Project Documentation

Project Report on

Predicting Life Expectancy using Machine Learning

Under

Remote Summer Internship Program 2020 by SmartInternz

Project by:

Sharath K.P.

BE 3rd year (CSE)

Babaria Institute of Technology

Email: sharathsivadas98@gmail.com

Mobile: +918469873001

Project Title: Predicting Life Expectancy using Machine Learning.

1. Introduction:

1.1 Overview: Life expectancy is a statistical measure of the average time a human being is expected to live. This project's aim is to predict the average life expectancy of a human being of a country which depends on various factors like Regional variations, Economic Circumstances, Sex Differences, Mental Illnesses, Physical Illnesses, Education, Year of their birth and other demographic factors.

1.2 Purpose and working: The purpose of this project is to predict the life expectancy of a person based on the inputs given by the user. In this project an UI is provided that is built using Node-Red application of IBM cloud in which the user will provide some values like year, adult mortality, alcohol, total expenditure, BMI, info regarding some diseases, thinness, schooling, etc. and the user will get the predicted life expectancy as output. The project uses a **Random Forest Regressor Machine Learning** model to predict the life expectancy which is deployed on IBM cloud using services like IBM Watson Studio and IBM Watson Machine Learning Service. I have also prepared a model using IBM AutoAI experiment which allows us to create a machine learning model without any knowledge of coding. I have deployed that model as well and integrated with a Node-Red UI also. The dataset used or the training of the model was downloaded from kaggle.com and Python is used to write the code for machine learning model.

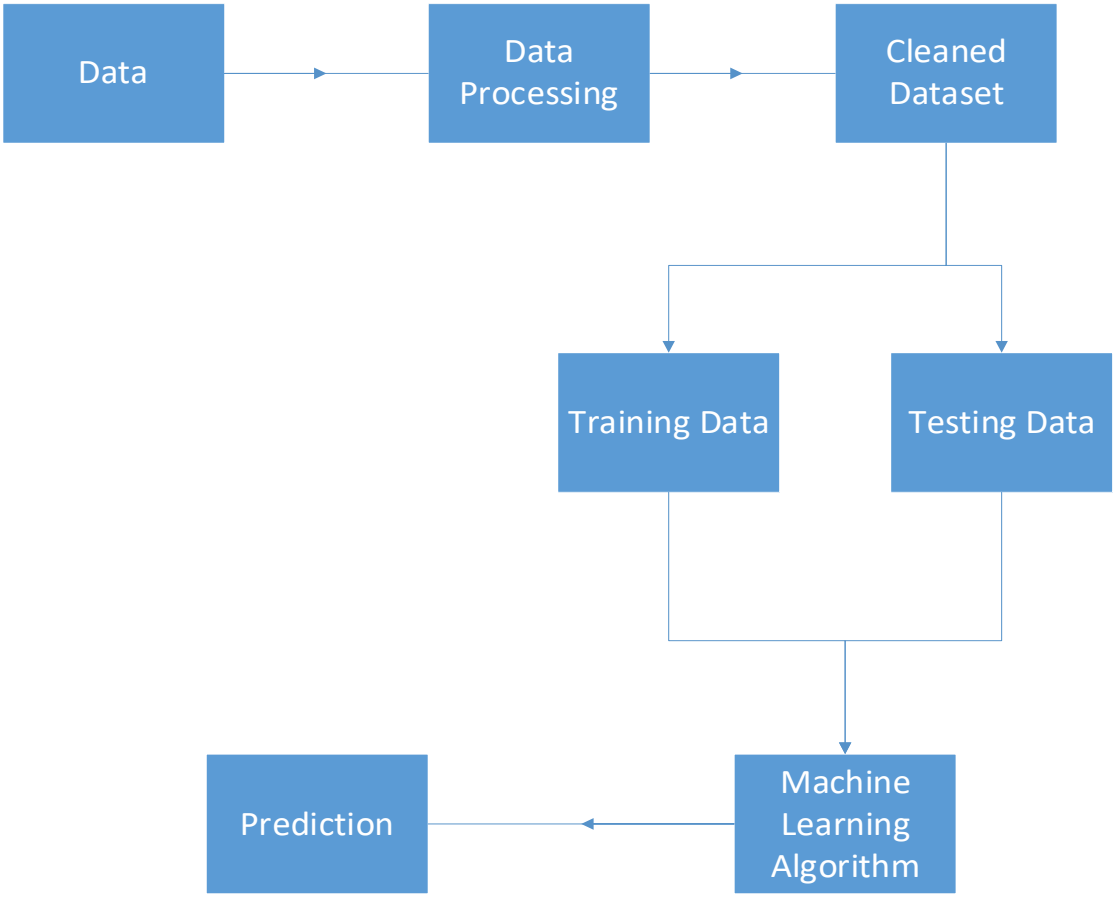
2. Literature Survey:

2.1 Existing problem: Currently the life expectancy of a person is calculated on the basis of the already collected data and the life expectancy of a person in future coming years cannot be known because of lack of technology. Due to which many problems may arise because the people are not prepared to tackle the issues that may affect the life expectancy as they don't know whether a factor may increase or decrease the life expectancy.

2.2 Proposed Solution: If the life expectancy of a person can be predicted for the coming years then we can be aware of the factors that may affect the life expectancy either in a positive way or in a negative way and accordingly take necessary actions or precautions.

3. Theoretical Analysis:

3.1 Block diagram:



3.2 Hardware/Software Design: The following were used to create the project:

Technical tools/methods/skills: The technical requirements of this project are:

- Datasets
- Knowledge of python
- Machine Learning
- Models
- Algorithms
- Libraries

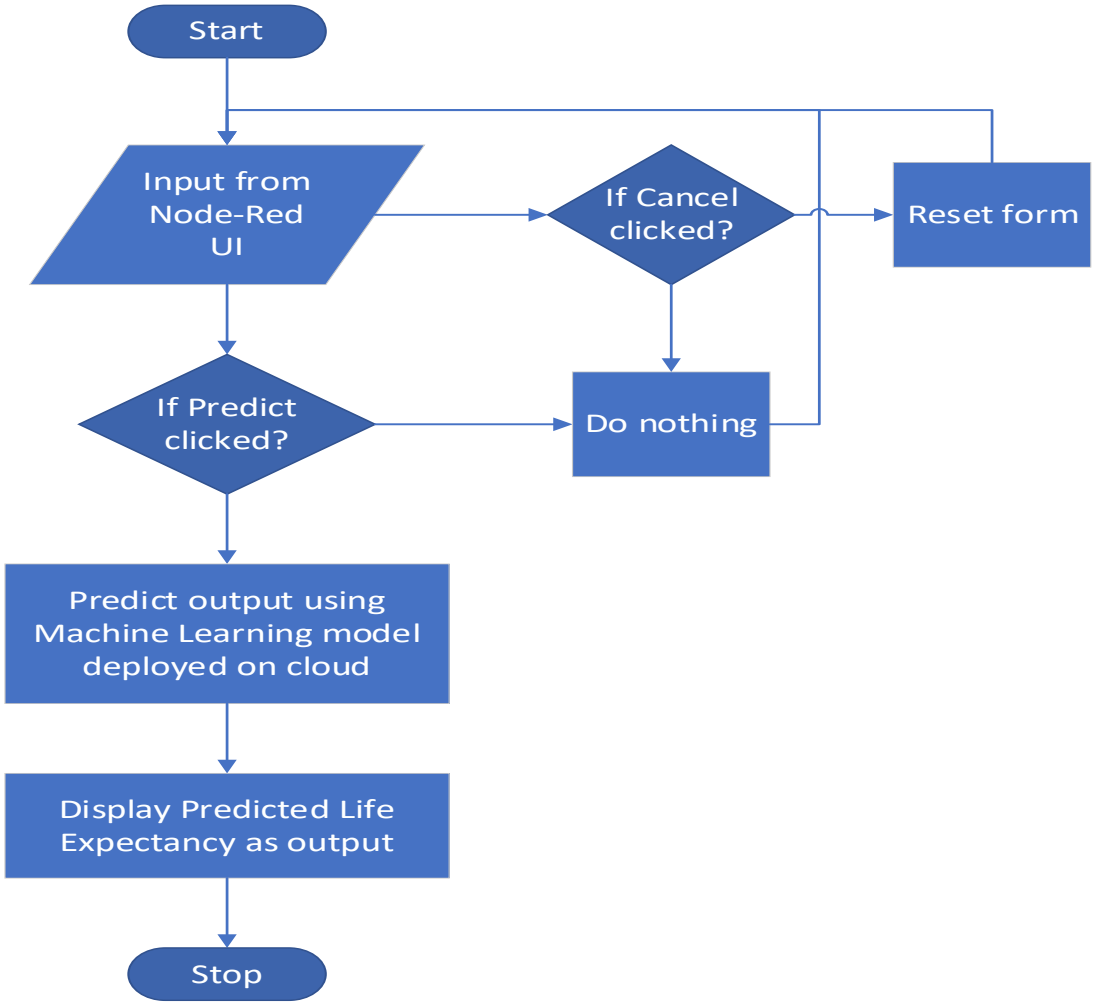
Softwares/services used: The software requirements of this project are:

- IBM cloud services
- IBM Watson services

- IBM Watson Studio
- IBM AutoAI experiment
- IBM Node-Red application
- SmartInternz Project Workspace
- Jupyter Notebook
- Github
- Slack
- Zoho document writer

4. **Experimental Investigations:** In experimental investigations I searched for the dataset. After the dataset was obtained, I analyzed the data and tried to find out the relation between each of the columns and which columns will be necessary for the project and how to preprocess or clean it. And after the I tried to find the best model that will fit my data to give the desired output. I also had done some research about the IBM cloud platform so that I could learn some basics about it.

5. **Flowchart:**



6. **Result:** The result of the project is the predicted life expectancy of a person based on the provided inputs.

7. **Advantages and Disadvantages:**

7.1 Advantages:

- Life Expectancy can be predicted
- Data can be analyzed
- Factors affecting Life expectancy can be analyzed

7.2 Disadvantages:

- Error in data can result in wrong prediction
- Accuracy is not 100%
- Error may occur due to inappropriate analysis of data

8. **Application:** To predict Life Expectancy.

9. **Conclusion:** We can conclude that this project will be able to predict the life expectancy of a person using machine learning and can thus we can analyze the data accordingly to take actions for future.

10. **Future Scope:** The machine learning model used for the project can be trained well using more and accurate data and more appropriate and trained model can be used to increase the accuracy and get more accurate result.

11. **Bibliography:**

- YouTube – https://www.youtube.com/
- IBM tutorials - https://developer.ibm.com/tutorials/
- https://developer.ibm.com/technologies/machine-learning/series/learning-path-machine-learning-for-developers/
- https://scikit-learn.org/stable/modules/preprocessing.html#preprocessing
- For Dataset - https://www.kaggle.com/kumarajarshi/life-expectancy-who

Appendix

Link to access the web-application:

https://node-red-zmrye.eu-gb.mybluemix.net/ui/

Link to the source code:

https://github.com/SmartPracticeschool/llSPS-INT-1696-Predicting-Life-Expectancy-using-Machine-Learning

Link to the project demonstration video:

https://youtu.be/wJ75ED6jOzs