Predicting Life Expectancy Using Machine Learning

1. ## **Introduction**

I am Akshita Sandip Khandelwal pursing BTech in Information Technology from Government College of Engineering, Aurangabad. Predicting life expectancy is analyzing the life span of an individual by considering various health factors. it helps the person to take an important decision about their phycisal health and all other life issues. We can predict the average lifespan of the people in the country.

a. ### **Overview**

Life expectancy enables us to predict the average life span of people in a specific region by considering various factors. In this project factors considered for the prediction of lifespan are adult mortality, infant deaths, alcohol, percentage expenditure, hepatitis B, measles, BMI, under-five deaths, polio, total expenditure, diphtheria, HIV, GDP, population, thinness, income composition of resources, schooling also considering the name of the country and its developing status.The model is built by considering the historical data from 2000 to 2015 for all the countries. Thus it helps to predict the average life span of the people in the country. It helps the country to control various factors that have a great impact on human life such as alcohol. It enables us to take many decisions based on human life. Any health disease that has a great impact on life will be taken into focus to minimize it.

b. ### **Purpose**

Predicting life expectancy for a country is important for the country to

take many economic and ealth decisions. Since life affects various factors of the country, the prediction of the life span helps the country to focus on the factors that affect the life of the people. The diseases that majorly affect life and which is still increasing rapidly and threatening people's life can be found with this model.

2. Literature Review

a. Existing Problem

After reviewing existing techniques in the prediction of humanLife Expectancy, and finally reached the conclusion that it is possible to predict an Average Life Expectancy for individuals using advancing technologies and devices such as big data, AI, machine learning techniques. It is noticed that the collection of data is a huge challenge due to the privacy and government policy considerations, which will require the collaboration of various bodies in the health industry. 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 predicting life expectancy 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.

b. Proposed Solution

The past research was done considering multiple linear regression based on a data set of one year for all the countries. Hence, 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. 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 a 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 developed with the help of node red presented to the users is a website url.

3. **Theoritical Analysis**

   a. **Block Diagram**
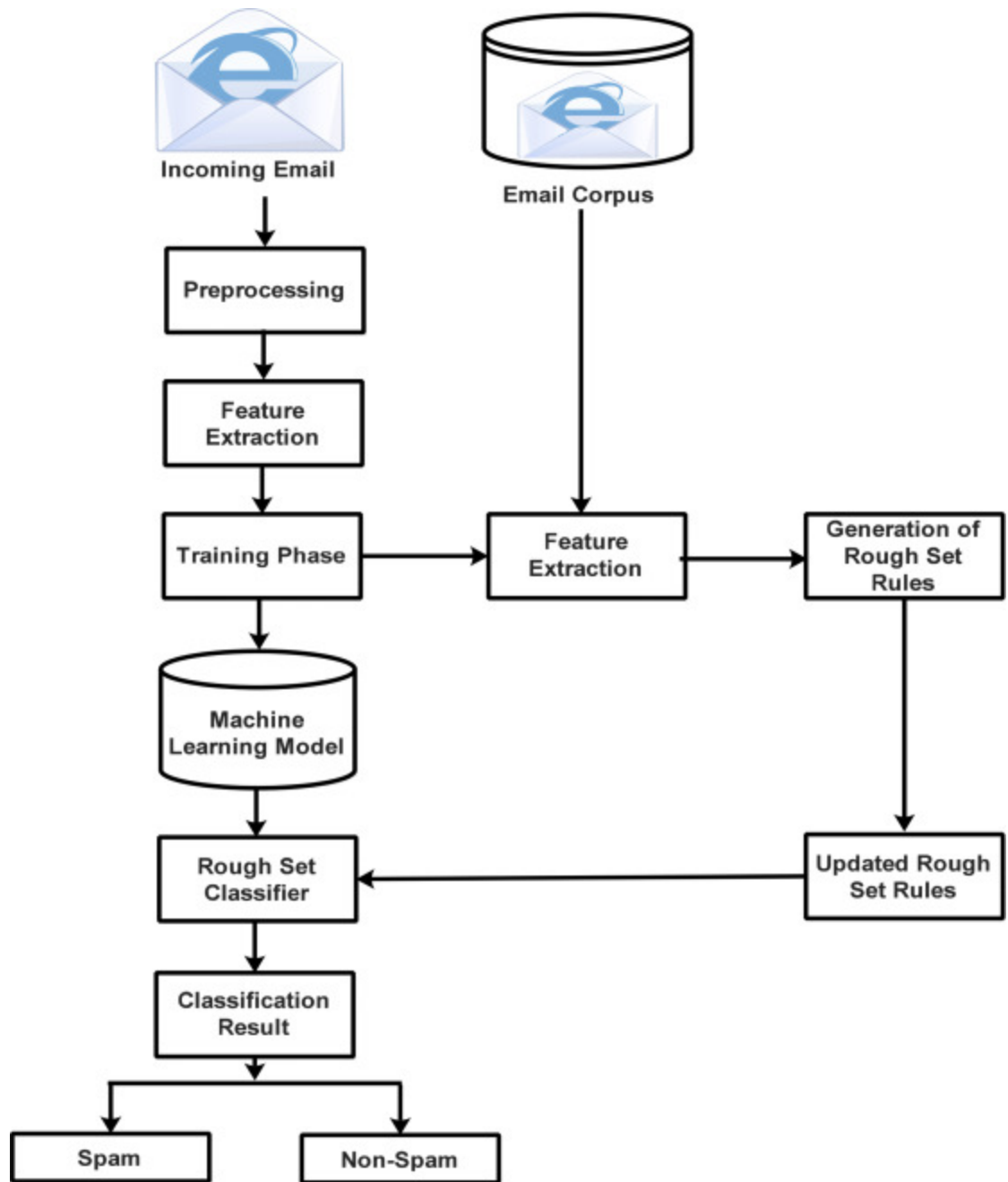
b. **Software Designing**

1. Create an IBM account.
2. Create an appropriate cloud and node red services.
3. Download the dataset(.csv file) from Kaggle and analyze it.
4. Train the regression model on different algorithms. Check for the best one

and finalize that algorithm to train our mode.

5. Build Node red flow for UI(web app).
6. Create a scoring end point for integrating our model to node red.
7. Provide the model with the input fields.
8. The model will return the output as the average predicted lifespan
9. Techincal requirement:
a. The GUI must be integrated with the backend trained model.
b. The model before training must be given with a dataset.
10. Software Requirements:
a. Python IDE
b. Excel
c. IBM Cloud Account
d. IBM Watson
e. Node Red

4. Experimental Investigations

It was my first time to interact with IBM services, I found it too much interesting and worth it. Inspite of my project, I explore many other features on IBM cloud. I also tried my project without python(Auto AI) was really interesting. Also I tried creating many flows on node-red and experimented on it. But at the last I deleted all the extra work despite of my project.

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.
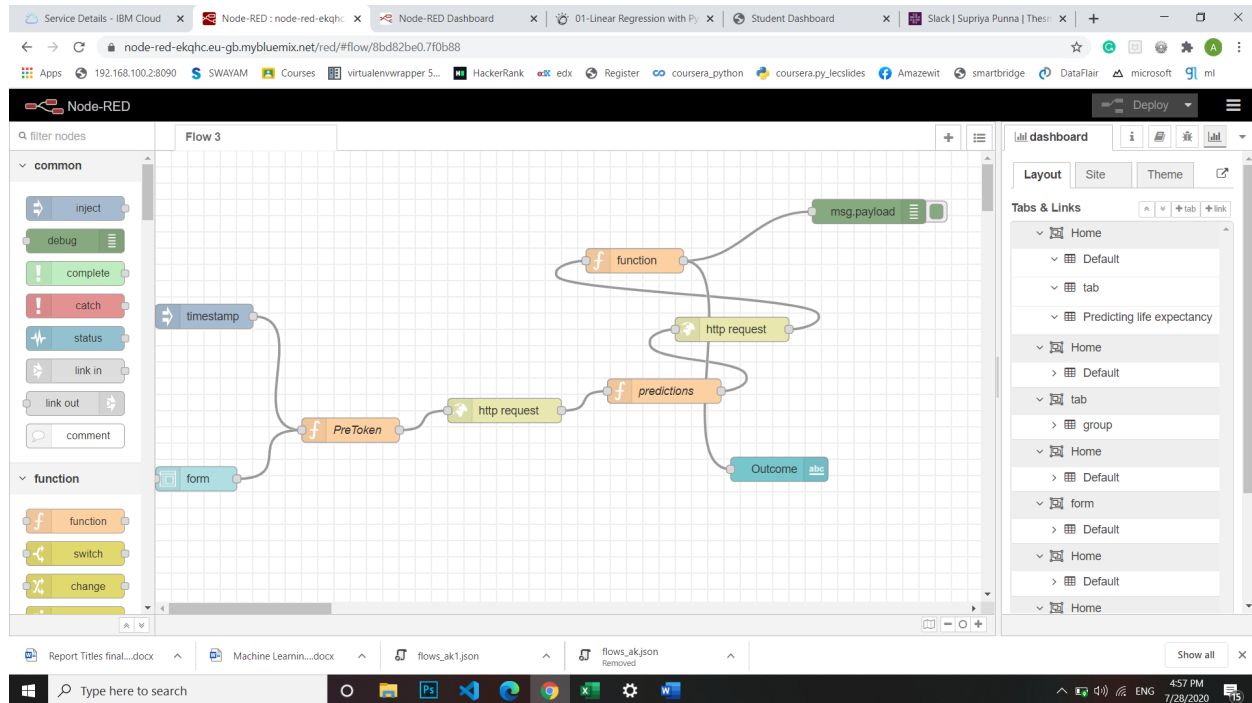
6. **Result**

The User interface is shown below. This UI 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 predicted label, based on the inputs provided as shown in:

7. Advantages and Disadvantages

Advantages:

1. Advantages of using IBM Watson:
   - Processes unstructured data
   - Overcome human limitations
   - Improves performances
   - Abilities by giving the best available dataImprove and transform customer service
   - Handle enormous quantities of data
   - Competitive Advantage
   2. Easy for users to interact with the model via UI.
   3. User-friendly.
   4. Easy to build and deploy.
   5. Doesn't require much storage space.

Disadvantages:

1. Disadvantages of using IBM Watson:
   - Maintenance and even requires an internet connection.
   - Doesn't process structured data directly.
   - Increasing rate of data, with limited resources.

8. **Applications**

   This project has various applications for an individual, company, organization, or country. We are also to distinguish different risk factors for life expectancy, such as smoking-status, occupation, socio-economic class, and others. More complex analyses for assessing cancer survival, that involves comparisons between two populations or a population in two points in time can also be undertaken. When used in biology, age specific fertility rates are also included in the calculations. When data has not been available, such as in low income countries, life tables have been modelled using what data are available, usually childhood mortality data. Life expectancy is the primary factor in determining an individual's risk factor and the likelihood they will make a claim. This project 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. this will not only 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 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 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 the 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.

11. **Bibilography**

I took refernce from youtube and the link provided in the project scope of the internship platform. There was huge support from my mentor.

12. **Appendix**

Dataset link:https://www.kaggle.com/kumarajarshi/life-expectancy-who

UI URL: https://node-red-ekqhc.eu-gb.mybluemix.net/red/#flow/8bd82be0.7f0b88

Source Code:

https://github.com/SmartPracticeschool/llSPS-INT-3261-Predicting-Life-Expectancy-using-Machine-Learning.git