EC930
Summer Project

Name: Uchenna Charles Onuoha

Registration No: 202186437

Course: EC930 Summer Project

Project Title: Understanding and Predicting Development: A Machine Learning Approach

Supervisor: David Comerford

2021/22

THE DEPARTMENT OF ECONOMICS

Table of Contents

Table of Figures

List of Tables

Tables of Equations

1.0 Introduction

Globally, countries are ranked as developed or developing based on various social, economic, and governance indicators. The World Economic Situation and Prospects (2019, 2022) report by the United Nations classifies countries into three broad categories; developed, transitioning, and developing. This classification was done based on fundamental economic conditions in each country; usually, the World Bank specified threshold levels of GNI per capita. This paper aims to apply ML methods to understand the economic development differences between developed, transitioning, and developing countries.

Economic growth models have been used to analyse the growth patterns in developed and developing countries. In the 1980s, the Neoclassical school of thought, which emphasised free-market fundamentalism, rational expectation theories, and supply-side macroeconomics, emerged as the dominant school of thought for economic growth and development theories (Todaro & Smith 2011). The primary argument of the neoclassical theories, of which Friedrich Hayek was a leading proponent, is that underdevelopment results from incessant government intervention in free markets and improper pricing policies (Jones & Jackson, 2016). Theories at this stage sought to enthrone free markets, privatise state-owned corporations and enthrone the free movement of goods and capital across countries.

During the neoclassical era, the Solow-Swan growth model represented the traditional neoclassical growth theory. Solow (1956) and Swan (1956) independently postulated that the level of three indicators could explain long-run economic growth (a component of economic development), capital accumulation (savings and investment), population growth (human capital), and productivity (technological progress). Solow built his theory on the classical Harrod-Domar model (Harrod, 1939) but rejected the assumption of fixed-proportion in substitutability of the factors enumerated. The vital aspect of this theory is its prediction that the income level of poorer countries will catch up to more prosperous countries, given that they have similar savings rates, labour force growth rates, and total factor productivity.

Authors such as Hamilton & Monteagudo (1998), Barossi-Filho et al. (2005), and Nkalu et al. (2018) have used traditional econometric approaches to test neo-classical growth models empirically. They investigated whether the hypothesis set forth by Solow-Swan could hold for countries worldwide. These authors employed traditional Ordinary Least Square (OLS) models and found that the Solow-Swan postulation was valid for some countries.

In recent times, however, Machine Learning (ML) techniques have been applied to analyse growth instead of traditional growth models, especially with the emergence of big data (Cogoljević et al. (2018); Herrera et al. (2019); Mele & Magazzino (2020); Wu et al. (2020); Otchia & Asongu (2021); Maccarrone et al. (2021)). This analysis revisits the issue using Machine Learning models to analyse, understand and predict the growth and development patterns across countries (economies) globally.

Machine Learning (ML) is a sub-field of Artificial Intelligence that broadly defines a machine's capability to imitate intelligent human behaviour (Brown, 2021). ML explores the analysis and construction of algorithms that can learn from and make predictions on data. ML has three subdivisions: supervised, unsupervised, and reinforcement learning. While unsupervised ML algorithms are used for analysis without a dependent variable, supervised ML algorithms are used for datasets with labelled target (dependent) variables. A significant similarity between ML and traditional econometric research is that both are based on the classical regression model estimation. However, while most estimations in econometrics are of the linear regression type, ML models are non-linear algorithms that can better find and fit the relationship between variables. These ML models are discussed subsequently in the methodology section of this paper.

ML has increasingly been applied to economic research since its first mention in the work of Lee and Lee (1974) as a means of pattern recognition. Since then, many authors have applied this technique to understanding economic growth and development. Wu et al. (2020) explored the nexus between financial development and economic growth in Asian economies using traditional Auto-regressive Distributed Lag (ARDL) models and ML classification algorithms. They employed traditional econometric approaches of unit root testing, ARDL for testing long-run relationships, and Granger causality test for bi-directional relation to evidence causality between the variables. For China and India, the supervised ML model employed had a predictive accuracy of 89.3% and 89.7%, respectively, predicting past economic growth and recession occurrences. These authors extended traditional econometric modelling with ML algorithms to help inform policy direction.

Similarly, Basuchoudhary et al. (2017) and Herrera et al. (2019) applied ML to economic growth and long-term forecasting of energy commodity prices. Herrera et al. (2019) compared a long-horizon forecast performance of traditional econometric models with ML methods for oil, coal, and gas prices worldwide using monthly data from the IMF. The predictive accuracy of each approach was measured using the root-mean-square error[1] (RMSE). The study showed that ML techniques provided more accurate results when compared with econometric models and even offered an additional advantage of predicting turning points in the analysis. Similarly, Basuchoudhary et al. (2017) recognised the ability of ML models to perform well using out-of-sample tests, an area in which econometric models tend to perform poorly. The result of the study categorised the most crucial policy levers for economic growth while also highlighting the importance of ML in eliminating policy levers that are not important.

Otchia and Asongu (2021) found that using ML techniques and nightlight data from African economies, the extent of economic development can be assessed in areas without sufficient data reporting. This practical use of visual data would not fit into traditional econometrics analysis.

The conclusions reached by the authors above evidence how ML algorithms' predictive power outperforms traditional econometric methods. Charpentier et al. (2019) assent to this conclusion noting that this development has occurred in recent studies but that ML models tend to trade increased accuracy for decreased explanatory powers. These authors have made a case for targeted policymaking by highlighting salient policy levers for the studied indicators.

The importance of econometric modelling for understanding the causal relationship between variables cannot be overestimated. The celebrated work of Angrist and Pischke (2015) provides insight into conducting both randomised and natural economic experiments using improved forms of OLS methodology such as Difference in Difference and Instrumental Variables. However, while such methods are essential for understanding the causal relationship between economic variables and economic growth, authors who have employed ML techniques for similar hypotheses have improved predictive accuracy. The inability of econometric models to generalise properly to new, unseen data and a tendency to overfit data

[1] Root Mean Square Error (RMSE) is the standard deviation of the residuals, the difference between the values predicted by a model and the observed values.

relegate their usefulness in predictive modelling. A relative advantage that ML algorithms offer is the range of options it gives a researcher in building the underlying pattern recognition between variables of interest. Such choices include cross-validation[2], feature engineering[3], and setting bias-variance tradeoff[4] in a model. ML models also offer econometric researchers the ability to select the best independent variables for econometric models from a much larger set.

The aims of this analysis are three-fold. First, the paper uses unsupervised ML to create a cluster for similar countries. This helps the reader understand the development stages of countries based on similar groups. It also fits supervised ML models using a panel dataset split into train[5] and test[6] datasets. The confusion matrix[7] is used to test the accuracy of the models employed in the analysis model. Second, using the fitted supervised ML models, a ranking is produced for the most significant predictor of a country's development status to help policymakers understand what variables present the most distinct differences between developing and developed nations. A traditional logistic regression is also estimated for comparison purposes with the ML approach.

The dataset used for this analysis was assessed programmatically from the World Bank Database. The World Bank Development Indicators (WDI) is the primary dataset published by the World Bank on economic development indicators, culled from recognised official sources. The dataset contains data for 217 economies worldwide, from 1960 to 2020, a panel data presentation. This work employs data from this database alone, iterating across 147 economies.

This work is important because it uses ML techniques to cluster countries based on specified economic variables. This will help policymakers understand which economies globally are similar, affirming the stage of economic development similarities across countries. The

[2] Cross validation is ML is a technique used to assess how a model will generalize to new unseen data during the fitting stage.

[3] Feature engineering in ML is the technique of using raw data to create new variables that can proxy an indicator of interest.

[4] This describes a feature during ML training where the model can either have high bias and low variance (underfitting the data) or it will have high variance and low bias(overfitting the data).

[5] This is the subset of the actual dataset that is fed into the machine learning model to discover and learn patterns

[6] This is also called validation datatset and is the remaining subset of the data used to evaluate the performance and progress of ML model training.

[7] In statistical classification problems, a confusion matrix is a table presentation used to describe the performance of a supervised algorithm in predicting the true classes of the test data.

variables employed to understand development are grouped into economic development indicators, social development indicators, industrial development indicators, and governance development indicators. Also, by using a supervised ML algorithm to learn development characteristics and classify countries into developing, transitioning, and developed, policymakers can predict an economy's development stage and understand which variables are most important to understanding (classifying) the stages of development globally.

2.0 Literature Review

2.1 Economic Growth and Development

While some authors use the terms economic growth and economic development synonymously, they are different concepts. Economic growth can be defined as the increase (decrease) in the inflation-adjusted market value of goods and services produced by an economy over a given period, usually a year. Real Gross Domestic Product (GDP) measures an economy's productive capacity. On the other hand, Economic Development has a greater scope and encompasses sustainable, long-term economic growth. According to Todaro & Smith (2011), Economic Development refers to improvements in quality of life and living standards. This implies the availability of life-sustaining goods, higher incomes, and freedom of choice.

> *"Development must therefore be conceived of as a multi-dimensional process involving major changes in social structures, popular attitudes, and national institutions, as well as the acceleration of economic growth, the reduction of inequality, and the eradication of poverty."* (Todaro & Smith 2011)

Traditionally, GDP per capita (real GDP/ population) is used to measure the standard of living in an economy. When output grows faster than the population, GDP will rise. When output grows slower than the population, the average standard of living will fall (Hall & Lieberman, 2013).

The literature on economic development can be subdivided into classical and contemporary models. The Classical Economic Development Theories started with the Linear Stage Theory, inspired by the Marshall Plan to reconstruct Europe after WWII, which conceptualises five stages that a developing country must pass through to attain a developed status; the traditional society, preconditions for take-off, take-off, drive to maturity, and the age of high mass consumption. Other classical theories include; the Structural Change theory popularised by the Dual Sector Model of Lewis (1954) and the Patterns of Development Model by Chenery (1960), the International Dependence Revolution Theory, and the false-paradigm model, and the Neo-classical growth theories featuring the Solow-Swan Growth theory.

Following the criticisms of the classical economic development theories, contemporary schools of thought emerged to explain the diverse development experiences worldwide. These theories recognise that economic development requires systematic effort, and better development strategies can be originated by understanding the underlying processes that spur development. Some of these theories include; Underdevelopment as a Coordination Failure and the Big Push Theory.

The classical and contemporary economic development theories enumerated above help clarify our understanding of development as a multi-dimensional issue. The literature review leads us to the macroeconomic questions appropriate to this study's purpose, questions which Mankiw (2004) also proposed; What factors explain the diverse growth experiences between developed and developing countries? What policies should the developing countries pursue to promote more rapid growth? The answers to such questions may be found by investigating the underlying differences between developing, transitioning, and developed economies as classified by the 2019 and 2022 World Economic Situation and Prospect (WESP) report, in this case, using machine learning models.

2.1 Understanding Machine Learning Models

Machine learning is a subset of Artificial Intelligence applied to computational problems to mimic human cognitive abilities of pattern recognition. The two major subsections of ML are Supervised and Unsupervised ML. Supervised ML is synonymous with traditional statistical and econometric estimations where a model is of the form:

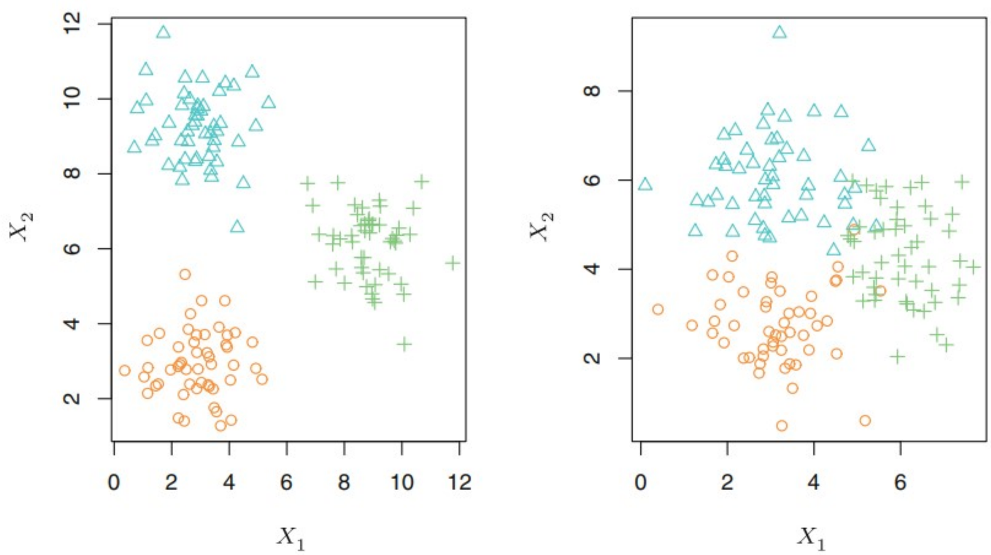
Equation 1: Example of a Typical Regression Estimation.

$$Y_i = \alpha + \beta X_i + \varepsilon$$

For the model above, a statistical or econometric model attempts to fit a model that relates the independent variable(s) to the dependent variable to accurately predict the dependent variable given future observation points (prediction) or better understand the relationship between the independent and dependent variables (inference). A supervised Machine Learning model also follows this estimation pattern.

On the other hand, an unsupervised ML model is a situation wherein no dependent variable (Y_i) is in the estimation procedure. We observe a vector of measurements X_i but no associated response Y_i. Such a situation is unsupervised because there is no dependent variable to supervise the estimation of a model. Thus, unsupervised ML models attempt to fit a relationship by finding clusters within the data. For example, in fig. 1 below, given two variables X_1 and X_2 plotted on a graph, an unsupervised ML model's objective is to separate the points into clusters based on how close they are to each other. In the left panel of fig 1 below, we expect our model to cluster the blue, green, and orange observations in three distinct clusters. In the right panel, it becomes more challenging for a model to cluster given overlaps.

Figure 2: A clustering data set involving three groups (culled from James et al., 2013)



In this work, for the unsupervised ML approach, we use the KMeans Model and the Gaussian Mixture Model to create clusters without a dependent variable. The K-means algorithm identifies k number of centroids and then allocates every data point to the nearest cluster while keeping the centroids as small as possible (Garbade, 2018). On the other hand, the Gaussian mixture model is a "probabilistic model for representing normally distributed subpopulations within an overall population. Mixture models do not require knowing which subpopulation a data point belongs to, allowing the model to learn the subpopulations automatically" (McGonagle, nd). The major difference between these two models is that while KMeans uses the distance from an observed centroid mean to cluster the observations, the Gaussian mixture model uses a probabilistic function to assign observations to clusters.

For the supervised ML, three models are employed, namely: Random Forest Classifier (RFC), Support Vector Machine Classifier (SVMC), and XGB Classifier (XGBC). It would also suffice to explain why the supervised ML models are classifier models, not regression ones. Variables can be characterised as either quantitative or qualitative. While quantitative variables take on numerical values such as GDP, qualitative variables take on values in one of K different classes or categories. Given that we are trying to fit a supervised model of the following form:

Equation 2: Supervised ML Classifier Model Form

$$DS = \alpha + \beta_1 E_{it} + \beta_2 S_{it} + \beta_3 I_{it} + \beta_4 G_{it}$$

Where DS is the Development Status of a nation in our dataset which can be one of the following three classes: Developing, Transitioning, and Developed, the proper model to be employed is a classification model. E_{it} = Economic Indicators of country i at time t, S_{it} = Social Indicators of country i at time t, I_{it} = Industrial Indicators of country i at time t, G_{it} = Governance Indicators of country i at time t.

RFC consists of many individual decision trees that operate as an ensemble. A decision tree is a map of the possible outcomes of a series of related choices (in our case, the development status of economies). RFC works by training multiple decision trees. Each tree in the random forest produces a class prediction, and the class with the most votes becomes our model's prediction (Yiu, 2021). A decision tree is built using each independent variable in the model and continuously iterating to adjust the branch in each tree. For example, at each node/ branch in a tree, the criteria could be that GDP per capita ≥ $30,000 is most related to class 2, representing developed economies. On the other hand, GDP per capita ≤ $30,000 is related to developing and transitioning economies. Another branch is further extended to separate developing and transitioning economies with GDP per capita ≤ the criterion value. When such a branch is built for each independent variable in the model, the decision tree is then able to classify countries based on the output of the nodes in the tree.

The SVMC aims to find a hyperplane in N-dimensional space (N - the number of independent variables) that distinctly classifies the data points. The SVMC model can produce highly accurate results without requiring extensive computation. It can also compute the results by assuming either a linear or non-linear relationship between the dependent and independent variables.

The third supervised ML model is the XGB Classifier which stands for Extreme Gradient Boosting. The XGBC model is very similar to the RFC model because it also employs decision trees to reach a class outcome. XGBC iteratively trains an ensemble of shallow decision trees, with each iteration using the error residuals of the previous model to fit the next model. A significant difference between the XGBC and RFC models is that while the latter builds trees parallel to each other and then selects the majority class output from each of the trees as the predicted class, the former builds trees sequentially with only one class output from the tree (boosting ensemble).

3.0 Methodology

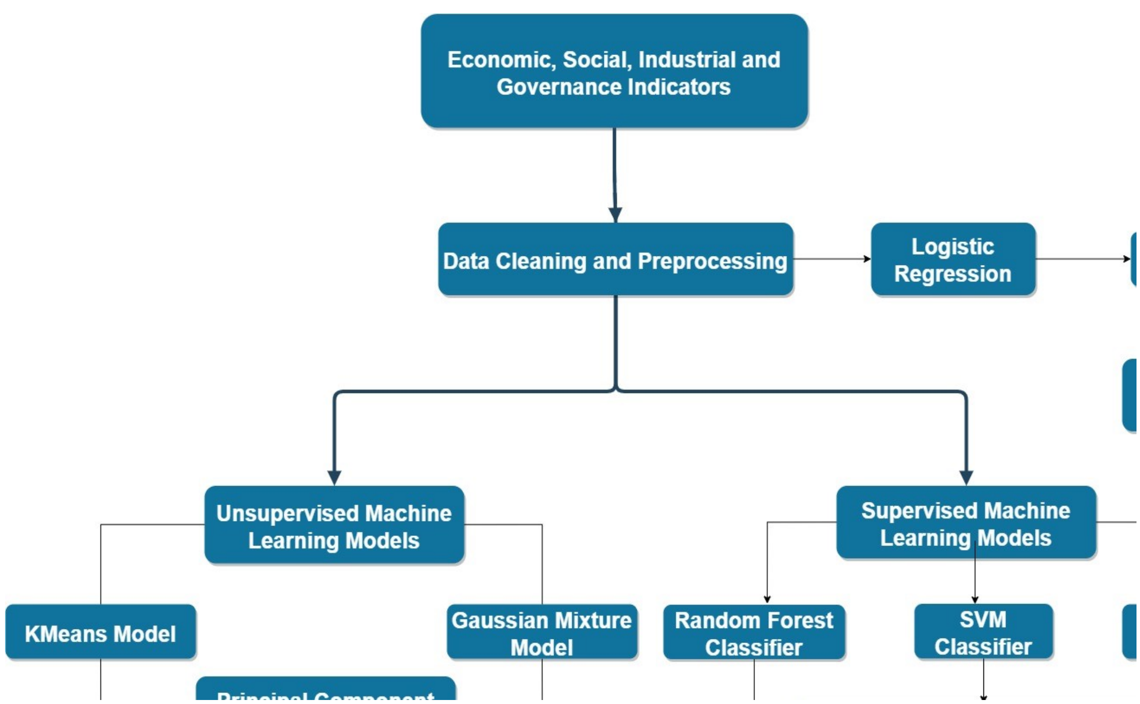
This study is conducted using a machine learning approach to research and logistic regression for comparison and explanation purposes. A logistic panel model is employed because the dependent variable in our supervised model has two classes; Developed Economies vs Developing/ Transitioning Economies.

The python programming language and relevant modules and libraries are used to execute the methodology proposed below. Also, the STATA application is used to run the econometric model, given that the data used is of the panel data form.

In this study, and based on previous development theories and research, we will employ four economic indicator groups to proxy a nation's economic development: Economic Indicators, Social Indicators, Industrial Indicators, and Governance Indicators. Table 1 in the appendices summarises the indicators (variables) employed in the ML and econometric models.

3.1 Analytical Framework

Figure 3: Research Analytical Framework Flowchart



Per the diagram above, at inception, the first stage entails accessing the relevant dataset from the World Bank database. This is followed by extensive data cleaning and preprocessing procedures for the accessed dataset using best practices in machine learning programming. These best practices are discussed below.

After cleaning the dataset, the second stage entails fitting two unsupervised machine learning models: the KMeans model and the Gaussian Mixture model, for the most recent years in our dataset. The unsupervised machine learning creates clusters based on the independent variables supplied (there is no dependent variable). After the unsupervised ML models produce clusters for the various countries in our dataset for a specific year, the Principle Component Analysis (PCA) is implemented to reduce the dimensionality of the variables to two variables, enabling us to examine the relationship between clusters visually. The KMeans and Gaussian mixture models are discussed subsequently and serve as comparisons for each other.

The third stage of the analysis involves fitting three supervised ML models, which serve as comparisons for each other. Per the World Bank classification of countries, a dependent variable is created with classes 0, 1, and 2, representing developing, transitioning, and developed economies, respectively. The dataset is split randomly in a 70/30 ratio. The former set is used to train each supervised machine learning model. In contrast, the latter set is used as a validation set to test the performance of each model in classifying each economy correctly. A confusion matrix is developed to compare performance across the three supervised ML models: Random Forest, Support Vector Model Classifier, and XGB Classifier. Subsequently, each model produces a variable importance listing which represents the relative importance of each independent variable to the model's ability to classify an economy correctly, representing the unique differences between economies. An out-of-sample test is also implemented for test countries.

The final stage of this study will involve implementing a logistic regression comparable to the supervised machine learning models specified above. The STATA software will be employed using the same variables as the supervised ML models. The top variables recommended by the ML models for classification will be used to fit the econometric model. This helps the researcher solve the independent variable selection conundrum. The Average Marginal Effects Output of the model will be used to interpret the logistic model.

3.2 Data Acquisition, Cleaning and Preprocessing

The dataset used in this study was accessed using the Python API[8] for the World Development Indicators (WDI). At initial access, data is available for 217 countries from 1960 to 2020.

A significant issue with the data from this source is missing variables. Please see Figure 1 in the appendices for the percentage of missing variables after the initial data pull. To maintain the originality of the data used, we drop any year for any economy with 50% of the observations missing. Subsequently, any economy that does not have at least 50% of the 60 years (1960 - 2020) is dropped entirely from the dataset. Please see Figure 2 in the appendices for the percentage of missing variables after the dropping years and countries with excessive missing data.

To account for the missing variables in the resulting dataset above, we implement the kNN imputation technique in ML. Brownlee (2020) refers to this approach as a best practice that predicts missing values in a dataset. In our case, for each country in the dataset and for each independent variable, a model is created which uses the three closest data points to a missing value to predict it. Figure 3 in the appendices presents a dataset summary after implementing the kNN imputation technique.

We perform feature engineering to create three variables: trade openness (import + export/ GDP), CO2 emission per capita (CO2 emission (kt)/ Total population), and Air Transportation per capita (Air Transportation (number of passengers)/ Total population).

We then proceed to standardise all the variables in the dataset. Standardising a dataset involves rescaling the distribution of values so that the mean of observed values is 0 and the standard deviation is 1. This process is essential primarily when the different variables used in ML models have different scales (Brownlee, 2016). This prevents variables that have higher orders from biasing the model. Tables 2 and 3 show a descriptive summary of the dataset before and after standardisation.

3.3 Performance Evaluation

A confusion matrix is a table used to define the performance of a classification model, and it summarises the prediction results of a machine learning model on the test dataset. Our three

[8] An Application Programming Interface (API) is a set of protocols used by software services to communicate with each other.

supervised ML models will be trained on 70% of the dataset, while the remaining test data will be used to test the performance of the models.

By comparing a supervised ML model's predicted class (outcome) to the actual class, four parameters are possible: True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN). With these parameters, four measures of performance can be calculated as follows:

Equation 3: Model Accuracy Metric

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN} + \mathrm{TN}}$$

Equation 4: Model Precision Metric

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$

Equation 5: Model Recall Metric

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

Equation 6: Model F1 Score Metric

$$F1\ Score = \frac{2 * (\mathrm{Recall} * \mathrm{Precision})}{(\mathrm{Recall} + \mathrm{Precision})}$$

Accuracy is the most intuitive performance measure. It is simply a ratio of correctly predicted observations to total observations. Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. The recall is the ratio of correctly predicted positive observations to all observations in the actual class. F1 score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. "Intuitively, it is not as easy to understand as accuracy, but F1 is usually more useful than accuracy, especially if you have an uneven class distribution" (Joshi, 2016). The confusion matrix is a presentation of all of these measures. We use this to measure both the performance of the models on the test data and against each other.

3.4 Econometric Model

The econometric model that will be fitted in this study will be a Panel Logistic Regression. This model fits our needs because it allows for a binary category-dependent variable. To present a simplified and explainable result, we will assume that developing and transitioning economies make up one class, whereas developed economies make up the other class. By fitting this traditional econometric model, we can investigate the relationship between an economy's development status and the economic, social, industrial, and governance indicators. This method is considered an attractive analysis because it does not make assumptions for normality, linearity, and homoscedasticity, but it does respond poorly to multicollinearity among the independent variables. Thus, we will estimate a multicollinearity test and drop highly correlated variables.

Since we will fit a logistic regression model on panel data with binary dependent variables, the Random Effect Model version will be employed. The Random Effects regression model estimates the effect of individual-specific characteristics that are inherently unmeasurable. This is appropriate for our model, given that our dataset has unobservable differences between the 147 countries. A random effect model allows each unique group in the dependent variable to have a regression fitted to its relationship. It also generalises the inferences that can be made beyond the sample used in the model. By estimating the Marginal Effect (ME) of the logistic regression coefficient, we can interpret the expected increase in the likelihood for a country to be placed as developing/transitioning or developed for a unit change in each independent variable.

FE models are used when one is interested in analysing the impact of variables that vary over time. FE explores the relationship between predictor and outcome variables within an entity (in our case, countries). The FE model controls for time-invariant differences between our countries, so coefficients cannot be biased because of time-invariant characteristics such as culture, race, and religion. The RE model, on the other hand, assumes that the variations between countries are random and uncorrelated with the independent variables in the model, an appropriate assumption for the dataset employed. Based on the above discussion, we will fit the following logistic model:

Equation 7: Logistic Regression for Developed vs Developing/ Transition Economies

$$\Pr[DS] = \alpha + \beta_1 E_{it} + \beta_2 S_{it} + \beta_3 I_{it} + \beta_4 G_{it} + \varepsilon_{it}$$

Where:

$\Pr[DS]$ = Probability of Outcome [Development Status (0,1)]

E_{it} = Economic Indicators of country i at time t.

S_{it} = Social Indicators of country i at time t.

I_{it} = Industrial Indicators of country i at time t.

G_{it} = Governance Indicators of country i at time t.

ε_{it} = Between-country error

4.0 Presentation of Results

Following the research methodology outlined above, the following are the results of the analysis performed. These results will be presented in three sections; Unsupervised Machine Learning Results, Supervised Machine Learning Results, and Econometric Regression Results. The explanation/ discussion of the findings will follow in the form of a blog article.

4.1 Unsupervised Machine Learning Results

The first objective of this paper was to use unsupervised ML to create a cluster for similar countries. This helps the reader understand the development stages of countries based on similar groups. We fitted two of such models: KMeans and Gaussian Mixture Models, for two years, 2000 and 2020. By examining these clusters, the reader should be able to identify significant changes that have taken place in the past decade in terms of similarity of economies globally. Figures 3 and 4 below present a graphical illustration of the clusters which both models generated for the years 2000 and 2020. It is critical to note that while the clusters produced for both models were precisely the same for 2020, there were few variations for the year 2000. The output of the KMeans model is thus presented below with a footnote about the variations for the year 2000.

The model results from fig 3 cluster the economy of the USA and Germany as being the most similar in 2000, grouped as cluster E. Western Europe, Australia, Japan, and Canada were grouped as most similar in cluster B. On the other hand, most of Eastern Europe, South America, and Southern Africa were grouped in Cluster D. Cluster C grouped most of the economies in Western, Central, and Eastern Africa.

By comparing the results in figure 4, for 2020, some divergence can be seen. While the economic development metrics of North America, Western Europe, and Japan (excluding the United Kingdom and Ireland) seemed to converge, and the economies in Eastern Europe and South America remained similar, there was a bit of divergence in the countries previously clustered together in Western, Central and Eastern Africa. These results will be discussed subsequently.

Figure 4: Year 2000 Unsupervised ML Cluster

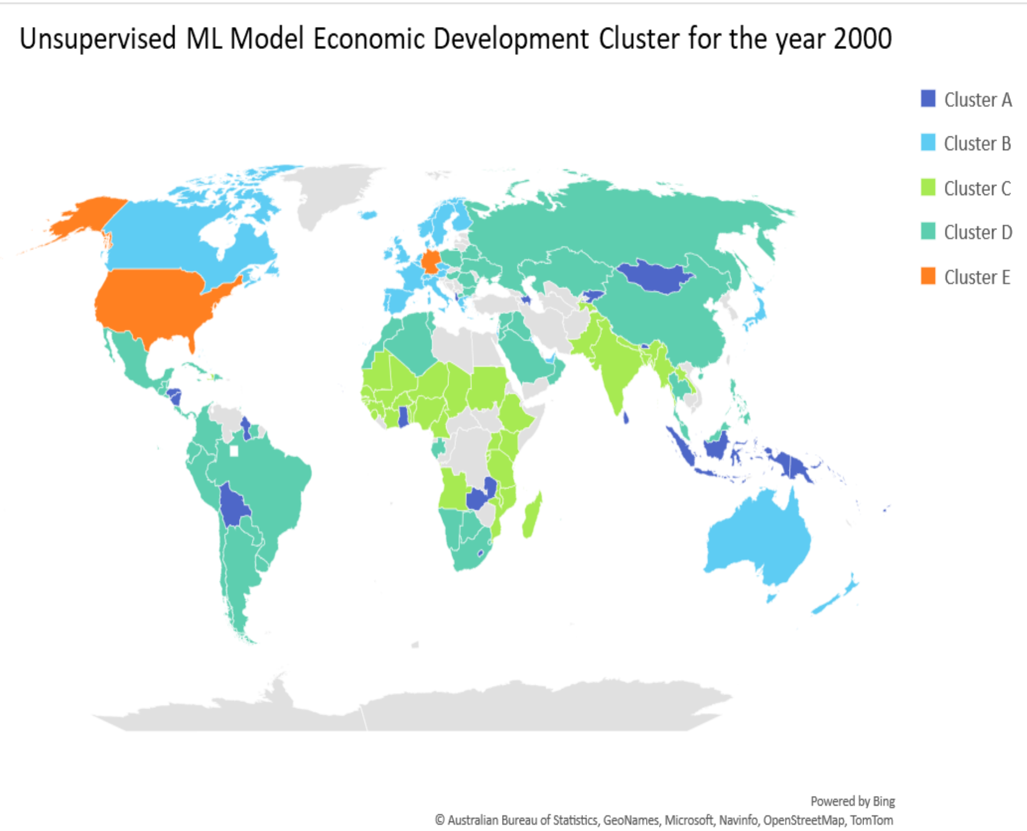
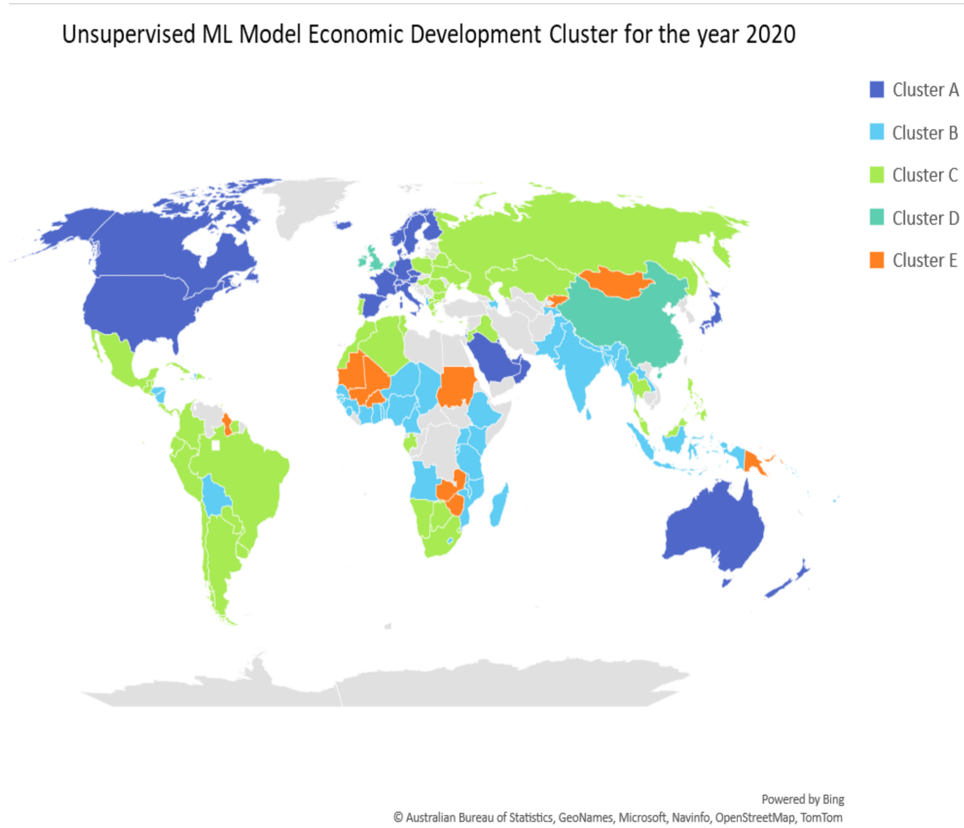


Figure 5: Year 2020 Unsupervised ML Cluster



Difference between Year 2000 GMM and KMeans models[9]

[9] The output of the Gaussian Mixture Model (GMM) clustered Oman and Saudi Arabia in Cluster B instead of Cluster D as shown above; clustered Bangledesh, India and Tajikistan in Cluster A instead of Cluster C as shown above; clustered The Solomon Islands in Cluster C, instead of Cluster A as shown above.

4.2 Supervised Machine Learning Results (Performance Evaluation & Out-of-Sample Testing)

The second aim of this study was to fit several supervised ML models and provide a breakdown of each model's performance and a list showing the importance of each independent variable to the ML model's attempt to classify economies. The results of these analyses are presented below.

Tables 1 to 3 below present the result of the performance evaluation tests carried out on each model.

Table 1: RFC Confusion Matrix

Model: Random Forest Classifier

Metric/ Class	Developing Economies	Transitioning Economies	Developed Economies
Precision	100%	100%	100%
Recall	100%	99%	100%
F1-Score	100%	99%	100%
Accuracy: 99.5%			

Table 2: SVMC Confusion Matrix

Model: Support Vector Model Classifier

Metric/ Class	Developing Economies	Transitioning Economies	Developed Economies
Precision	99%	71%	90%
Recall	95%	98%	96%
F1-Score	97%	82%	93%
Accuracy: 95.4%			

Table 3: XGBC Confusion Matrix

Model: XGBoost Classifier

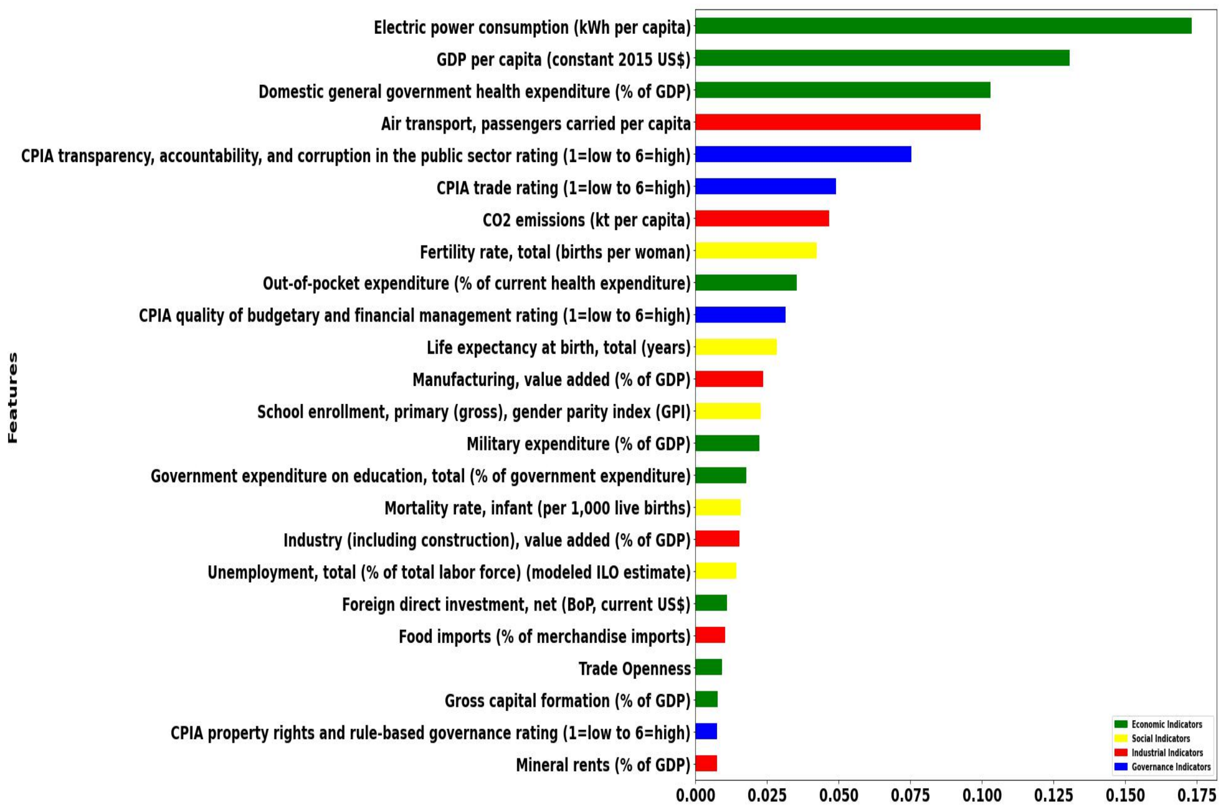
Metric/ Class	Developing Economies	Transitioning Economies	Developed Economies
Precision	100%	99%	100%
Recall	100%	100%	100%
F1-Score	100%	99%	100%
Accuracy: 99.9%			

To know how important each variable was to each of the three supervised ML models trained in this study, we output the feature importance presented in Table 7 in the appendices. Since each model had sufficiently high performance in terms of accuracy, even with a 95% confidence interval, we can take a simple average of the feature importance across the tree models, resulting in figure 5 below.

The XGBC model had the highest accuracy of the three models. Thus, it was used to perform the out-of-sample testing. Out-of-sample testing implies that a country is wholly excluded from a model's fitting. The model is then used to predict the development status of the country. For India and Nigeria, the model consistently classified them as developing economies. For Canada and France, the model consistently classified them as developed economies. For Albania (WB ranked as a transitioning economy), pre-2008, the model ranked it as developing and subsequently as developed.

Similarly, for China, post-2002, the model predicted its status as developing and subsequently as developed (with 2010 and 2012 returned as developing). Also, the Russian Economy (WB ranked as transitioning) presented mixed results. Post-1999, the economy was ranked as developed. However, from 2000 onwards, the model predicted its development status as transitioning.

Figure 6: Features Importance Listing (Average across RFC, SVMC, and XGBC models)



4.3 Econometric Regression Result

To implement the econometric regression as specified in equation 7 above, a correlation test was first conducted on the top eight most important variables as ranked by the ML output in fig. 5 above. The correlation result is presented in fig. 10 in the appendices. Two governance variables namely, CPIA Transparency and CPIA Trade Rating, were highly correlated and could not be used in the logistic model. Thus, Electricity consumption, GDP per capita, Government Health Expenditure, Air Transportation, Co2 emissions and Fertility rate were used as the independent variables, while the dependent variable was two classes; Developing/ Transitioning Economies vs Developed Economies.

The result of the panel logistic regression is presented in the appendices as fig 9. The coefficients are individually and jointly significant at the 5% confidence interval. Given that the logistic regression coefficient is not directly interpretable, we employ the average marginal effects estimation of the results of the model. The results are presented in figure 11, but summarised in the table below:

Table 4: Logistic Regression Result - Average Marginal Effect (AME)

Variable	Coefficient	P-value
Electric power consumption (kWh per capita)	0.021	0.024
GDP per capita (constant 2015 US$)	0.039	0.000
Domestic general government health expenditure (% of GDP)	0.086	0.000
Air transport, passengers carried per capita	-0.012	0.000
CO2 emissions (kt per capita)	0.016	0.011
Fertility rate, total (births per woman)	-0.049	0.022

As shown in table 4 above, the coefficients of the (AME) are all significant at the 5% level. Also the interpretation of the signs fall in line with apriori expectation. Interpretation are as follows:

- A 1% increase in the electric power consumption of a typical country on average increases the probability that the country is a developed economy by 2.1%.
- A 1% increase in the GDP per capita of a typical country on average increases the probability that the country is a developed economy by 3.9%.
- A 1% increase in the domestic government health expenditure of a typical country on average increases the probability that the country is a developed economy by 8.6%.
- A 1% increase in the air transportation per capita of a typical country on average decreases the probability that the country is a developed economy by 1.2%.
- A 1% increase in the Co2 emission of a typical country on average increases the probability that the country is a developed economy by 1.6%.
- A 1% increase in the fertility rate of a typical country on average decreases the probability that the country is a developed economy by 4.9%.

WORD COUNT: 5,830

References

Angrist, J.D. and Pischke, J.-S. (2015) *Mastering 'metrics: the path from cause to effect*. Princeton ; Oxford: Princeton University Press.

Barossi-Filho, M., Silva, R.G. and Diniz, E.M. (2005) 'The Empirics of the Solow Growth Model: Long-Term Evidence', *Journal of Applied Economics*, 8(1), pp. 31–51. Available at: https://doi.org/10.1080/15140326.2005.12040617.

Basuchoudhary, A., Bang, J.T. and Sen, T. (2017) *Machine-learning Techniques in Economics*. Cham: Springer International Publishing (SpringerBriefs in Economics). Available at: https://doi.org/10.1007/978-3-319-69014-8.

Brown, S. (2021) *Machine learning, explained*, *MIT Sloan*. Available at: https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained (Accessed: 27 May 2022).

Brownlee, J. (2016) 'How to Normalise and Standardise Time Series Data in Python', *Machine Learning Mastery*, 11 December. Available at: https://machinelearningmastery.com/normalize-standardize-time-series-data-python/ (Accessed: 21 June 2022).

Brownlee, J. (2020) 'kNN Imputation for Missing Values in Machine Learning', *Machine Learning Mastery*, 23 June. Available at: https://machinelearningmastery.com/knn-imputation-for-missing-values-in-machine-learning/ (Accessed: 21 June 2022).

Charpentier, A., Flachaire, E. and Ly, A. (2019) 'Econometrics and Machine Learning', *Economie et Statistique / Economics and Statistics*, (505d), pp. 147–169. Available at: https://doi.org/10.24187/ecostat.2018.505d.1970.

Chenery, H.B. (1960) 'Patterns of Industrial Growth', *The American Economic Review*, 50(4), pp. 624–654.

Cogoljević, D. *et al.* (2018) 'A machine learning approach for predicting the relationship between energy resources and economic development', *Physica A: Statistical Mechanics and its Applications*, 495, pp. 211–214. Available at: https://doi.org/10.1016/j.physa.2017.12.082.

Garbade, M. (2018) *Understanding K-means Clustering in Machine Learning*, *Medium*. Available at: https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1 (Accessed: 22 June 2022).

Hall, R.E. and Lieberman, M. (2013) *Economics: principles & applications*. 6th ed. Australia ; Mason, OH: South-Western Cengage Learning.

Hamilton, J.D. and Monteagudo, J. (1998) 'The augmented Solow model and the productivity slowdown', *Journal of Monetary Economics*, 42(3), pp. 495–509. Available at: https://doi.org/10.1016/S0304-3932(98)00036-1.

Harrod, R.F. (1939) 'An Essay in Dynamic Theory', *The Economic Journal*, 49(193), p. 14. Available at: https://doi.org/10.2307/2225181.

Herrera, G.P. *et al.* (2019) 'Long-term forecast of energy commodities price using machine learning', *Energy*, 179, pp. 214–221. Available at: https://doi.org/10.1016/j.energy.2019.04.077.

Jones, D.S. and Jackson, B. (2016) 'The Neoclassical Counter-Revolution and the Ascendancy of Business', in R. Skidelsky and N. Craig (eds) *Who Runs the Economy?* London: Palgrave Macmillan UK, pp. 73–84. Available at: https://doi.org/10.1057/978-1-137-58017-7_6.

Joshi, R. (2016) *Accuracy, Precision, Recall & F1 Score: Interpretation of Performance Measures*, *Exsilio Blog*. Available at: https://blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures/ (Accessed: 22 June 2022).

Lee†., S.C. and Lee, E.T. (1974) 'Fuzzy Sets and Neural Networks', *Journal of Cybernetics*, 4(2), pp. 83–103. Available at: https://doi.org/10.1080/01969727408546068.

Lewis, W.A. (1954) 'Economic Development with Unlimited Supplies of Labour', *The Manchester School*, 22(2), pp. 139–191. Available at: https://doi.org/10.1111/j.1467-9957.1954.tb00021.x.

Maccarrone, G., Morelli, G. and Spadaccini, S. (2021) 'GDP Forecasting: Machine Learning, Linear or Autoregression?', *Frontiers in Artificial Intelligence*, 4, p. 757864. Available at: https://doi.org/10.3389/frai.2021.757864.

Mankiw, N.G. (2004) *Principles of economics*. 3rd ed. Mason, Ohio: Thomson/South-Western.

McGonagle, J. (nd) *Gaussian Mixture Model*. Available at: https://brilliant.org/wiki/gaussian-mixture-model/ (Accessed: 22 June 2022).

Mele, M. and Magazzino, C. (2020) 'A Machine Learning analysis of the relationship among iron and steel industries, air pollution, and economic growth in China', *Journal of Cleaner Production*, 277, p. 123293. Available at: https://doi.org/10.1016/j.jclepro.2020.123293.

Nkalu, C.N., Edeme, R.K. and Chukwuma, Q.O. (2018) 'Testing the Validity of the Solow Growth Model: Empirical Evidence from Cross-Country Panel Data', *American Economic & Social Review*, 3(1), pp. 18–22. Available at: https://doi.org/10.46281/aesr.v3i1.197.

Otchia, C. and Asongu, S. (2021) 'Industrial growth in sub-Saharan Africa: evidence from machine learning with insights from nightlight satellite images', *Journal of Economic Studies*, 48(8), pp. 1421–1441. Available at: https://doi.org/10.1108/JES-05-2020-0201.

Solow, R.M. (1956) 'A Contribution to the Theory of Economic Growth', *The Quarterly Journal of Economics*, 70(1), pp. 65–94. Available at: https://doi.org/10.2307/1884513.

Swan, T.W. (1956) 'Economic Growth and Capital Accumulation', *Economic Record*, 32(2), pp. 334–361. Available at: https://doi.org/10.1111/j.1475-4932.1956.tb00434.x.

Todaro, M.P. and Smith, S.C. (2011) *Economic development*. 11th ed. Harlow Munich: Pearson, Addison Wesley.

United Nations Department for Economic and Social Affairs (2019) *World Economic Situation and Prospects 2019.* United Nations.

United Nations Department for Economic and Social Affairs (2022) *World Economic Situation and Prospects 2022.* S.l.: United Nations.

Wu, C.-F. *et al.* (2020) 'The nexus of financial development and economic growth across major Asian economies: Evidence from bootstrap ARDL testing and machine learning approach', *Journal of Computational and Applied Mathematics*, 372, p. 112660. Available at: https://doi.org/10.1016/j.cam.2019.112660.

Appendices

Table 5: List of Variables to Model Economic Development

Economic Indicators	Industrial Indicators	Social Indicators	Governance Indicators
Trade Openness (Import + Export / GDP)	Industry Value Added (% of GDP)	Unemployment Rate (% of Labour Force)	CPIA transparency, accountability, and corruption in the public sector
Health Expenditure (% of GDP)	Manufacturing, value added (% of GDP)	Fertility Rate	CPIA quality of budgetary and financial management rating
Out-of-pocket expenditure (% of current health expenditure)	Air Transportation (passengers carried per capita)	Infant Mortality Rate	CPIA property rights and rule-based governance rating
Education Expenditure (% of Govt. Exp.)	CO_2 emissions (kt per capita)	Life expectancy at birth	CPIA trade rating
Electricity Consumption per capita	Mineral rents (% of GDP)	School enrollment, gender parity index (GPI)	
Gross Capital Formation (% of GDP)	Food imports (% of merchandise imports)		
Foreign Direct Investment, net (BoP current USD)			
Military expenditure (% of GDP)			
GDP per capita (constant 2015 USD)			

Figure 7: Summary of Missing Variables for Accessed Dataset (217 economies)

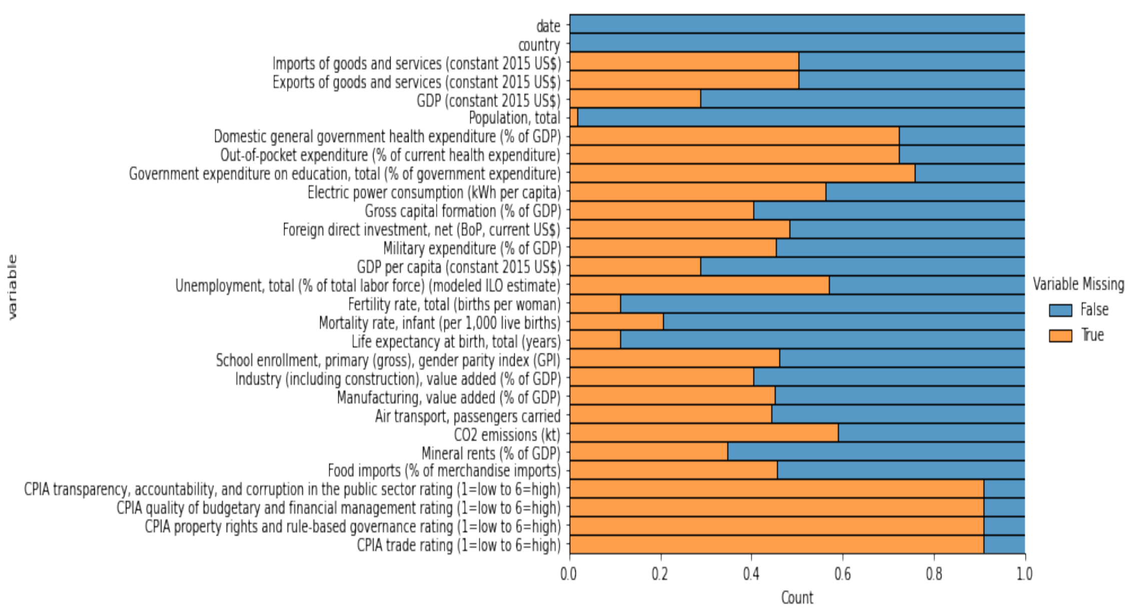


Figure 8: Summary of Missing Variables After Implementing 50% Cutoff (there are 147 resulting economies)

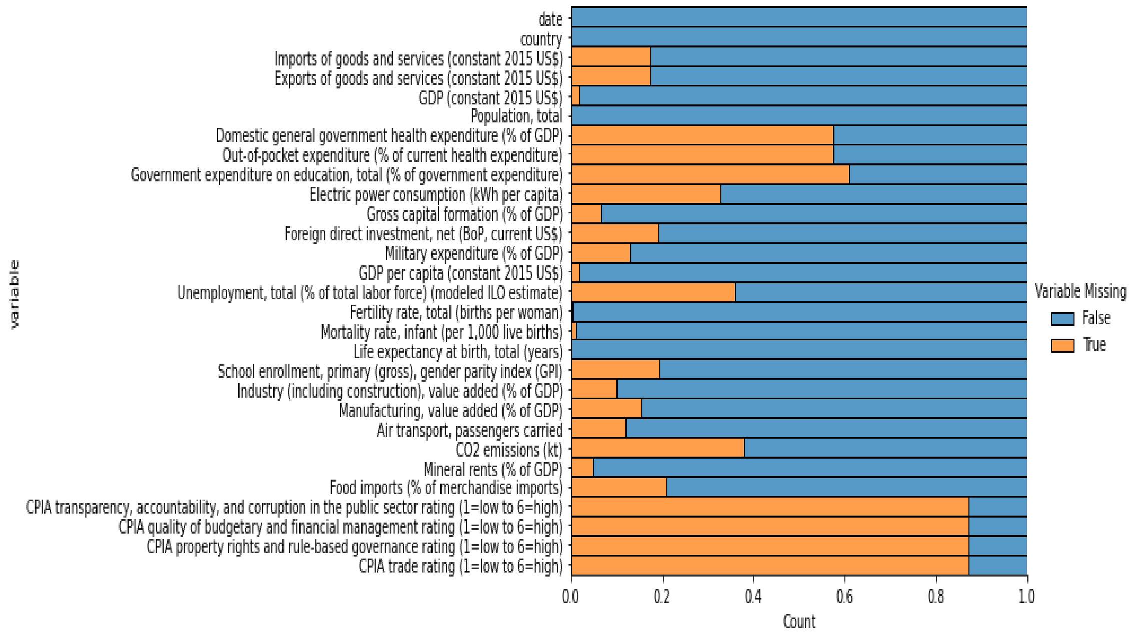


Figure 9: Summary of Missing Variables After Implementing kNN Imputation (147 economies)

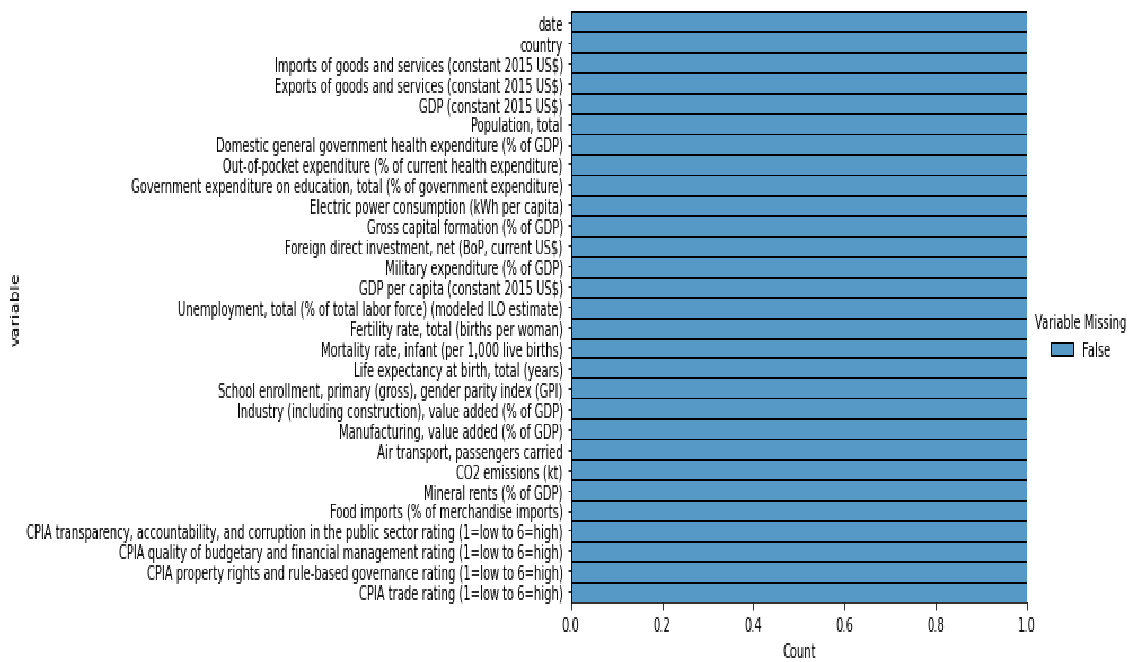


Table 6: Table of Descriptive Statistics for Variables Employed

Variables	count	mean	std	min	25%	50%	75%	max
Trade Openness	6716	0.766	0.797	0	0.34	0.587	0.944	15.055
Domestic general government health expenditure (% of GDP)	6762	3.037	2.11	0	1.331	2.47	4.353	12.063
Out-of-pocket expenditure (% of current health expenditure)	6762	35.098	18.935	0	18.342	33.652	50.555	84.125
Government expenditure on education, total (% of government expenditure)	6762	13.974	5.509	0	10.527	13.829	17.438	44.802
Electric power consumption (kWh per capita)	6762	2666.026	4653.813	0	77.502	755.934	3441.543	54799.18
Gross capital formation (% of GDP)	6762	22.917	9.277	-13	17.951	22.473	27.332	89.381
Foreign direct investment, net (BoP, current US$)	6762	-2.1E+08	1.68E+10	-3.4E+11	-7.9E+08	-9E+07	-495586	2.18E+11
Military expenditure (% of GDP)	6762	2.438	2.96	0	1.063	1.765	2.976	117.35
GDP per capita (constant 2015 US$)	6762	11063.71	16248.76	0	1266.481	3606.824	14327.33	112417.9
Unemployment, total (% of total labor force) (modeled ILO estimate)	6762	7.476	6.101	0	3.46	5.727	9.633	38.8
Fertility rate, total (births per woman)	6762	3.687	1.925	1	1.936	3.168	5.311	8.462
Mortality rate, infant (per 1,000 live births)	6762	43.486	39.532	0	11.3	29.7	67.4	210.9
Life expectancy at birth, total (years)	6762	66.281	10.646	26	58.466	68.957	74.52	85.388
School enrollment, primary (gross), gender parity index (GPI)	6762	0.925	0.146	0	0.912	0.98	0.999	1.481
Industry (including construction), value added (% of GDP)	6762	27.594	11.857	0	19.937	25.979	32.691	84.796
Manufacturing, value added (% of GDP)	6762	13.66	7.1	0	8.63	13.508	18.051	49.879
Air transport, passengers carried per capita	6762	0.592	1.535	0	0.031	0.129	0.545	34.485
CO2 emissions (kt per capita)	6762	0.004	0.006	0	0.001	0.002	0.007	0.07
Mineral rents (% of GDP)	6762	0.808	2.418	0	0	0.019	0.335	36.853
Food imports (% of merchandise imports)	6762	14.508	7.892	0	8.854	13.298	18.268	62.416
CPIA transparency, accountability, and corruption in the public sector rating (1=low to 6=high)	6762	1.104	1.43	0	0	0	2.5	4.5
CPIA quality of budgetary and financial management rating (1=low to 6=high)	6762	1.257	1.618	0	0	0	3	4.5
CPIA property rights and rule-based governance rating (1=low to 6=high)	6762	1.118	1.444	0	0	0	2.5	4
CPIA trade rating (1=low to 6=high)	6762	1.484	1.891	0	0	0	3.662	5

Table 7: Table of Descriptive Statistics for Variables Employed After Standardisation

Variables	mean	std	min	25%	50%	75%	max
Trade Openness	0	1	-0.961	-0.535	-0.225	0.222	17.92
Domestic general government health expenditure (% of GDP)	0	1	-1.439	-0.808	-0.269	0.624	4.277
Out-of-pocket expenditure (% of current health expenditure)	0	1	-1.854	-0.885	-0.076	0.816	2.589
Government expenditure on education, total (% of government expenditure)	0	1	-2.537	-0.626	-0.026	0.629	5.596
Electric power consumption (kWh per capita)	0	1	-0.573	-0.556	-0.41	0.167	11.203
Gross capital formation (% of GDP)	0	1	-3.916	-0.535	-0.048	0.476	7.165
Foreign direct investment, net (BoP, current US$)	0	1	-20.468	-0.034	0.007	0.013	12.976
Military expenditure (% of GDP)	0	1	-0.824	-0.465	-0.227	0.182	38.821
GDP per capita (constant 2015 US$)	0	1	-0.681	-0.603	-0.459	0.201	6.238
Unemployment, total (% of total labor force) (modeled ILO estimate)	0	1	-1.226	-0.658	-0.287	0.354	5.135
Fertility rate, total (births per woman)	0	1	-1.481	-0.91	-0.27	0.844	2.481
Mortality rate, infant (per 1,000 live births)	0	1	-1.1	-0.814	-0.349	0.605	4.235
Life expectancy at birth, total (years)	0	1	-3.768	-0.734	0.251	0.774	1.795
School enrollment, primary (gross), gender parity index (GPI)	0	1	-6.32	-0.084	0.378	0.506	3.801
Industry (including construction), value added (% of GDP)	0	1	-2.327	-0.646	-0.136	0.43	4.825
Manufacturing, value added (% of GDP)	0	1	-1.924	-0.708	-0.021	0.619	5.101
Air transport, passengers carried per capita	0	1	-0.386	-0.366	-0.302	-0.031	22.078
CO2 emissions (kt per capita)	0	1	-0.782	-0.68	-0.415	0.421	11.9
Mineral rents (% of GDP)	0	1	-0.334	-0.334	-0.326	-0.195	14.905
Food imports (% of merchandise imports)	0	1	-1.838	-0.716	-0.153	0.476	6.071
CPIA transparency, accountability, and corruption in the public sector rating (1=low to 6=high)	0	1	-0.772	-0.772	-0.772	0.976	2.375
CPIA quality of budgetary and financial management rating (1=low to 6=high)	0	1	-0.777	-0.777	-0.777	1.078	2.005
CPIA property rights and rule-based governance rating (1=low to 6=high)	0	1	-0.775	-0.775	-0.775	0.957	1.996
CPIA trade rating (1=low to 6=high)	0	1	-0.785	-0.785	-0.785	1.151	1.859

Table 8: Feature Importance Listing for Each Model, along with an Average.

Feature	RFC	SVMC	XGBC	Average
Electric power consumption (kWh per capita)	0.16705	0.09453	0.2583	0.17329
GDP per capita (constant 2015 US$)	0.12001	0.17941	0.09242	0.13061
Domestic general government health expenditure (% of GDP)	0.12519	0.01339	0.17038	0.10299
Air transport, passengers carried per capita	0.07099	0.2082	0.01901	0.0994
CPIA transparency, accountability, and corruption in the public sector rating (1=low to 6=high)	0.01491	0.10776	0.10329	0.07532
CPIA trade rating (1=low to 6=high)	0.01511	0.07342	0.05892	0.04915
CO2 emissions (kt per capita)	0.05272	0.04965	0.03736	0.04657
Fertility rate, total (births per woman)	0.06631	0.03575	0.02486	0.04231
Out-of-pocket expenditure (% of current health expenditure)	0.05457	0.03103	0.02018	0.03526
CPIA quality of budgetary and financial management rating (1=low to 6=high)	0.01316	0.04486	0.03614	0.03139
Life expectancy at birth, total (years)	0.05913	0.01666	0.00923	0.02834
Manufacturing, value added (% of GDP)	0.0285	0.00983	0.03258	0.02364
School enrollment, primary (gross), gender parity index (GPI)	0.02701	0.03319	0.00786	0.02269
Military expenditure (% of GDP)	0.01608	0.0268	0.02389	0.02226
Government expenditure on education, total (% of government expenditure)	0.02614	0.00964	0.01781	0.01786
Mortality rate, infant (per 1,000 live births)	0.02729	0.01872	0.0015	0.01584
Industry (including construction), value added (% of GDP)	0.02041	0.00666	0.01928	0.01545
Unemployment, total (% of total labor force) (modeled ILO estimate)	0.02138	0.00358	0.01817	0.01437
Foreign direct investment, net (BoP, current US$)	0.01194	0.01189	0.00899	0.01094
Food imports (% of merchandise imports)	0.01786	0.00349	0.00959	0.01031
Trade Openness	0.01625	0.0018	0.00953	0.00919
Gross capital formation (% of GDP)	0.00516	0.00449	0.01335	0.00766
CPIA property rights and rule-based governance rating (1=low to 6=high)	0.0108	0.01196	0	0.00759
Mineral rents (% of GDP)	0.01203	0.00327	0.00738	0.00756
Summation	**1**	**1**	**1**	**1**

Figure 10: Panel Logistic Regression STATA Output

```
Random-effects logistic regression              Number of obs     =      6,736
Group variable: Country                         Number of groups  =        147

Random effects u_i ~ Gaussian                   Obs per group:
                                                              min =         30
                                                              avg =       45.8
                                                              max =         61

Integration method: mvaghermite                 Integration pts.  =         12

                                                Wald chi2(6)      =     319.79
Log likelihood  = -51.971054                    Prob > chi2       =     0.0000

------------------------------------------------------------------------------------
        Development_Status |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------------------+--------------------------------------------------------
electricpowerconsumptionkwhperca |   4.595168   2.204968     2.08   0.037     .2735101    8.916825
    gdppercapitaconstant2015us |   8.566262    1.45887     5.87   0.000     5.706929     11.4256
domesticgeneralgovernmenthealthe |   18.60471   1.282848    14.50   0.000     16.09037    21.11905
airtransportpassengerscarriedper |  -2.506121   .5166999    -4.85   0.000    -3.518834   -1.493407
      co2emissionsktpercapita |   3.393295   1.560667     2.17   0.030     .3344438    6.452146
fertilityratetotalbirthsperwoman |  -10.68116    5.24066    -2.04   0.042    -20.95266    -.409653
                         _cons |  -44.00117   5.783935    -7.61   0.000    -55.33748   -32.66487
---------------------------+--------------------------------------------------------
                      /lnsig2u |   6.693163   .2921177                      6.120623    7.265703
---------------------------+--------------------------------------------------------
                       sigma_u |   28.40546   4.148869                       21.3342    37.82051
                           rho |   .9959392   .0011814                      .9928237    .9977053
------------------------------------------------------------------------------------
LR test of rho=0: chibar2(01) = 1518.32                Prob >= chibar2 = 0.000
```

Figure 11: Variables Correlation Matrix STATA Output

```
. correlate electricpowerconsumptionkwhperca gdppercapitaconstant2015us domesticgeneralgovernmenthealthe airtransportpassengersca
> rriedper co2emissionsktpercapita fertilityratetotalbirthsperwoman cpiatransparencyaccountabilityan cpiatraderating1lowto6high
(obs=6,736)

             | electr~a gdpper~s domest~e airtra~r co2emi~a fertil~n cpiatr~n cpiatr~h
-------------+------------------------------------------------------------------------
electricpo~a |   1.0000
gdppercapi~s |   0.7554   1.0000
domesticge~e |   0.5215   0.5680   1.0000
airtranspo~r |   0.4773   0.5237   0.2609   1.0000
co2emissio~a |   0.6242   0.6962   0.4128   0.2982   1.0000
fertilityr~n |  -0.4798  -0.5205  -0.5607  -0.2988  -0.4527   1.0000
cpiatransp~n |  -0.4113  -0.4614  -0.4634  -0.2408  -0.5166   0.5444   1.0000
cpiatrader~h |  -0.4145  -0.4701  -0.5068  -0.2509  -0.5234   0.5649   0.9661   1.0000
```

Figure 12: Average Marginal Effect STATA Output

```
. margins, dydx(*)

Average marginal effects                          Number of obs     =      6,736
Model VCE    : OIM

Expression   : Pr(Development_Status=1), predict(pr)
dy/dx w.r.t. : electricpowerconsumptionkwhperca gdppercapitaconstant2015us domesticgeneralgovernmenthealthe
               airtransportpassengerscarriedper co2emissionsktpercapita fertilityratetotalbirthsperwoman

------------------------------------------------------------------------------------------------
                                 |            Delta-method
                                 |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------------------------+--------------------------------------------------------------
electricpowerconsumptionkwhperca |   .0211305    .0093311     2.26   0.024     .0028418    .0394192
      gdppercapitaconstant2015us |   .0393912    .0104445     3.77   0.000     .0189203    .0598621
domesticgeneralgovernmenthealthe |   .0855521    .0129936     6.58   0.000     .0600852    .1110191
airtransportpassengerscarriedper |  -.0115242    .0029977    -3.84   0.000    -.0173995   -.0056488
         co2emissionsktpercapita |   .0156038    .0061581     2.53   0.011     .0035341    .0276734
fertilityratetotalbirthsperwoman |  -.0491164    .0215049    -2.28   0.022    -.0912651   -.0069676
------------------------------------------------------------------------------------------------
```