

## Project Report

# "One Year Life expectancy post thoracic surgery Using IBM Watson"

Category- Machine Learning

SubmittedBy- Sagar Nagalgawe

Email-ID: : [sagarnagalgawe43@gmail.com](mailto:sagarnagalgawe43@gmail.com)

Submitted to: SmartInternz.com



# INDEX

## **1) INTRODUCTION**

1.1 Overview

1.2 Purpose

## **2) LITERATURE SURVEY**

2.1 Existing problem

2.2 Proposed solution

## **3) THEORITICAL ANALYSIS**

3.1 Block diagram

3.2 Hardware / Software designing

## **4) EXPERIMENTAL INVESTIGATIONS**

## **5) FLOWCHART**

## **6) RESULT**

## **7) ADVANTAGES & DISADVANTAGES**

## **8) APPLICATIONS**

## **9) CONCLUSION**

## **10) FUTURE SCOPE**

## **11) BIBILOGRAPHY**

## **12) APPENDIX**

A. Source code

## 1) INTRODUCTION

Life expectancy, an estimate of the number of remaining years of life a person has, is an important consideration for making clinical decisions in primary care. Predicting Life Expectancy helps analyze the average lifespan of the person which helps in making crucial health decisions.

One of the common cause of cancer deaths in worldwide today is Lung Cancer. The incidence of lung cancer has increased considerably and has turned to be the most widely recognized cancer in men in most of the nations. Illnesses are caused when the cells develop without any control. These illnesses are known as cancer. Lung cancer occurs when these uncontrolled cells develops in lungs. It can start with one or both the lungs. Lung cancer can bring troubles with vision and weakness on one side of the body if it is spread to the brain. Symptoms of lung cancer include blood cough, wheezing, fever, weight loss, chest pain, bone pain and clubbing of fingernails. One of the important cause of lung cancer is smoking. It consists of 4,000 chemicals or more, where most of them have been identified as causing cancer. " Person who smokes more than one pack of cigarettes per day has a 20-25 times greater risk of developing lung cancer than someone who has never smoked". Approximately 85% of lung cancer emerge due to usage of tobacco. However other factors like radon gas, air pollution, asbestos may be helping for the cause of lung cancer. Thoracic surgery has existed as a particular surgical train for over a century. At first, its primary concentration was surgery for tuberculosis and bronchiectasis. Fast advance has been made in surgery for lung cancer. Thoracic surgery speaks to the surgical part in treating ailments of the lungs and the thorax.

## 1.1 Overview

Integrating computer applications into the medical field have directly affected the productivity and accuracy of doctors nowadays.

Measuring health outcomes is one of these applications. Clearly, there is a growing role for health outcomes in the purchasing and management of healthcare. These days cancer is one of the major causes of death in the most countries. Currently, lung cancer is the most frequent augury for thoracic surgery. Researchers applied different strategies, such as examination in early stage, to identify the type of cancer before the emergence of symptoms. Furthermore, new methods for the early prediction of cancer therapy outcome have been developed . With the raise of new techniques in the field of medicine, massive datasets of cancer have been collected and now available to researchers in the medical field .However, the most challenging task is predicting a disease outcome accurately .So, the current research efforts examine the use of machine learning techniques for discover and identify models and relationships between them, from large datasets, the data is analyzed to extract useful information that supports disease augury, and to improve models that predict patients health more accurately.

## 1.2 Purpose

Patients who receive thoracic surgery for lung cancer do so with the expectation that their live will be prolonged for a sufficient amount of time afterwards , and wheather they survived within a one year time frame . The problem to solve is wheather there is a way to determine post-operative life expectancy of lung cancer patients attributes in the data set.

If there is a pattern to be recognized with the attributes and wheather the patients do not survive the one year mark , this would help physicians and patients make a more educated decisions on wheather they should proceed forward with surgery. If physicians feel the surgery will only hinder the patients quality of life with a recognized high risk of death within one year time frame , then both parties can make a decision to follow through on surgery or decide to find alternative treatment methods of palliative care.

Not only would this influence physicians and patients, this information could be utilized by health insurance companies and national health organizations when it comes to making decisions on finances for thoracic surgery involving lung cancer. Also clinical researchers could consolidate any useful findings with other data research findings to search for new research areas.

## **2) LITERATURE REVIEW**

### **2.1 Existing Problem**

Past studies have revealed a lot of work in the field of predicting life expectancy of a human being. After reviewing existing works and techniques in the prediction of human Life Expectancy, and finally reached a conclusion that it is possible to predict a Average Life Expectancy for individuals using advancing technologies and devices such as big data, AI, machine learning techniques, and PHDs, wearables and mobile health monitoring devices, IOT. It is noticed that the collection of data is a huge challenge due to the privacy and government policy considerations, which will require collaboration of various bodies in the health industry. The interworking of a heterogeneous health network is also a challenge for data collection. Despite these challenges, a possibility of predicting Life by proposing an approach of data collection and application by smartphone, in which users can enter their information to access the cloud server to obtain their own predicted Lifespan based on the given inputs.

To verify the accuracy of PLE prediction and validation of data quality, big data techniques and analysis algorithms need to be developed and tested in a real-life situation with several sample groups. As artificial intelligence technology is evolving and being applied rapidly, feasibility may be increasing to collect health data from the public as well as existing health agencies such as centralized health servers.

## 2.2 Proposed Solution

Machine learning is a branch of artificial intelligence which utilizes statistical, optimization and probabilistic techniques that allows computers to “learn” from past examples and to detect hard-to-discern patterns from large, noisy or complex data sets. These techniques have become a popular tool in medical diagnosis, which can find and identify models and relationships between them from large, noisy or complex datasets. The inputs are the information about the patient’s age, gender, past medical history, past medical procedures, family medical history and current symptoms , while labels are the illnesses.

More recently, it has been widely applied in the field of cancer prediction and prognosis which are differ from cancer detection and diagnosis. There are three types of cancer prediction and prognosis: One of them is prediction of cancer receptivity. In this type,

one is trying to predict the probability of cancer progression before occurrence of the disease.

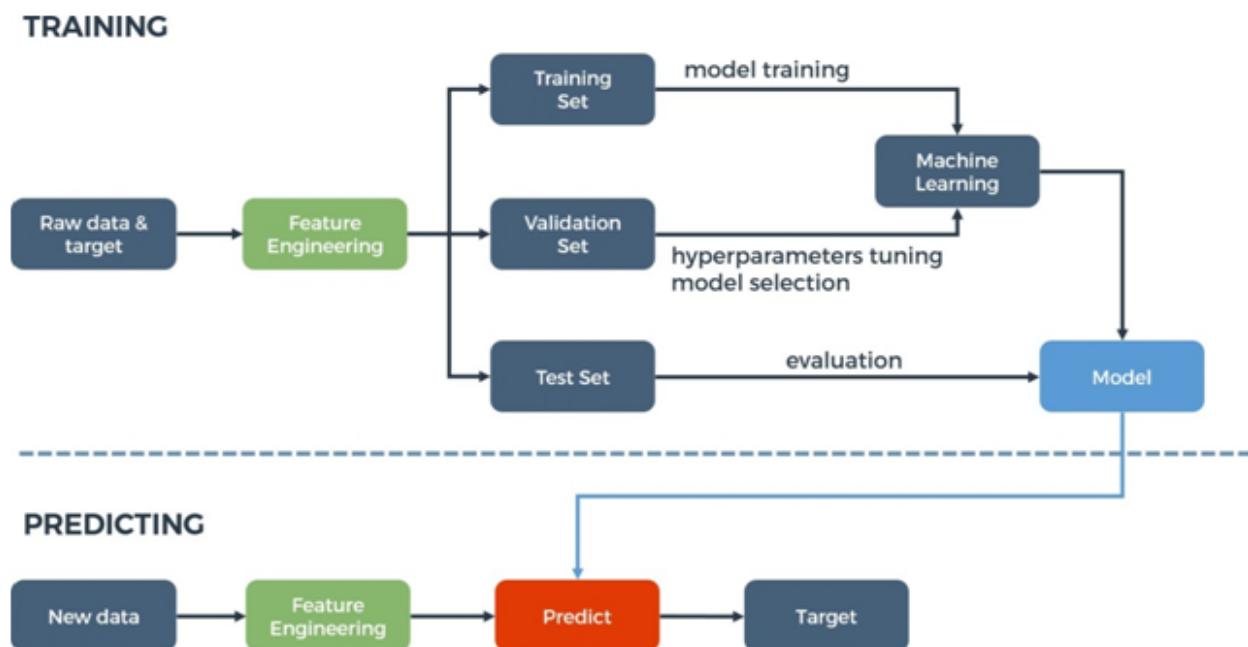
Second type is the prediction of cancer recurrence by trying to predict the probability of redeveloping cancer after treatment and after a period of time during which the cancer cannot be detected.

Third type is the prediction of cancer survivability by trying to predict an outcome which usually refers to life expectancy, survivability, progression and tumor-drug sensitivity.

The model of ”One Year Life expectancy post thoracic surgery using IBM Watson” uses IBM Cloud services, which helps to avoid any storage issues. The UI Presented to the users is a website url i.e. on users fingertips.

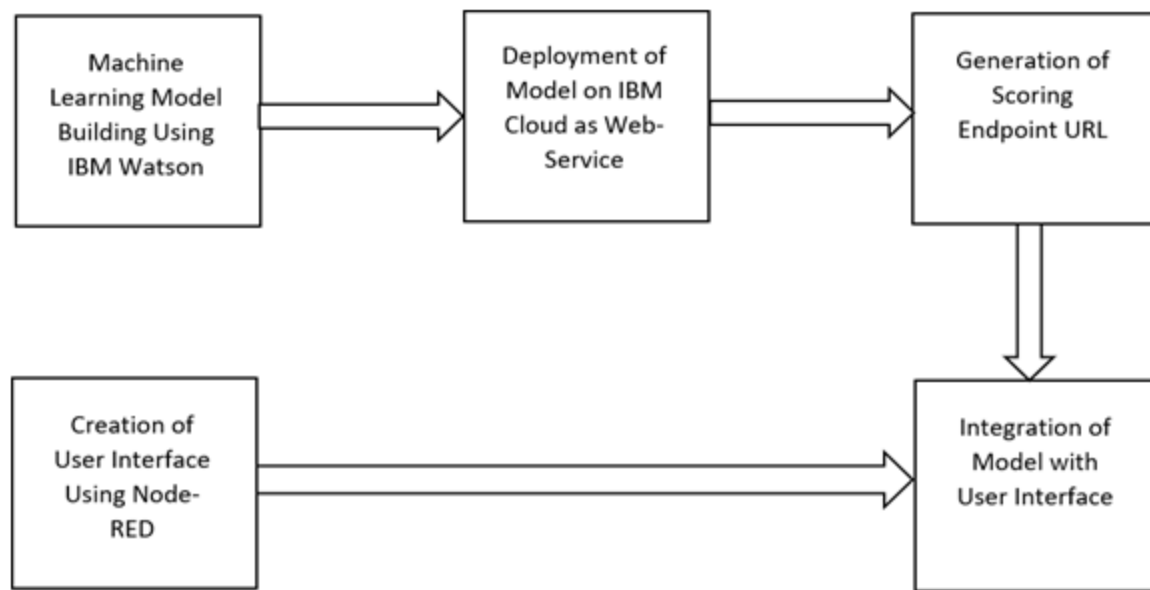
### 3 ) THEORITICAL ANALYSIS

#### 3.1 Block diagram



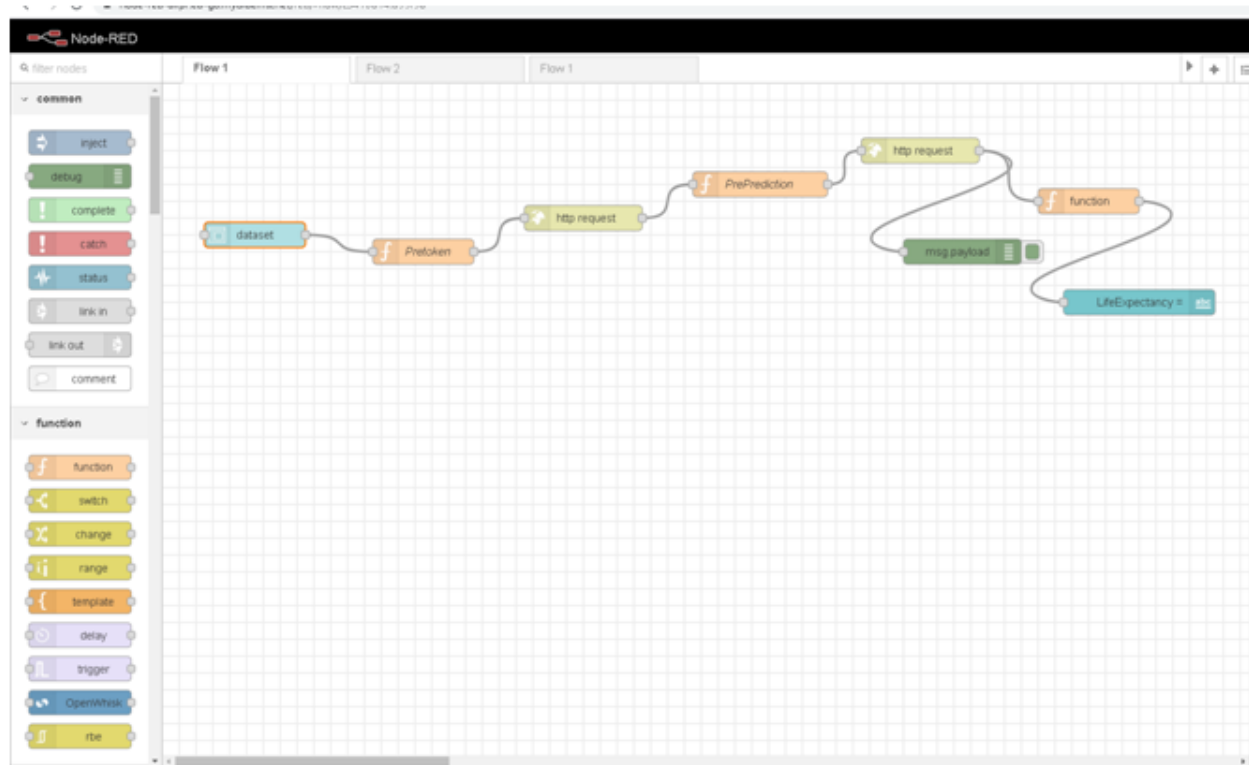
### 3.2 Hardware / Software designing

There is no Hardware involved as such. The web-app has IBM Watson Machine Learning as its backend service and Node-RED for the User Interface development. The web-app has been deployed on IBM Cloud as a Web Service. Basic Flow is as follows:





A node red flow is made for the prediction of life expectancy. Which is as shown below,



#### 4) EXPERIMENTAL INVESTIGATIONS

On IBM Watson studio machine learning using auto AI build a model to predict life expectancy.

To do so first create account on IBM Watson studio.

Using Add to project choose auto ai.

Then upload data set

Choose best way to predict.

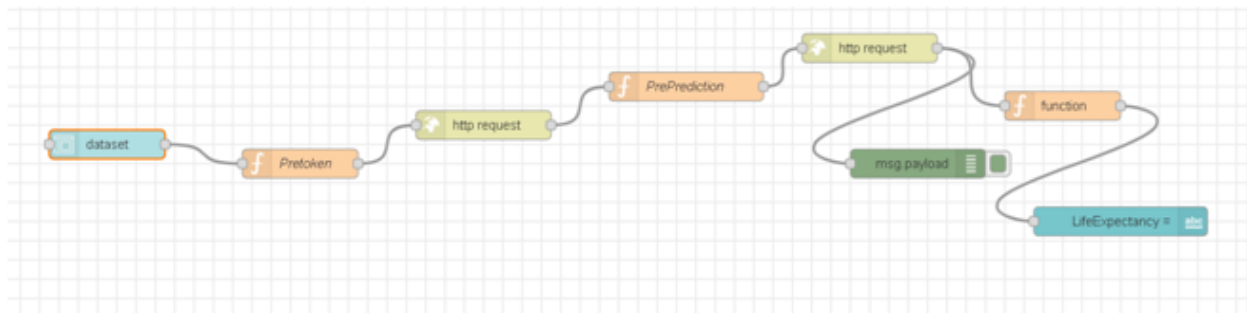
Save as a model which is on the top

Deploy the model.

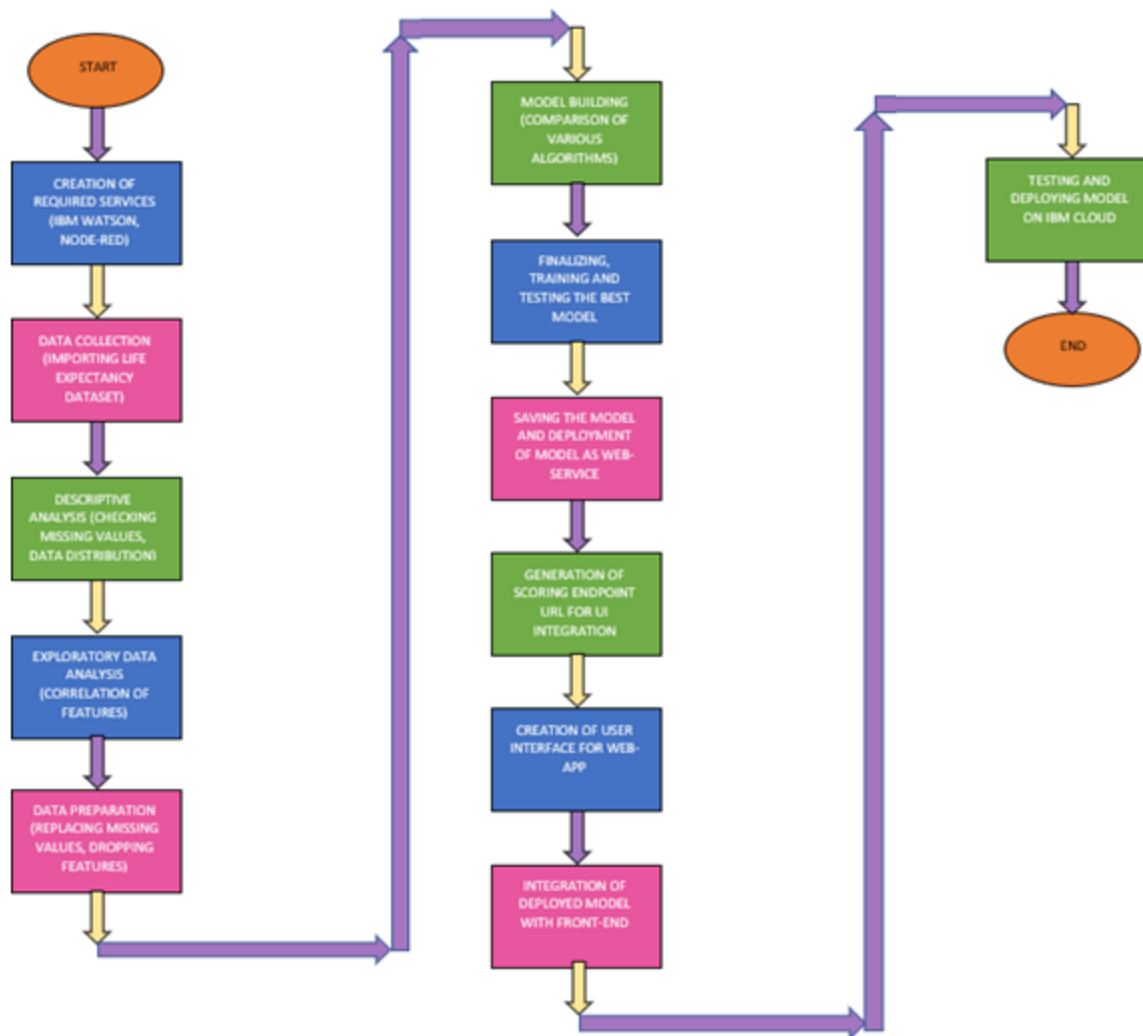
Test the model.

Create service credential

Build node red flow

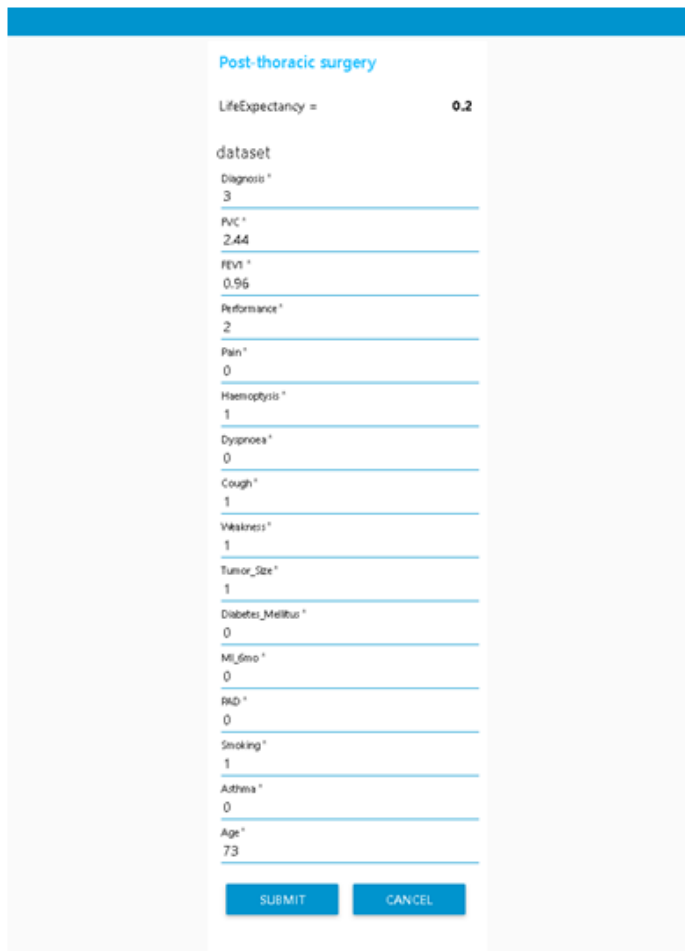


## 5) FLOWCHART



## 6 ) RESULT

The user friendly Graphical User interface is shown in Figure. This GUI is connected to the trained machine learning model present in the backend. The user has to fill in the inputs accordingly and click on the “Predict” button present at the end of the form. On clicking the “Predict” button, the user will be displayed the predicted life expectancy at the predict label, based on the inputs provided as shown in Figure.



The screenshot displays a web application titled "Post-thoracic surgery". It features a form with various input fields, each with a label and a value. The values are as follows:

Field Label	Value
LifeExpectancy =	0.2
dataset	
Diagnosis *	3
PVC *	2.44
FEV1 *	0.96
Performance *	2
Pain *	0
Haemoptysis *	1
Dyspnoea *	0
Cough *	1
Weakness *	1
Tumor_Size *	1
Diabetes_Mellitus *	0
Mt_gmo *	0
RAD *	0
Smoking *	1
Asthma *	0
Age *	73

At the bottom of the form, there are two buttons: "SUBMIT" and "CANCEL".

## **7) ADVANTAGES AND DISADVANTAGES**

### **Advantages:**

#### **7.1.1 Advantages of using IBM Watson:**

1. Processes unstructured data
2. Fills human limitations
3. Acts as a decision support system, doesn't replace humans
4. Improves performance + abilities by giving best available data
5. Improve and transform customer service
6. Handle enormous quantities of data
7. Sustainable Competitive Advantage
8. Easy for users to interact with the model via the UI.
9. User-friendly.
10. Easy to build and deploy
11. Doesn't require much storage space.

### **Disadvantages:**

#### **7.2.1 Disadvantages of using IBM Watson:**

1. Only in English (Limits areas of use)
2. Seen as disruptive technology

3.Maintenance

4.Doesn't process structured data directly

5.Increasing rate of data, with limited resources

6.The web-app is not handy to use for users as it is not a mobile app.

7. The user has to input all the fields and only then the prediction will be given to the user. However, inputting all these values is a tiresome job, and the user cannot input these values through speech.

## **8) APPLICATIONS:**

The application of machine learning techniques for predicting post-operative life expectancy in the lung cancer patients is an area with little research and few concrete recommendations.

In order to use machine learning techniques effectively, attribute ranking and selection is an integral component to successful health outcome prediction.

## **9) CONCLUSION:**

By the end of all of our iterations and improvements, we were able to achieve fairly good 4 results with the random forest .These results have large implications in the medical field. An analysis similar to ours could be performed before a patient goes in for surgery to see how high risk they are, which could be crucial information. Users can interact with the system via a simple Graphical user interface which is in the form of a form with input spaces which the user needs to fill the inputs into and

then predict outcomes.

#### 10) **FUTURE SCOPE:**

1. We can make the model more accurate and do more relevant featureextraction.
2. We can make a mobile app as now-a-days, it is morepreferred.
3. We can connect the model to the database which can predict the life Expectancy of not for one medical research but for other research areas too. This will help us analyze the trends in the lifespan.

#### 11) **BIBILOGRAPHY:**

- 1) Learn how to add, edit, delete text using Writer:

<https://www.zoho.com/writer/help/working-with-text.html>

- 2) GitHub account creation:

<https://github.com/>

- 3) Data visualization, preparation, and transformation using IBM

Watson Studio:

<https://developer.ibm.com/tutorials/watson-studio-data-visualization-preparation-transformation/>

4) Automate model building in IBM Watson Studio:

<https://developer.ibm.com/tutorials/watson-studio-auto-ai/>

5) Create a Node-RED starter application:

<https://developer.ibm.com/tutorials/how-to-create-a-node-red-starter-application/>

<https://www.youtube.com/watch?v=s7wmiS2mSXY&feature=youtu.be>

6) Life Expectancy Post Thoracic Surgery:

<http://cs229.stanford.edu/proj2014/Adam%20Abdulhamid,%20Ivaylo%20Bahtchevanov,%20Peng%20Jia,Life%20Expectancy%20Post%20Thoracic%20Surgery.pdf>

<https://archive.ics.uci.edu/ml/datasets/Thoracic+Surgery+Data>



## **12) APPENDIX:**

### **A. Source code**

The source code for IBM Watson Project and Node-RED Flow can be found at:

-

<https://github.com/SmartPracticeschool/II-SPS-INT-3470-One-Year-Life-expectancy-post-thoracic-surgery-Using-IBM-Watson>

The web-app can be used at:

<https://node-red-ulfpr.eu-gb.mybluemix.net/ui/#!/0?socketid=Hg3zRQjO1hPIveGVAAAAA>