Project Name:-Predicting Life Expectancy Using Machine Learning.
Date:-05/24/2020.

- **Project Summary:-** Birth rate is total number of live births per 1000 in a population in a period and life expectancy is measure of how long an organism expected to live.The problem of processing datasets such as electronic medical records (EMR), and their integration with genomics, environmental factors, socioeconomic factors and patient behavior variations have posed a problem for researchers in the health industry. Due to the evolution of data science technologies such as big data virtualization and analytics, data wrangling and with the cloud, health workers now have an improved way of processing and developing meaningful information from huge datasets that have been accumulated over many years.it shows that big data and machine learning techniques can benefit public health researchers with analyzing thousands of variables to obtain data regarding life expectancy and anxiety disorders. They used the demographics of selected regional areas and multiple behavioral health disorders across regions to find correlations between individual behavior indicators and behavioral health outcomes.

Project Requirements:- Following things are required for project:-

➤ ***Functional Requirement:-***

1. IBM Cloud Services
2. Watson Studio
3. Machine learning algorithms.

➤ ***Software Requirement:-***

1. jupyter notebook

2. Node red application

➤ ***Technical Requirement:-***

1. Dataset
2. Processor- 10 core processor, 2.2 GHz with Turboboost upto 3.1 GHz. 25 MB Cache.
3. RAM- 8GB

- **Project Background:-** Although there have been lot of studies undertaken in the past on factors affecting life expectancy considering demographic variables, income composition and mortality rates. It was found that affect of immunization and human development index was not taken into account in the past. Also, some of the past research was done considering multiple linear regression based on data set of one year for all the countries. Hence, this gives motivation to resolve both the factors stated previously by formulating a regression model based on mixed effects model and multiple linear regression while considering data from a period of 2000 to 2015 for all the countries. Important immunization like Hepatitis B, Polio and Diphtheria will also be considered. In a nutshell, this study will focus on immunization factors, mortality factors, economic factors, social factors and other health related factors as well. Since the observations this dataset are based on different countries, it will be easier for a country to determine the predicting factor which is contributing to lower value of life expectancy. This will help in suggesting a country which area should be given importance in order to efficiently improve the life expectancy of its population.

- **Scope of Project:-**

Few works have been done to provide an individually customized life expectancy prediction. we know that it is feasible to predict a PLE for individuals using evolving technologies and devices such as big data, AI,

machine learning techniques, and PHDs, wearables and mobile health monitoring devices. We also knows that the collection of data will be a huge challenge due to the privacy and government policy considerations, which will require collaboration of various bodies in the health industry. The inter working of a heterogeneous health network is also a challenge for data collection. Despite these challenges, we showed a possibility of a LE prediction by proposing an approach of data collection and application by smartphone, with which users can enter their information to access the cloud server to obtain their own LE. No attempt has been made to create this novel idea of using smartphone integrating cloud servers for real-time data entry. We investigated obstacles and barriers that can be resolved by future works described below. It is proposed that this can be extended to a lifetime prediction by using big data to generate a generic data, which can be used to create a LE based on training data as a future solution. Building a generic database will take a considerable amount of time for data collection and analysis, taken from birth to death for various demographic groups to be useful and accurate in representing each attribute classifications. Whilst current applications attempt to show PLE(personalized life expectancy) for smartphone users, they are complicated and difficult to collect technical data requested by the questionnaire, as users are unlikely to be able to provide these data themselves. This can be resolved by connecting the app to the central cloud server with the mHealth networks which provide other health related applications. A centralized cloud server plays a key role in collecting, processing, and creating meaningful value using big data, which forms the input of the solution as well as creates generic data against each user's PLE requirements. Service providers shall envisage challenges and hurdles to obtain consent of personal health 'of heterogeneous health networks across developed countries. This will lead to classification of data based on processing big data and each group's traits, which can be used as

personalized threshold ranges. When this has been completed in a cloud, it can be connected to a smart device app that can provide questionnaires developed by health specialists and collect answers to customize the user's PLE.Optimization of the generic groups data is done by developing an algorithm using machine learning for continuously building and optimizing the user's generic data. As the proposed solution requires processing and transmitting health information of users, information security is a key aspect to consider such as privacy as well as ethical requirements recommenced by regulation bodies, such as the Australian national health and medical research (NHMRC). The scope of security and ethical requirements need to be clearly defined and specified for future work as challenges are expected to build a centralized database with incorporation of health networks. For example, North America, Asia, and Europe may have their own unique requirements to satisfy in dealing with health data with different health research guidelines.To verify the accuracy of LE prediction and validation of data quality, big data techniques and analysis algorithms need to be developed and tested in a real-life situation with several sample groups. As artificial intelligence technology is evolving and being applied rapidly, feasibility may be increasing to collect health data from the public as well as existing health agencies such as centralized health servers.