Project Scope Document

Project Title: Predicting Life Expectancy Using Machine Learning.

Project Team:

Name: Kunal Sukhadev Kokare. (Indivisual)
Email: kunalkokare789@gmail.com

Project Schedule:

Task	Date	Status
Project Scope Document	22/05/2020	Applied
Explore IBM platform	22/05/2020	Applied

Project Scope:

- **Project Summary:**

  "Life Expectancy Prediction" is the project which is developing for measuring the average time of human being expected to live. average time will be calculated on the basis of different kind of factors stated as follows.
  -Variation in region
  -Mental illness
  -Physical illness
  -Alcohol intake
  -Sex differences
  -Year of birth etc.

  On the basis of above all features we could predict the average time of life.

- **Project Requirements:**

1)Machine Learning Algorithm:

Before starting to build the project in machine learning we should have to take a look on different algorithms and choose algorithm which will giving highest accuracy.

2)Dataset:

To working with machine learning algorithm and for prediction something, we take a reference from previously generated data called dataset. it can help us to build a model for prediction. dataset is a basically collection of different attributes & features.

For this project we required dataset containing attributes with values for country, mental illness, alcohol intake percentage and many more.

3)Project Report & Deliverables:

A deliverable is any unique and verifiable product, result, that must be produced to complete a process, phase, or project. Before project start project manager, customer should agree upon result.

- **Technical Requirements:**

As a part of machine learning project we have to perform different steps stated as follows:

1)Loading Packages:

To build this project we required different packages like Numpy, Panda, sklearn. these packages help us to visualize and prediction of result in better manner. Sklearn is the most widely used package for the machine learning process.

2)Reading Data:

As we are going to use dataset for training and testing data we have to read data from dataset which will be most probabaly csv file format. after we can perform different operation on data imported from csv file.

3)Exploratory Data Analysis (EDA):

It is an approach to analyzing data sets to summarize their main characteristics, often with visual methods. A statistical model can be used or not, but primarily EDA is for seeing what the data can tell us beyond the formal modeling or hypothesis testing task.

4)Data Preprocessing:

It is a process of preparing the raw data and making it suitable for a machine learning model.It is an first & crucial step while building a machine learning model. also while building model it is mandatory to clean the data and put in a formatted way.

5)Feature Scalling:

It basically helps to normalise or generalize the data within a particular range. Sometimes, it also helps in speeding up the calculations in an algorithm.

- **Software Requirements:**

1)Jupyter Notebook:

The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text.

2) Anaconda:

Anaconda is popular because it brings many of the tools used in data science and machine learning with just one install, so it's great for having short and simple setup.

3) IBM watson studio:

Watson Studio provides you with the environment and tools to solve your business problems by collaboratively working with data. You can choose the tools you need to analyze and visualize data, to cleanse and shape data.

4) GitHub:

GitHub is a web based graphical interface which provide facility to store, push different conetent like source code, reports , and many more.

- **Hardware Requirements**:

1)GPU:

As we required plot different graphs as well as need to perform different visualization we required hardware called GPU to handle all these tasks. GPU have large accumulated large processing power.

2)CPU:

To training the dataset we require fast processing power .A CPU such as i7–7500U can train an average of ~115 examples/second. if our task is not so big then we don't need to use GPU

3)RAM & Storage:

As of now as per our problem statement we need maximum 4gb RAM and 10 gb storage.

- **Functional Requirements:**

Predicting the life expectancy rate of a country.

- **Project Deliverables:**

1)Project Documentation
2)ML Prediction Model
3)Node red flow diagram
4)Watson auto ml model