PREDICTING LIFE EXPECTANCY USING MACHINE LEARNING

1.Project Summary

This project aims at creating a new model based on the data provided is to evaluate the life expectancy of the country.
The data set from WHO offers a time frame from 2000 to 2015. The algorithm that produces the best result is ELASTIC NET regression algorithm

2. Project Requirement

Life expectancy is one of the most important factors in end-of-life decision making. Good prognostication for example helps to determine the course of treatment and helps to anticipate the procurement of health care services and facilities, or more. By using supervised machine learning techniques. we can extract a model that will be able to predict the life expectancy of future years. One method of approach is to use LSTM models to achieve this task.

3. Functional Requirement

1. Create a data model present on the database.
2. The data set are made available to the public to the purpose of health data analysis.
3. It is related to the different countries depending on the different countries while finding the data set in different countries might be difficult and hence some countries are excluded from the final data set.

4. Technical Requirements

1. The merged data set by using the databases in the .csv formats from Kaggle
2. Datasets need to be integrated into the Python IDE.

5. Project Deliverables Expected

The project creates a model which gives life expectancy of a country depending on various factors like schooling, GDP, BMI etc.

6. Team

The project is done individually with the help of IBM Cloud and the project can be written in Watson Studio and deployed using Node-Red Apps.

7. Software Requirements

- Python IDE
- Excel
- IBM Cloud
- IBM Watson
- IBM Node-Red Service

7.1 Algorithm for Project

Machine Learning:
Machine learning is an application of artificial intelligence (AI) that provides systems the ability to learn automatically 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.

The machine learning algorithm is classified into Supervised, unsupervised and reinforced learning . For this project a type of supervised model is used.

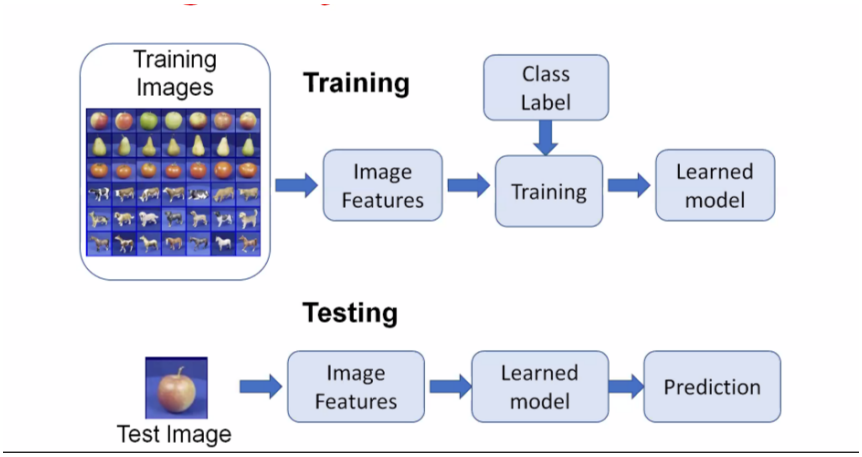
7.1.1 Supervised Model

Supervised machine learning algorithms can apply what has been learned in the past to new data using labeled examples to predict future events. Starting from the analysis of a known training data set, the learning algorithm produces an inferred function to make predictions about the output values. The system can provide targets for any new input

after sufficient training. Supervised learning problems can be further grouped into **Regression** and **Classification** problems

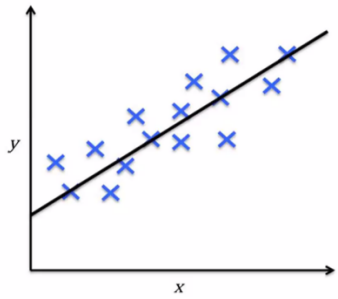
7.1.1.1 Classification

A classification problem is when the output variable is a category, such as “red” or “blue” or “disease” and “no disease”. A classification model attempts to draw some conclusion from observed values. Given one or more inputs a classification model will try to predict the value of one or more outcomes.



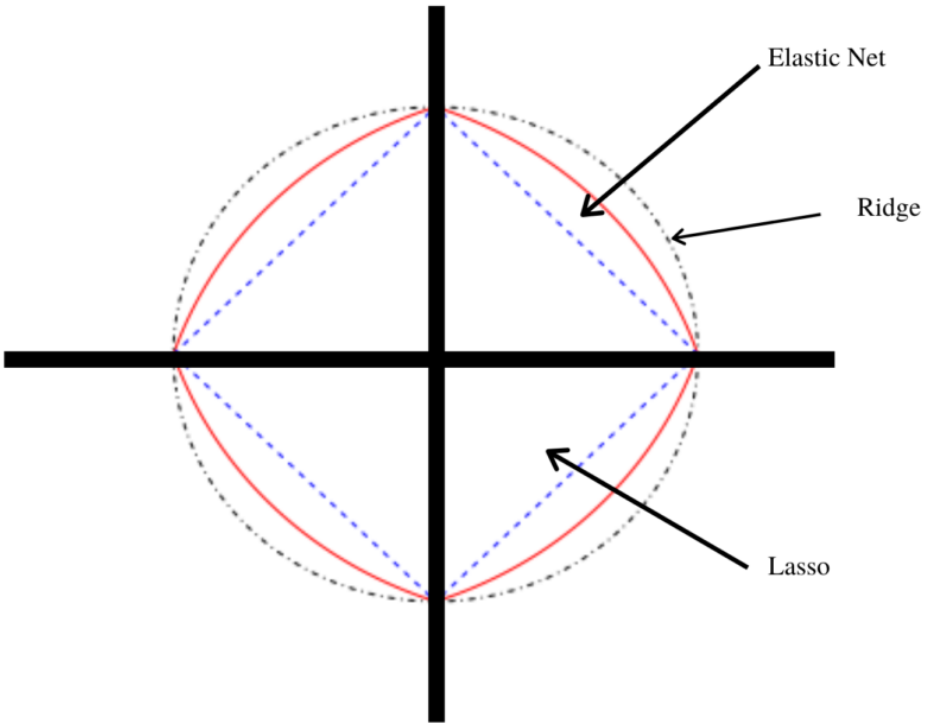
7.1.1.2 Regression

A regression problem is when the output variable is a real or continuous value. Many different models can be used, the simplest is the linear regression. It tries to fit data with the best hyper-plane which goes through the points.



In this project we are using Elastic Nest Regression as it produces the least error rate among other algorithms available.

Elastic net linear regression uses the penalties from both the lasso and ridge techniques to regularize regression models. The technique combines both the lasso and ridge regression methods by learning from their shortcomings to improve on the regularization of statistical models.



The elastic net method improves on lasso's limitations, i.e., where lasso takes a few samples for high dimensional data, the elastic net procedure provides the inclusion of "n" number of variables until saturation. In a case where the variables are correlated groups, lasso tends to choose one variable from such groups and ignore the rest entirely.

To eliminate the limitations found in lasso, the elastic net includes a quadratic expression in the penalty, which, when used in isolation, becomes ridge regression. The

quadratic expression in the penalty elevates the loss function toward being convex. The elastic net draws on the best of both worlds – i.e., lasso and ridge regression.

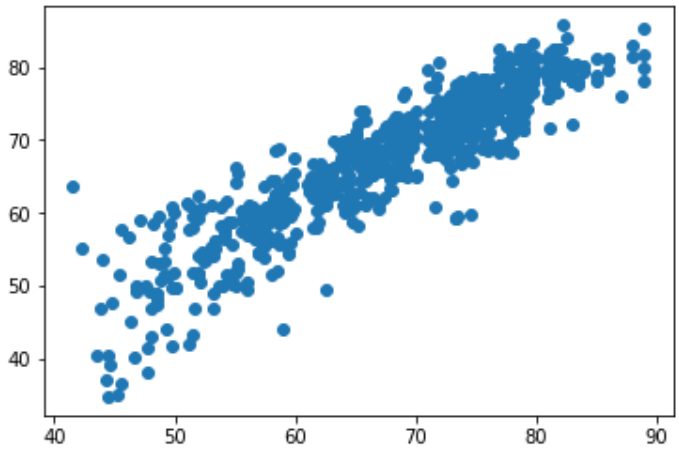
8. Project Model

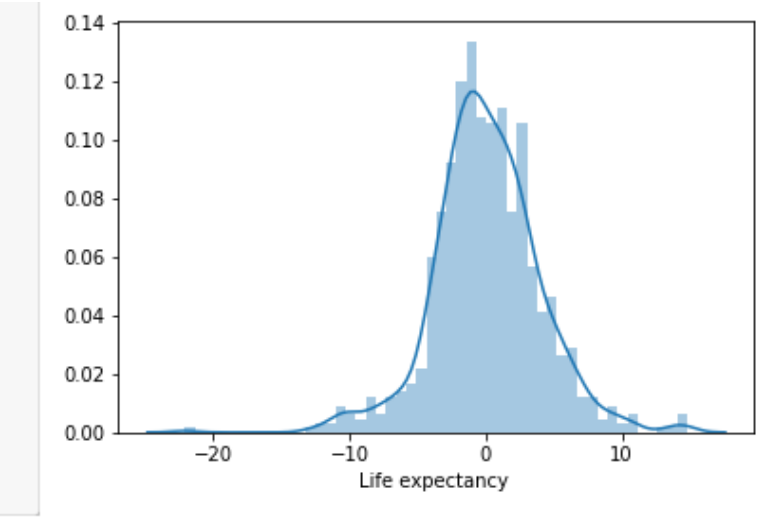
The model was created in Watson studio and Jupyter notebook.
The data set contains 22 features including countries,adult mortality etc.
Since regression can only use number values the data set is grouped using countries, to avoid conflict.

The null values were filled with their mean and the data was then split into training and testing data. The following heat map was generated from the data

8.1 End Result- Project output

After training the data using Elastic net regression the prediction models were obtained



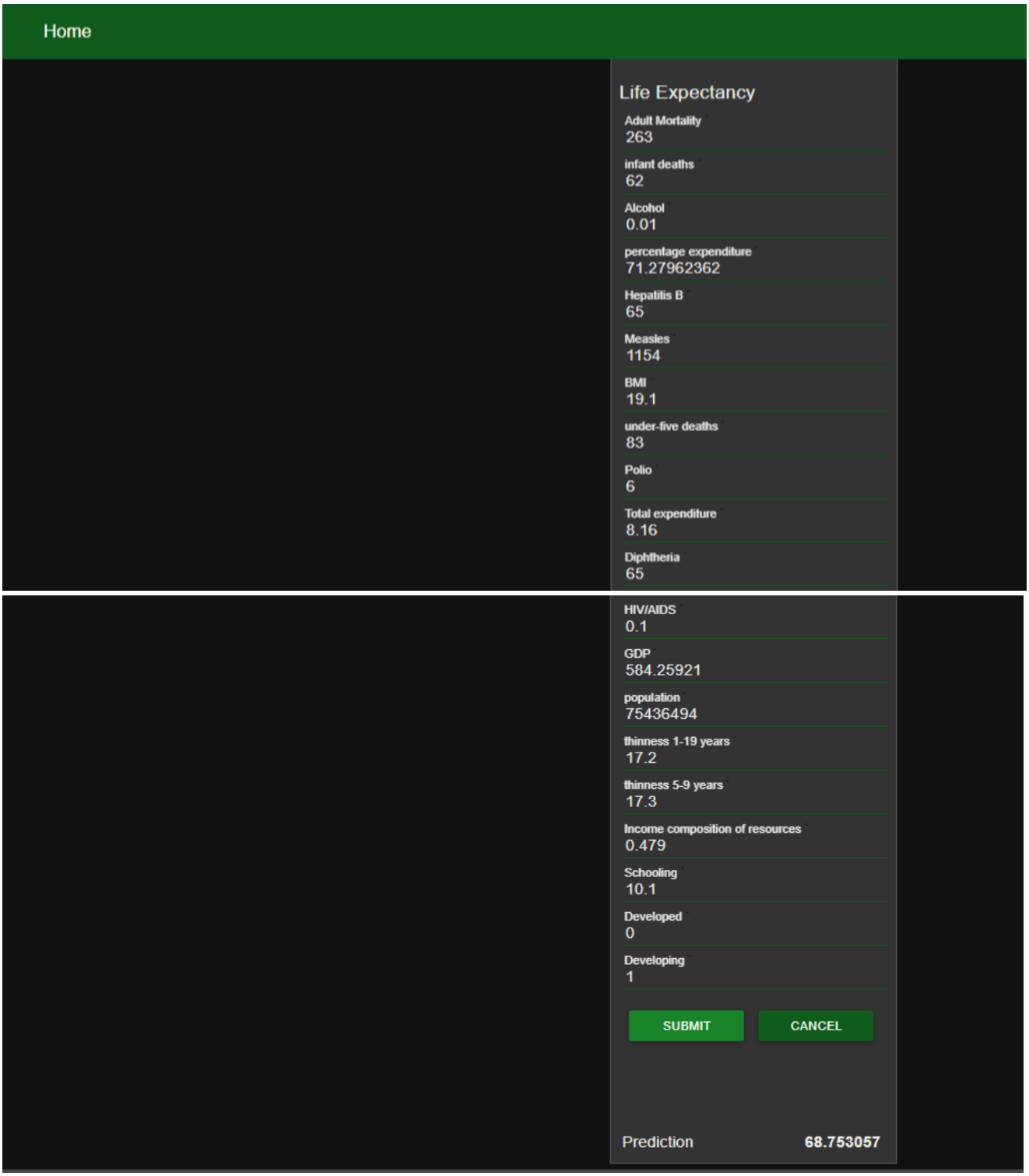


The most favorable parameters and least error rates obtained are

Best Parameters: {'alpha': 0, 'l1_ratio': 0.1, 'max_iter': 1000}
R^2 score on training data: 0.81
R^2 score: 0.83
Mean squared error: 15.57
Mean absolute error: 2.94

9. Project Delivarables

The UI output of the project is shown below with sample values



10. Project Schedule

Project Start : May 16,2020
Project End : June 14,2020

11. References

- Sebastian Raschka ,Python Machine Learning, PACKT Publishing, UK, 2015
- https://www.datasciencesociety.net
- https://developer.ibm.com/tutorials/how-to-create-a-node-red-starter-application/
- https://in.mathworks.com/