

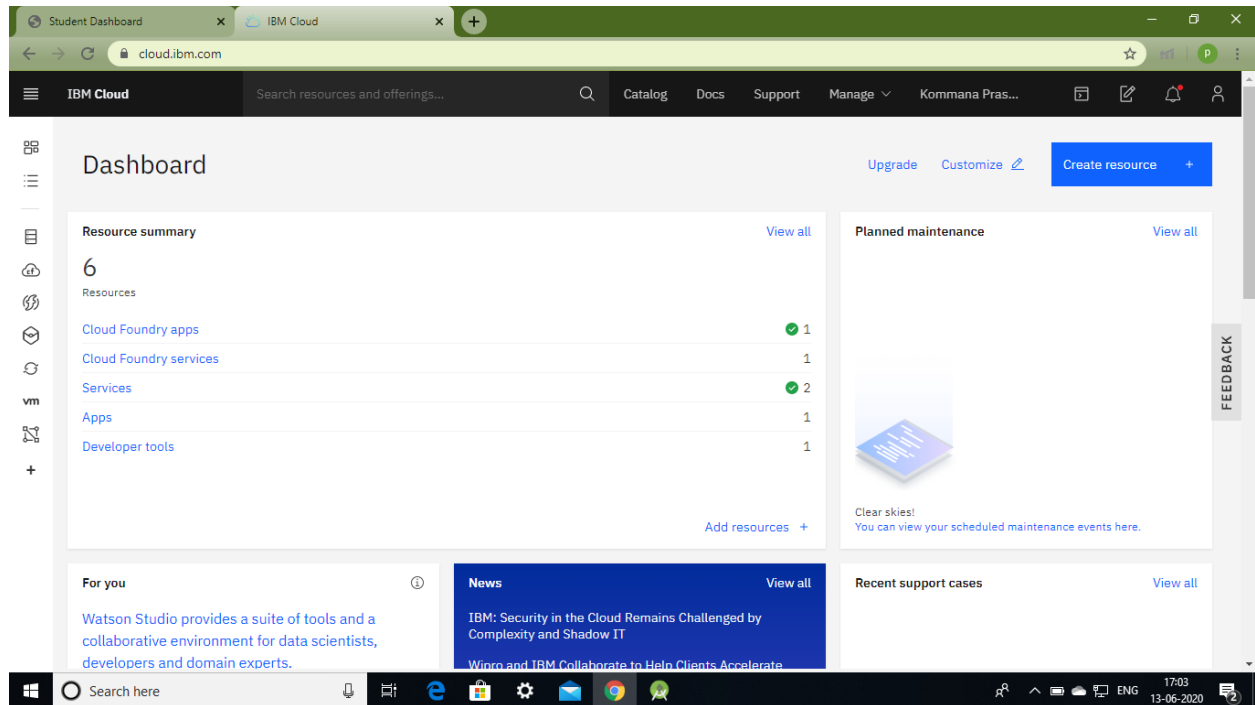
PROJECT SCOPE DOCUMENT

Project Name : Predicting Life Expectancy using Machine Learning

Project Summary	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.
Project Requirements	This project aims at using a reference dataset to predict life expectancy of a given based on a statistical measure of the average time a human being is expected to live
Functional Requirements	Using Various factors like Regional variations, Economic Circumstances, Sex Differences, Mental Illnesses, Physical Illnesses, Education, Year of their birth and other demographic factors training and build a ML model
Technical Requirements	Building ML model that can be usable,maintainable,adaptable and predict accurate results
Software Requirements	Python,IBM Cloud,IBM Watson,Node-Red
Project Team	Individual Project
Project Schedule	Four weeks (started on 8th June 2020)

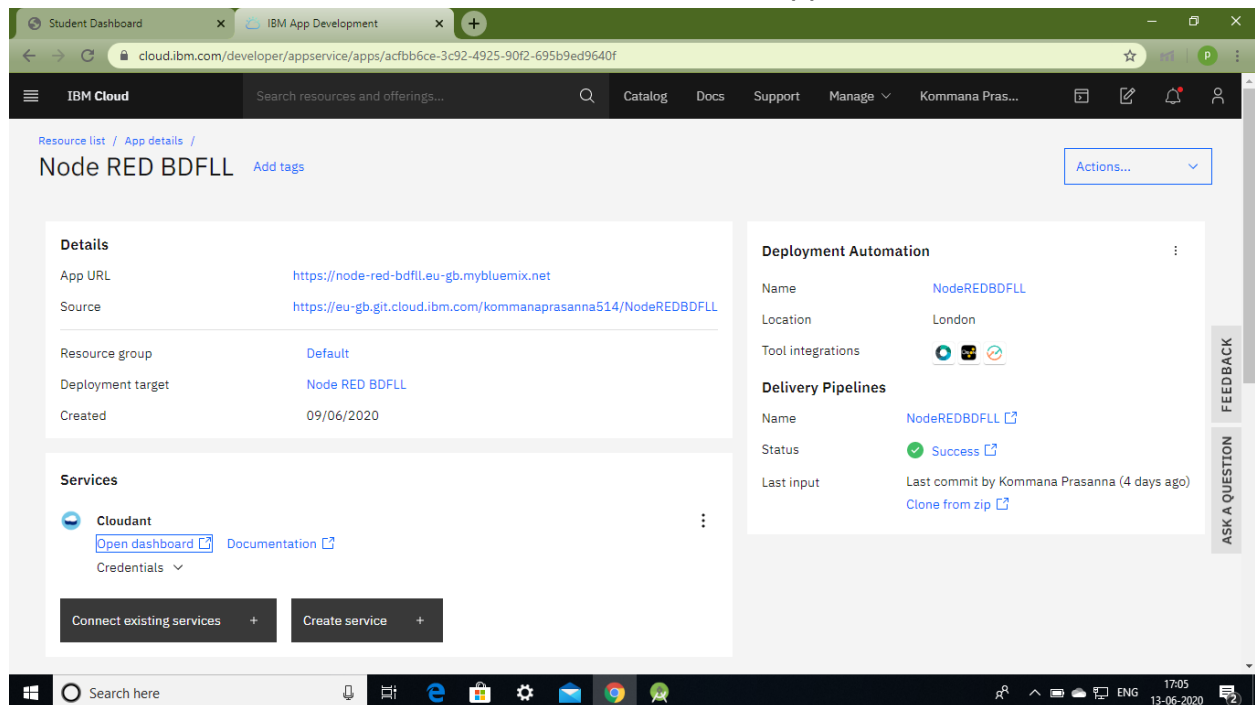
2) EXPLORE IBM CLOUD PLATFORM:

1.CREATE IBM CLOUD ACCOUNT:



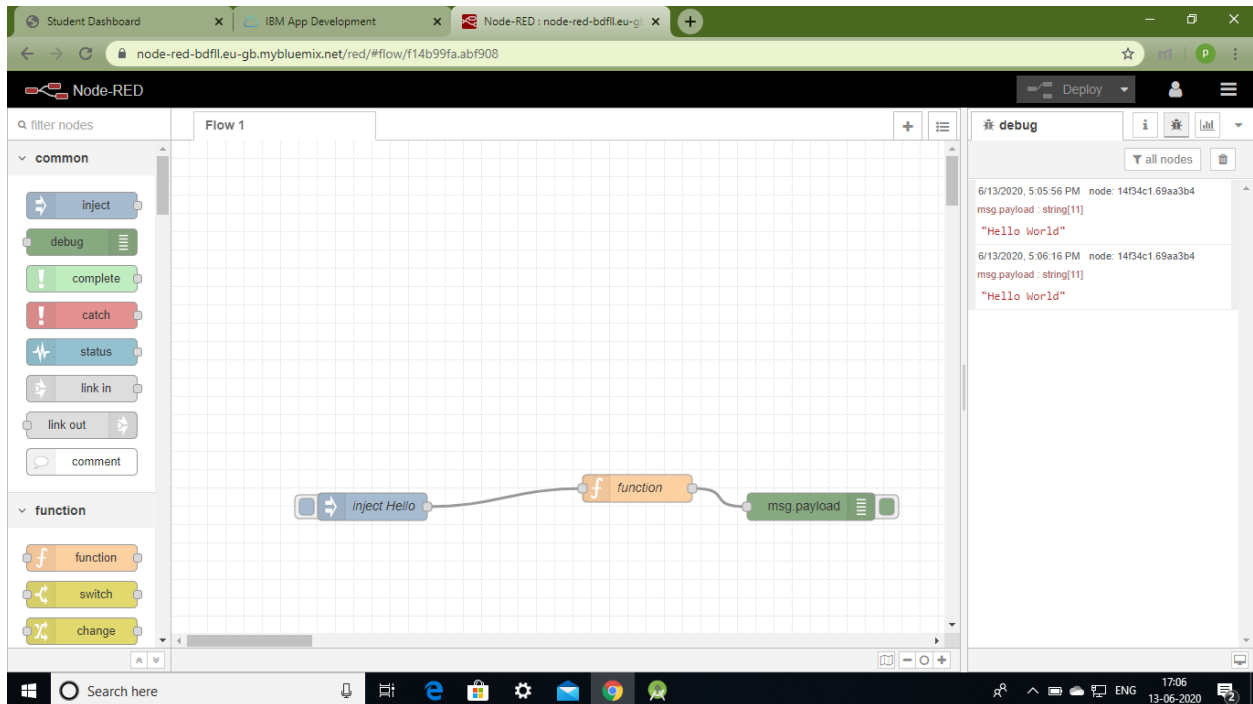
The screenshot shows the IBM Cloud Dashboard in a web browser. The browser tabs include 'Student Dashboard' and 'IBM Cloud'. The address bar shows 'cloud.ibm.com'. The dashboard header includes a search bar and navigation links for Catalog, Docs, Support, Manage, and Kommana Prasanna. The main content area is titled 'Dashboard' and features a 'Create resource' button. Below this, there are several sections: 'Resource summary' showing 6 resources (Cloud Foundry apps, Cloud Foundry services, Services, Apps, Developer tools), 'Planned maintenance', 'For you' (Watson Studio), 'News' (IBM Security in the Cloud), and 'Recent support cases'. The Windows taskbar at the bottom shows the search bar and various application icons.

2.Create a Node-RED Starter application



The screenshot shows the IBM Cloud Developer App Service page for a Node RED BDFLL application. The browser tabs include 'Student Dashboard' and 'IBM App Development'. The address bar shows 'cloud.ibm.com/developer/appservice/apps/acfbb6ce-3c92-4925-90f2-695b9ed9640f'. The page header includes a search bar and navigation links for Catalog, Docs, Support, Manage, and Kommana Prasanna. The main content area is titled 'Node RED BDFLL' and features an 'Actions...' dropdown. Below this, there are several sections: 'Details' (App URL, Source, Resource group, Deployment target, Created), 'Services' (Cloudant), 'Deployment Automation' (Name, Location, Tool integrations), and 'Delivery Pipelines' (Name, Status, Last input). The Windows taskbar at the bottom shows the search bar and various application icons.

3. Node-red Hello world app



3) Explore IBM Watson Services

Explore IBM Watson Machine Learning

Introduction to Machine Learning (Lab ibm watson studio URL)

[To predict whether the customer will churn.](#)

4) Introduction To Watson Studio

Build Your Own ML Model In IBM Watson Studio Using ML service:

The screenshot displays the IBM Watson Studio web interface. At the top, there's a navigation bar with tabs for 'Chapter 4 Build and Deploy mode', 'Service Details - IBM Cloud', and 'IBM Watson Studio'. The main header shows the project name 'Just_sample_creation' and buttons for 'Launch IDE', 'Add to project', 'Upgrade', and user account information. Below the header, there's a 'Storage' section indicating '0 Byte used' on 'Cloud Object Storage'. The 'Associated services' section lists 'WatsonMachineLearning' under the 'Watson - Machine Learning' plan. The 'Access tokens' section shows a table with columns for Name, Role, Created, and Last used, with a 'New token' button. The bottom of the image shows a Windows taskbar with various application icons and the system clock.

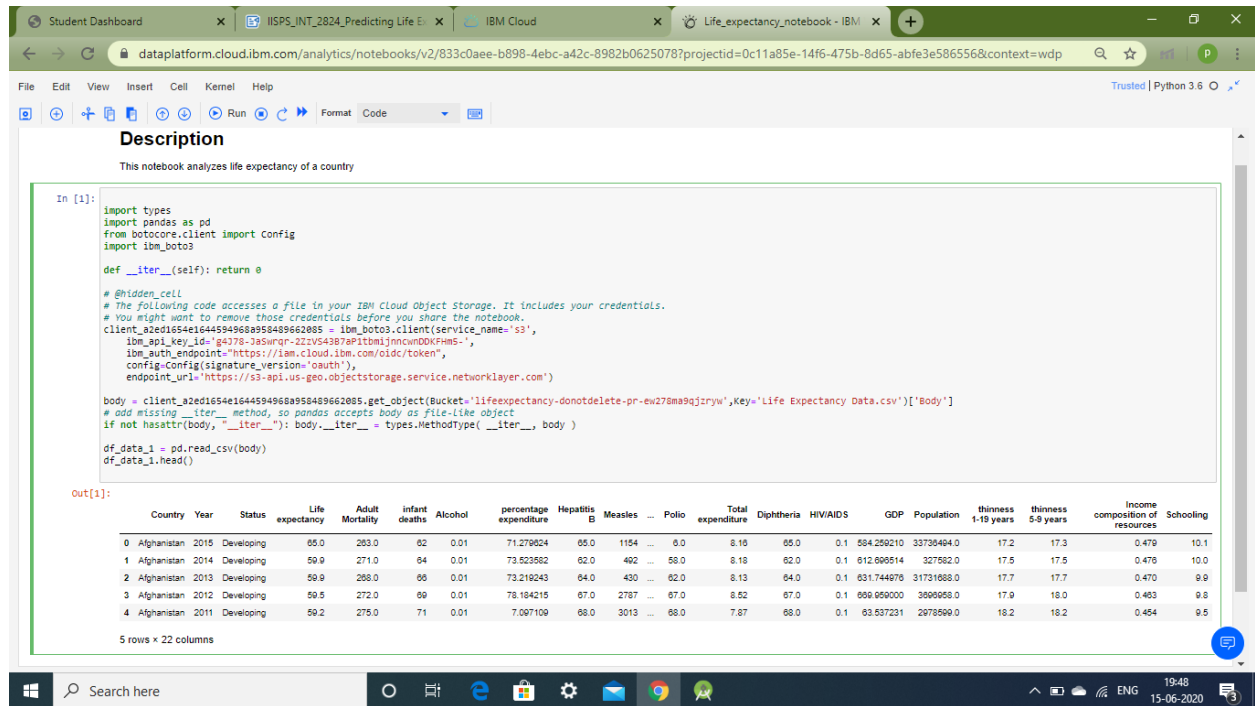
5) Predicting Life Expectancy With Python

Collect The Dataset For The Project from kaggle into our system

The screenshot shows a Microsoft Excel spreadsheet titled 'Life Expectancy Data'. The spreadsheet contains a dataset with columns for Country, Year, Status, Life expectancy, and various health indicators. The data is organized into rows, with the first row being the header. The following table represents the data shown in the spreadsheet:

Country	Year	Status	Life expect	Adult Mor	infant dea	Alcohol	percentag	Hepatitis	Measles	BMI	under-five	Polio	Total exp	Diphtheri	HIV/AIDS	GDP	Populatio	thinness	thinness	Income
Afghanistan	2015	Developing	65	263	62	0.01	71.27962	65	1154	19.1	83	6	8.16	65	0.1	584.2592	33736494	17.2	17.3	0.4
Afghanistan	2014	Developing	59.9	271	64	0.01	73.52358	62	492	18.6	86	58	8.18	62	0.1	612.6965	327582	17.5	17.5	0.4
Afghanistan	2013	Developing	59.9	268	66	0.01	73.21924	64	430	18.1	89	62	8.13	64	0.1	631.745	31731688	17.7	17.7	0.4
Afghanistan	2012	Developing	59.5	272	69	0.01	78.18422	67	2787	17.6	93	67	8.52	67	0.1	669.959	3696958	17.9	18	0.4
Afghanistan	2011	Developing	59.2	275	71	0.01	7.097109	68	3013	17.2	97	68	7.87	68	0.1	63.53723	2978599	18.2	18.2	0.4
Afghanistan	2010	Developing	58.8	279	74	0.01	79.67937	66	1989	16.7	102	66	9.2	66	0.1	553.3289	2883167	18.4	18.4	0.4
Afghanistan	2009	Developing	58.6	281	77	0.01	56.76222	63	2861	16.2	106	63	9.42	63	0.1	445.8933	284331	18.6	18.7	0.4
Afghanistan	2008	Developing	58.1	287	80	0.03	25.87393	64	1599	15.7	110	64	8.33	64	0.1	373.3611	2729431	18.8	18.9	0.4
Afghanistan	2007	Developing	57.5	295	82	0.02	10.91016	63	1141	15.2	113	63	6.73	63	0.1	369.8358	26616792	19	19.1	0.4
Afghanistan	2006	Developing	57.3	295	84	0.03	17.17152	64	1990	14.7	116	58	7.43	58	0.1	272.5638	2589345	19.2	19.3	0.4
Afghanistan	2005	Developing	57.3	291	85	0.02	1.388648	66	1296	14.2	118	58	8.7	58	0.1	25.29413	257798	19.3	19.5	0.3
Afghanistan	2004	Developing	57	293	87	0.02	15.29607	67	466	13.8	120	5	8.79	5	0.1	219.1414	24118979	19.5	19.7	0.3
Afghanistan	2003	Developing	56.7	295	87	0.01	11.08905	65	798	13.4	122	41	8.82	41	0.1	198.7285	2364851	19.7	19.9	0.3
Afghanistan	2002	Developing	56.2	3	88	0.01	16.88735	64	2486	13	122	36	7.76	36	0.1	187.846	21979923	19.9	2.2	0.3
Afghanistan	2001	Developing	55.3	316	88	0.01	10.57473	63	8762	12.6	122	35	7.8	33	0.1	117.497	2966463	2.1	2.4	0
Afghanistan	2000	Developing	54.8	321	88	0.01	10.42496	62	6532	12.2	122	24	8.2	24	0.1	114.56	293756	2.3	2.5	0.3
Albania	2015	Developing	77.8	74	0	4.6	364.9752	99	0	58	0	99	6	99	0.1	3954.228	28873	1.2	1.3	0.7
Albania	2014	Developing	77.5	8	0	4.51	428.7491	98	0	57.2	1	98	5.88	98	0.1	4575.764	288914	1.2	1.3	0.7
Albania	2013	Developing	77.2	84	0	4.76	430.877	99	0	56.5	1	99	5.66	99	0.1	4414.723	289592	1.3	1.4	0.7
Albania	2012	Developing	76.9	86	0	5.14	412.4434	99	9	55.8	1	99	5.59	99	0.1	4247.614	2941	1.3	1.4	0.7
Albania	2011	Developing	76.6	88	0	5.37	437.0621	99	28	55.1	1	99	5.71	99	0.1	4437.179	295195	1.4	1.5	0.7
Albania	2010	Developing	76.2	91	1	5.28	41.82276	99	10	54.3	1	99	5.34	99	0.1	494.3588	291321	1.4	1.5	0.7
Albania	2009	Developing	76.1	91	1	5.79	348.056	98	0	53.5	1	98	5.79	98	0.1	4114.137	2927519	1.5	1.6	0.7
Albania	2008	Developing	75.3	1	1	5.61	36.62207	99	0	52.6	1	99	5.87	99	0.1	437.5396	2947314	1.6	1.6	0.7

Create A Jupyter Notebook In IBM Watson And Import Data



Description

This notebook analyzes life expectancy of a country

```
In [1]: import types
import pandas as pd
from botocore.client import Config
import boto3

def __iter__(self): return 0

# @hidden_cell
# The following code accesses a file in your IBM Cloud Object Storage. It includes your credentials.
# You might want to remove those credentials before you share the notebook.
client_azed1654e1644594968a958489662085 = boto3.client(service_name='s3',
ibm_api_key_id='84278-Ja5urqr-22iv54387aP1tbn1jnnwnDDKPHms-',
ibm_auth_endpoint='https://iam.cloud.ibm.com/oidc/token',
config=Config(signature_version='oauth')),
endpoint_url='https://s3-api.us-geo.objectstorage.service.networklayer.com')

body = client_azed1654e1644594968a958489662085.get_object(Bucket='lifeexpectancy-donotdelete-pr-ew278ma9qjzryw',Key='Life Expectancy Data.csv')['Body']
# add missing __iter__ method, so pandas accepts body as file-like object
if not hasattr(body, "__iter__"): body.__iter__ = types.MethodType(__iter__, body)

df_data_1 = pd.read_csv(body)
df_data_1.head()
```

Out[1]:

	Country	Year	Status	Life expectancy	Adult Mortality	Infant deaths	Alcohol	percentage expenditure	Hepatitis B	Measles ...	Polio	Total expenditure	Diphtheria	HIV/AIDS	GDP	Population	thinness 1-19 years	thinness 5-9 years	Income composition of resources	Schooling	
0	Afghanistan	2015	Developing	65.0	263.0	62	0.01	71.279624	65.0	1154	...	6.0	6.18	65.0	0.1	584.259210	33739494.0	17.2	17.3	0.479	10.1
1	Afghanistan	2014	Developing	59.9	271.0	64	0.01	73.523582	62.0	492	...	58.0	6.18	62.0	0.1	612.699514	327582.0	17.5	17.5	0.476	10.0
2	Afghanistan	2013	Developing	59.9	268.0	66	0.01	73.216243	64.0	430	...	62.0	6.13	64.0	0.1	631.744976	31731868.0	17.7	17.7	0.470	9.9
3	Afghanistan	2012	Developing	59.5	272.0	69	0.01	78.194215	67.0	2787	...	67.0	6.52	67.0	0.1	690.659000	3696659.0	17.9	18.0	0.463	9.8
4	Afghanistan	2011	Developing	59.2	275.0	71	0.01	7.097109	68.0	3013	...	68.0	7.67	68.0	0.1	63.537231	2979599.0	18.2	18.2	0.454	9.5

5 rows x 22 columns

Predicting Life Expectancy Using Machine Learning

Prepared by: Kommana Prasanna

1.INTRODUCTION

The term "life expectancy" refers to number of years a person can expect to live. By definition, life expectancy is based on an estimate of the average age that members of a particular population group will be when they die.

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 countrymen which helps in making crucial health decisions

1.1 Overview

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.

This project is to build a model while considering historical data from a period of 2000 to 2015 for all the countries. The model trained in this project will be able to predict the average lifetime of a human being given some input factors. With the help of this project any country is able to predict the expected lifetime of their countrymen and then accordingly take preventive measures to improve on their healthcare measures. This will also help countries in improving a particular field such as GDP, alcohol intake, etc which have a high impact on a country's life expectancy. Good prognostication 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. So this problem statement is aimed at predicting Life Expectancy rate of a country given various features. It predicts the average lifetime of a human being and predicts on the basis of various factors like Regional variations, Economic Circumstances, Sex Differences, Mental Illnesses, Physical Illnesses, Education, Year of their birth and other demographic factors. So the end product will predict the future life expectancy of the person with the help of prior given appropriate matrix of features by the user like current 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.

1.2 Purpose

The average life Expectancy of a certain country says many things about that particular country. It ultimately helps in predicting the health conditions and the development of the health sector in that particular country. This ultimately helps the nation to find the area which needs attention in an urge to improve its contribution in average lifespan of a human being. The expectancy obviously depends upon the country's population, GDP, the economy of the country and many more factors. It is not enough to have a long life , Instead with having a long life one should have a fit life as well.

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.

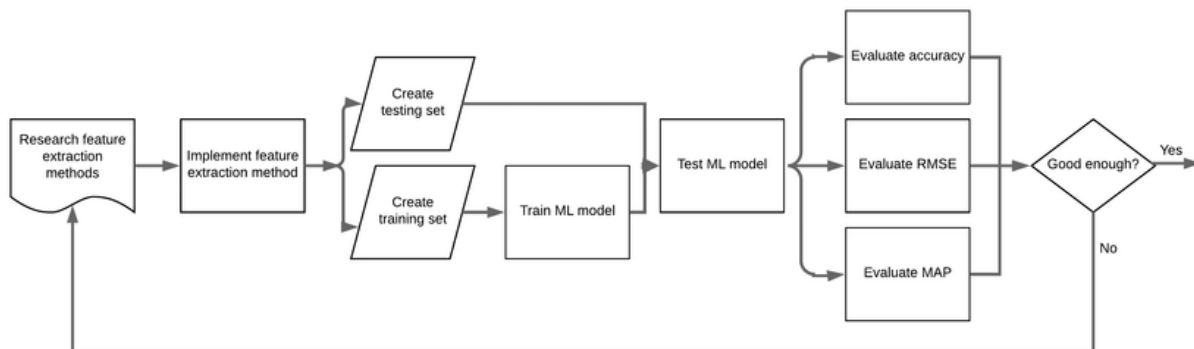
2.2 Proposed Problem

Although there have been a lot of studies undertaken in the past on factors affecting life

expectancy considering demographic variables, income composition and mortality rates. It was found that the effect of immunization and human development index was not taken into account in the past. Also, some of the past research was done considering multiple linear regression based on a data set of one year for all the countries. Hence, this gives motivation to resolve both the factors stated previously by formulating a regression model based on mixed effects model and multiple linear regression while considering data from a period of 2000 to 2015 for all the countries. Important immunization like Hepatitis B, Polio and Diphtheria will also be considered. In a nutshell, this study will focus on immunization factors, mortality factors, economic factors, social factors and other health related factors as well. Since the observations in this dataset are based on different countries, it will be easier for a country to determine the predicting factor which is contributing to lower value of life expectancy. The model of "Predicting Life Expectancy using Machine Learning" 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 Software designing

Software Requirements

- Python IDE
- Excel
- IBM Cloud Account
- IBM watson studio

- Node Red
- Watson Machine Learning

Milestones

- Create IBM cloud account
- Create a node-Red application
- Create IBM Watson studio
- Create Watson Machine learning services
- Collect the dataset for the project from kaggle.com
- Create a watson studio project
- Create a jupyter notebook in IBM watson and import data
- Build a Machine learning model and create endpoints for Node-Red integration
- Build Node-Red flow to integrate ML services

4. EXPERIMENTAL INVESTIGATIONS

This project mainly aims at predicting life expectancy. The basic requirement of the project is the availability of the suitable dataset which will aid the prediction. So in this project I have used the standard WHO dataset on kaggle. The machine learning model is trained on the basis of the data provided, such that it could predict the average lifespan of an individual in the coming years.

Technical Requirements

- ◇ The GUI must be integrated with the backend trained model.
- ◇ The model before training must be given with clean dataset (done by preprocessing)

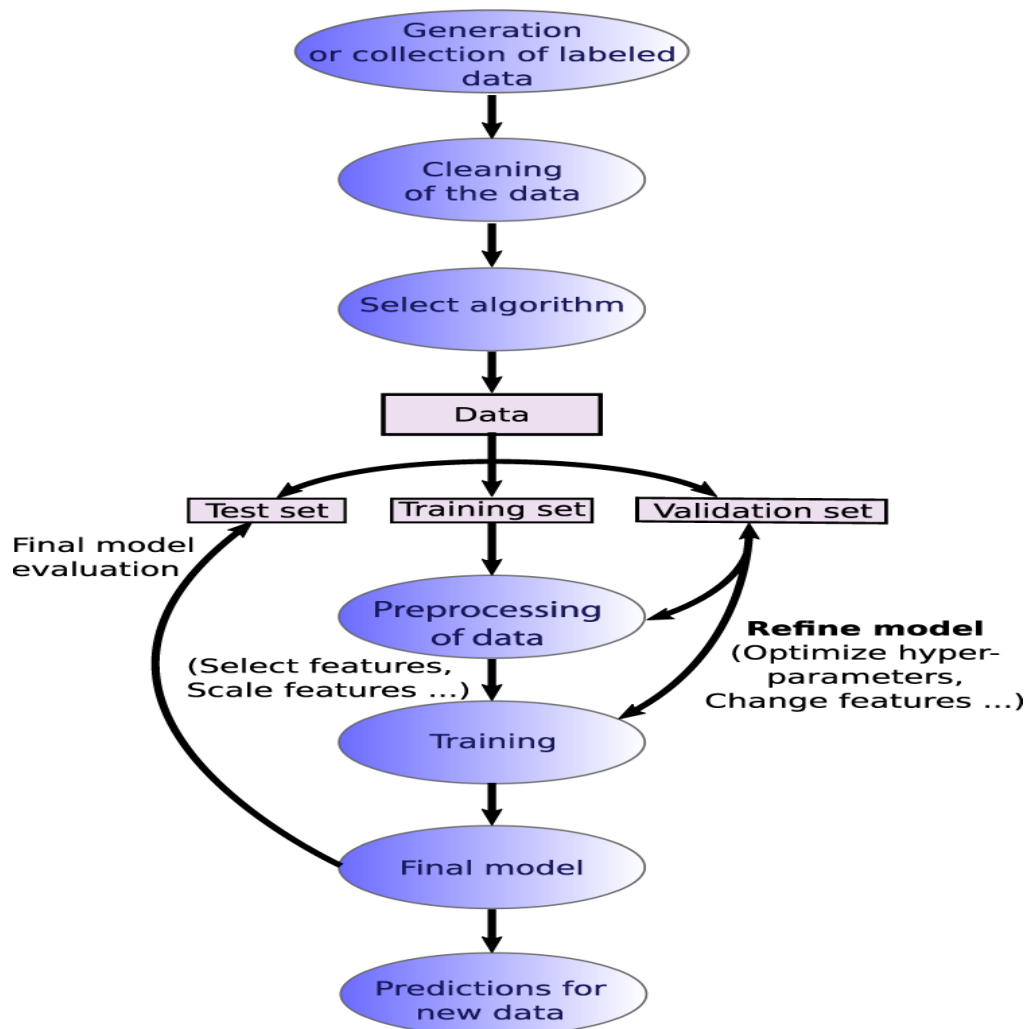
Process

- ◇ Download the dataset of WHO
- ◇ Analyze it and clean the dataset
- ◇ select the best features that gives good prediction
- ◇ Train the regression model on different algorithms
- ◇ Check for the best one and finalize that algorithm to train our mode
- ◇ Build Node red flow for GUI(web app)
- ◇ Create scoring end point for integrating our model to node red
- ◇ Provide the model with the inputs fields

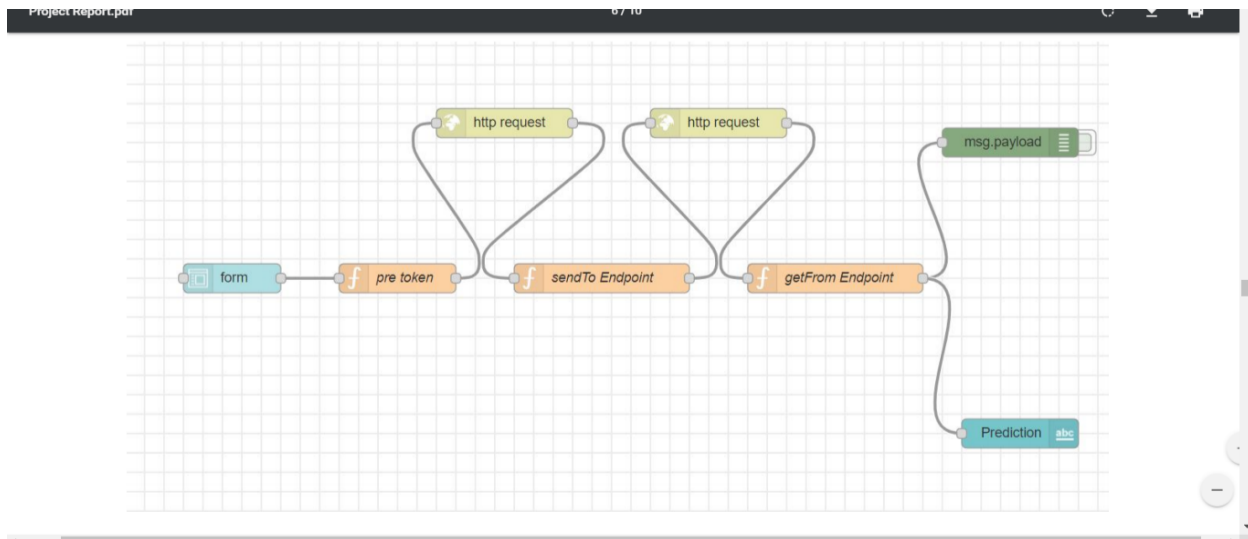
- ◇ The model will return the output as the average predicted lifespan

5. FLOWCHART

A flowchart is a diagram that depicts a flow of process, system or computer algorithm. They are widely used in multiple fields to document, study, plan, improve and communicate complex processes in clear, easy-to-understand diagrams. Flowcharts, sometimes spelled as flow charts, use rectangles, ovals, diamonds and potentially numerous other shapes to define the type of step, along with connecting arrows to define flow and sequence.



Node-Red Flow Diagram



6. RESULT

20:08

-bdfll.eu-gb.mybluemix.net

Home

Life Expectancy Model

Prediction **61.66399999999999**

Year *
2021

Status *
Developed

Adult Mortality *
265

Infant deaths *
66

Alcohol *
0.01

percentage expenditure *
73.214

Hepatitis B *
63

Measles *
466

BMI *
14.7

under-five deaths *
96

Polio *
6

Total expenditure *
8.33

Diphtheria *

The user friendly Graphical User interface is shown in Figure . This GUI is connected to the trained machine learning model present in the backend(IBM watson notebook). 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 the above figure.

7. ADVANTAGES AND DISADVANTAGES

7.1 Advantages

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

7.2 Disadvantages

1. Disadvantages of using IBM Watson:
 - Only in English (Limits areas of use)
 - Seen as disruptive technology
 - Maintenance and even requires internet connection.
 - Doesn't process structured data directly
 - Increasing rate of data, with limited resources

8. APPLICATIONS

Life expectancy is the primary factor in determining an individual's risk factor and the

likelihood they will make a claim. This project/idea is useful for Insurance companies as they consider age, lifestyle choices, family medical history, and several other factors when determining premium rates for individual life insurance policies. The principle of life expectancy suggests that you should purchase a life insurance policy for yourself and your spouse sooner rather than later. Not only will you save money through lower premium costs, but you will also have longer for your policy to accumulate value and become a potentially significant financial resource as you age. It can be used by researchers to make meaningful research out of it and thus, bring something that will help increase the expectancy considering the impact of a specific factor on the average lifespan of people in a specific country.

9. CONCLUSION

Thus, we have developed a model that will predict the life expectancy of a specific demographic region based on the inputs provided. Various factors have a significant impact on the life span such as Adult Mortality, Population, Under 5 Deaths, Thinness 1-5 Years, Alcohol, HIV, Hepatitis B, GDP, Percentage Expenditure and many more. 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 press the “predict” button.

10. FUTURE SCOPE

As future scope, we can connect the model to the database which can predict the life Expectancy of not only human beings but also of the plants and different animals present on the earth. This will help us analyze the trends in the life span. A model with country wise bifurcation can be made, which will help to segregate the data demographically.

APPENDIX

A. Dataset Reference: <https://www.kaggle.com/kumarajarshi/life-expectancy-who>

B. Project Explanation video:

<https://drive.google.com/file/d/1opxkUXwZJe8txKHN6WxJoLQvfuEoqu-L/view?usp=sharing>

C. GUI link:

<https://node-red-bdf11.eu-gb.mybluemix.net/ui/#!/0?socketid=HymPG6P-X2971W6YAAAh>

D. Source code(github link):

<https://github.com/SmartPracticeschool/IISPS-INT-2824-Predicting-Life-Expectancy-using-Machine-Learning>