**Life Expectancy: Prediction & Analysis using ML**

**Overview**

Life Expectancy is an analytical as well as a statistical

measure of the longevity of the population depending

upon distinct factors. Over the years, Life expectancy

observations are being vastly used in medical, healthcare

planning, and pension-related services, by concerned

government authorities and private bodies. Advancements

in forecasting, predictive analysis techniques, and data-

science technologies have now made it possible to

develop accurate predictive models. In many countries, it

is a matter of political debate about how to decide the

retirement age and how to manage the financial issues

related to the public matter. Life expectancy predictions

provide solutions related to these issues in many

developed countries. With the advancement in new

systematic, accurate, efficient, and result-oriented

techniques in the field of Data Science, now predictions

of the Life Expectancy of the selected region are

becoming more prominent in demand of the government

authorities and the private bodies and their policy-

making

Life Expectancy is an analytical as well as a statistical

measure of the longevity of the population depending

upon distinct factors. Over the years, Life expectancy

observations are being vastly used in medical, healthcare

planning, and pension-related services, by concerned

government authorities and private bodies. Advancements

in forecasting, predictive analysis techniques, and data-

science technologies have now made it possible to

develop accurate predictive models. In many countries, it

is a matter of political debate about how to decide the

retirement age and how to manage the financial issues

related to the public matter. Life expectancy predictions

provide solutions related to these issues in many

developed countries. With the advancement in new

systematic, accurate, efficient, and result-oriented

techniques in the field of Data Science, now predictions

of the Life Expectancy of the selected region are

becoming more prominent in demand of the government

authorities and the private bodies and their policy-

making

Life Expectancy is an analytical as well as a statistical

measure of the longevity of the population depending

upon distinct factors. Over the years, Life expectancy

observations are being vastly used in medical, healthcare

planning, and pension-related services, by concerned

government authorities and private bodies. Advancements

in forecasting, predictive analysis techniques, and data-

science technologies have now made it possible to

develop accurate predictive models. In many countries, it

is a matter of political debate about how to decide the

retirement age and how to manage the financial issues

related to the public matter. Life expectancy predictions

provide solutions related to these issues in many

developed countries. With the advancement in new

systematic, accurate, efficient, and result-oriented

techniques in the field of Data Science, now predictions

of the Life Expectancy of the selected region are

becoming more prominent in demand of the government

authorities and the private bodies and their policy-

making

Life Expectancy is an analytical as well as a statistical measure of the longevity of the population depending upon distinct factors. Over the years, Life expectancy observations are being vastly used in medical, healthcare planning, and pension-related services, by concerned

government authorities and private bodies. Advancements in forecasting, predictive analysis techniques, and data-science technologies have now made it possible to develop accurate predictive models. In many countries, it is a matter of political debate about how to decide the

retirement age and how to manage the financial issues related to the public matter. Life expectancy predictions provide solutions related to these issues in many developed countries. With the advancement in new systematic, accurate, efficient, and result-oriented techniques in the field of Data Science, now predictions of the Life Expectancy of the selected region are becoming more prominent in demand of the government authorities and the private bodies and their policymaking.

**Content**

The project relies on accuracy of data. The Global Health Observatory (GHO) data repository under World Health Organization (WHO) keeps track of the health status as well as many other related factors for all countries. The data-sets are made available to public for the purpose of health data analysis. The data-set related to life expectancy, health factors for 193 countries has been collected from the same WHO data repository website and its corresponding economic data was collected from United Nation website. Among all categories of health-related factors only those critical factors were chosen which are more representative. It has been observed that in the past 15 years, there has been a huge development in health sector resulting in improvement of human mortality rates especially in the developing nations in comparison to the past 30 years. Therefore, in this project we have considered data from year 2000-2015 for 193 countries for further analysis. The individual data files have been merged together into a single data-set. As the data-sets were from WHO, we found no evident errors. Finding all data for these countries was difficult and hence, it was decided that we exclude these countries from the final model data-set. The final merged file (final dataset) consists of 22 Columns and 2938 rows which meant 20 predicting variables. All predicting variables was then divided into several broad categories:​Immunization related factors, Mortality factors, Economical factors and Social factors.

**Features**

* Country
* Year
* Status-Developed or Developing status
* Life expectancy-Life Expectancy in age
* Adult Mortality-Adult Mortality Rates of both sexes (probability of dying between 15 and 60 years per 1000 population)
* infant deaths-Number of Infant Deaths per 1000 population
* Alcohol-Alcohol, recorded per capita (15+) consumption (in litres of pure alcohol)
* percentage expenditure-Expenditure on health as a percentage of Gross Domestic Product per capita(%)
* Hepatitis B-Hepatitis B (HepB) immunization coverage among 1-year-olds (%)
* Measles-Measles - number of reported cases per 1000 population
* BMI-Average Body Mass Index of entire population
* under-five deaths-Number of under-five deaths per 1000 population
* Polio-Polio (Pol3) immunization coverage among 1-year-olds (%)
* Total expenditure-General government expenditure on health as a percentage of total government expenditure (%)
* Diphtheria-Diphtheria tetanus toxoid and pertussis (DTP3) immunization coverage among 1-year-olds (%)
* HIV/AIDS-Deaths per 1 000 live births HIV/AIDS (0-4 years)
* GDP-Gross Domestic Product per capita (in USD)
* Population-Population of the country
* thinness 1-19 years-Prevalence of thinness among children and adolescents for Age 10 to 19 (% )
* thinness 5-9 years-Prevalence of thinness among children for Age 5 to 9(%)
* Income composition of resources-Human Development Index in terms of income composition of resources (index ranging from 0 to 1)
* Schooling-Number of years of Schooling(years).

**References**

* Noorhannah Boodhun, Manoj Jayabalan, “Risk prediction in life insurance industry using supervised learning algorithms,” *Complex & Intelligent Systems,* vol. 4, no. 2, pp. 145-154, 2018.
* Mahumud, R.A., Hossain, G., Hossain,R., Islam, N. and Rawal, L.,;, “Impact of Life Expectancy on Economic Growth and Health Care Expenditures in Bangladesh.,” *Universal Journal* *of Public Health,* vol. 1, no. 4, pp. 180-186, 2013.
* Bhosale, A.A. and Sundaram, K.K., “Life prediction equation for humanbeings,” *International Conference on Bioinformatics and BiomedicalTechnology,* vol. IEEE, pp. 266-268, 2019.
* <https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-who>