PROJECT REPORT

on

### Life Expectancy

submitted by

### SHUBHAM KUMAR(T.L) PRINCE KUMAR ASHUTOSH KUMAR ABHISHEK KUMAR

*SmartBridge in collaboration with IBM offers Internship ,Machine Learning with Python IBM Watson Studio, Internship Certificate.*



DIVISION OF COMPUTER SCIENCE AND ENGINEERING SMARTBRIDGE IN COLLABORATION WITH IBM

JUNE 2019

## Declaration

*W* e, SHUBHAM KUMAR, PRINCE KUMAR, ASHUTOSH KUMAR,ABHISHEK KUMAR hereby

declare that this project is the record of authentic work carried out by us using the year 2018-2019 and has not been submitted to any other Institute towards the award of any Summer certificate.

‘

# Contents

1. Introduction 1
   1. [Objectives of Research 1](#_TOC_250011)
   2. [problem statement 2](#_TOC_250010)
2. Abstract 3
   1. [Life Expectancy 3](#_TOC_250009)
3. Data collection 4
   1. [data collection source 4](#_TOC_250008)
   2. [data subset df 4](#_TOC_250007)
   3. [Libraries Used 4](#_TOC_250006)
   4. [Data Set 4](#_TOC_250005)
4. Methodology 5
   1. [Exploratory Data Analysis 5](#_TOC_250004)
5. System Design 8
   1. [Introduction 8](#_TOC_250003)
   2. [Data Flow Diagram 8](#_TOC_250002)
   3. [Modular Design 8](#_TOC_250001)
6. System Implementation 12
   1. [Platform and Tools 12](#_TOC_250000)
   2. Sample Code 14
7. Testing 21
8. Future Scope 25
9. Conclusion 26
10. References 27

**List of Figures**

4.1 A typical sentiment analysis model 6

* 1. Level 0 Data Flow Diagram 9
  2. Level 1 Data Flow Diagram 10
  3. Use Case Diagram 11
  4. Sample code for python 14
  5. Sample code for python 15
  6. Sample code for python 16
  7. Sample code for python 17
  8. Sample code for python 18
  9. Sample code for python 19
  10. Sample code for python 20
  11. Window to enter the ASIN code of product 22
  12. Report for the analysis of required product 23
  13. PDF report for the same product 24

## Chapter 1

**Introduction**

*M* achine learning is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of Computer Programs that can change when exposed to new data.

Machine learning involves computer to get trained using a given data set, and use this training to predict the properties of a given new data. For example, we can train computer by feeding it 1000 images of cats and 1000 more images which are not of a cat, and tell each time to computer whether a picture is cat or not. Then if we show the computer a new image, then from the above training, computer should be able to tell whether this new image is cat or not.

Python community has developed many modules to help programmers implement machine learn- ing.we will be using numpy, scipy and scikit-learn modules. We can install them using cmd command:

pip install numpy scipy scikit-learn

A better option would be downloading miniconda or anaconda packages for python, which come prebundled with these packages. Follow the instructions given here to use.

### Objectives of Research

The project tries to create a model based on data provided by the World Health Organization (WHO) to evaluate the life expectancy for different countries in years. The data originates from here: h[ttps://www.kaggle.com/kumarajarshi/lif](http://www.kaggle.com/kumarajarshi/life-expectancy-who/data)e-exp[ectancy-who/data](http://www.kaggle.com/kumarajarshi/life-expectancy-who/data) The output algorithms have been used to test if they can maintain their accuracy in predicting the life expectancy for data they havent been trained.Based on the data we can predict the average mortility of human life.

Life expectancy Review

### problem statement

the nature of age-related health problems, the current trend of population ageing, the social impact of HIV and AIDS, These factors might have negative or positive impact on the quality of life of the older people and influence their perceptions about the ageing phenomenon. Governments and health professionals are challenged to develop programs and interventions that would enable the older persons to enhance or maintain their quality of life. The World Health Organisation (WHO) states that Investing in health and promoting it throughout the life span is the only way to ensure that more people will reach old age in good health and capable of contributing to societ intellectually,spiritually and physically

Division of Computer Science Page 2

## Chapter 2

**Abstract**

### 2.1 Life Expectancy

This study was designed to demonstrate the feasibility of forecasting functional health for the elderly. Using life-table techniques, we analyzed the expected remaining years of functional well-being, in terms of the activities of daily living, for noninstitutionalized elderly people living in world. The expected years, or active life expectancy, showed a decrease, from 10 years for those aged 65 to 70 years to 2.9 for those 85 or older. Active life expectancy was shorter for the poor than for others, and women had a longer average duration of expected dependence than men. The measure of active life expectancy provides important information about health at a given population level, in terms other than death. This information can be used for actuarial purposes in planning and policy making. It is also useful in identifying high-risk populations for which preventive health care and medical care can compress morbidity during the last years of life.

## Chapter 3

**Data collection**

### data collection source

*h*[ttps://www.k](http://www.kaggle.com/tags/web-sites)aggle.c[om/tags/web-sites](http://www.kaggle.com/tags/web-sites)

### data subset df

***I***nfant deaths Alcohol,Hepatitis B,Measles ,BMI , Polio, Diphtheria ,HIV/AIDS, GDP,Life ex- pectancy

### Libraries Used

Matplotlib,seaborn,pandas,Basemap:for visualization -pandas,Numpy:for manipulation Sklearn,scipy:for modeling

### Data Set

Life Expectancy Data.csv

## Chapter 4

**Methodology**

*M* ethodology is the systematic, theoretical analysis of the methods applied to a field of study. It comprises the theoretical analysis of the body of methods and principles associated with a branch of knowledge. Typically, it encompasses concepts such as paradigm, theoretical model, phases and quantitative or qualitative techniques.

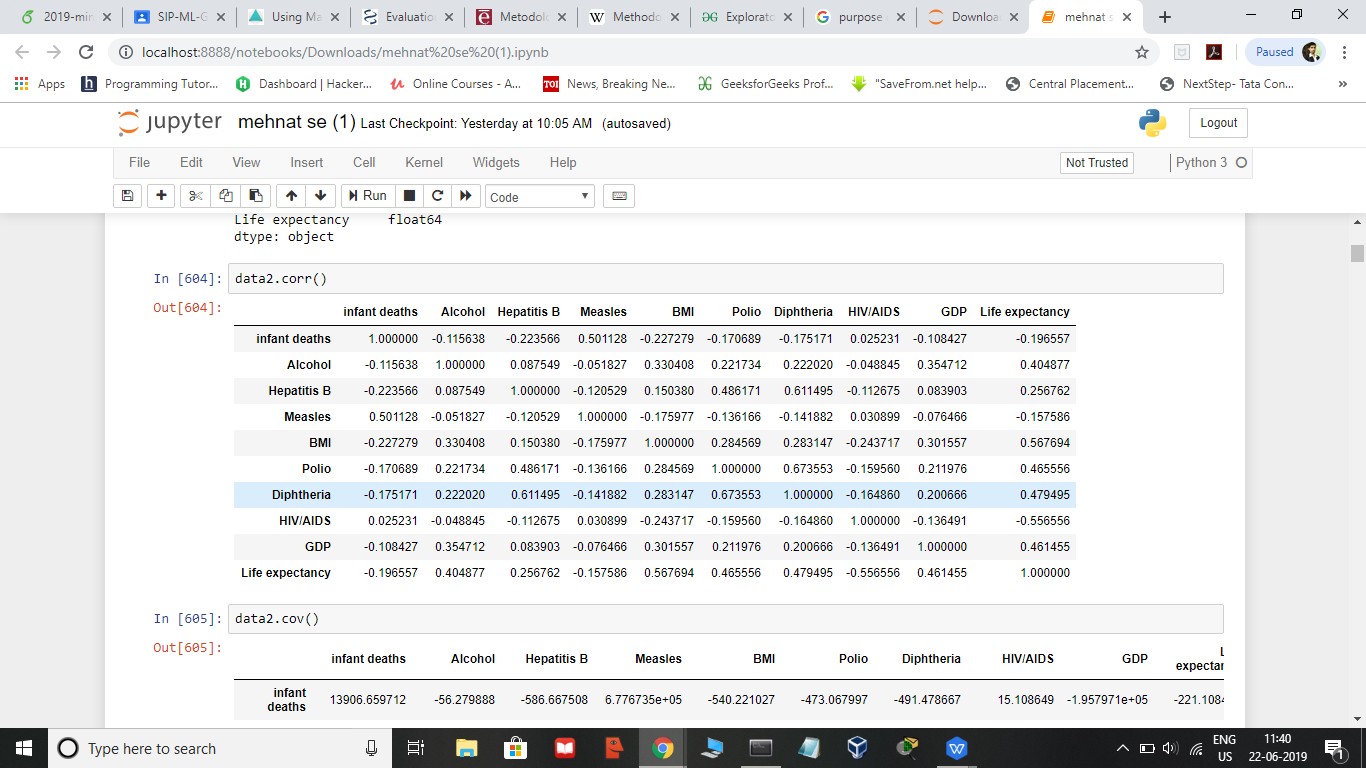
### Exploratory Data Analysis

*E* DA is a phenomenon under data analysis used for gaining a better understanding of data aspects like: main features of data, variables and relationships that hold between them, identifying which variables are important for our problem,

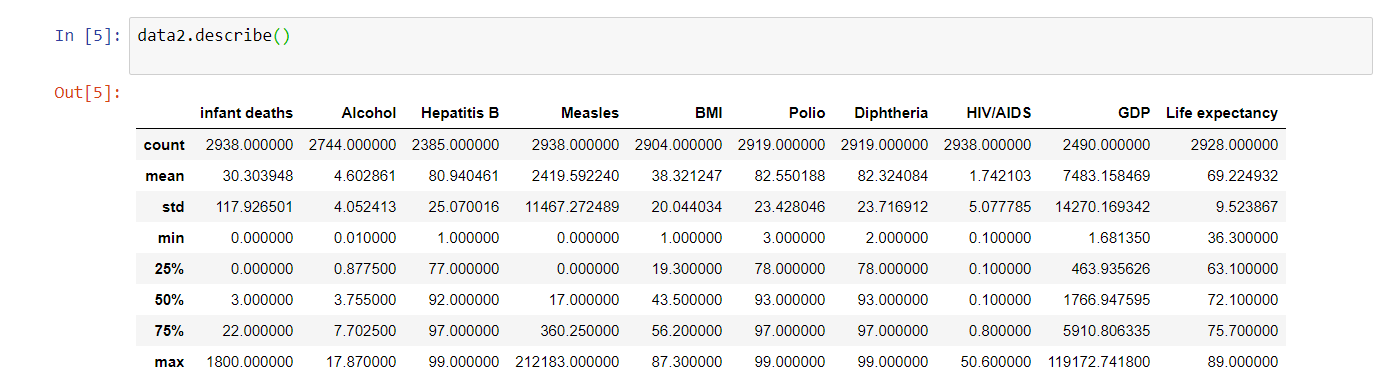
### Purpose

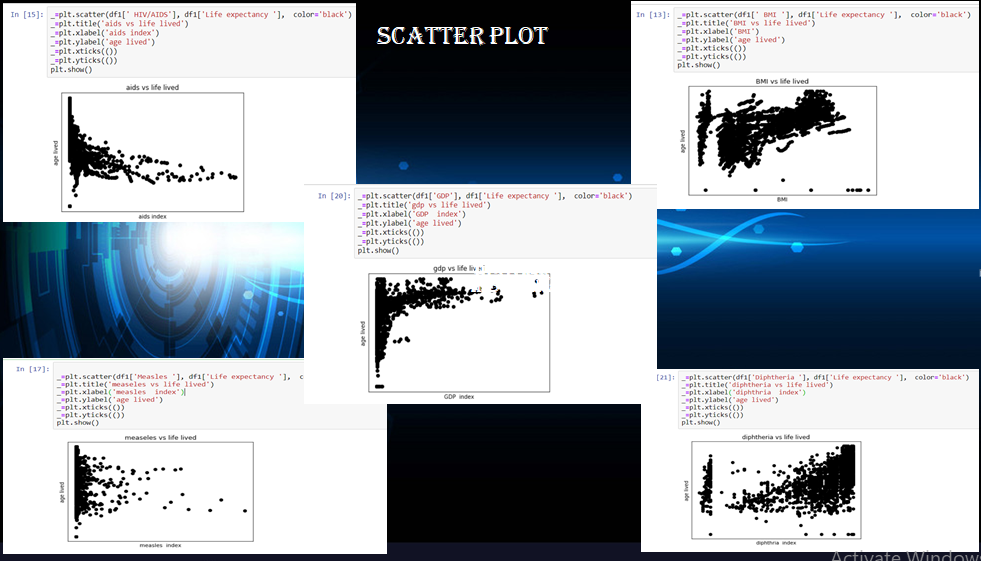
exploratory data analysis (EDA) is an approach to analyzing data sets to summarize their main characteristics, often with visual methods. A statistical model can be used or not, but primarily EDA is for seeing what the data can tell us beyond the formal modeling or hypothesis testing task.

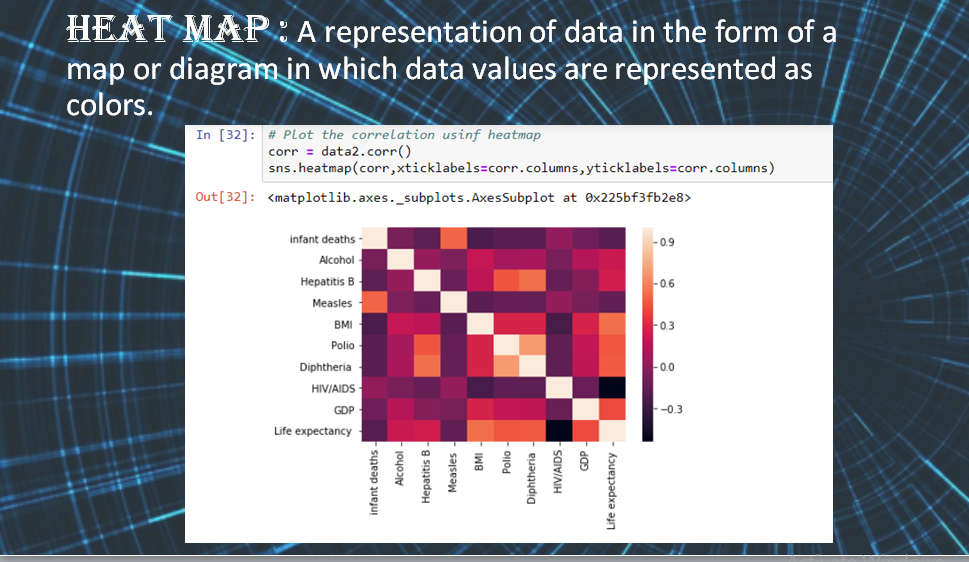
### Figues and Tables

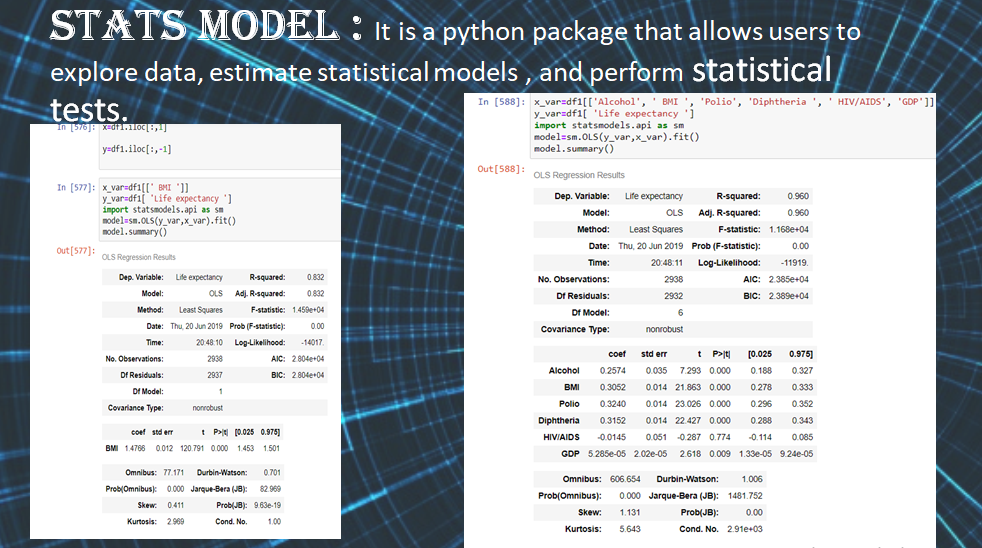
 Correlation (means how the two variable related to each other strongly or weakly.

Description of data

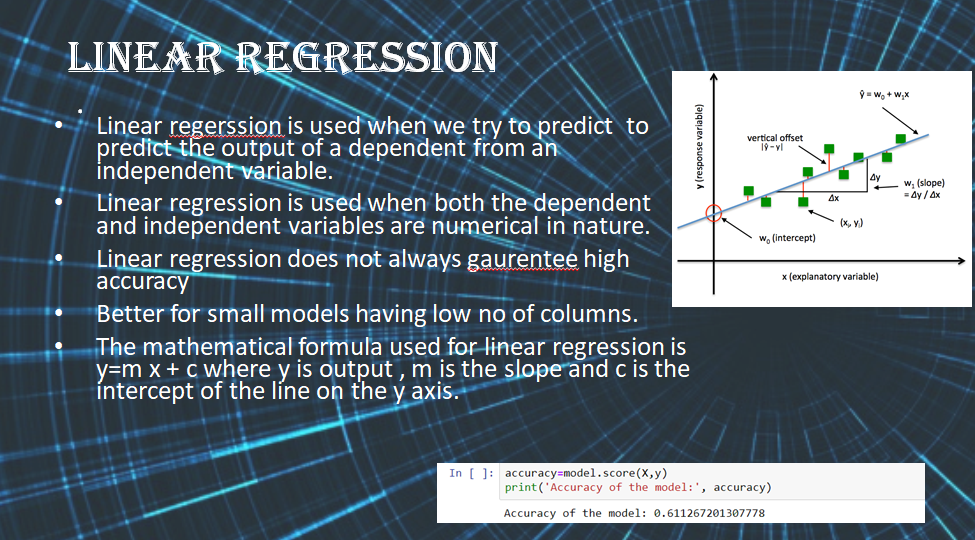


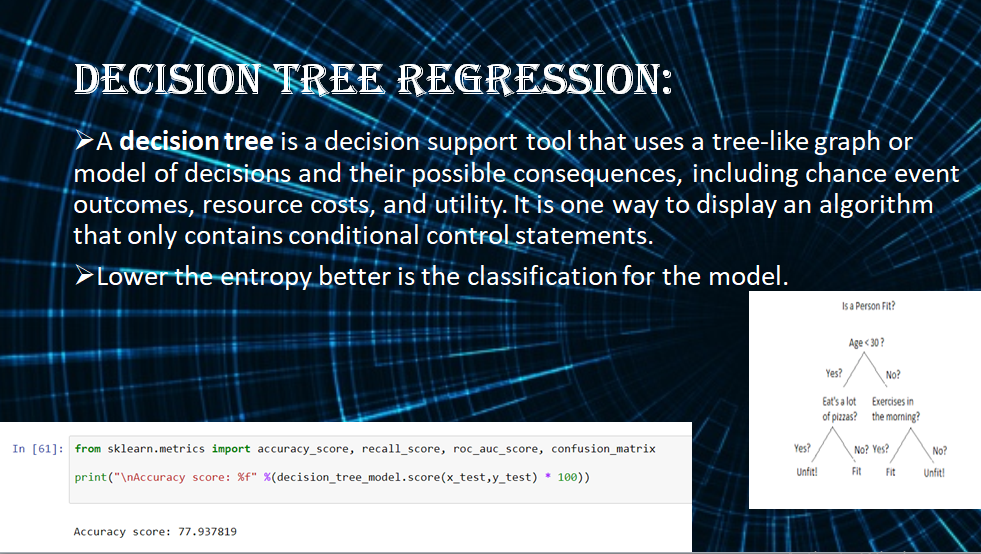


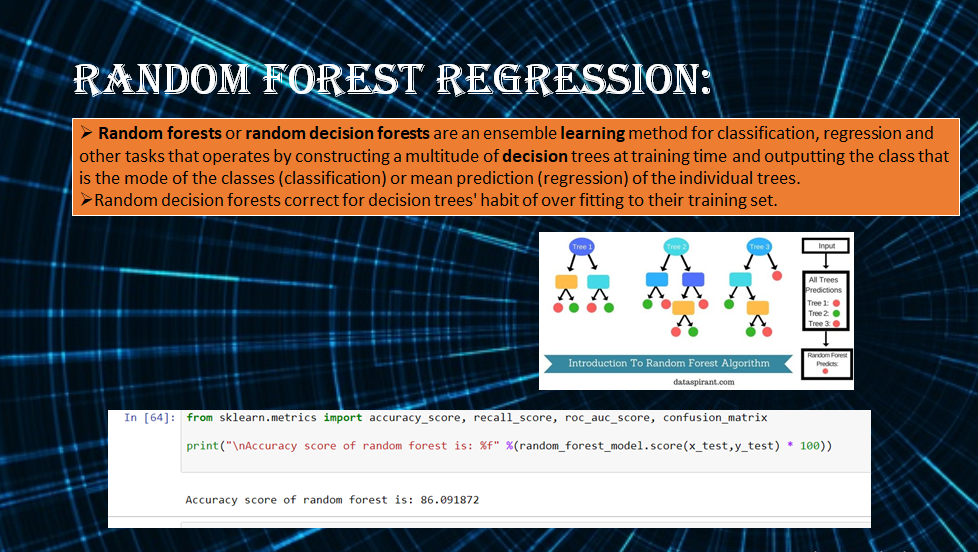


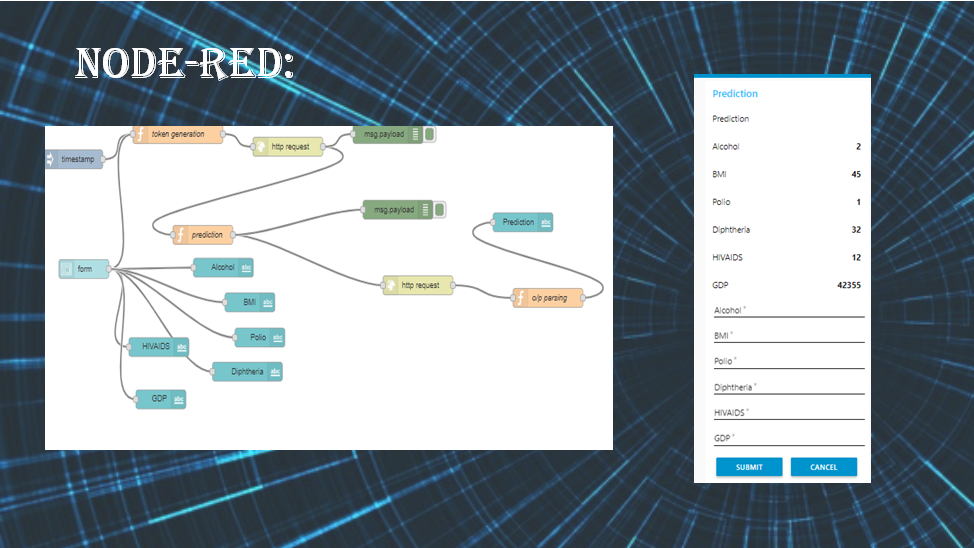


Algorithm:









Conclusion:

Hereby, we conclude that Random Forest is giving highest accuracy of 86% when taken into consideration by important factors like AIDS, alcohol index, BMI, GDP and dipth area index.