**Exploring Global Socio-Economic Indicators using Predictive Analytics and Classification**

Submitted to:

Professor Ceni Babaoglu, PhD

Submitted by:

Mir Russell

Student ID:

501211762

Date of submission: 25 January 2024

**Abstract:**

Understanding the historical evolution of the dynamics of global socio-economic indicators is critical for informed policy decisions. This research, employs a longitudinal analysis of global data spanning several decades, addresses the feasibility of predicting future trends in key indicators such as education, income, GDP per capita, life expectancy, education levels, and health variables.

The problem being considered for this project is determining whether classification method, the technique used to answer the hypotheses in United Nations Development Program­ (UNDP). (2020). "Human Development Report.", will produce almost similar results from the Gapminder data. What will be gained from this project are How have socio-economic indicators evolved globally over the years? Can we predict future trends in key indicators, such as life expectancy and GDP per capita, using historical data? What factors contribute to the classification of countries into different development categories?

The project uses the data from “Gapminder”, a combination of data from multiple sources into unique coherent time-series that you cannot find anywhere else. Dataset:<https://raw.githubusercontent.com/BME1478H/Winter2022class/master/data/world-data-gapminder.csv> This dataset contains14 attributes and 39202 observations, both numerical and categorical data. This project will highlight the understanding the patterns and relationships within the provided dataset and investigate the global socio-economic indicators using a combination of Predictive Analytics and Classification.

The research project aims to uncover the key determinants shaping the classification of countries and provides insights into the complexities of global development categorization patterns which will provide a practical tool for policymakers and researchers to assess and address the multi-dimensional nature of development.

We begin by carrying out data preprocessing and exploratory analysis where we impute the missing values by using techniques like Mean, k- Nearest Neighbors. Employing machine learning classification algorithms ­such as Decision Trees and Support Vector Machines (SVM) to categorize countries based on their developmental status. Through regression analysis and time-series forecasting techniques, we unravel the intricate relationships within the data as well as to predict future values of key indicators as well as pattern mining. We will use Python with Pandas, Scikit-learn as well as Matplotlib and Seaborn for visualizing trends, patterns, and findings.

**Reference:**

Easterly, W. (2000). "Economic Growth and Human Development" World Development

Blumenstock, J., Cadamuro, G., & On, R. (2015). "Predicting Poverty and Wealth from Mobile Phone Metadata" Science Journal

Roser, M., & Ortiz-Ospina, E. (2020). "World Population Growth." Our World in Data.

Murray, C. J. L., & Lopez, A. D. (1996). "The Global Burden of Disease: A Comprehensive Assessment of Mortality and Disability from Diseases, Injuries, and Risk Factors in 1990 and Projected to 2020." World Health Organization.

World Bank. (2020). "World Development Indicators."

Barro, R. J., & Lee, J. W. (2013). "A New Data Set of Educational Attainment in the World, 1950–2010." Journal of Development Economics.

United Nations Development Program (UNDP). (2020). "Human Development Report."

Lipton, Z. C., Berkowitz, J., & Elkan, C. (2017). "Time Series Prediction with Deep Learning: A Summary and New Challenges" Journal of Machine Learning Research