**Life Expectancy: A Data-Driven Exploration and Prediction**

**Milestone: Performance Evaluation and Interpretation**

Group 13

Sai Varun Kumar Namburi

Prajwal Srinivas

[namburi.sai@northeastern.edu](mailto:namburi.sai@northeastern.edu)

[srinivas.pra@northeastern.edu](mailto:srinivas.pra@northeastern.edu)

**Percentage of Effort Contributed by Student1: 50%**

**Percentage of Effort Contributed by Student2: 50%**

**Signature of Student 1: Sai Varun**

**Signature of Student 2: Prajwal**

**Date of Submission: 16th Apr 2023**

**Data Source:** This dataset is taken from kaggle.com

<https://www.kaggle.com/code/philbowman212/life-expectancy-exploratory-data-analysis/data>

**Data Description:**

The World Health Organization's Global Health Observatory (GHO) maintains records of all countries' health statuses and related factors. The data concerning life expectancy and health factors for 193 countries were obtained from the WHO's Global Health Observatory data repository. It was noted that over the past 15 years, there has been significant progress in the health sector, leading to a significant improvement in human mortality rates, particularly in developing nations compared to the last 30 years. In this project, data from the years 2000 to 2015 for 193 countries were selected for further analysis. In this dataset, we have 22 columns as below

(Country, Year, Status, Life expectancy, Adult Mortality, infant deaths, Alcohol, percentage expenditure, Hepatitis B, Measles, BMI, under-five deaths, Polio, Total expenditure, Diphtheria, HIV/AIDS, GDP, Population, thinness 1-19 years, thinness 5-9 years, Income composition of resources, Schooling)

**Linear Regression:**

Linear Regression is a simple and widely used statistical model for predicting a continuous outcome variable based on one or more predictor variables. It assumes a linear relationship between the outcome variable and the predictors, which can be limiting in cases where the relationship is more complex.

**Text

Description automatically generated**

It is important to note that linear regression assumes that the relationship between the outcome variable and the predictor variables is linear. Additionally, it assumes that the errors or residuals are normally distributed and have constant variance. Therefore, you should assess the assumptions of the linear regression model before using it to make predictions.

**Mixed Effect Model:**

A mixed-effects model can be a useful approach to analyze life expectancy data when there are repeated measurements on the same individuals, or when individuals are nested within groups (e.g., cities, countries, or families).

Text

Description automatically generated

Text

Description automatically generated

Chart, scatter chart

Description automatically generated

**Neural Networks Model:**

A neural network can be a powerful approach for predicting life expectancy based on a wide range of predictor variables. Neural networks are a type of machine learning model that can learn complex non-linear relationships between input and output variables.

To build a neural network model for life expectancy, you would need a dataset that includes observations of individuals along with their ages, sex, education, income, health behaviors, and life expectancy. You would use this dataset to train the neural network to predict life expectancy based on the predictor variables.

Text

Description automatically generated with medium confidence

A screenshot of a computer screen

Description automatically generated with medium confidence

Text

Description automatically generated

After comparing accuracy with the three models

Text

Description automatically generated

Text

Description automatically generated with medium confidence

In this particular case, the mixed effect model is superior to both normal linear regression and the used neural network architecture, since it takes into account the dependence of the data.