INTRODUCTORY REPORT

ON

PREDICTING LIFE EXPECTANCY USING MACHINE LEARNING

BY

IRISH MEHTA

Prepared in partial fulfilment of the
Smartbridge internship procedure,
23rd May, 2020

Table of Contents

Project Summary

The project is aimed at creating a model based on data provided by the World Health Organization (WHO) to evaluate the life expectancy for different countries in years, implemented on IBM cloud using NODE RED integration for the endpoints and using AUTO AI for training without the code. The data offers a timeframe from 2000 to 2015. The data being used to create the model is available on kaggle at the following link. There are many existing models based on the same dataset which utilize different supervised learning algorithms like regression and xgboost.

Some documentation on the data:
The final dataset consists of 22 Columns and 2938 rows. All predicting variables was then divided into several broad categories:Immunization related factors, Mortality factors, Economical factors and Social factors.

The documentation for each column is given:
country: country
year: year
status: Developed or Developing
Life expectancy*:* Life Expectancy in age
Adult Mortality*:* Adult Mortality Rates of both sexes (probability of dying between 15 and 60 years per 1000 population)
infant deaths: Number of Infant Deaths per 1000 population
Alcohol*:* Alcohol, recorded per capita (15+) consumption (in litres of pure alcohol)
percentage expenditure*:* Expenditure on health as a percentage of Gross Domestic Product per capita(%)
Hepatitis B: Hepatitis B (HepB) immunization coverage among 1-year-olds (%)
Measles*:* Measles - number of reported cases per 1000 population
BMI*:* Average Body Mass Index of entire population
under-five deaths*:* Number of under-five deaths per 1000 population
Polio*:* Polio (Pol3) immunization coverage among 1-year-olds (%)
Total expenditure*:* General government expenditure on health as a percentage of total government expenditure (%)
Diphtheria: Diphtheria tetanus toxoid and pertussis (DTP3) immunization coverage among 1-year-olds (%)
HIV/AIDS: Deaths per 1 000 live births HIV/AIDS (0-4 years)
GDP: Gross Domestic Product per capita (in USD)
Population*:* Population of the country
thinness 1-19 years*:* Prevalence of thinness among children and adolescents for Age 10 to 19(%)

thinness 5-9 years: Prevalence of thinness among children for Age 5 to 9(%)
Income composition of resources: Human Development Index in terms of income composition of resources (index ranging from 0 to 1)
Schooling: Number of years of Schooling(years)

Project Requirements

Functional Requirements

- Working internet connection with access to IBM cloud

Technical Requirements

- A working knowledge of the entire ML pipeline from data cleaning to deploying the model on IBM cloud

Software Requirements

- Jupyter Notebook (with Python 3.6+)
- IBM Cloud services
- Slack (for communication)

Project Deliverables

1. A working model which predics the life expectancy of population across the world
2. A comparison between perfomances of different algorithms
3. Some basic visualisations suiting the data

Project Team

Irish Mehta

Project Schedule

Week1: Discussing project scope and getting access to software requirements
Week2: Cleaning data & training the model using IBM Watson and creating a NODE RED integration
Week3: Completing Report as per the submission norms