PROJECT REPORT

Project Title: Predicting Life Expectancy using Machine Learning

1) Project Summary

1.1 Project Description

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.

2) Project Requirements

The project requires:

> A diverse and detailed dataset containing Regional variations, Economic Circumstances, Sex Differences, Mental Illnesses, Physical Illnesses, Education, Year of their birth and other demographic factors as its variable input columns and life expectancy as its variable output column.

> A machine learning model (basically a regression model) that takes in various inputs and predicts the life expectancy. It can be made in two ways - using AUTO AI service of IBM cloud and by making a machine learning model using python manually. Creating both will help compare and improve the final model.

> An appropriate computing platform such as IBM cloud containing our dataset and

model.

> A URL for making our model available online (Scoring end point).

> A tool to generate URL for the model to be accessible globally i.e., a Node-Red Application that helps to deploy that model on a web page very efficiently.

3) Functional Requirements

The project requires the following functionalities:

> A front end web application that can interact with the users of our model/application. It should take inputs from the user and return a valid prediction by implementing our model.

> A back end model that takes inputs that were taken from the user to make an accurate and reliable prediction of life expectancy.

> A scoring end point url (In other words, a link to connect our model to the front end) to make our model to be available online on the web for use.

4) Technical Requirements

> The application should be available for everyone to access.

> It should not be very complex to use. Any person who is aware of the basic terminology of life expectancy and knows how to use a computer and internet must be able to use it.

> The predictions should be reliable and accurate so that the companies and organisations that work on such data can use the application.

Our web application is easy to use and manageable. It can work on any operating system with internet connection. However, it is not 100% accurate and the accuracy will differ largely depending on the data that various countries might put into it. Also, problems may occur if a large amount of requests are made at the same time.

5) Software Requirements

The following is required to create a Machine Learning Regression model -
> Python Language (easy and efficient language to write code in)
> Libraries like
> > Pandas (to manage dataframe),
> > Numpy (for numerous calculations),
> > Scikit-Learn (for our regression model, to predict accuracy etc.),
> > Matplotlib & Seaborn (for better visualisations of data as well as results in the form of graphs and pictures).
> An IDE to work on like spyder, jupyter notebook etc.
> IBM cloud account (It provides the IDE, storage space and various other features altogether at a single platform)
> An IBM Watson service (To make a model manually or use the AUTO AI feature)
> An IBM Machine Learning service (To make our Regression Machine Learning Model)
> A Node-Red Application (Also provided on IBM Cloud to deploy our model on web using a .json flow)

6) Project Deliverables

The project delivers an easy way to know the expected life span of a being, given some factors like GDP, Population of the country, any mental or physical health issues the person might be having and the health facilities of the region etc.

7) Project Team

The internship is given to individuals, hence, the team consists of only one person.
Name: Chinmayata Kaushal
Project: Predicting Life Expectancy using Machine Learning

8) Project Schedule

Start: The project was given on 15th of May, 2020.
Duration: 1 month

Schedule:

15th - 17th May:
(3 days)
Familiarizing with the project and creating environments and accounts that will be required throughout the internship.

18th - 24th of May:
(6 days)
Exploring various services provided by IBM - IBM cloud, IBM Watson Studio, IBM Machine Learning Service, Auto AI, Red-Node Application etc. and creating a basic project so that they become easy to use while doing the main project.

25th to 26th May:
(2 days)
To make a linear regression machine learning model on IBM Watson Studio, Machine Learning service manually and changing the various hyper parameters to get the best possible accuracy. Also, to make a Watson Studio service to use its Auto AI feature.

27th to 28th May:
(2 days)
To create a Node-Red cloud foundry app to deploy our model. To create the flow and link the web application and ML model.

29th May:
(1 day)
To write a detailed project report and make a video demonstration.

9) Project Scope

Life expectancy is one of the most important factors in end-of-life decision making. Good prognostication for example helps to determine the course of treatment and helps to anticipate the procurement of health care services and facilities, or more broadly: facilitates Advance Care Planning. Advance Care Planning improves the quality of the final phase of life by stimulating doctors to explore the preferences for end-of-life care with their patients, and people close to the patients. Physicians, however, tend to overestimate life expectancy, and miss the window of opportunity to initiate Advance Care Planning. Machine learning and natural language processing techniques offer a

feasible and promising approach to predicting life expectancy. The research has potential for real-life applications, as in supporting timely recognition of the right moment to start Advance Care Planning.