

Project Report

**PREDICTING LIFE EXPECTANCY USING MACHINE LEARNING**

**INTRODUCTION**

Overview

With the advancement in technology, the field of health is progressing. A lot of devices like wearable bands and smartphone apps are available which help to keep a daily track of our fitness. By using machine learning these technologies can be extended to predict the life expectancy. Life expectancy is a statistical measure of the average time an organism is expected to live, based on the year of its birth, its current age and other including gender. The WHO consists of a repository of data from many countries for the purpose of health status tracking. It consists of various related factors. This data set can be processed and then used to predict life expectancy.

Purpose

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 broadly: facilitates Advance Care Planning. Advance Care Planning improves the quality of the final phase of life by stimulating doctors to explore the preferences for end-of-life care with their patients, and people close to the patients.

**LITERATURE SURVEY**

Existing problem

Although medical records are increasingly available in the form of electronic medical records (EMRs), they remain underutilized for developing clinical decision support systems, and improving health care in general. EMRs are characterized by irregularly-sampled time-series data, missing values, long-term dependencies involving symptoms, diagnoses and interventions, and are prone to documentation errors. Moreover, they contain important information in the form of unstructured, textual data, from which information cannot be extracted straightforwardly. These challenges lead to suboptimal use and even waste of large portions of data, especially when the data is unstructured and noisy.

### Proposed Solution

The task of predicting life expectancy is done using machine learning approach. Machine learning algorithms can be trained to learn which pieces of information are important to execute a task, and which patterns are indicative for producing correct output. A dataset was imported and by performing data processing and analytics, a machine learning model was built and trained to predict life expectancy.

## **THEORETICAL ANALYSIS**

### Block Diagram



### Hardware/Software Designing

#### Backend:

The backend consists of a python code which was written in Jupyter notebook. For designing backend without python, AutoAI tool can also be used. Both were provided by IBM Cloud(Watson Studio).

#### Frontend:

Application was designed using Node-Red provided by IBM Cloud.

#### Connection:

The frontend and backend were connected using Watson Machine Learning Service.

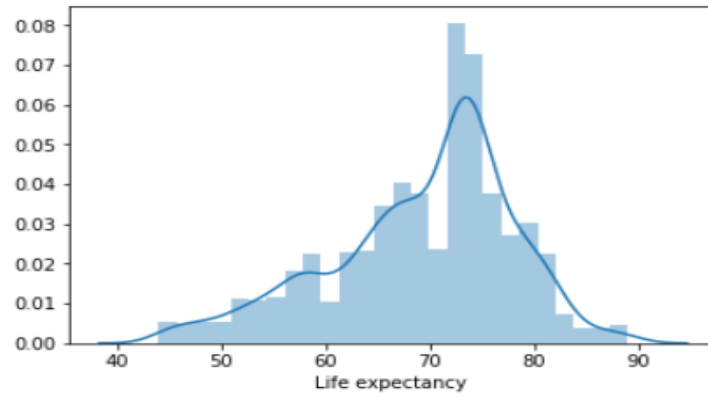
## **EXPERIMENTAL INVESTIGATIONS**

1. Compared accuracy of different machine learning models.
2. Performed descriptive analysis of data.
3. Used pairplot for visualising data.
4. Plotted Heatmap to find the effects of different factors on life expectancy.

## Distplot

```
In [10]: sns.distplot(data['Life expectancy '])
```

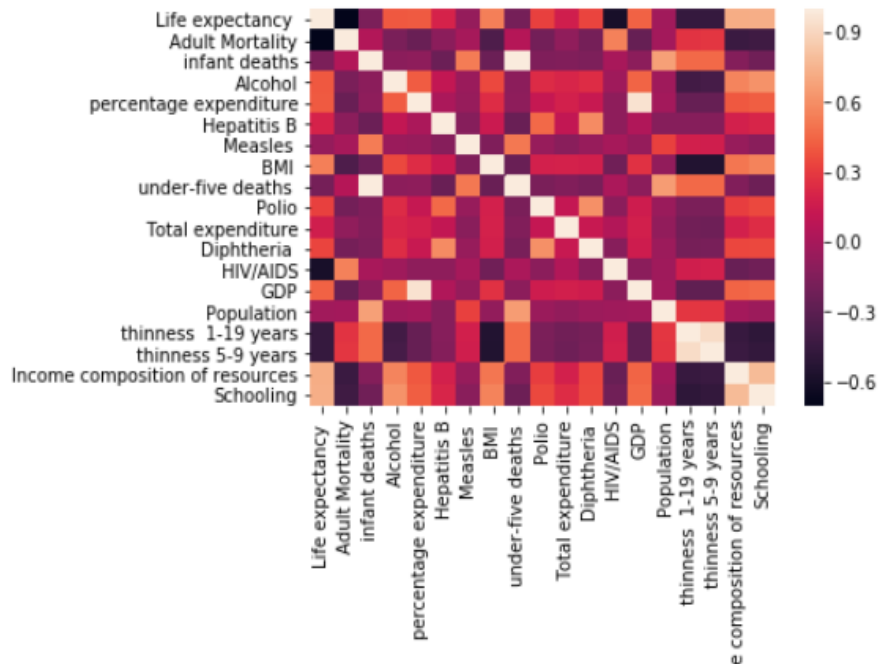
```
Out[10]: <matplotlib.axes._subplots.AxesSubplot at 0x7fee8defc2e8>
```



## Heatmap

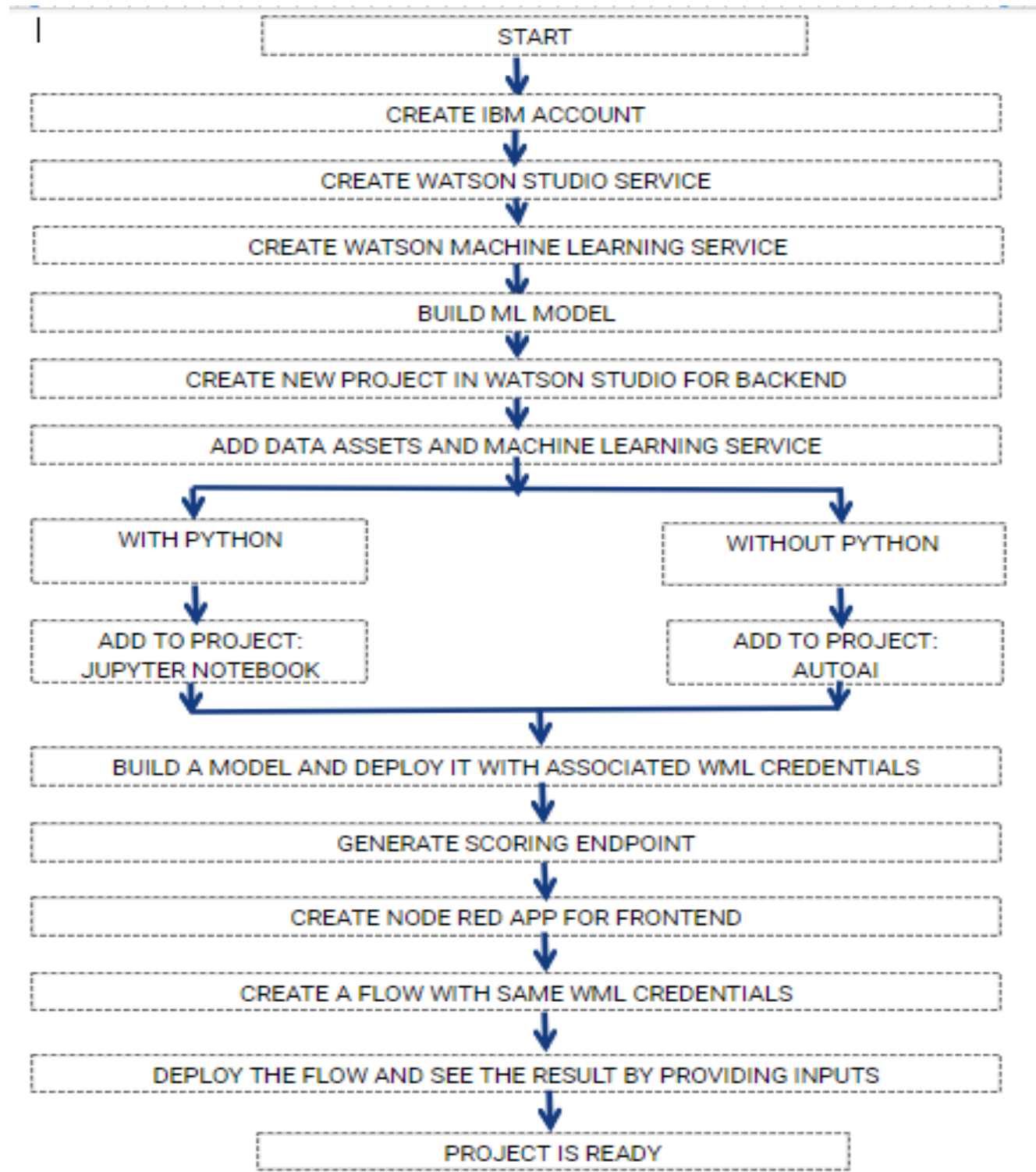
```
In [11]: sns.heatmap(data.corr())
```

```
Out[11]: <matplotlib.axes._subplots.AxesSubplot at 0x7fee8d701f60>
```

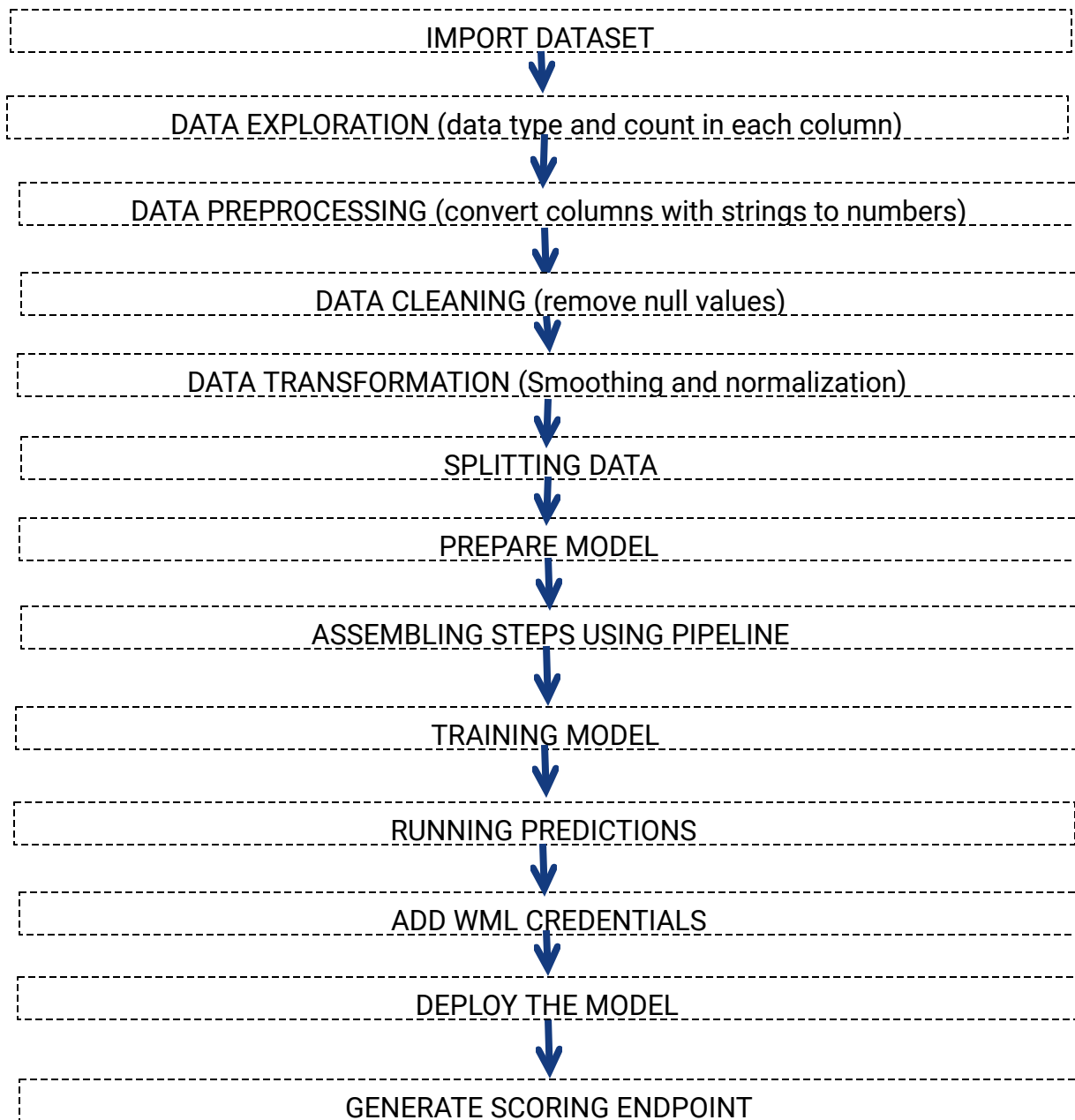


## FLOWCHART

### Project Flow



## Machine Learning Model Flow



## **RESULT**

1. Model used is RandomForestRegressor.
2. It has highest accuracy of 98.25%.
3. The Status column doesnot have any effect on life expectancy.
4. The top two factors which affect life expectancy are schooling and income composition of resources.
5. Adult Mortality has the least effect on life expectancy.

## **ADVANTAGES**

1. Machine learning helps us to have a lot of models with different degrees and choices. We can choose the model with highest accuracy.
2. It helps to know about which fators affect upto which extent on life expectancy.
3. With more experience, accuracy increases

## **DISADVANTAGES**

1. ML requires large amounts of data.
2. More time required for improved efficiency.
3. Difficult to choose correct algorithm.
4. High risk of error if data set is biased.

## **APPLICATIONS**

1. Medical field
2. Research field

## **CONCLUSION**

1. Machine learning is useful to determine which indicators are statistically significant, and to predict life expectancy.
2. Classifying countries based on income levels, we can see how economics impact health.

## **FUTURE SCOPE**

1. The model can be integrated with an application or wearable devices
2. It can be made open to health centres from which accurate data can be obtained and hence more accurate results can be obtained.

## **BIBILOGRAPHY**

1. <https://cloud.ibm.com/catalog>
2. <https://developer.ibm.com/technologies/machine-learning/series/learning-path-machine-learning-for-developers/>
3. [https://developer.ibm.com/tutorials/how-to-create-a-node-red-starter-application /](https://developer.ibm.com/tutorials/how-to-create-a-node-red-starter-application/)
4. <https://developer.ibm.com/tutorials/watson-studio-auto-ai/>
5. <https://medium.com/swlh/predicting-life-expectancy-w-regression-b794ca457cd4>
6. <https://medium.com/vickdata/a-simple-guide-to-scikit-learn-pipelines-4ac0d974bdcf>

## **APPENDIX**

Link of Python code

<https://drive.google.com/file/d/1Ek233OjfGf0afBAUSjkrC9mrq8f9vn6q/view?usp=sharing>