

Project Proposal: Predicting Life Expectancy Using Regression Techniques

Introduction: Life expectancy is a crucial indicator of a nation's health and development. In this project, we aim to analyze the factors influencing life expectancy using regression techniques. We will utilize the "Life Expectancy (WHO)" dataset from Kaggle, which contains health-related data for 193 countries. We aim to build predictive models to understand the relationships between various features and life expectancy.

Methodology: We will begin with **Exploratory Data Analysis (EDA) and Data Cleaning**, exploring the dataset to identify missing values, outliers, and inconsistencies. We will visualize key relationships between variables and perform necessary data transformations to ensure data quality and integrity.

Next, we will apply **Feature Selection with Lasso Regression** to identify the most significant predictors of life expectancy. Lasso regression applies a penalty to the regression coefficients, effectively shrinking less important feature weights to zero. This approach will help us focus on the most relevant features and reduce model complexity.

Following feature selection, we will **implement Polynomial Regression** to capture nonlinear relationships in the data. Polynomial regression extends linear models by including higher-order terms, allowing for a better representation of complex patterns in life expectancy trends. Additionally, we will **explore Regression Splines**, which use piecewise polynomial functions to provide flexible modeling of non-linear relationships, improving predictive accuracy and interpretability.

Finally, we will evaluate our models and discuss Insights and Discussion. We will analyze the performance of different regression techniques by comparing their predictive accuracy and assessing their real-world applicability. By interpreting our results, we aim to provide meaningful conclusions about the factors influencing life expectancy and their broader implications on global health.

Expected Outcomes: Through this project, we expect to gain a better understanding of the key factors influencing life expectancy. Our analysis will provide insights into how various health and socioeconomic indicators contribute to longevity. Additionally, we will evaluate the effectiveness of different regression techniques in modeling real-world health data.

Conclusion: This project will enhance our understanding of regression techniques while applying them to a meaningful real-world problem. By systematically analyzing the data and leveraging advanced regression methods, we aim to derive valuable insights that can contribute to policy discussions on global health and longevity.

References:

1. Dataset: <https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-who>

Submitted By:

1. Tanjina Zaman (tanjina.zaman@marquette.edu)
2. Dewan Imran Ahmed (dewanimran.ahmed@marquette.edu)
3. Sajjad Islam (sajjad.islam@marquette.edu)