Data-Oriented Programming Paradigms: Exercise 3 Report

Aaron Abebe 01463783, Gabriel Breiner 01447920, Carlos Carlos Vargas 11823257, Sanja Priselac 01428180

January 2020

1 Research Questions

The aim of this project is to answer the questions considering poverty:

- What percentage of the world population lives in extreme poverty?
- Which characteristics are predictive for countries living in extreme poverty?
- Which characteristics are predictive for populations emerging from extreme poverty?

The goal of this work is to get a better overview of how the world wide socio-economics measures influence extreme poverty.

2 Methodology

Data used for this project were taken from *UNESCO Institute for Statistics*. It contains different demographic and economic parameters for 233 countries starting from the year 1970 to 2015.

This data set did not contain the target variable (living in extreme poverty / not living in extreme poverty, so we had to calculate it ourselves. We decided for the World Bank's Definition for "extreme poverty", which is defined by defining a threshold for the average daily GNI (per capita) at 1.9$ per day - as of 2020. This threshold changed through out the years: Starting with 1$ to 1.25$ in 2005 and then finally 1.9$ in 2015. Also the way of measuring the GNI changed throughout the years. We accounted for all of the changes by labeling extreme poverty according to the standard's of the considered year.

In order to find out which characteristic are important for poverty and countries emerging from the extreme poverty we used only a subset of features available from the datasets - only demographic features that have less than 50% missing values were taken. We used the following parameters:

TIME - year, Fertility rate - total (births per woman), Life expectancy at birth - total (years), Mortality rate - infant (per 1,000 live births), Population growth (annual %), Rural population (% of total population), Total population, Population aged 65 years or older, Population aged 25-64 years, Population aged 14 years or younger, Population aged 15-24 years.

In order to get predictive characteristics for the countries living in extreme poverty we firstly had to solve the problem of missing data. Missing data were handled with the linear interpolation for each feature and each country separately. Afterwords we used the Ridge Classifier for a classification task with poverty as response variable (0: country not extremely poor, 1:country extremely poor) on the scaled data and checked the coefficients of the model. Ridge Classifier is a linear classifier that finds the coefficients for each feature and introduces penalties for the less important features. Since our data is scaled, we can use the coefficient as a measure of feature importance for the model.

To identify the countries emerging from poverty we identify first the countries which have been classified as in extreme poverty at least once until 2015. Then, we identify the countries that are not classified as extreme poverty in 2015 and from those two datasets, we took the intersection, we called this dataset emerging countries. Here we found that from 233 countries, 68 have emerged from poverty along the time frame.

After identifying the countries emerging from poverty, we built a model that identify the most important features that impact the classification of those countries as in extreme poverty. Here we tried different techniques: 1) The variability of the features within the classification groups (ANOVA F-values), 2) computing chi-squared test to identify whether there is a statistically significance of the features regarding to imply the classification of a country as emerging from poverty, 3) We identify the importance of the features by modelling a decision tree a decision tree (impurity) and finally 4) we computed a correlation heat-map to see the relation between the features (correlation matrix).

3 Findings

As an answer to the second research question, *Which characteristics are predictive for countries living in extreme poverty?* we used the weights from the Ridge Classifier. The important feature when predicting extreme poverty are: **fertility rate**, **population growth** and **Life expectancy at birth**. The feature that present population (*Total population, Population aged 65 years or older, Population aged 25-64 years, Population aged 14years or younger, Population aged 15-24 years*) are not important at all.

We found that 68 out of 233 countries have emerged from extreme poverty from 1970 until 2015. The features that impact more for a country to emerge from poverty are in a descending order: Fertility, Mortality rate and Life expectancy at birth and the Rural population percentage.

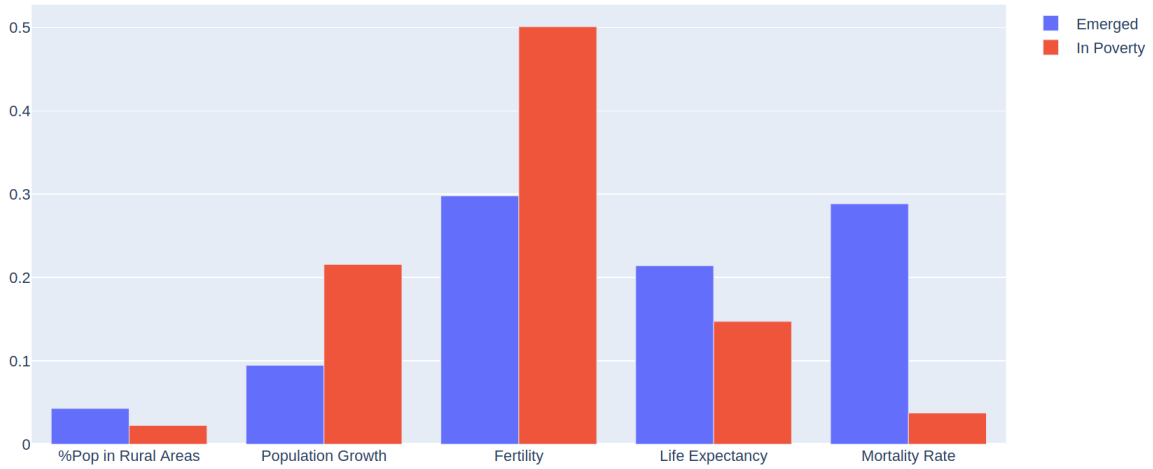

Figure 1: Comparison of the results for Question 2 and 3

Interestingly, questions 2 and 3 come to a similar conclusion about the importance of the features. From the figure 1, which shows the weights of the most important features for both questions, we can see that weights from the 2nd Question (in poverty) have much more variability than the weights from the 3rd Question (Emerged). When predicting the countries in poverty, the fertility rate seems to be the dominant feature. Moreover, the mortality rate seems to be pretty important for the Question 3 but not important for the Question 2.