Predicting Life Expectancy Using Machine Learning

FINAL PROJECT REPORT

COMPANY NAME: SMARTINTERNZ

VERSION 1.0.0

PROJECT MANAGER: SMIT S. PATEL

DURATION: 21/5/2020 TO 15/6/2020

UNDER THE GUIDANCE OF: PRASHANTH &
CHARAN REDDY

1. ACKNOWLEDGEMENT

First, I would like to thank Mr. Prashanth and Mr. Charan for giving me an opportunity to do an internship within a 'ML Engineer' discipline in smartinternz. Although quite short, for me this was a great experience I can learn from. It helped me to learn & explore my skills and increased my interest in Data science.

I would like to thank the team of smartinternz for mentoring me throughout the internship & I appreciate the patience and help which I received from them.

2. ABSTRACT

For the successful and accurate prediction, one should be aware of all the latest algorithm in field of machine learning so he can provide best for the stakeholder. While there are always some new algorithms are going to discover, one should also keep historic conventional algorithm in back of his mind. When there is a need of ease in working on project, cloud services are coming handy.

During the internship program I get my hands on some of the latest technology available for machine learning on cloud-based platform (IBM cloud) and apply them on my project. Working on back-end always as a ML engineer, in this internship I also learnt front-end using one of IBM service. So different technology and its future scope explained in this report.

TABLE OF CONTENTS

1. INTRODUCTION

1.1 OVERVIEW

It is a standard problem that answers the life expectancy of a human by their various habits, health related data, demographic data and other pertinent factors that affect avg. life. The goal of this project is to deliver a model that can predict life expectancy accurately. In the process of developing the model, there's huge role of IBM services that ease the process of exploring, creating and deploying the model.

1.2 PURPOSE

Life expectancy is a measure that is often used to gauge the overall health of a community. Life expectancy at birth measures health status across all age groups. Shifts in life expectancy are often used to describe trends in mortality. As this project gives the insights of future, it can be used by person itself or some company based on their needs. Like an example, Government of certain region can obviate some factor that cause regional life expectancy to be low. Insurance company can used to pass/reject its customer's request or they can increase the premium. It can be used in various way depending on profession.

2. LITERATURE SURVEY

2.1 EXISTING PROBLEM

In the world where everywhere, problem related to food and providing basic needs is still going on, amid of that crisis the government of all country needs to know their country's life expectancy to calculate per capita expenses on its people. Increase or decrease in some region's life expectancy can affect its economy. The factor affecting the life expectancy can be foreseen and prevent.

2.2 PROPOSED SOLUTION

To solve this problem, we are going to use machine learning to predict the life expectancy. We are going to use regression model that can provide highly accurate and reliant result to end customers.

The model's inputs are going to entered by UI form-based page in web browser, so non-technical person can also access it easily.

3. THEORITICAL ANALYSIS

3.1 SOFTWARE/MODEL DESIGNING

Software/Model comprises of various tools and techniques that helps throughout the project.

- IBM Cloud

  IBM Cloud offers the most open and secure public cloud for business, a next-generation hybrid multicloud platform, advanced data and AI capabilities, and deep enterprise expertise across 20 industries. IBM Cloud is a set of cloud computing services for business offered by the information technology company IBM. It combines platform as a service (PaaS) with infrastructure as a service (IaaS). The platform scales and supports both small development teams and organizations, and large enterprise businesses. IBM cloud has 170 service that cover almost all major fields in today's world and we're going to use some of them in our project.

  - IBM Watson Studio: IBM Watson Studio helps data scientists and analysts prepare data and build models at scale across any cloud. With its open, flexible multicloud architecture, Watson Studio provides capabilities that empower businesses to simplify enterprise data science and AI. Predict and optimize your business outcomes.
  - Node-Red: Node-RED 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. With over 225,000 modules in Node's package repository, it is easy to extend the range of palette nodes to add new capabilities.

- MODEL DESIGNING
    - Data Collection and Pre-processing:

      In this project data for the training is taken from Kaggle. If the data contains too many out-of-range values or impossible combination values or Null values, it became arduous task for model to get accurate result. The phrase "garbage in, garbage out" is particularly applicable machine learning projects. So, in preprocessing with different methods we can remove or fill the missing data to get prediction with less deviation.

    - Feature Extraction:

      Feature extraction is a process of dimensionality reduction by which an initial set of raw data is reduced to more manageable groups for processing. The factor which is having the most importance in predicting life expectancy is going to be extracted in model as input. Feature extraction reduce the load on process without compromising the accuracy.

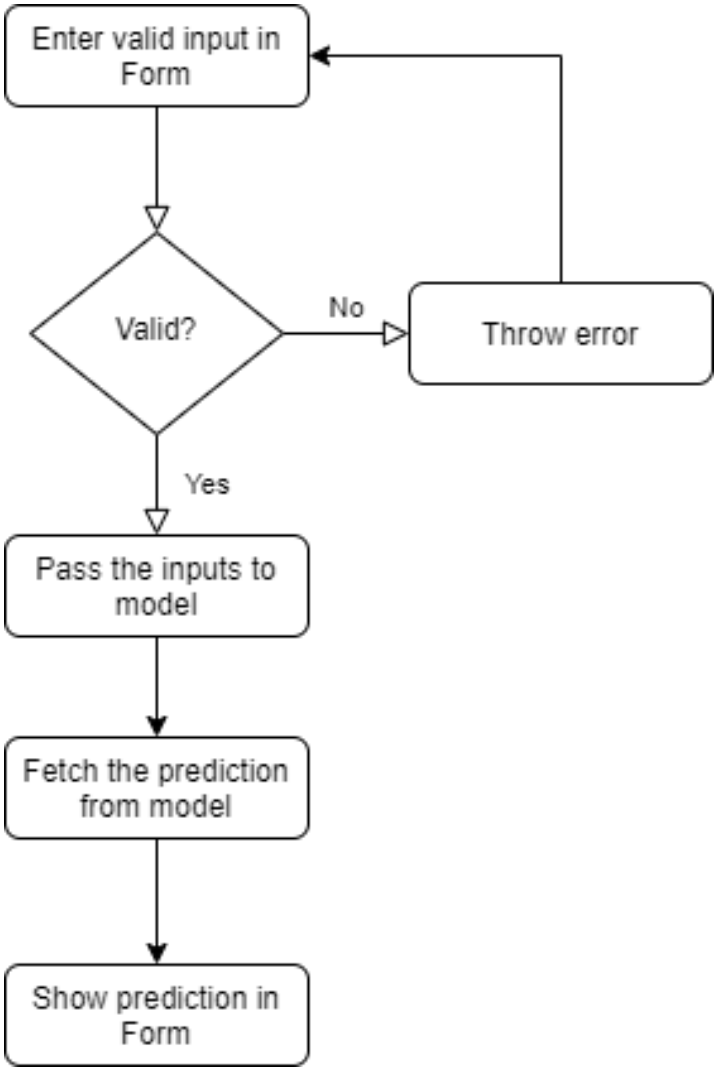
    - Model building and Testing:

      Model building, algorithm selection is one of the skills that a machine learning engineer must have it. Here we have to predict a number so we are going to use a regression model. There are also various types of regression models are available, because we have multiple inputs, we're going to use multivariable regression model to predict. the data we use is usually split into training data and test data. The training set contains a known output and the model learns on this data in order to be generalized to other data later on. We have the test dataset (or subset) in order to test our model's prediction on this subset.

    - Front-End (Node-Red):

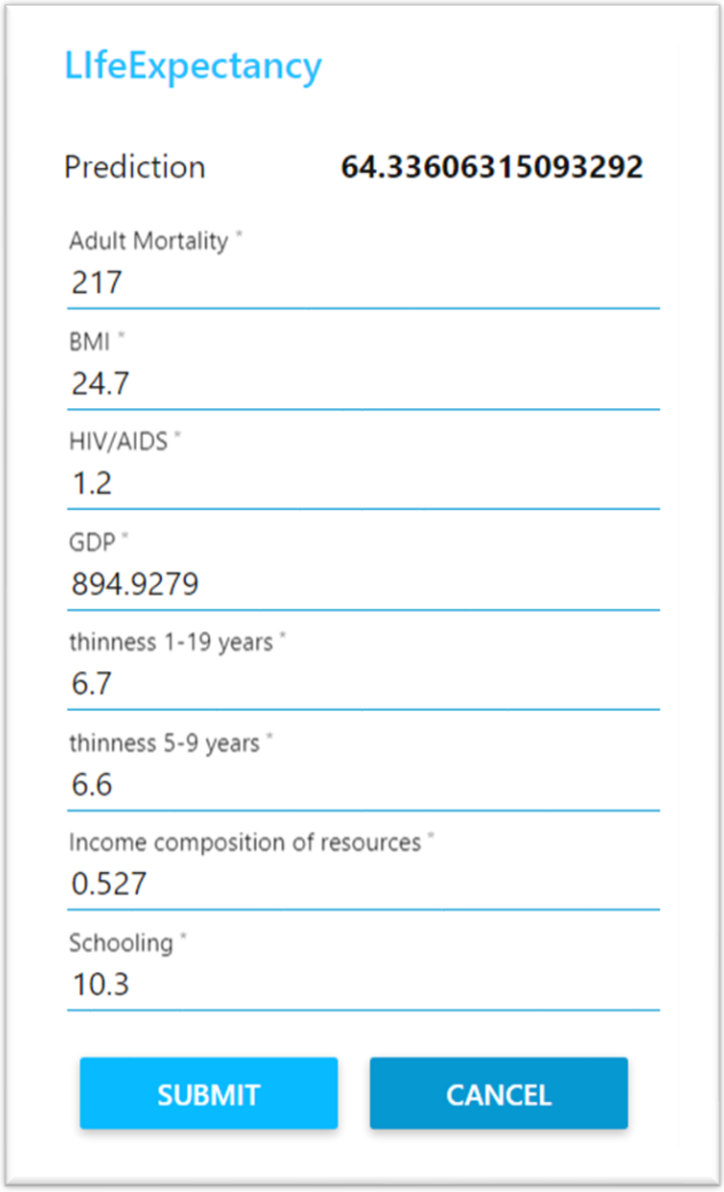
      After successfully created model that can predict life expectancy, we have to provide our model in user friendly way so, everyone can access it. We will use Node -Red to link our model to front-end. In this project, I created form to interface with user that send data back to model and prediction fetched back to UI page.

4. FLOWCHART


Enter valid input in Form
Valid?
No
Throw error
Yes
Pass the inputs to model
Fetch the prediction from model
Show prediction in Form

5. RESULT

As at the end of the project we are able to deliver the product as accurate as required by end customer. Result of this project is shown in picture below and during the finding of this project we're able acquire the importance of different factor that affect our prediction and accordingly we trained our model and created front-end that make process easy.



6. APPLICATION

- Known life expectancy helps government to derive the problem that are going to cause by exploit in population if expectancy is high.
- Life expectancy helps health domains and other insurance company to set their policy according to customer's demographical and medical data.
- Life expectancy measure of mortality providing an overall picture of mortality, allowing countries and regions to be compared.

7. CONCLUSION

Hence conclude that machine learning is emerging field in today's world and it will be in future too. With new emerging requirement of solution that can be solved with Computer, machine learning is great way to put end to that problems. As tech enthusiastic person got their mind in basic and root about machine learning by learning all conventional algorithm, this internship helps to acknowledge us to newer technology like Cloud service and front-end that is also an important factor for machine learning engineer that make him/her to stand out in crowd.

REFERENCE

1. Node-Red Tutorial
   https://developer.ibm.com/tutorials/how-to-create-a-node-red-starter-application/

2. IBM Cloud Tutorial
   https://www.ibm.com/cloud/get-started

3. Machine learning Tutorial
   https://developer.ibm.com/technologies/machine-learning/series/learning-path-machine-learning-for-developers/

4. Auto AI reference
   https://developer.ibm.com/tutorials/watson-studio-auto-ai/