LIFE EXPECTANCY PREDICTION

DOCUMENTATION

Darshana S

Smartinternz intern

IISPS-INT-2105

Table of Contents

CHAPTER 1 2
INTRODUCTION 2
CHAPTER 2 3
LITERATURE SURVEY 3
CHAPTER 3 4
THEORETICAL ANALYSIS 4
CHAPTER 4 5
EXPERIMENTAL INVESTIGATIONS 5
CHAPTER 5 6
FLOWCHART 6
CHAPTER 6 7
RESULTS 7
CHAPTER 7 8
ADVANTAGES AND DISADVANTAGES 8
CHAPTER 8 9
APPLICATIONS 9
CHAPTER 9 10
CONCLUSION AND FUTURE SCOPE 10
BIBLIOGRAPHY 11
APPENDIX 12

CHAPTER 1

INTRODUCTION

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 project is to predict the life expectancy rate of a country given various features.

1.2 Purpose

Life expectancy is one of the most important indicators of socio-economic development. Mortality analyses are of widespread interest among academics, policymakers, medical researchers, and others in order to direct the flow of funds in the most effective way possible to the population groups in most need. Due to the fact that the crude death rate is not a precise indicator of the mortality level or of the health conditions and living standards in a country, international publications and researchers nowadays regularly make use of life expectancy at birth in the analysis and the description of the level of mortality.

CHAPTER 2

LITERATURE SURVEY

2.1 Existing 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 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 data set of one year for all the countries.

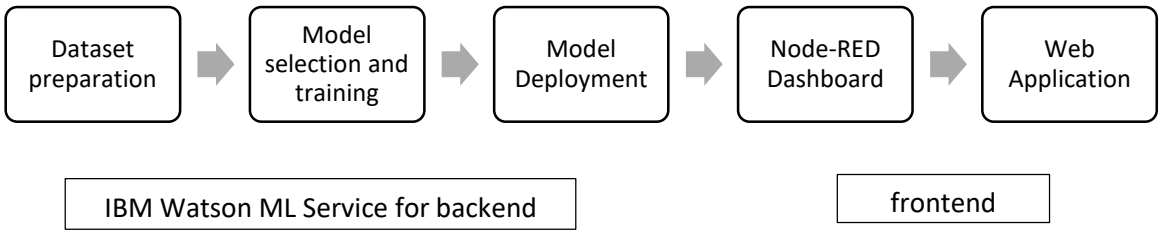
2.2 Proposed solution

This problem can be solved by incorporating both the factors stated previously and formulating a regression model 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 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. This will help in suggesting a country which area should be given importance in order to efficiently improve the life expectancy of its population.

CHAPTER 3

THEORETICAL ANALYSIS

3.1 Block Diagram



3.2 Software Design

The project has been designed entirely using IBM Cloud Services. The backend consists a regression model made using IBM Watson Machine Learning Service. Data cleaning, feature engineering, model selection and training was done using AutoAI. The frontend consists a web application made using Node-RED.

- IBM Watson Machine Learning Service:

  Helps data scientists and developers accelerate AI and machine-learning deployment. It has the capability to deploy models using IBM Watson studio and other open source tools, dynamically retrain models and automatically generate AI-Powered applications.

- AutoAI:

  This an AI Lifecycle management tool which automates the process of data preparation, model development, feature engineering and hyper-parameter optimisation.

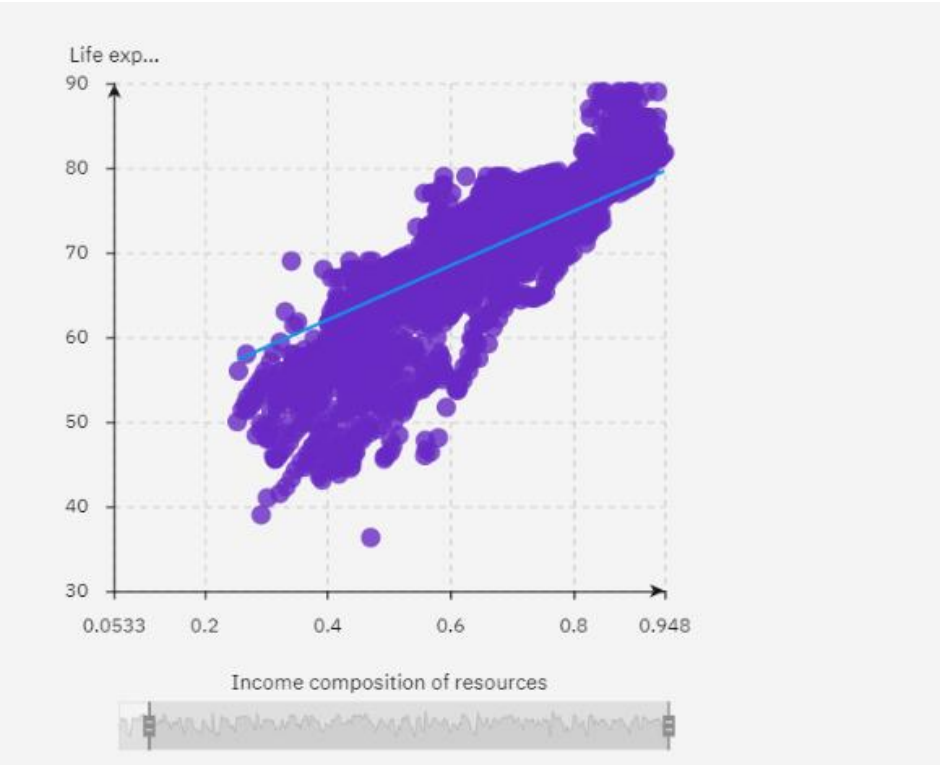
- Node-RED:

  This is a programming tool for wiring together hardware devices, APIs and online services in new and interesting ways. It provides a browser-based editor that makes it easy to wire together flows using the wide range of nodes in the palette that can be deployed to its runtime in a single-click. The light-weight runtime is built on Node.js, taking full advantage of its event-driven, non-blocking model. This makes it ideal to run at the edge of the network on low-cost hardware such as the Raspberry Pi as well as in the cloud. The flows created in Node-RED are stored using JSON which can be easily imported and exported for sharing with others.

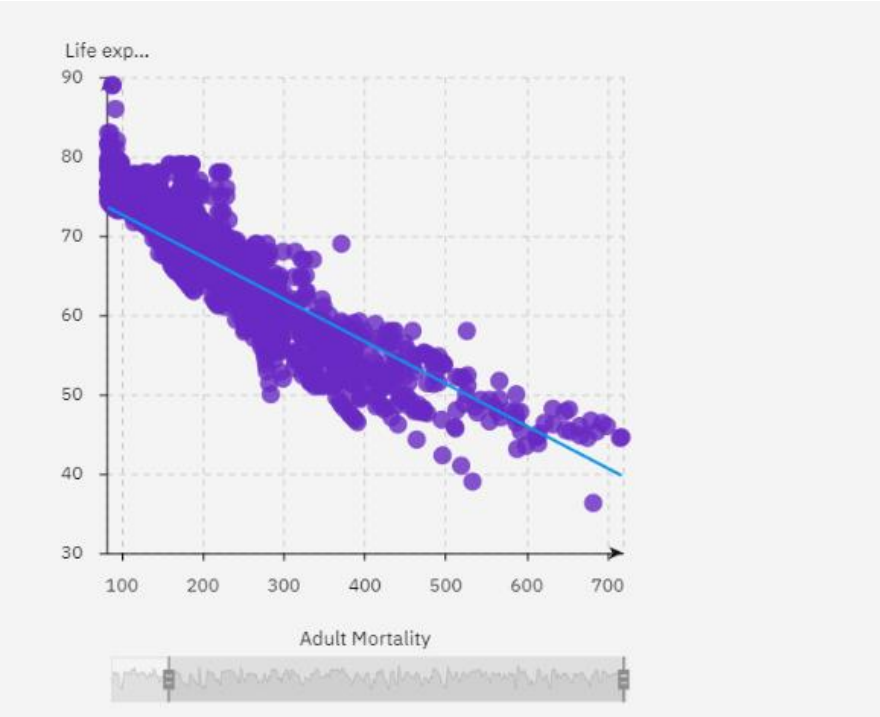
CHAPTER 4

EXPERIMENTAL INVESTIGATIONS

1. Life expectancy and Income composition seem to be highly co-related



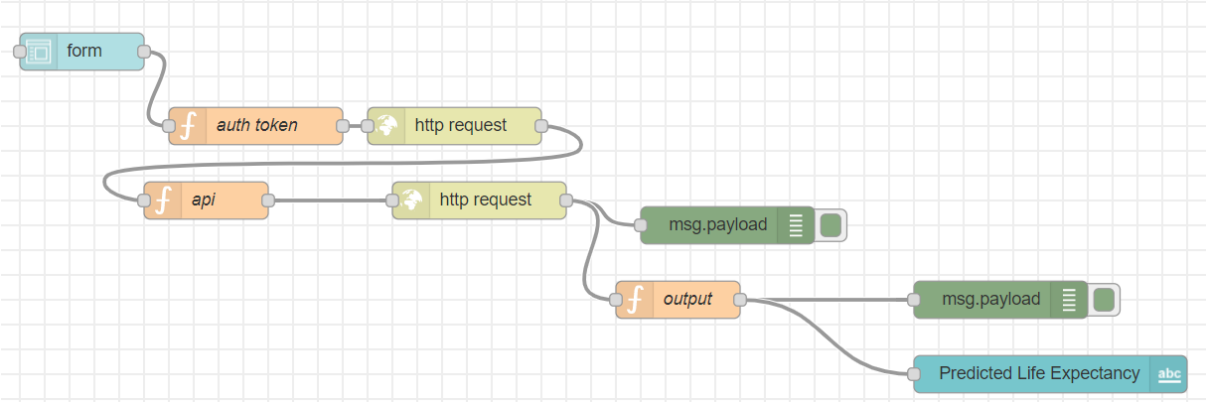
2. Life expectancy and Adult Mortality have seem to inversely related



CHAPTER 5

FLOWCHART

Node-RED flow



CHAPTER 6

RESULTS

The dataset was split into two for training and testing (hold out). 90% of the data was used for training and 10% for testing. The method adopted for cross validation is 3-fold cross validation.

Metrics	Holdout score	Cross validation score
Root Mean Squared Error (RMSE)	1.830	2.010
R^2	0.961	0.956
Explained Variance	0.961	0.956
Mean Squared Error (MSE)	3.347	4.057
Mean Squared Log Error (MSLE)	0.001	0.001
Mean Absolute Error (MAE)	1.182	1.282
Median Absolute Error (MedAE)	0.740	0.747
Root Mean Squared Log Error (RMSLE)	0.028	.031

CHAPTER 7

ADVANTAGES AND DISADVANTAGES

7.1 Advantages

- The Web-application is user-friendly.
- The indicator is well understood, widely used and reported. Regularly updated national data is available for almost half of countries, and there is now substantial global momentum to improve civil registration and vital statistics, including death registration data, in countries without reasonable national coverage at present. There are reasonably acceptable methods to estimate child mortality and adult mortality from other sources, though regularity of data availability and time delays remain a problem, as does the problems of assessing levels of under-reporting.

7.2 Disadvantages

- Does not directly address non-fatal health outcomes, disability etc except through the proxy of mortality risks

CHAPTER 8

APPLICATIONS

Some potential applications for this project is listed below:

- Governments may be able to use predictions to more efficiently allocate limited resources, such as social welfare assistance and health care funding, to individuals and areas of greater need.
- Can be used by insurance companies to provide individualised services, such as how some car insurance companies use black-box technology to reduce premiums for more cautious drivers.
- Can help people be more aware of their general health, and its improvement or deterioration over time. This may motivate them to make healthier lifestyle choices.

CHAPTER 9

CONCLUSION AND FUTURE SCOPE

Life expectancy at birth is largely affected by the population health and socioeconomic development in the country; in other words, when population health and socioeconomic development in a country are getting better, infant mortality rate has decreased; accordingly, the life expectancy at birth appears to have increased. GDP per capita increases the life expectancy at birth through increasing economic growth and development in a country and thus leads to the prolongation of longevity. Hence it is vital to be able to calculate Life Expectancy in a reliable manner.

Further enhancements can be done by improving the quality and quantity of the dataset. Countries can further divided into regions and predictions can be done for every region. Personal life expectancy prediction can also be done in a similar manner if relevant data for individuals can be obtained.

BIBLIOGRAPHY

https://www.who.int/healthinfo/indicators/hsi_indicators_SDG_TechnicalMeeting_December2015_BackgroundPaper.pdf?ua=1

Miladinov, G. Socioeconomic development and life expectancy relationship: evidence from the EU accession candidate countries. Genus 76, 2 (2020). https://doi.org/10.1186/s41118-019-0071-0

https://en.wikipedia.org/wiki/Life_expectancy

APPENDIX

A. Link to Source code

https://github.com/SmartPracticeschool/llSPS-INT-2105-Predicting-Life-Expectancy-using-Machine-Learning.git