Project Report

1. INTRODUCTION

1.1 Overview

The project is about developing and deploying a machine learning model which will predict the average life expectancy of various countries by providing the various factors affecting it. It will be integrated with Node-RED to create a web application which will display the average life expectancy of a country when the factors affecting it are provided as input to the application.

1.2 Purpose

The purpose of the project is to make government and private organizations as well as people aware about the general health conditions by predicting the average life expectancy.

1. LITERATURE SURVEY

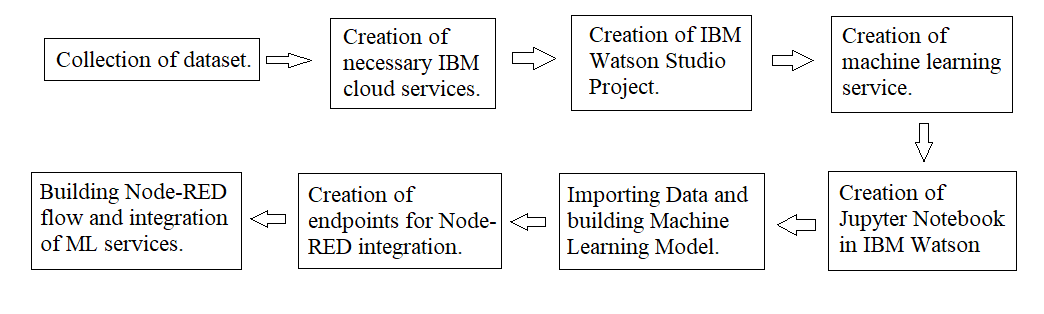
2.1 Existing problem

Many countries are not disposing proper healthcare services and the citizens are unaware about various health issues. They are least concerned about maintaining personal and social health.

2.2 Proposed Solution

Predicting life expectancy and providing the factors which majorly affect it will help the government and healthcare institutions to put emphasis on improving those amenities and facilities for the citizens and inspire the citizens to adopt healthy lifestyle habits.

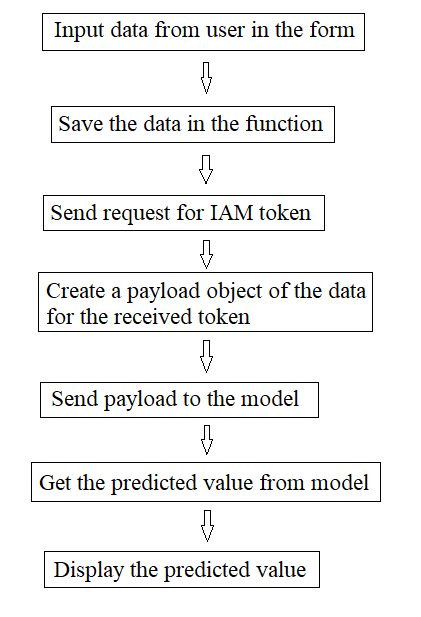
1. THEORETICAL ANALYSIS

3.1 Block Diagram

3.2 Hardware / Software Designing

1. Loading libraries: The libraries required for creating the model such as pandas, matplotlib, seaborn, numpy, sklearn, etc are imported.
2. Loading dataset: The csv file containing the dataset is stored in a dataframe.
3. Data Preprocessing:
   1. Checking data types: The data type of each column is checked whether it is suitable for the analysis, if not the data type is changed.
   2. Checking empty cells: If values are missing in the dataset, then they’re filled.
   3. Plotting graphs: Graphs are plotted on the data to gain a better understanding of the data.
   4. Checking correlation: Correlation between different columns and the column to be predicted is checked. If the correlation between any column and the prediction column is negligible, then that column can be neglected.
   5. Checking outliers: If there are outliers in the dataset, they can be eliminated.
   6. Scaling data: It is done to normalize the data in a particular range if it varies too much in order to speed up the calculations.
4. Splitting the data: The data is split into training and testing parts.
5. Creating a model: A model is created using the suitable algorithm.
6. Training and testing the model: The model is trained by providing the data segregated for training and tested by using the data preserved for testing.
7. Predicting the result: The result column is predicted. The predicted column is compared with the original column.
8. Finding accuracy: The accuracy and the error in the model is calculated.
9. Saving the model: The model is saved and endpoints are created for Node-RED integration.
10. Deploying the model: The model is deployed using IBM Watson Studio.
11. Integrating the model with Node-RED: A Node-RED flow is created and with the help of endpoints, the model is integrated with Node-RED to create a web application.

1. EXPERIMENTAL INVESTIGATIONS
2. Life Expectancy is positively related to the factors like alcohol consumption, expenditure on health as percent of GDP per capita, Hepatitis-B immunization coverage, average BMI of the population, polio immunization coverage, percentage government expenditure on health, DPT immunization coverage, GDP of the country, human development index based on income composition of resources and number of years of schooling.
3. Life Expectancy is negatively related to the factors like adult mortality rate, number of infant deaths, number of reported measles cases, number of under five deaths, number of deaths caused due to HIV, population of the country, prevalence of thinness among young and adolescent children.
4. Life Expectancy of different countries vary depending on their social, economical, demographic and financial factors.
5. According to the data, the average life expectancy of the world is around 69 years of age.
6. FLOWCHART



1. RESULT

The Machine Learning Model is able to predict the average life expectancy of a nation. A web application is created using Node-RED which is integrated with the Machine Learning Model which displays the life expectancy of a particular nation when the various factors affecting the life expectancy are provided via input.

1. ADVANTAGES AND DISADVANTAGES

Advantages:

1. The project makes it possible for any person or organization to find out the life expectancy of any country provided that he/she has the appropriate data.
2. Life Expectancy prediction is useful to governments, health service providers, insurance companies, etc.
3. It will make people more aware about their health and the need to improve it.

Disadvantages:

1. The project does not take into consideration factors like deaths caused due to natural and man-made disasters including accidents which also affect the average life expectancy of a nation.
2. The project gives us a prediction of the average life expectancy of a nation and not of individuals.
3. The life expectancy of individuals may vary according to their personal, economic and financial conditions.
4. APPLICATIONS
5. It will help insurance companies to offer individualised services.
6. Government can use the predictions to provide better health care services.
7. Life expectancy is an indicator of growth and development of a country, hence the prediction will give an insight of the current development status of the nation, so that it can take steps which will promote the development rate of the country.
8. CONCLUSION
9. The average life expectancy of a nation depends upon its various socio-economic, financial, immunization, health, mortality and other factors.
10. Life expectancy is an essential factor which describes the current development status of a nation.
11. It is a significant element in end-of-life decision making.
12. Predicting the average life expectancy of a nation will help in the development of the nation by providing an insight about the development index of the nation.
13. FUTURE SCOPE
14. The project can be used by the government, healthcare firms and insurance companies to plan their activities.
15. The project can be expanded in the future by adding more features in the dataset which affect the life expectancy.
16. The project can be considered as a basis for calculating development index by the government, developing insurance plans by the insurance companies, etc.
17. BIBLIOGRAPHY

<https://my15.digitalexperience.ibm.com/b73a5759-c6a6-4033-ab6b-d9d4f9a6d65b/dxsites/151914d1-03d2-48fe-97d9-d21166848e65/>

<https://github.com/watson-developer-cloud/node-red-labs>

<https://www.ibm.com/watson/products-services>

<https://developer.ibm.com/technologies/machine-learning/series/learning-path-machine-learning-for-developers/>

<https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-019-0775-2#:~:text=Life%20expectancy%20is%20one%20of,broadly%3A%20facilitates%20Advance%20Care%20Planning.>

<https://bookdown.org/caoying4work/watsonstudio-workshop/jn.html>

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

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

<https://www.youtube.com/watch?v=DBRGlAHdj48&list=PLzpeuWUENMK2PYtasCaKK4bZjaYzhW23L>

<https://www.youtube.com/watch?v=-CUi8GezG1I&list=PLzpeuWUENMK2PYtasCaKK4bZjaYzhW23L&index=2>

<https://www.youtube.com/watch?v=Jtej3Y6uUng>

<https://towardsdatascience.com/data-preprocessing-concepts-fa946d11c825>

<https://towardsdatascience.com/ways-to-detect-and-remove-the-outliers-404d16608dba>

<https://bookdown.org/caoying4work/watsonstudio-workshop/jn.html#deploy-model-as-web-service>

<https://bookdown.org/caoying4work/watsonstudio-workshop/auto.html#add-asset-as-auto-ai>

<https://dataplatform.cloud.ibm.com/docs/content/wsj/analyze-data/github-integration.html>

1. APPENDIX

Source Code

https://github.com/SmartPracticeschool/llSPS-INT-2074-Predicting-Life-Expectancy-using-Machine-Learning.git