PROJECT REPORT

PREDICTING LIFE EXPECTANCY USING MACHINE LEARNING

Steve Vinod Mathew

INDEX

1.INTRODUCTION

1.1 Overview

Life expectancy is the key metric for assessing population health. Life expectancy from birth is a frequently utilized and analysed component of demographic data for the countries of the world. It represents the average life span of a new born and is an indicator of the overall health of a country. Broader than the narrow metric of the infant and child mortality, which focus solely at mortality at a young age, life expectancy captures the mortality along the entire life course. Life expectancy can fall due to problems like famine, war, disease and poor health. It tells us the average age of death in a population. Improvements in health and welfare increase life expectancy. The higher the life expectancy, the better shape a country is in.

1.2 Purpose

The aim of this project is to create machine learning model which will be able to predict the life expectancy given various factors such as GDP, year, population, adult mortality etc. The dataset from The World Health Organization gives the values of the various factors including the life expectancy for different countries from 2000 to 2015, this data can be used to train and test the model.

2.LITERATURE SURVEY

2.1 Existing Problem

The World Health Organization (WHO) began producing annual life tables for all Member States in 1999. These life tables are a basic input to all WHO estimates of global, regional and country-level patterns and trends in all-cause and cause-specific mortality. After the publication of life tables for years to 2009 in the 2011 edition of World Health Statistics, WHO has shifted to a two year cycle for the updating of life tables for all Member States. Even still the model is not really updated in every fields. WHO applies standard methods to the analysis of Member State data to ensure comparability of estimates across countries. This will inevitably result in differences for some Member States with official estimates for quantities such as life expectancy, where a variety of different projection methods are used.

2.2 Proposed Solution

The proposed solution is to create a machine learning model that is able to predict the value of life expectancy by using supervised learning. This will make the task more simple and easy than using complex mathematics to find the value. This method will be more effective and time efficient.

3.THEORITICAL ANALYSIS

Machine Learning

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.

The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.
Machine learning algorithms are classified into supervised, unsupervised and reinforcement learning. In this project we will be using supervised learning.

Consider yourself as a student sitting in a math class wherein your teacher is supervising you on how you're solving a problem or whether you're doing it correctly or not. This situation is similar to what a supervised learning algorithm follows, i.e., with input provided as a labelled dataset, a model can learn from it. Labelled dataset means, for each dataset given, an answer or solution to it is given as well. This would help the model in learning and hence providing the result of the problem easily. A labelled dataset of animal images would tell the model whether an image is of a dog, a cat, etc.. Using which, a model gets trained, and so, whenever a new image comes up to the model, it can compare that image with the labelled dataset for predicting the correct label.

There are two types of problems supervised learning deals with classification and regression.
Classification problems ask the algorithm to predict a discrete value that can identify the input data as a member of a particular class or group. Taking up the animal photos dataset, each photo has been labelled as a dog, a cat, etc., and then the algorithm has to classify the new images into any of these labelled categories.
Regression problems are responsible for continuous data, e.g., for predicting the price of a piece of land in a city, given the area, location, etc.. Here, the input is sent to the machine for predicting the price according to previous instances. And the machine determines a function that would map the pairs. If it is unable to provide accurate results, backward propagation is used to repeat the whole function until it receives satisfactory results. Our current project is an example of a regression problem.

4. EXPERIMENTAL INVESTIGATIONS

Various regression algorithms such as decision tree regressor, extra trees regressor, gradient boosting regressor, LGBM regressor, random forest regressor, linear regression, ridge and XGB were tested along with various versions of them being optimized by hyperparameter optimization and feature engineering. Considering overfitting, underfitting and the least error rate the most suitable algorithm was found to be the extra trees regressor.

5. RESULT

5.1 Project Model

The dataset was formatted to have 21 features and was split into two , one for training the model and the other to test the model. The model was created and deployed using the AUTO AI service provided in the IBM cloud platform. Various pipelines for the various algorithms and their optimizations was used to make the model and the most suitable was selected from among them.

5.2 User Interface

The User interface was created using the node red service in the IBM cloud platform.

Home Page

Machine Learning Model

Prediction **63.87999992370605**

Year *
2021

Status *
1

Adult Mortality *
900

infant deaths *
6

Alcohol *
9.86

percentage expenditure *
29.86

Hepatitis B *
67

Measles *
6590

BMI *
13

under-five deaths *
122

Polio *
52

Total expenditure *
8.9

Diphtheria *
5

HIV/AIDS *
0.1

GDP *
875

Population *
5000

thinness 1-19 years *
17.2

thinness 5-9 years *
20

Income composition of resources *
0.725

Schooling *
15.2

SUBMIT CANCEL

6. ADVANTAGES & DISADVANTAGES

Advantages

- Simple to use
- Time efficient
- Resource efficient

Disadvantages

- Data set used was not up to date
- Not completely error free

7. CONCLUSION

Life expectancy, estimate of the average number of additional years that a person of a given age can expect to live. The most common measure of life expectancy is life expectancy at birth. Life expectancy is a hypothetical measure. It assumes that the age-specific death rates for the year in question will apply throughout the lifetime of individuals born in that year. The estimate, in effect, projects the age-specific mortality (death) rates for a given period over the entire lifetime of the population born (or alive) during that time. The measure differs considerably by pandemics, geographic location and much more. This project can helps to predict life expectancy in a simple, easy and efficient way.

8. BIBLIOGRAPHY

- Chen, Jonathan H., and Steven M. Asch. "Machine learning and prediction in medicine—
  beyond the peak of inflated expectations." The New England journal of medicine 376.26
  (2017): 2507.
- https://www.britannica.com/science/life-expectancy
- https://developer.ibm.com/tutorials/how-to-create-a-node-red-starter-application/
- https://nodered.org/
- https://developer.ibm.com/technologies/machine-learning/series/learning-path-machine-learning-for-developers/

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

APPENDIX

A. Source Code


```
[
  {
    "id": "2f44c401.a1d40c",
    "type": "tab",
    "label": "Flow 1",
    "disabled": true,
    "info": ""
  },
  {
    "id": "b1b11140.4e4ef",
    "type": "inject",
    "z": "2f44c401.a1d40c",
    "name": "",
    "topic": "",
    "payload": "Hello Node-RED!",
    "payloadType": "string",
    "repeat": "",
    "crontab": "",
    "once": false,
    "x": 141,
    "y": 61,
    "wires": [
      [
        "f2f2649a.0d0d98"
      ]
    ]
  },
  {
    "id": "f2f2649a.0d0d98",
    "type": "debug",
    "z": "2f44c401.a1d40c",
    "name": "",
    "active": true,
    "console": "false",
    "complete": "false",
    "x": 401,
    "y": 87,
```

```
        "wires": []
    }
]
```