

PROJECT REPORT

on

PREDICTING LIFE EXPECTENCY

Using

Machine Learning

Project ID:SPS_PRO_215

Project Platform:SMART BRIDGE

BY: Deepak Raj Diwan

deepakrajdiwano6@gmail.com

CONTENTS

1)Introduction.....	4
1.1)Overview.....	4
1.2)Purpose.....	4
2)Literature Survey.....	5
2.1)Existing Problem.....	5
2.2)Proposed Solution.....	5
3)Theoretical Analysis.....	7
3.1)Block Diagram.....	7
3.2)Hardware Design.....	9

4)Experimental Investigations.....	12
5)Flowchart.....	12
6)Result.....	13
7)Advantages & Disadvantages.....	13
8)Applications.....	15
9)Conclusion.....	16
10)Future Scope.....	16
11)Bibliography.....	17
Appendix.....	17
a)Source Code.....	

1)INTRODUCTION

1.1)OVERVIEW:

Life expectancy is a statistical measure of the average time a person is expected to live, based on the year of its birth, its current age, and other demographic factors including gender. In mathematical terms, life expectancy refers to the expected number of years remaining for an individual at any given age. It is very important to predict average life expectancy of a country to analyze further requirements to increase its rate of growth or stabilize the rate of growth in that country.

To predict the life expectancy rate of a particular country, we will be using machine learning to draw inferences from the given data set and give a prediction. We will also be creating a UI using Node-RED for making the model accessible to general users.

1.2)PURPOSE:

1)Predicting life expectancy plays a vital role in judging the growth and development of the economy. Across countries, high life expectancy is associated with high income per capita. Increase in life expectancy also leads to an increase in the “manpower” of a country.

2)Helps the government bodies take appropriate measures to control the population growth and also direct the utilization of

the increase in human resources and skill set acquired by people over many years.

3) This project would also help an individual assess his/her lifestyle choices and alter them accordingly to lead a longer and healthier life. It would make them more aware of their general health and its improvement or deterioration over time.

2) LITERATURE SURVEY

2.1) EXISTING PROBLEM:

In the context of life expectancy, we are often devoid of ideas of where to concentrate resources on to increase the average life expectancy of a community. This is because in the modern world there are so many other factors other than medical factors that our lifespan depends on like our Lifestyle, our choice of food, etc. Thus it is necessary to find the right areas to pour resources and spent time working on so that the results may improve in our favor.

2.2) PROPOSED SOLUTION:

Machine learning, natural language processing, and data mining in general have grown to be increasingly popular methods for processing data within the medical domain. Given examples, machine learning algorithms can be trained to learn which pieces of information are important to execute a task, and which patterns are indicative for producing correct output. Machine learning and language processing techniques have

been applied to a broad range of tasks, including medical decision support and decision making, automatic disease detection, automatic diagnostics, identifying the role of genes in the onset of diseases, adverse event detection, identifying interactions between drugs and side-effects of drugs, and phenotyping.

*Solution for this problem is that we can create a machine learning regression model which can predict the life expectancy of people based on different affecting factors which will help in knowing what to implement for betterment of humans. In order to do so we would take the help of IBM cloud services, Watson studio and Node – red for deploying the model.

1)Create IBM cloud Services

- Watson Studio
- Machine Learning resource
- Node-Red

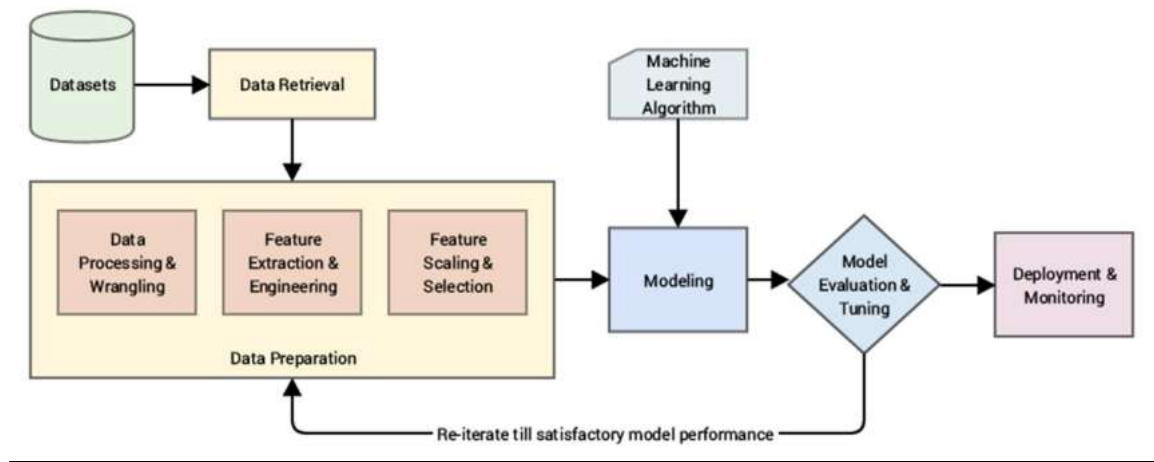
2)Configure Watson Studio

3)Create Node-Red Flow to connect all services together.

Deploy and run Node Red app. Deploy the Node Red flow. Then copy the link url upto .net/ and paste at a new tab by ui at the end of the url like this .

3)THEORETICAL ANALYSIS:

3.1)BLOCK DIAGRAM:



The above block diagram summarizes the entire process in a simple manner. The blocks include:

Datasets:

A data set (or dataset) is a collection of data. In the case of tabular data, a data set corresponds to one or more database tables, where every column of a table represents a particular variable, and each row corresponds to a given record of the data set in question.

*The data set lists values for each of the variables, such as height and weight of an object, for each member of the data set. Each value is known as a datum. Data sets can also consist of a collection of documents or files.

Data Preparation:

Data preparation is the act of manipulating (or pre-processing) raw data (which may come from disparate data sources) into a form that can readily and accurately be

analysed, e.g. for business purposes.

*Data preparation is the first step in data analytics projects and can include many discrete tasks such as loading data or data ingestion, data fusion, data cleaning, data augmentation, and data delivery.

Modelling:

The process of modeling means training a machine learning algorithm to predict the labels from the features, tuning it for the business need, and validating it on holdout data.

Machine Learning Algorithms:

Machine Learning algorithm is an evolution of the regular algorithm. It makes your programs “smarter”, by allowing them to automatically learn from the data you provide. The algorithm is mainly divided into:

1) Training Phase

2) Testing phase

Model Evaluation and Tuning:

Methods for evaluating a model's performance are divided into 2 categories: namely, holdout and Cross-validation. Both methods use a test set (i.e data not seen by the model) to evaluate model performance. It's not recommended to use the data we used to build the model to evaluate it. This is because our model will simply remember the whole training set, and will therefore always predict the correct label for any point in the training set.

Deployment and Monitoring:

Deployment is the method by which you integrate a machine learning model into an existing production environment to make practical business decisions based on data. It is one of the last stages in the machine learning life cycle and can be one of the most cumbersome.

3.2) HARDWARE / SOFTWARE DESIGNING :

- *IBM Cloud

- *Node-RED

- *Watson Studio

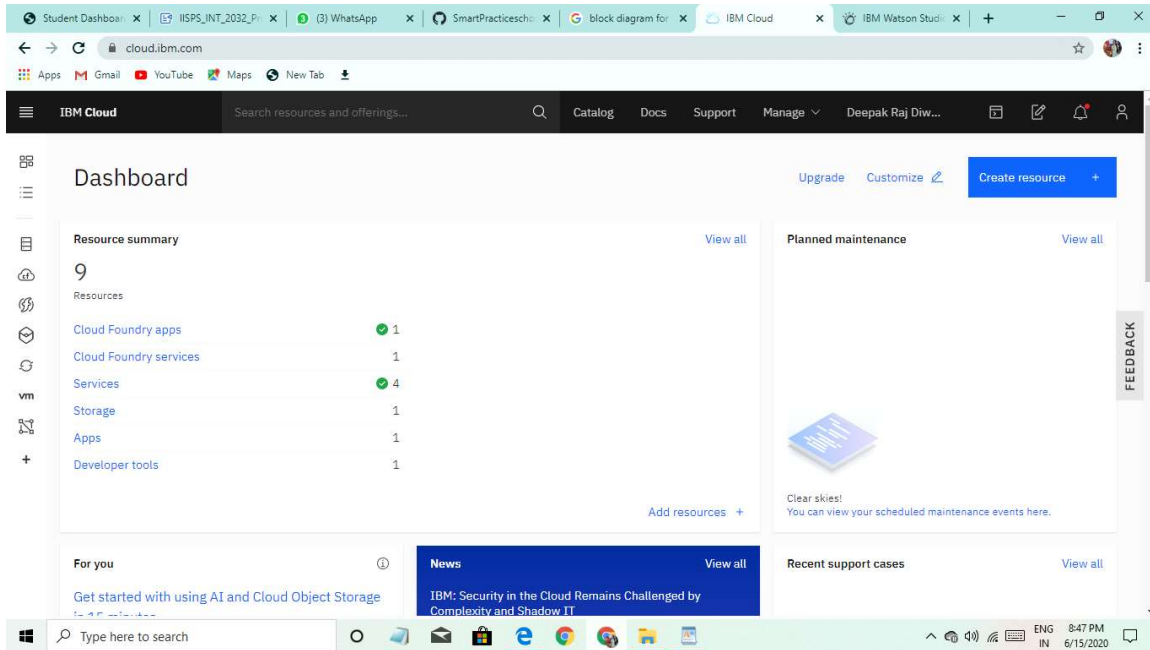
- *Watson Machine Learning

IBM Cloud:

The IBM cloud platform combines platform as a service (PaaS) with infrastructure as a service (IaaS) to provide an integrated experience. The platform scales and supports both small development teams and organizations, and large enterprise businesses. Globally deployed across data centers around the world, the solution you build on IBM Cloud spins up fast and performs reliably in a tested and supported environment you can trust.

The platform is built to support your needs whether it's working only in the public cloud or taking advantage of a multi-cloud deployment model. With our open-source technologies, such as Kubernetes, Red Hat Open Shift, and a full range of compute options, including virtual machines, containers, bare metal, and server less, you have as much control and flexibility

as you need to support workloads in your hybrid environment. You can deploy cloud-native apps while also ensuring workload portability.

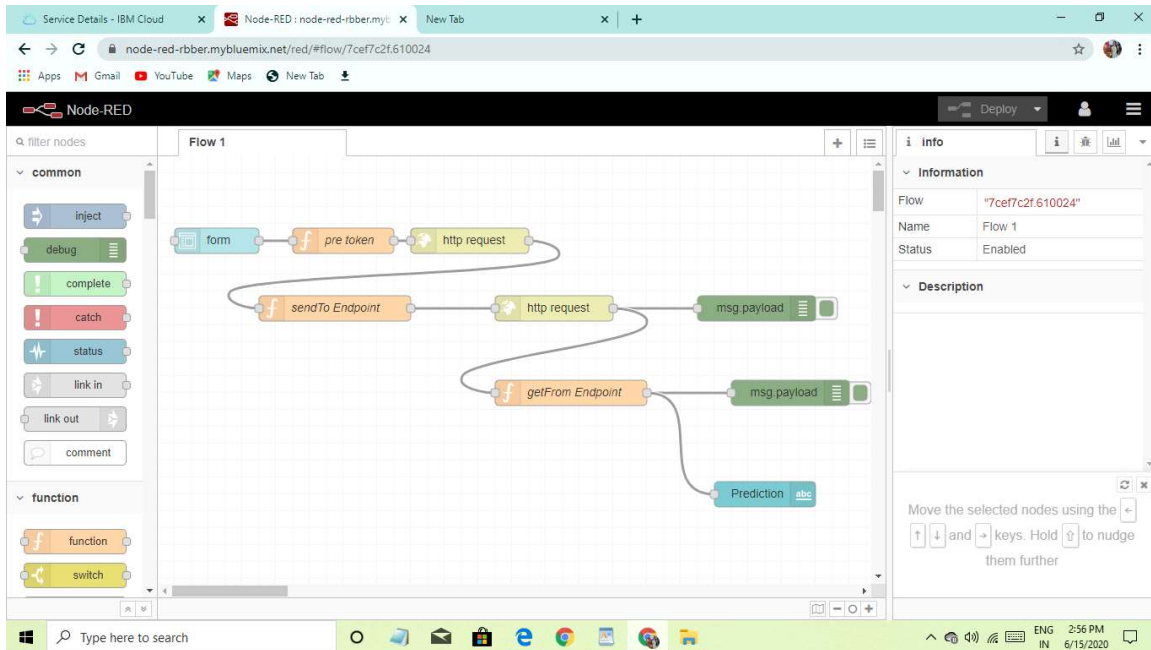


Node-RED:

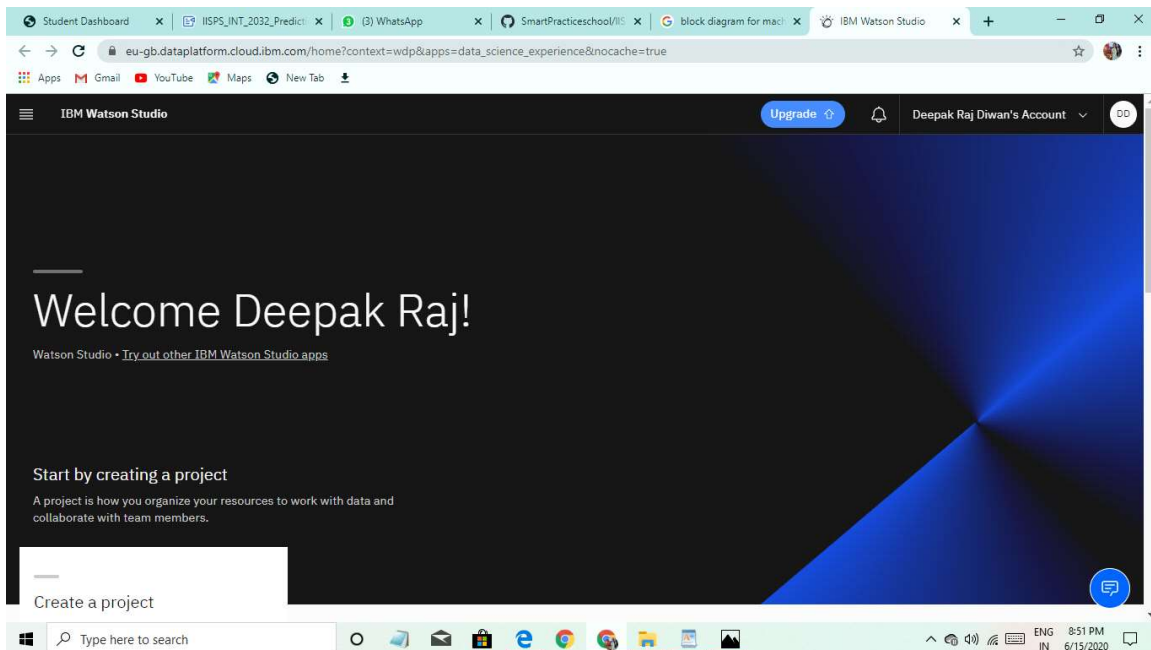
Node-RED is a programming tool for wiring together hardware devices, APIs and online services in new and interesting ways. It provides a browser-based editor that makes it easy to wire together flows using the wide range of nodes in the palette that can be deployed to its run time in a single-click.

*Node-RED provides a browser-based flow editor that makes it easy to wire together flows using the wide range of nodes in the palette. Flows can be then deployed to the run time in a single-click. JavaScript functions can be created within the

editor using a rich text editor. A built-in library allows you to save useful functions, templates or flows for re-use.



Watson Studio:

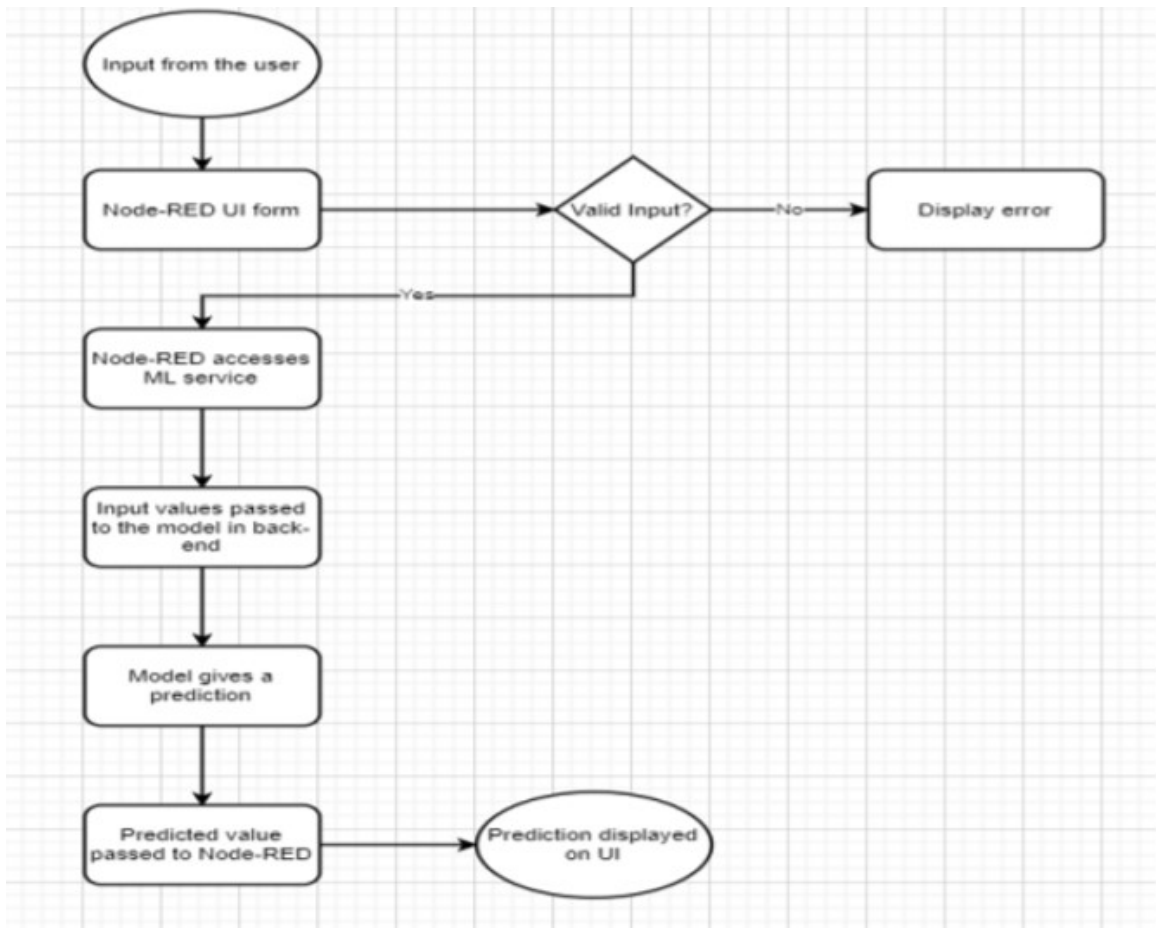


Watson Machine Learning:

4)EXPERIMENTAL INVESTIGATIONS:

The WHO data set for life expectancy contains 22 columns in total. The data set has historical data of life expectancy for the period between 2000 to 2015. There are numerous rows with null values for some of the columns and outliers as well. These null values will be replaced with the mean for the respective columns and the outliers will be adjusted. All the columns except the "Country" and "Status" are of integer type. The country column is not considered into the model training because of its less relation to the "Life Expectancy" column. The "Status" column is changed to integer type such that "Developing" is mapped to 1 and "Developed" is mapped to 0. After these steps, the data is ready for the model to be trained.

5)FLOW CHART:



6)RESULT:

A machine learning model developed using Linear Regression and IBM Cloud and its services. The model has a high accuracy.

An interactive UI deployed using the Node-RED service for the user to interact with the model in the back-end and get predictions for a given set of input values.

7)ADVANTAGES AND DISADVANTAGES:

ADVANTAGES:

1.) Easily identifies trends and patterns-

Machine Learning can review large volumes of data and discover specific trends and patterns that would not be apparent to humans.

2) No human intervention needed -

The biggest advantage of using Machine Learning for predictions is that it learns by itself. It could evolve over time as more data becomes available. The algorithm analyzes the trends in the data and adjusts itself to increase the accuracy.

DISADVANTAGES:

With all those advantages to its powerfulness and popularity, Machine Learning isn't perfect. The following factors serve to limit it:

1) Data Acquisition:

Machine Learning requires massive data sets to train on, and these should be inclusive/unbiased, and of good quality. There can also be times where they must wait for new data to be generated.

2) Time and Resources:

ML needs enough time to let the algorithms learn and develop enough to fulfill their purpose with a considerable amount of accuracy and relevancy.

8.APPLICATIONS:

- 1) Useful in the medical field for studying and isolating causes of deviation in human life expectancy over an observed period of time.
- 2) Supporting timely recognition of the right moment to start Advance Care Planning.
- 3) Individuals can predict their own life expectancy by inputting values in the corresponding fields. This could help make people more aware of their general health, and its improvement or deterioration over time. This may motivate them to make healthier lifestyle choices.
- 4) It could help the government bodies take appropriate measures to control the population growth and also direct the utilization of the increase in human resources and skill set acquired by people over many years. Across countries, high life expectancy is associated with high income per capita. Increase in life expectancy also leads to an increase in the “manpower” of a country. The knowledge asset of a country increases with the number of individuals in a country.
- 5) Based on the factors used to calculate life expectancy of an individual and the outcome, health care will be able to fund and provide better services to those with greater need.
- 6) Insurance sector will be able to provide individualized services to people based on the life expectancy outcomes and factors.

9)CONCLUSION:

Prognostication of life expectancy is difficult for humans. 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.

*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 pursuing. 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.

* Life Expectancy prediction can play a vital role in numerous sectors of the industry and machine learning can be used to achieve this with a high level of accuracy. .

10)FUTURE SCOPE:

This research should be considered to be exploratory. In order to replicate and extend this research, we can expand the data set substantially, by collecting additional data of both deceased and active patients. This will allow us to zoom in on specific illness trajectories, and to rephrase the task in such a way that it will match clinical settings more closely,

for example by aiming to make predictions about patients while they are still active.

******This research should be considered to be exploratory. In order to replicate and extend this research, we can expand the data set substantially, by collecting additional data of both deceased and active patients. This will allow us to zoom in on specific illness trajectories, and to rephrase the task in such a way that it will match clinical settings more closely, for example by aiming to make predictions about patients while they are still active.

*****We will be able to compare a range of predictive models, alternative patient representations, and (interpretations of) output variables in future work. To provide a better comparison between automatic and human prognostication, we can investigate the prediction accuracy of both the system and general practitioners by presenting them with the same task and test data.

11)BIBLIOGRAPHY:

APPENDIX:

<https://cloud.ibm.com>

<https://www.ibm.com/in-en/cloud/watson-studio>

<https://nodered.org>

NODE-RED Flow:

```
[  
{  
  "id":"e555521.6fd29b",  
  "type":"ui_text",  
  "z":"a8ec5994.bedba8",  
  "group":"1bac549b.37e41b",  
  "order":1,  
  "width":0,  
  "height":0,  
  "name":"",  
  "label":"Prediction",  
  "format":"{{msg.payload}}",  
  "layout":"row-spread",  
  "x":720,  
  "y":400,  
  "wires":[]
```

```
},  
  
{  
  "id": "1bac549b.37e41b",  
  "type": "ui_group",  
  "z": "",  
  "name": "Machine Learning Model",  
  "tab": "af1e5839.0cfac8",  
  "order": 1,  
  "disp": true,  
  "width": "6",  
  "collapse": false  
},  
  
{  
  "id": "af1e5839.0cfac8",  
  "type": "ui_tab",  
  "z": "",  
  "name": "Home Page",  
  "icon": "dashboard",  
  "disabled": false,
```

```
"hidden":false
```

```
}
```

```
]
```

My Github Repository Link:

<https://github.com/SmartPracticeschool/llSPS-INT-2032-Predicting-Life-Expectancy-using-Machine-Learning>