1_MAIN_Life_Expectancy_WHO_UN_Analysis_Modeling_MAIN

January 29, 2023

Table of Contents

1 1_MAIN_Life_Expectancy_WHO_UN_Analysis_Modeling

To: Magnimind

From: Matt Curcio, matt.curcio.ri@gmail.com

Date: 2023-01-29

Re: NOTEBOOK #1

1.1 Executive Summary

Problem Statement

This report investigates modifiable environmental factors that may contribute to life expectancy in the hopes of finding a mathematical model. The primary questions answered here:

1. *Which variables are significant and important in modeling and predicting Life Expectancy?*

Impetus

Information on Life Expectancy-Kaggle from the World Health Organization and United Nations was gathered for the years 2000-2015 for 193 countries in three categories:

1. *Communicable diseases*: Hepatitis B, Measles, Polio, Diphtheria, HIV/AIDS.
2. *Country-specific monetary data*: Gross domestic product, Governmental expenditure rates, Average personal income, Developmental status, Population, Education.
3. *Health-related information*: Infant and adult mortality rates, Number of deaths less than 5 yr, Alcohol consumption rates, Child-related malnutrition, Average body mass index per capita.

Important Features

21 Features were culled to six (6) features that were deemed most important then used in the final modeling of life expactancy:

1. Income
2. Education
3. HIV
4. DTP
5. Polio
6. AdultMort

Results

Four algorthms were tested on the data and the accuracy alone was used as a benchmark:

Model	Average % Accuracy (cv=5)
Gradient Boosting Regressor	94.7
Decision Tree Regressor	88.5
Linear Regression	81.0
Support Vector Regressor	19.0

1.2 Introduction

In 2006, The World Health Organization published a report entitled *Preventing disease through healthy environments*. >[It] "confirms that approximately one-quarter of the global disease burden, and more than one-third of the burden among children, is due to modifiable environmental factors." > >WHO-Preventing Disease Through Healthy Environments

The dataset was obtained from Kaggle.com: https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-who

1.2.1 Project Planning

- Initial Data Analysis
- Exploratory Data Analysis
- Exploratory Data Analysis #2
- Pandas Profiling Library
- Feature Engineering
- Recursive Feature Elimination
- Modeling
- Discussion of Linear Model
- Results

Response Variable	**Description**
LifeExpectancy	Life Expectancy (Yr)

**

Green indicates feature used in final model

Blue indicates feature of questionable importance

Orange indicates dropped due to high collinearity

Red indicates dropped due to missing values greater than 15%
**

List Of Features	Description

| **
Income
** | Income composition of resources, Human Development Index | | **
Education
** | Years of Education | | **
HIV
** | HIV/AIDS: Deaths per 1,000 | | **
DTP
** | Diphtheria, tetanus toxoid & pertussis % immunization coverage among 1-year-olds | | **
Polio
** | Pol3: % immunization coverage among 1-year-olds | | **
AdultMort
** | Adult Mortality Rates of both sexes (probability of dying between 15 and 60 years per 1000 population) | | **
Year
** | Year | | **
Country
** | Country | | **
Status
** | Developed(1) or Developing(0) | | **
BMI
** | Average Body Mass Index of entire population | | **
lt5yD

** | Number of under-five deaths per 1,000 population | | **

Thin1_19y

** | % Prevalence of thinness among children and adolescents 10 < Age < 19 | | **

TotalExpen

** | Total Expenditure | | **

PercExpen

** | Percent Expenditure | | **

EtOH

** | Alcohol consumption, litres of pure alcohol per capita | | **

Measles

** | Number of reported cases per 1,000 population | | **

Thin5_9y

** | % Prevalence of thinness among children and adolescents 5 < Age < 9 | | **

InfD

** | Number of Infant Deaths per 1,000 population | | **

HepB

** | Hepatitis B: % immunization coverage among 1-year-olds | | **

Population

** | Population of country | | **

GDP

** | Gross Domestic Product per capita (in USD) |