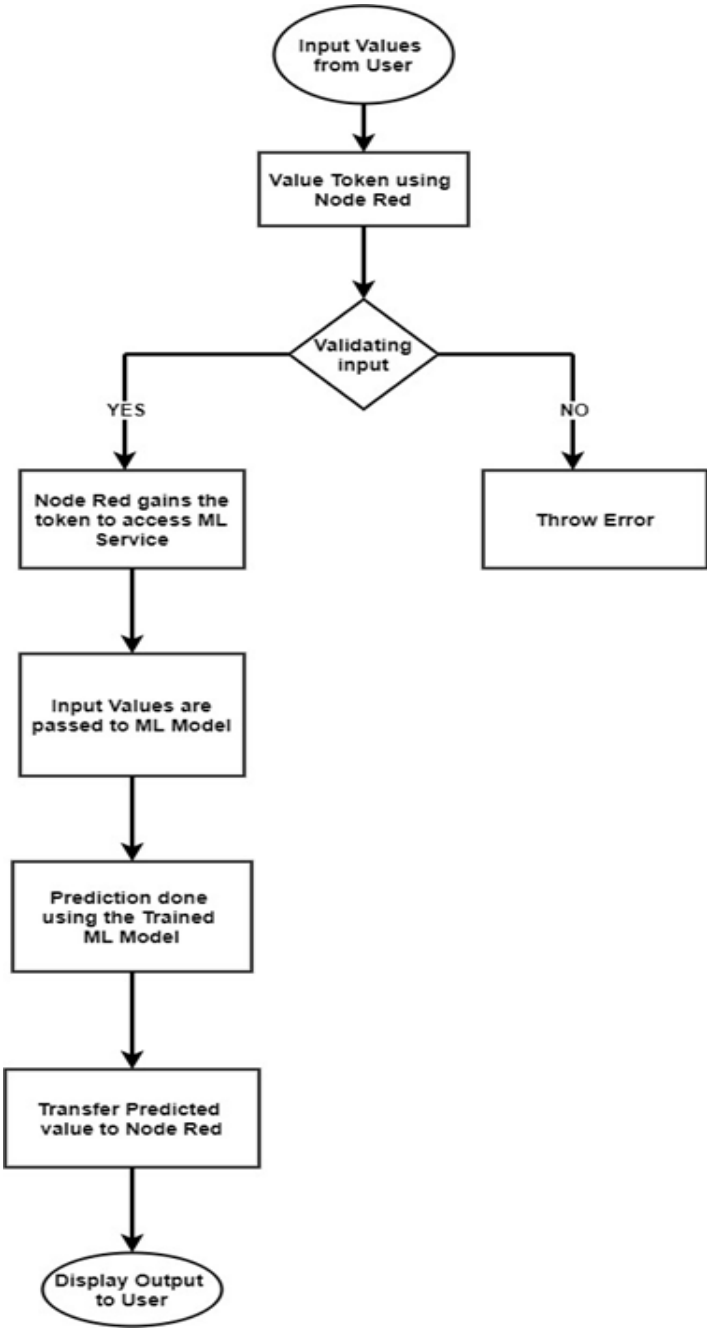
b) IBM Cloud

c) IBM Watson

Model Designing (Watson Studio) :

Steps: Open Watson studio => New Project => Create an empty Project => Give project name => Click Create => Add to Project => Notebook

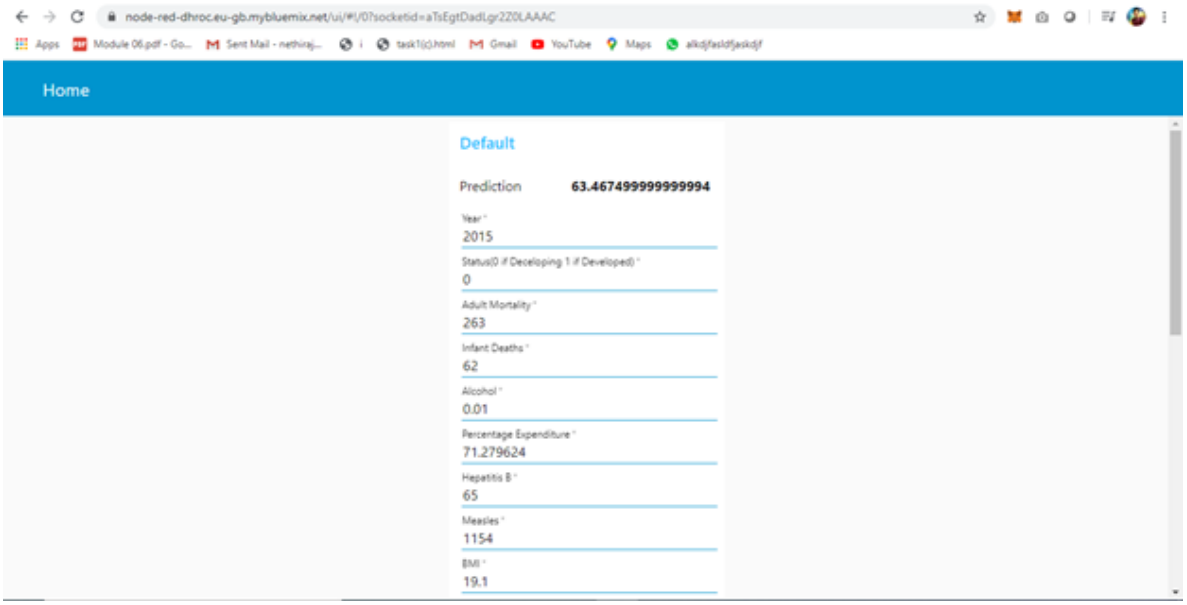
4. FLOW CHART

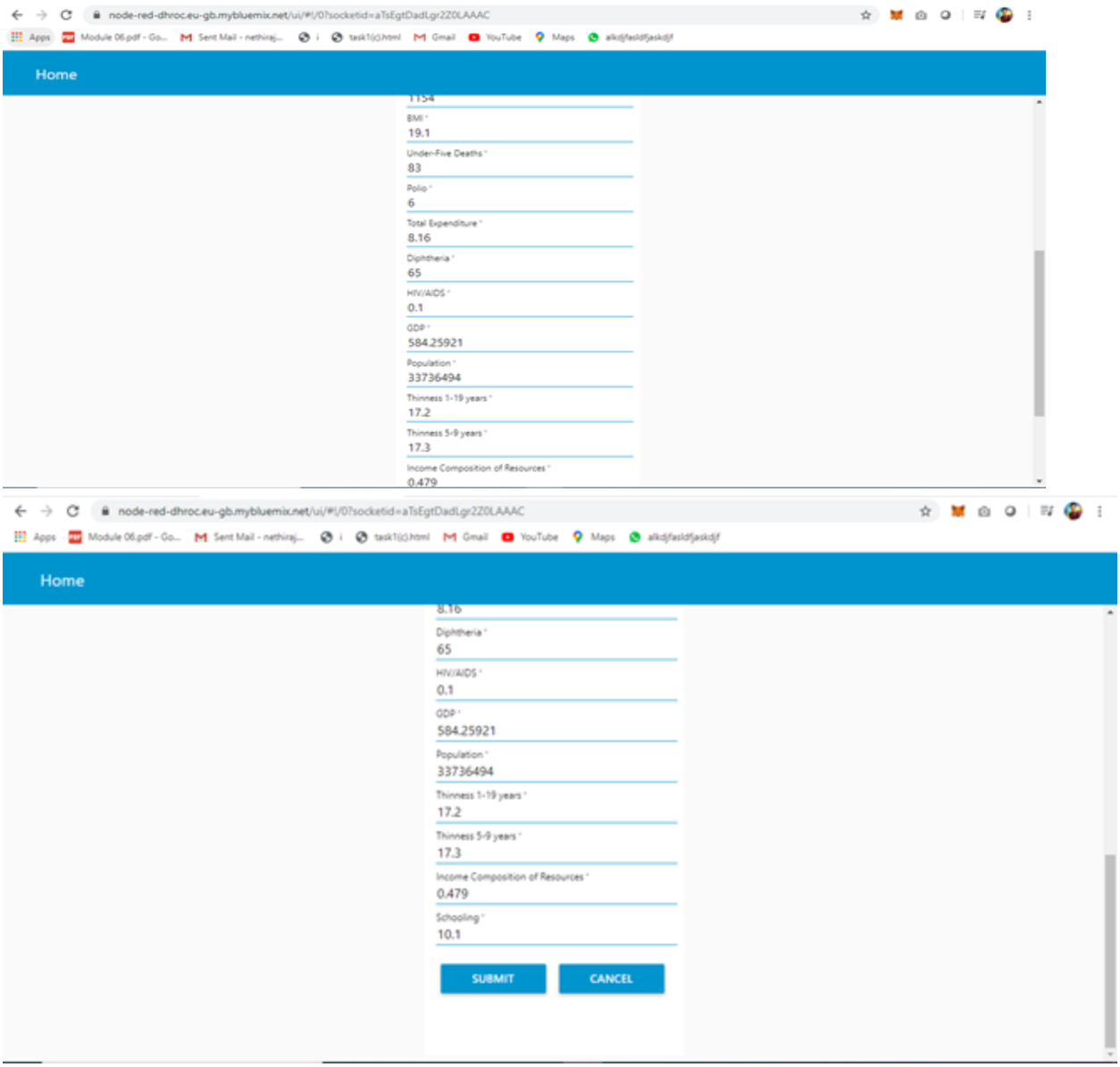


5.Cost and Budget

The costs of production includes infrastructure costs (cloud compute, data storage), integration costs (data pipeline development, API development, documentation) and maintenance costs. We will estimate these costs for a 12 months period.

6. **Result**





7.Limitation

If you feed a **model** poorly, then it will only give you poor results. This can manifest itself in two ways: lack of data, and lack of good data. Many machine learning algorithms require large amounts of data before they begin to give useful results.

8. Conclusion

Our research shows that machine learning techniques offer a feasible and promising approach to predicting life expectancy. The research has potential for real-life applications, such as supporting timely recognition of the right moment to start Advance Care Planning.
Even though the model's performance is far from perfect, we consider this work to be among the first steps in a line of research that has much potential for clinical applications, for several reasons: good prognostication has the potential to contribute significantly to end-of-life decision making, therefore we believe that any increase in prognostic accuracy is worth persuing. Additionally, human prognostication is costly, time-consuming, requires medical expertise, and is a subjective task.