**SMARTBRIDGE**

**Summer Internship Report**

**Predicting Life Expectancy using Machine Learning**

Report written By: Pattela Dhanush

Qualification: Pursuing B.tech 3rd year CSE

College: TKR College of Engineering and Technology

Project Id: SPS\_PRO\_215

Project Title: Predicting Life Expectancy using Machine Learning - SB52850

Organization: SmartBridge

Starting Date: 19-05-2020

Completion Date: 18-06-2020

**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 AND DISADVANTAGES**

**8 APPLICATIONS**

**9 CONCLUSION `**

**10 FUTURE SCOPE**

**11 BIBILOGRAPHY**

**APPENDIX**

A. Source code

1. **Introduction:**
   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.

* 1. **Purpose**

We aimed to develop a predicting life expectancy tool for clinicians counselling patients. The tool would provide estimates of patient life expectancy based on the factors year, GDP, education, alcohol intake ,expenditure and some diseases, etc.

**2 LITERATURE SURVEY**

###### 2.1Existing Problem

Although there have been lot of studies undertaken in the past on factors affecting life expectancy considering demographic variables, income composition and mortality rates.

It was found that 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 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.

**2.2 Proposed Solution:**

Predicting life expectancy is not a new concept. Experts do this at a population level by classifying people into groups, often based on region or ethnicity.

Also, tools such as deep learning and artificial intelligence can be used to consider complex variables, such as biomedical data, to predict someone’s biological age. Biological age refers to how “old” their body is, rather than when they were born. A 30-year-old who smokes heavily may have a biological age closer to 40.

Calculating a life expectancy reliably would require a sophisticated system that considers a breadth of environmental, geographic, genetic and lifestyle factors – all of which have influence.

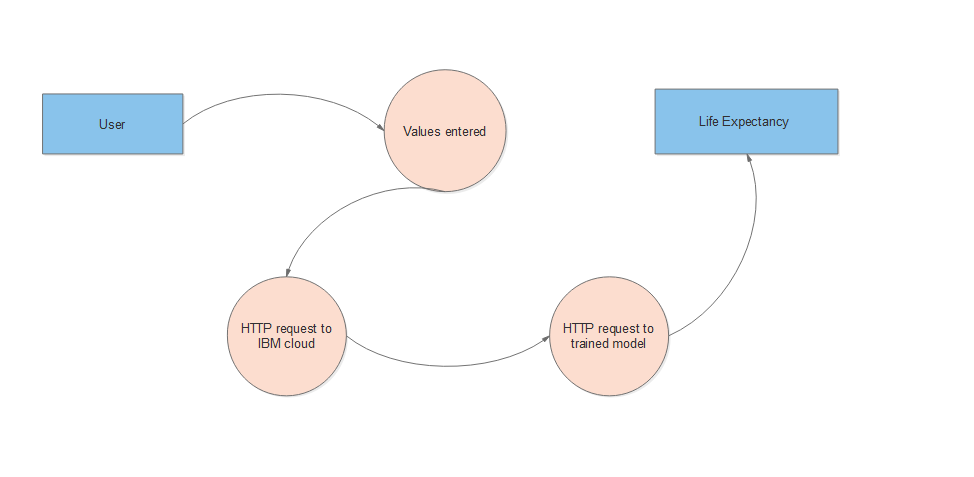
With machine learning and artificial intelligence, it’s becoming feasible to analyse larger quantities of data. The use of deep learning and cognitive computing, such as with IBM Watson, helps doctors make more accurate diagnoses than using human judgement alone.

This, coupled with predictive analytics and increasing computational power, means we may soon have systems, or even apps, that can calculate life expectancy.

IBM cloud provide platform to make model using AUTO AI. I have made model without using a line of code then deployed it. Then using API key, instance key and URL a node-red flow was created. By sharing the URL of node-red flow it can be user friendly and can also use as fun.

1. **THEORITICAL ANALYSIS:**

**3.1 Block diagram:**



##### **3.2 Hardware/Software Design**

##### **Hardware**

1. laptop with 4-GB RAM

2. 2-GB CPU

##### **Software**

1. IBM Cloud Account
2. Jupyter Notebook
3. Watson Studio Service
4. Machine Learning Service
5. Node Red Flow Application

##### **Experimental Investigations**

* Collection of data set from Kaggle. <https://www.kaggle.com/c/predicting-life-expectancy/data>
* On IBM Watson studio machine learning using auto ai build a model to predict life expectancy.

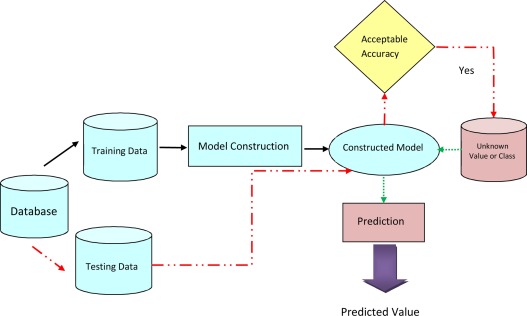
1. To do so first create account on IBM Watson studio.
2. Using Add to project choose auto AI.
3. Then upload data set.
4. Choose best way to predict.
5. Save as a model which is on the top
6. Deploy the model.
7. Test the model.
8. Create service credential.

• Create cloud foundry app https://node-red-shubhangi.eu-gb.mybluemix.net/red/#flow/c8ce04dc.84f0c8

• Make node red flow.

* Then add API key, instance Id and url.
* After deploying the model from dashboard UI can be seen.

1. **Flowchart**



##### **RESULT**

The user should enter the credentials required for prediction. The average life expectancy prediction will be displayed on the screen.

##### **Advantages and Disadvantages**

##### Advantages:

**1.** Identifies trends and patterns easily- Machine learning involves reviewing large volumes of data to discover specific trends and patterns that would most often not be apparent to humans.

1. Automation- Machine learning does not require human intervention. It gives machines the ability to learn. It helps machines make predictions and improve the algorithms by themselves. Anti-virus software is a common example of this as they automatically filter new threats as & when they are recognized.
2. Continuous improvement- Machine learning algorithms improve in accuracy and efficiency as they gain experience. This helps them take better decisions.

##### **Disadvantages**

1.Time and resources- Machine learning requires massive resources to function. It may demand additional computing power. Machine learning requires enough time to let the algorithms learn & develop to fulfil their intended purpose with a considerable amount of accuracy and relevancy.

2.Interpretation of results- Accurately interpreting the results generated by the algorithms is a challenging task. One needs to exercise caution while choosing algorithms for their specific purpose.

3. Data acquisition- Machine learning needs massive datasets to train on. These must be unbiased/inclusive and of good quality. In certain situations, they may need to wait for new data to be generated.

##### **Applications**

1. The project can be used as a basis to develop personalized health applications.

2. As the model uses a wide range of features for prediction, it will be easier for a country to determine the predicting factor which is contributing to lower value of life expectancy. This will help in suggesting a country, which area should be given importance in order to efficiently improve the life expectancy of its population.

3. The governments can plan and develop their health infrastructures by keeping the most correlated factors in mind.

4. The project can help governments to keep track of their countries’ health status so they can plan for the future accordingly.

##### **Conclusion**

The project makes use of electronic medical records for predicting life expectancy. The potential use of project is not limited to health care in practice, but could also be useful in other clinical applications such as clinical trials. In clinical trials, outcomes often depend on prognostic factors. Automatic processing of medical records would enable quick and systematic stratification of patients based on their prognoses, which could be used to further reduce biases. The project makes a good use of machine learning in predicting life expectancy of a country that can help respective government in making policies that will serve for the benefit of the nation and entire humankind.

##### **Future Works**

Some future improvements can be made on current project. They are as follows:

1. As more data comes, that can be fed to the model for more accurate predictions.
2. Currently, the project is just a web application. It can be developed to support other platforms like Android, IOS and Windows Mobile.
3. Other regression models can also be used for prediction and later the best among them should be chosen.
4. User interface can be modified for various countries according to their regional languages.

##### **Bibilography**

[1] IBM Cloud setup [Online]. Available:

<https://www.ibm.com/cloud/get-started> .

[2] IBM Developer, “Node-RED starter tutorial” [Online]. Available: <https://developer.ibm.com/tutorials/how-to-create-a-node-red-starter-application/> .

[3] “Node-RED labs on the use of the Watson Developer Cloud services - watson-developer-cloud/node-red-labs.” [Online]. Available: <https://github.com/watson-developer-cloud/node-red-labs> .

[4] “Infuse AI into your applications with Watson AI to make more accurate predictions”. [Online]. Available: <https://www.ibm.com/watson/products-services> .

[5] IBM Watson, “Intro to IBM Watson”, 2018 [Online]. Available: <https://www.youtube.com/watch?v=W3iPbFTAAds&feature=youtu.be> .

[6] “Get an understanding of the principles of machine learning.” [Online]. Available: <https://developer.ibm.com/technologies/machine-learning/series/learning-path-machine-learning-for-developers/> .

[7] IBM Developer, “IBM Watson Machine Learning: Get Started in IBM Cloud”, 2020 [Online]. Available: <https://www.youtube.com/watch?v=NmdjtezQMSM> .

[8] Watson Studio Workshop, “Chapter 4 Build and Deploy models in Jupyter Notebooks” [Online]. Available: <https://bookdown.org/caoying4work/watsonstudio-workshop/jn.html> .

[9] Kumar Rajarshi, “Life Expectancy (WHO) Statistical Analysis on factors influencing Life Expectancy”, 2018. [Online]. Available: <https://www.kaggle.com/kumarajarshi/life-expectancy-who>

[10] IBM Developer, “IBM Watson: Sign up for Watson Studio and Watson Knowledge Catalog”, 2019. [Online]. Available: <https://www.youtube.com/watch?v=DBRGlAHdj48&list=PLzpeuWUENMK2PYtasCaKK4bZjaYzhW23L> .

[11] IBM Developer, “IBM Watson Studio: Create a project”, 2019. [Online]. Available: <https://www.youtube.com/watch?v=-CUi8GezG1I&list=PLzpeuWUENMK2PYtasCaKK4bZjaYzhW23L&index=2>

[12] IBM Developer, “IBM Watson Studio: Jupyter notebook basics”, 2019 [Online]. Available: <https://www.youtube.com/watch?v=Jtej3Y6uUng>

##### **APPENDIX:**

##### **Source Code**

The experiment is done under AutoAI. So the source code is present in the experiment itself.

**Node Red Flow is:**

