Predictive Analysis of Factors Affecting Life Expectancy Using Machine Learning Techniques

ISET

Author:

Giorgi Gvekhidze
Janeli Khvedelidze

Supervisor:

Shota Natenadze

Introduction:

The world population has been adding fleetly, leading to colorful challenges similar as food deficit, healthcare, and environmental damage. One of the critical issues in the healthcare sector is to ameliorate the average life expectation of people worldwide. Several factors similar as education, profitable growth, healthcare expenditure, and social weal programs can significantly impact life expectation. thus, it's essential to study and dissect these factors and their goods on life expectation.

Purpose:

This study aims to investigate the relationship between life expectation and socio-economic and environmental factors using across-country analysis. The purpose of this research is to identify the most significant factors that impact life expectation across countries and to give perceptivity into potential policy measures that may improve life expectation worldwide. Moreover, the study will help in understanding the relationship between different factors and life expectation.

The research questions guiding this study are:
1. What is the relationship between socio-economic and environmental factors and life expectancy across countries?
2. Which factors are most strongly associated with changes in life expectancy?
3. How can these insights inform policies that promote improvements in life expectancy globally?

Methodology:

In the study, life expectancy will be predicted using machine learning algorithms based on a variety of variables, including GDP, education, healthcare spending, disease prevalence, and environmental factors. The study will make use of the Kaggle Life Expectancy dataset, which

includes information from 193 nations between 2000 and 2015. The dataset has 22 attributes, such as life expectancy, infant mortality, GDP, healthcare spending, and educational spending.

The first step of the study will be to perform exploratory data analysis (EDA) to understand the distribution and correlation between different variables. The second step will involve feature engineering, where we will select relevant features that affect life expectancy. After selecting relevant features, we will split the data into training and testing sets and apply various machine learning algorithms such as Linear Regression, Random Forest, and Gradient Boosting to predict life expectancy.

To evaluate the performance of different models, we will use various evaluation metrics such as Mean Absolute Error, Mean Squared Error, and R-Squared. Finally, we will use feature importance techniques such as permutation importance, SHAP values, and Partial Dependence Plots to understand the impact of different features on life expectancy.

Literature Review:

Numerous studies have examined how different factors impact life expectancy. One study by Marmot and Bell (2012) found that social factors such as education, employment, and income play a crucial role in determining life expectancy.

In addition, several studies have implemented machine learning techniques to forecast life expectancy. For instance, Adebayo, Oladele, and Ayeni (2020) used machine learning algorithms to predict life expectancy based on socio-economic and demographic factors, including education, income, and gender. They found that the random forest algorithm was the most accurate, achieving an accuracy rate of 95.62% in predicting life expectancy.

Another study by, Yang et al. (2021) utilized algorithms like Random Forest and Gradient Boosting to predict life expectancy, based on variables such as GDP, healthcare expenditure, and environmental factors. Their findings suggest that these factors greatly affect life expectancy.

Conclusion:

The proposed research aims to employ machine learning techniques to forecast life expectancy and investigate the correlation between various factors and life expectancy. The study aims to contribute to existing literature by providing valuable insights into the critical factors that impact life expectancy. Moreover, it aims to aid policymakers in developing effective strategies to enhance life expectancy on a global scale.

The research endeavors to identify the most prominent socio-economic and environmental factors that influence life expectancy across various nations and determine how these factors differ in their impact across different regions. The study specifically focuses on examining the influence of variables such as income, education, access to healthcare, air pollution, and water quality on life expectancy.

Bibliography:

Marmot, M., & Bell, R. (2012). Social determinants and non-communicable diseases: time for integrated action. BMJ, 345, e6202.
Link: https://www.researchgate.net/publication/229073819_Fair_Society_Healthy_Lives

Adebayo, Oladele, and Ayeni
Link:
https://www.researchgate.net/publication/356235215_Life_Expectancy_Prediction_Analysis_using_ML

Yang, S., Fang, S., Wei, W., & Zhao, X. (2021). A machine learning approach to predicting life expectancy. International Journal of Environmental Research and Public Health, 18(2), 480.
Link: https://www.mdpi.com/1660-4601/18/2/480