Machine Learning for Policy Guidance

ABSTRACT

This paper leverages machine learning algorithms and techniques to create models that can assist in a country's policy guidance. The machine learning process used to conduct research is discussed with steps such as preprocessing, feature selection, model selection, and model interpretation. Specifically, using datasets from the CIA's World Factbook and the United Nations' Human Development Index (HDI), machine learning models are created that use select features from several counties (e.g., real gross domestic product (GDP), population, and area). Then, the models make predictions on the countries' HDI scores. Model interpretation methods are used to find the most important features in predicting a country's score. This paper argues that important features can be derived through machine learning and guide government policy relevant to human development. Supply-side policies are discussed based on the results from the machine learning models. The use of machine learning with other indexes is also explored.

Introduction

Machine learning is a field where humans try to get computers to learn how to perform specific tasks instead of following detailed instructions given by humans. For example, if humans programmed responses to every possible sentence that could ever be curated, then the computer would be able to converse with humans. However, having a computer learn how to converse is much more efficient than giving a rule for every scenario. That is machine learning. Machine learning can even surpass the capabilities of humans in specific tasks. For instance, computers have shown to be as good or better than humans at recognizing lung cancer in computerized tomography (CT) scans (a type of X-ray image) (Ardila et al., 2019) and have shown better video game performance than professional Esports competitors (Vinyals et al., 2019). With such evidence of ability, machine learning stands poised to make significant changes to society in the future.

The type of machine learning that is employed is supervised learning. Supervised learning entails using data to train a machine learning model and then using the error between the predictions and the ground truth to improve the model. Like how a baby learns new words from the parent's speech, machine learning models learn to make predictions using data given by humans. In supervised learning, a machine learning model takes in inputs and outputs. The inputs can be considered context, and the outputs the answer. The model then uses the answers to see how wrong its predictions were and learns to get better answers.

The inputs used are from the CIA's World Factbook, and the outputs are from the UN's HDI. The World Factbook has data comparing countries based on specific statistics (Central Intelligence Agency, 2022). These statistics can be combined to become the input for the machine learning models. On the other hand, the UN's HDI is a metric for human development in countries where the UN measures three goals: a long and healthy life for citizens, functional and easy access to education, and a decent standard of living (United Nations, 2020). Thus, the UN gives each country an HDI score based on its achievements in human development.

This paper aims to show that by using an index (e.g., HDI, World Happiness Report, and Happy Planet Index) and the data from the World Factbook, machine learning can help guide policymakers by revealing the most relevant statistics in improving a country's index score.

Dataset

As mentioned before, this paper used the World Factbook and the HDI data. Data was downloaded from websites run by the CIA and UN. This data was read and combined into a Python coding environment (Python is the choice programming language for machine learning) (Elliott, 2019) using a popular data processing library called Pandas (McKinney, 2010). The World Factbook separated its statistical data into several downloads. Thus, Pandas was used to combine the data into a single table (see Table 1).

Table 1. Combined data from the World Factbook and HDI; 180 rows x 55 columns

Name	Region	Reserves	…	HDI
China	East and Southeast Asia	$3,236,000,000,000	…	0.761
Saudi Arabia	Middle East	$496,400,000,000	…	0.854
…	…	…	…	…
Andorra	Europe	NaN[1]	…	.868

[1]NaN stands for Not a Number in computer science and represents missing values in the data

Features

In machine learning, features are properties of the data. For instance, each column other than the name and HDI columns in Table 1 is considered a feature. What region a country belongs to and how many reserves they have are features. The features are all part of the input data from the World Factbook and will be described based on clusters of features defined by the World Factbook.

Geography

There is only one feature in the geography cluster: area. The area of each country is in square kilometers.

People and Society

The people and society cluster has features that help describe the living experience of the country's citizens. For example, some of the features in the cluster include life expectancy at birth, HIV/AIDS deaths, education expenditures, and tobacco use. This cluster helps show the general conditions and societal standards of the citizens in the country.

Environment

The environment cluster has two features relating to the revenue the relevant country gains from the natural resources in its environment: forest resources and coal.

Economy

The economy cluster has several economic indicators. This cluster includes major economic indicators such as real GDP, the inflation rate, and the unemployment rate. Other economic indicators provide more context about the country's economic state: Gini Index coefficient (which helps measure inequality in an economy), real GDP growth rate (which helps show whether an economy is growing or shrinking), labor force size, and others.

Energy

The energy cluster includes the usage and trading of energy products. For example, each country has features attributed to petroleum products and carbon emissions.

Communications

The communications cluster includes features that describe the country's communication network. For instance, the number of telephone lines, broadband subscriptions, and internet users are included as features.

Transportation

A country's transportation network is also recorded as features. The number of airports, railways, roadways, waterways, and merchant marine ships are recorded.

Military and Security

Military expenditures is the only feature recorded in the military and security cluster.

Methods

The steps undertaken in this machine learning project are listed as follows.

1. Preprocessing
2. Exploratory data analysis
3. Feature selection
4. Model selection
5. Error analysis
6. Model Interpretation

Preprocessing

The steps taken to prepare data for machine learning are called preprocessing, as many flaws in data can interfere with the performance of models. Some features have too many missing values for machine learning models to get meaningful information. As seen in Table 1, some values are missing or NaN due to some countries not having recorded statistics in the World Factbook (e.g., Andorra does not have their reserves recorded). Thus, certain features were removed from the data if more than thirty percent of the features were missing (see Figure 1) (Dhaduk, 2021).

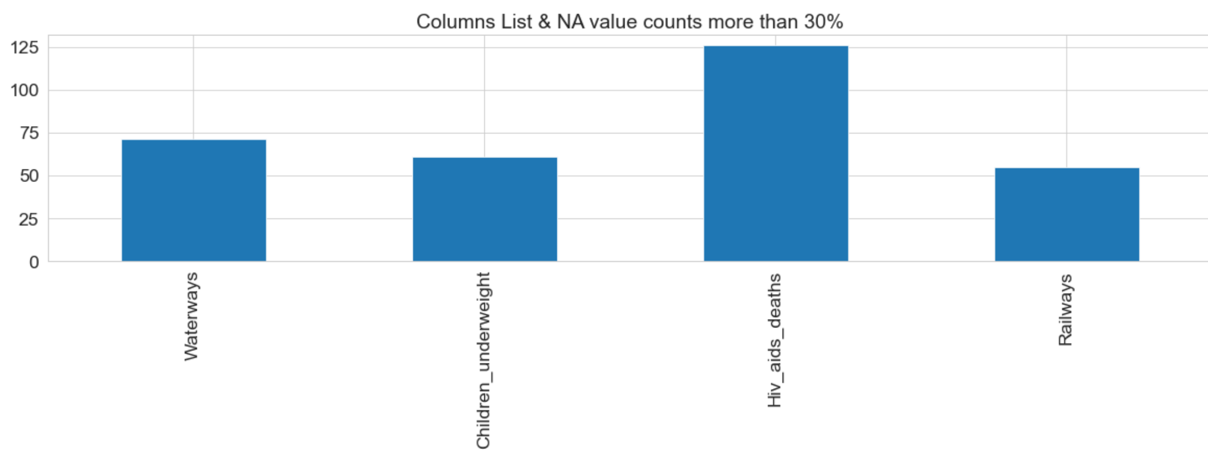

Figure 1. Features that had more than 30% of missing data: waterways, children underweight, HIV/AIDS deaths, and railways. These features were removed from the data.

Next, some features from the World Factbook contained data that machine learning models could not use. Machine learning models require numbers to run and cannot understand words. Features such as the reserves column in Table 1 contained characters like dollar signs and commas that machine learning models could not understand. Thus, the characters were removed, and the features were converted into numeric type features.

Some features still contained missing values. So, k-nearest-neighbors (KNN) imputation was utilized. KNN imputation uses Euclidean distance, essentially the shortest straight-line distance between two coordinates, to find a certain number of nearest neighbor values. Five nearest neighbors were used to impute data. Thus, because Andorra is missing a reserves value, the five most similar countries are found using the features and Euclidean distance. Then, the average reserves values of the five nearest countries fill the missing reserves value for Andorra. This process repeats until KNN imputation fills every missing value.

Equation 1: The Euclidean distance formula in n-dimensional space where p and q are coordinates for two different points. In a two-dimensional plane, this process is the same as finding the hypotenuse of a triangle using the Pythagorean theorem.

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_n - q_n)^2}$$

The final step of preprocessing is data standardization. Some machine learning models assume that the input data is normally distributed. In addition, different features are on different scales. For example, real GDP typically has data in the billions of dollars, while the inflation rate is closer to numbers around zero. Differently scaled data can make it difficult for machine learning models to learn relationships between features (Radečić, 2022). Thus, data is standardized so that machine learning models can easily make use of the data. The method used to standardize the data is called robust scaling. Robust scaling was chosen out of other methods because it is resistant to outliers in the data (Brownlee, 2020a). Due to the nature of data from countries, there are many instances of outliers. Thus, robust scaling is appropriate in this research. As illustrated by Equation 2, data is scaled by calculating the data's median and interquartile range (IQR). Then, the original data is subtracted by the median and then divided by the IQR of the data. The result of such data scaling is a scaled dataset with a mean and median of zero and a standard deviation of one, which machine learning models can easily use.

Equation 2: Robust scaling formula using the median and IQR of the data.

$$Scaled\ data = (data - median) / IQR$$

Exploratory Data Analysis (EDA)

Exploratory data analysis helps machine learning practitioners understand the data they are working with and find interesting relationships between variables. Several types of EDA were used: univariate, bivariate, and multivariate. Univariate EDA looks for information from one variable, bivariate two, and multivariate many. Through univariate analysis, the features are shown to have a primarily normal distribution, as shown in Figure 2B. However, some features have outliers, as seen in Figure 2A. The use of univariate analysis helps validate the usage of robust scaling and the machine learning model metric: mean absolute error, which will be clarified later.

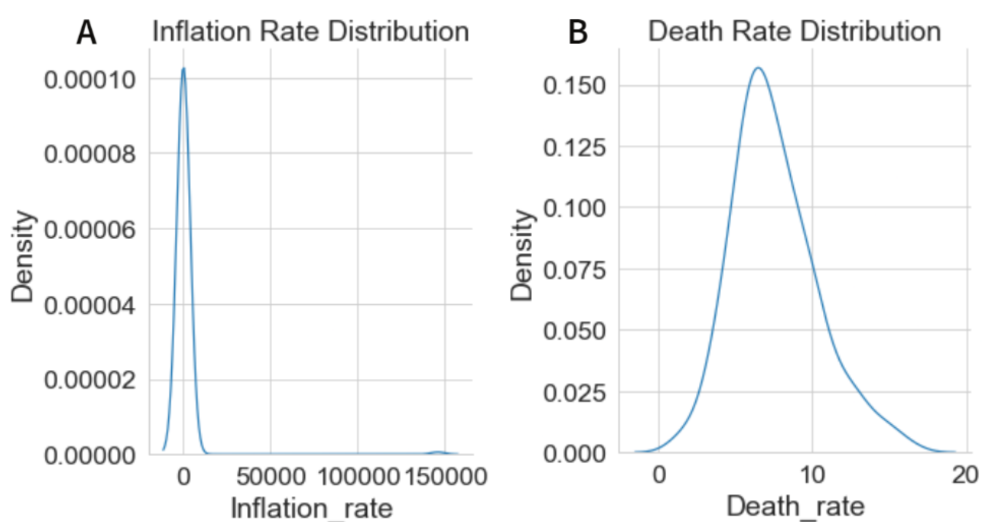

Figure 2. Univariate visualizations of the features' distributions. A shows the inflation rate distribution. B shows the death rate distribution.

The next step is looking at bivariate relationships between the input features and the HDI output. Many different relationships can be seen, as shown in Figure 3. The three main types of relationships revealed from bivariate analysis are no relation, logarithmic relation, and linear relation, as shown in Figures 3A, 3B, and 3C, respectively.

Figure 3A validates the understanding that real GDP does not relate well to human well-being (Kapoor & Debroy, 2019). Figure 3B shows that real GDP per capita stops having as much of an effect on the HDI after a certain point. Finally, the linear relationship between life expectancy at birth and the HDI in Figure 3C validates that life expectancy is one of the three main pillars of how the HDI is calculated (United Nations, 2020).

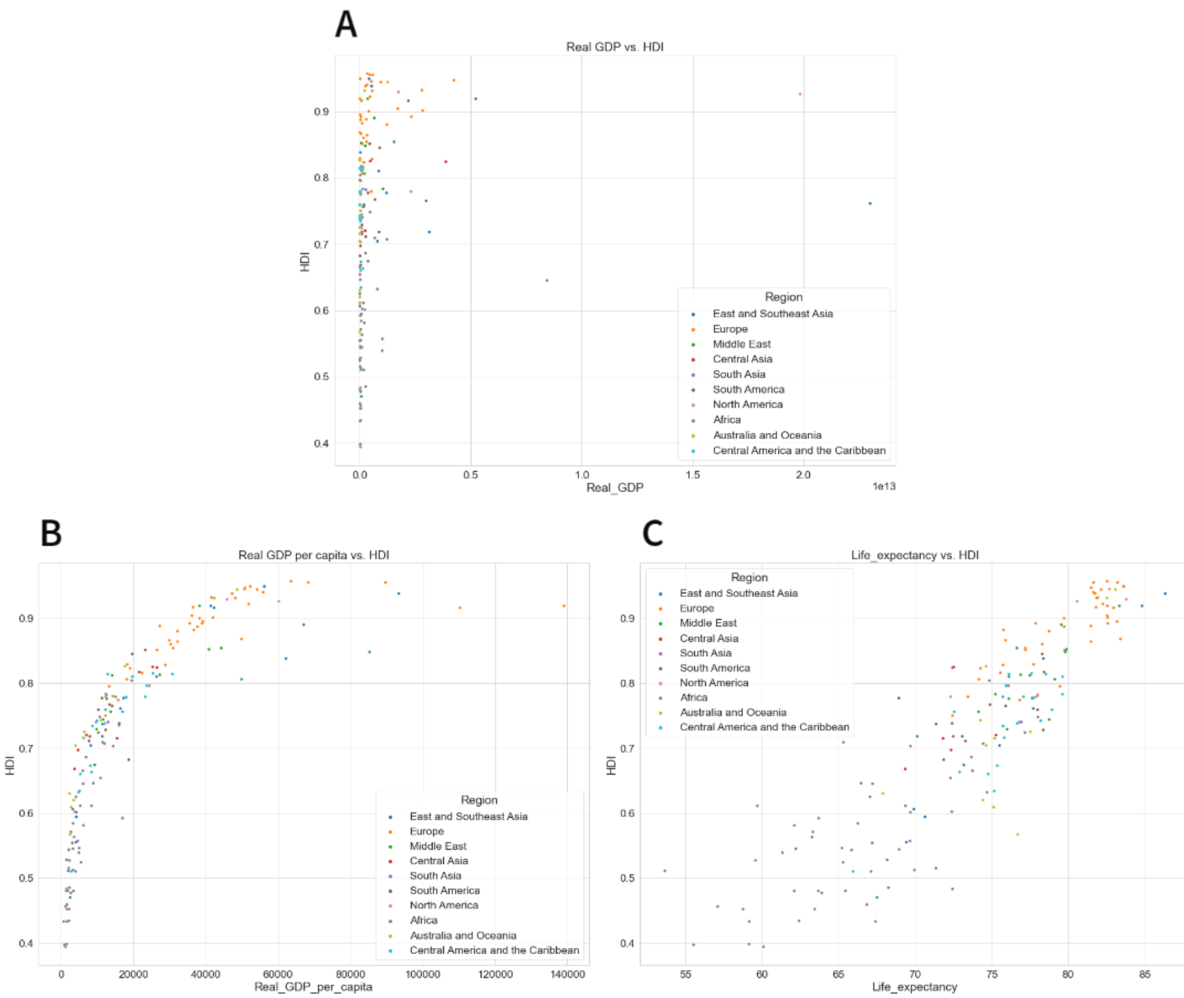

Figure 3. Bivariate visualizations using scatter plots of features vs. HDI. A shows real GDP vs. HDI. B shows real GDP per capita vs. HDI. C shows life expectancy at birth vs. HDI.

Multivariate analysis was the last type of EDA utilized. The visualization created for such analysis was a correlation heatmap. The correlation heatmap measures the Pearson correlation coefficient (PCC) between all combinations of features.

Equation 3: The Pearson correlation coefficient formula where n is the sample size, x_i and y_i are indexed samples of two features, and $\bar{x}$ and $\bar{y}$ are means.

$$r_{XY} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

The PCC helps show how well a straight line can fit the relationships between two variables. Thus, we can see what variables have a high correlation. For example, population is highly correlated to the number of internet users and people in the labor force (see Figure 4). These correlations make sense because an increase in population increases the number of people who can use the internet and work. The following section explains that these correlations will assist in feature selection.

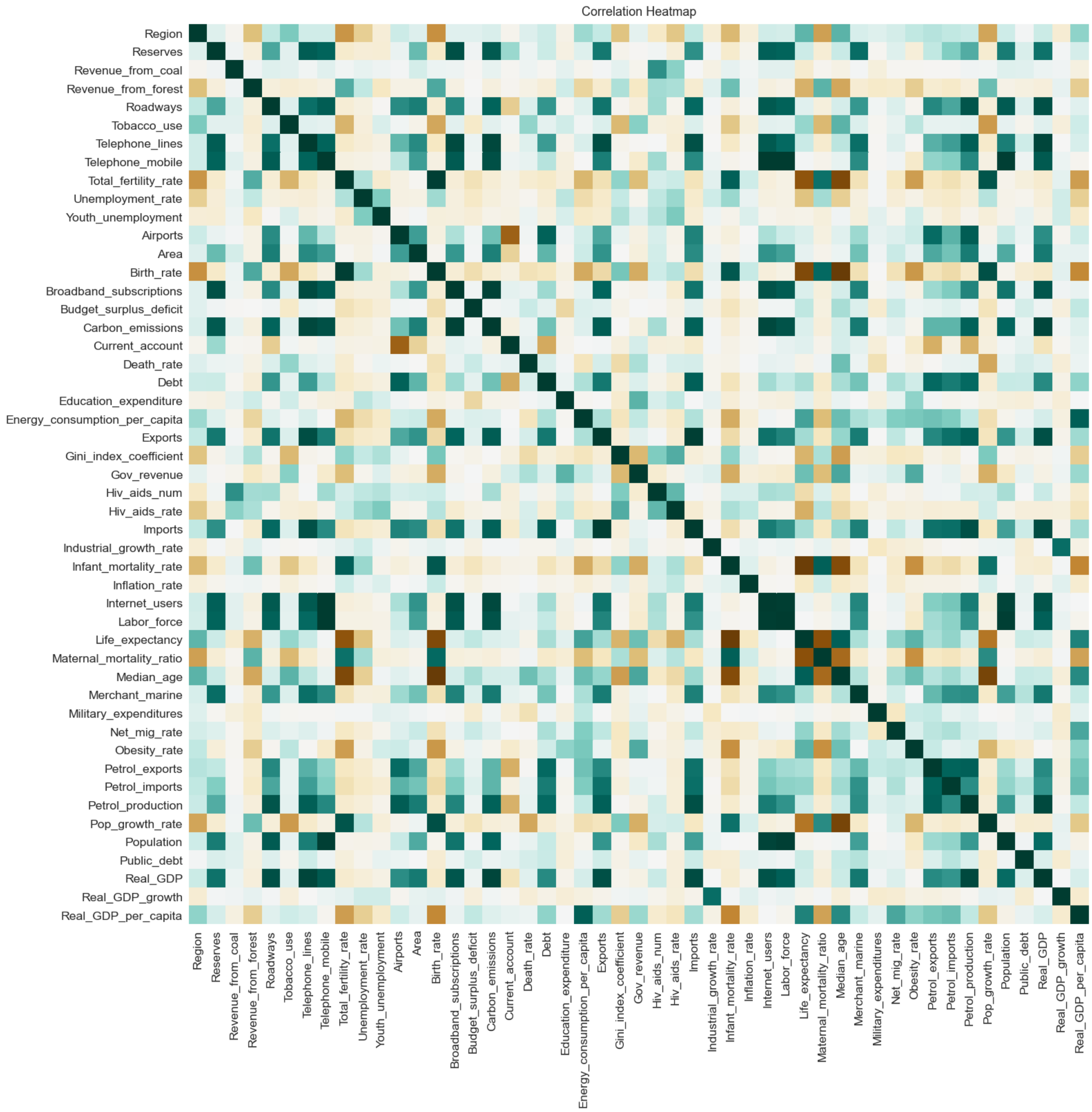

Figure 4. Multivariate visualization using a correlation heatmap between all features. The darker the color, the higher the correlation between two features. Green indicates a positive correlation while brown indicates negative.

Feature Selection

Feature selection is choosing specific features to feed into the machine learning algorithm as input. Feature selection was employed using variance inflation factor (VIF). The main goal of feature selection was to decrease multicollinearity. Multicollinearity is when three or more features are all related to each other. This relation between multiple features leads to all such features explaining the same part of the output. For instance, because most of the population are internet users in most countries, they both explain the general number of people in a country and how that relates to the HDI. This concept can happen with several variables, thus introducing multicollinearity to a dataset. Multicollinearity is a problem because when analyzing machine learning models for which features had the most impact on predicting the output, the effect is spread among several multicollinear variables. Other variables that are not as important take the spotlight instead (Khanna, 2020).

VIF helps combat multicollinearity by revealing which variables have multicollinearity problems (see Table 2). Features with high VIF values are removed one at a time until all VIF values are below five and multicollinearity is mostly gone (Tavares, 2017). Thus, we have a dataset with a few chosen features void of multicollinearity and ready to be run through machine learning models.

Table 2. VIF values of features before and after feature selection

Before (49 features)		After (17 features)	
VIF	**Feature**	**VIF**	**Feature**
583677.0	Population growth rate	2.0	Real GDP per capita
431664.3	Birth rate	1.6	Military expenditures
48166.7	Net migration rate	1.4	Total fertility rate
38109.3	Death rate	1.4	Net migration rate
…	…	…	…
7.2	Budget surplus or deficit	1.1	Infant mortality rate
6.3	Inflation rate	1.1	Population
5.7	Youth unemployment rate	1.0	Birthrate

Several feature sets with different features can be created by keeping one important feature and removing other multicollinear features. In Table 2, the one feature that was kept was real GDP per capita. Three other features were chosen for feature sets: maternal mortality ratio, industrial growth rate, and education expenditure. These features were chosen by training an early machine learning model and finding the most important features (this process will be explained further for model interpretation).

Model Selection

Several machine learning algorithms could be successful with the data at hand. Thus, several models were tried to find the best model for this paper's research. The model with the lowest error was chosen. The error metric used to find the best model was mean absolute error. Mean absolute error (MAE) was chosen because it is resistant to outliers (present in the dataset as seen during EDA) (Swalin, 2018).

Equation 4: Mean absolute error equation where m is the amount of data, y is the actual output value, and $\hat{y}$ is the predicted output value from a machine learning model.

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|y_i - \hat{y}_i|$$

Another decision to test the models was what form of cross-validation should be used. In machine learning, while a model needs data to learn, it also needs data to test its success at generalizing to unseen data. Thus, the dataset is split into train and test sets. However, more data is needed to improve the model until it is ready to use the test set for a final evaluation using the test set. K-fold cross-validation was used due to the relatively small amount of available data. K-fold cross-validation splits the training data into K folds (typically ten), using one fold as a proxy test set and the rest as a proxy train set. This process is repeated so that each fold becomes a proxy test set, and the error metrics are averaged for a single cross-validation score (Brownlee, 2020b). The models used were improved using K-fold cross-validation to reach optimal performance. For a good visualization of K-fold cross-validation, see one of the schematics in Scikit-Learn's user guide (*3.1. Cross-Validation*, n.d.).

Models

Multiple Linear Regression: Multiple linear regression is a model that attempts to map the relationship between the input features of the data and the output using weights. The HDI is on a scale from zero to one. Input features are changed by the weights so that all multiplications add up to an HDI value. A bias value is added to account for missing information not captured by the features and weights (Ng, 2022).

Equation 5: Linear regression equation where p is the number of features, x are the feature values, w are the learned weights given to the feature values, b is the bias term, and $\hat{y}$ is the prediction made.

$$\hat{y} = w_1 x_1 + w_2 x_2 + \cdots + w_p x_p + b$$

Ridge Regression: Ridge regression is very similar to linear regression. Linear regression can fail to learn the relationships we want. This is due to a concept known as overfitting. Overfitting is when a machine learning model can predict on the data it learns with very high accuracy. However, when the model is used on unseen data, it fails to generalize and performs poorly. Ridge regression addresses this issue by adding a regularization term to linear regression. Essentially, Ridge regression adjusts the weights of the multiple linear regression model while learning so that the model overfits less (Ng, 2022).

Bayesian Ridge Regression: Bayesian ridge regression is a variant of ridge regression that uses Bayesian statistics to make predictions. Bayesian regression uses prior assumptions about the data to learn the weights. Bayesian regression aims to find the highest probability that the weights it learns are correct given the data and distribution assumptions (Koehrsen, 2018). Overfitting is avoided through ridge regularization, and Bayesian regression works well with small amounts of data due to assumptions about the data's distribution (Gonfalonieri, 2019). Thus, Bayesian regression is applied to the relatively small dataset used for this research.

Decision Tree Regression: Decision tree regression is an algorithm that uses features to make defining splits in the input data. These splits help segment the data so predictions can eventually be made (see Figure 5). The features for each split, called decision nodes, are chosen based on which features can make more accurate splits. For example, take the example decision tree in Figure 5. Three features are used to split between higher and lower HDI scores. The top decision node (called the root node) was most likely chosen due to being the most accurate at splitting between higher and lower scores. The health and education decision nodes also help split the data further and create predictions (Singh Chauhan, 2022).

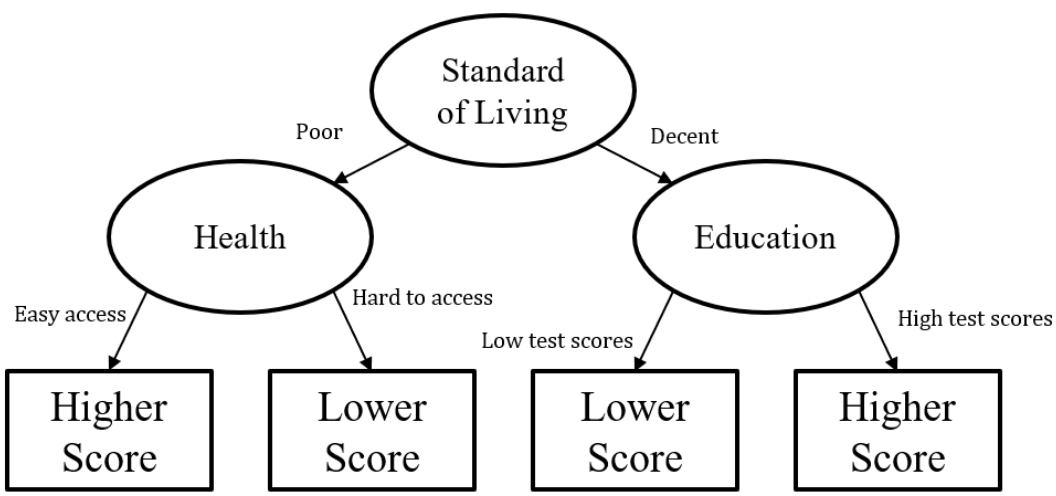

Figure 5. A visualization of a decision tree where the ellipses are decision nodes and the boxes are predictions.

Random Forest Regression: Random forests are essentially an extension of decision trees. The random forest algorithm is a type of ensemble model. An ensemble model uses several models and creates a prediction based on an aggregate of all the models in the ensemble. Thus, random forests use several versions of decision trees in an ensemble to make predictions. Different versions are created by using random features. While most of the flawed decision trees cancel out with each other in the aggregation of the ensemble, some of the decision trees find meaningful relationships in the data and contribute to making a robust ensemble prediction (Breiman, 2001).

Hyperparameter Tuning

With all the models chosen, they need to be optimized through hyperparameter tuning. Hyperparameter tuning can be thought of as tuning an engine. While learning from the training data constructs the engine, hyperparameter tuning squeezes extra performance out of the engine. A hyperparameter is an aspect of a machine learning model that affects the parameters and, thus, how the model learns. Machine learning practitioners choose hyperparameters, while on the other hand, parameters, such as the weights mentioned in linear regression, are learned directly from the data. For example, the weights of a ridge regression model are learned from the training data, while the regularization hyperparameter is set manually.

The process of finding the optimal hyperparameters includes two steps. The first step is to randomly try hyperparameters over a set distribution of values in which the most optimal hyperparameters are likely to be. The second step is to exhaustively search for the best hyperparameters around where the random search found its hyperparameters. The models can produce their best results through this process, and the model with the lowest error will be chosen for analysis.

Error Analysis

Once a model has been chosen, additional ideas for improvement can be revealed by analyzing the model's errors. Learning curves were used to analyze training and cross-validation errors. Learning curves show how well a model learns based on how much data it is fed. Typically, more data leads to a higher training error and lower cross-validation error. Thus, as the number of training examples or data increases, the training score decreases and the cross-validation score increases. Due to the lack of data, overfitting is likely to occur. Steps such as further feature selection, increasing the amount of available data, and increasing regularization can all decrease overfitting and improve a model (Ng, 2022).

In addition, all error scores are observed, whether from the training set, cross-validation, or the test set. Observing the errors and their differences can also reveal the next steps in the machine learning process. Once again, due to the high probability of overfitting, cross-validation error is likely to be closer to test error than to train error. This analysis would once again confirm the existence of overfitting.

Model Interpretation

The machine learning process's final step is understanding how the final model reached its predictions. Permutation importance was used to interpret the model. Permutation importance finds features that significantly affect the model's accuracy (*4.2. Permutation Feature Importance*, n.d.). Significant effects are found by removing features and observing the decrease in accuracy in the model: the more significant the decrease in accuracy, the more important the removed feature. Thus, the features that impact the HDI the most can be used to guide policy decisions related to increasing a country's HDI.

Results

Feature Selection

Through VIF feature selection, several feature sets were created. Keeping one important feature and removing other highly multicollinear variables leads to several feature sets. The four feature sets created used real GDP per capita, education expenditure, industrial growth rate, and maternal mortality ratio. The four feature sets were run through a multiple linear regression model to see which feature set had the lowest error. The feature set that used real GDP per capita had the best performance and was thus used for all future models (see Table 3).

Table 3. MAE of the feature sets.

Feature Set	Mean Absolute Error
All Features	0.0724
Real GDP per Capita	0.0427
Education Expenditure	0.0683
Industrial Growth Rate	0.0688
Maternal Mortality Ratio	0.0553

Model Selection

After conducting hyperparameter tuning, the model with the lowest cross-validation error was the random forest regression model (see Table 4).

Table 4. Cross-validation MAE of the machine learning models.

Model	Mean Absolute Error
Ridge	0.0418
Bayesian Ridge	0.0418
Decision Tree	0.0384
Random Forest	0.0264

Error Analysis

Through error analysis, the problem of overfitting was confirmed. The difference between the test and cross-validation errors was less than that of the cross-validation and train errors (see Table 5). These error differences show that the model fits the training data better than unseen data. Thus, overfitting occurred in the data. However, the error was at an acceptable level, so the model was left as is.

Table 5. MAE of the train set, cross-validation, and test set.

Data	Mean Absolute Error
Train	0.0099
Cross-validation	0.0264
Test	0.0284

Additionally, it seemed that not much more performance could be gained from the model. The learning curves of the random forest model reveal diminishing returns after around 20 to 30 training examples are used to train the model (see Figure 6).

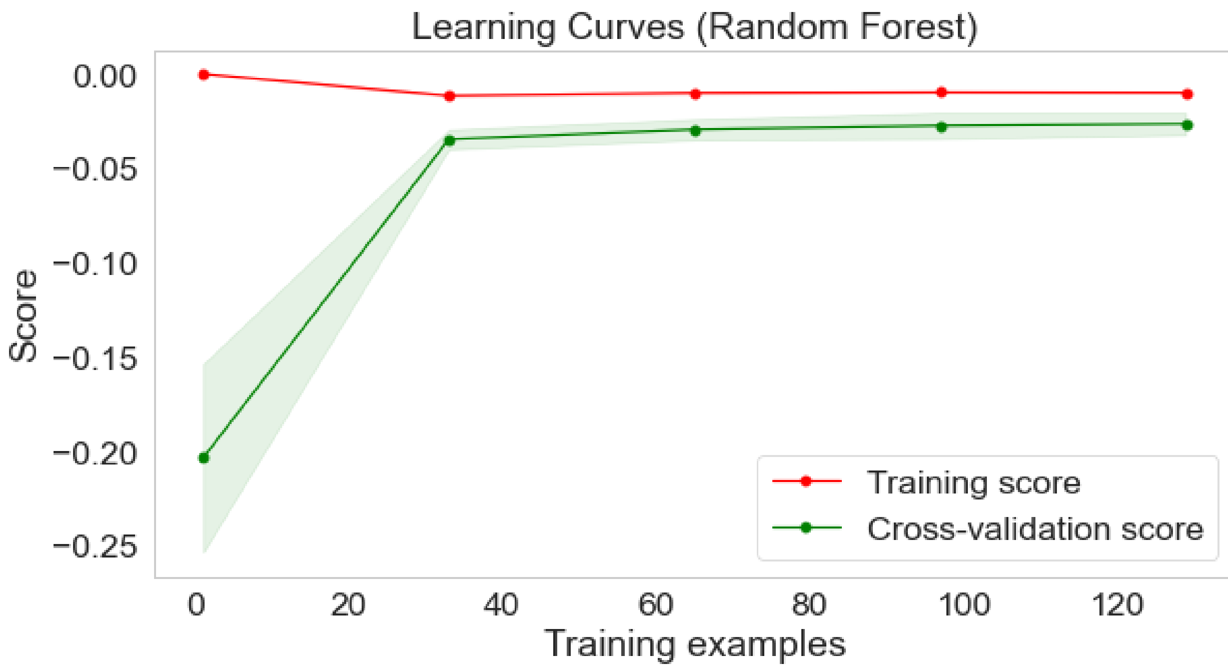

Figure 6. The learning curve of the final random forest model.

Model Interpretation

Features that impacted prediction accuracy the most can be seen by extracting permutation importance scores from the final random forest model. The most important feature of the random forest model was GDP per capita. Other features that got high scores include infant mortality rate, birth rate, and total fertility rate (see Figure 7).

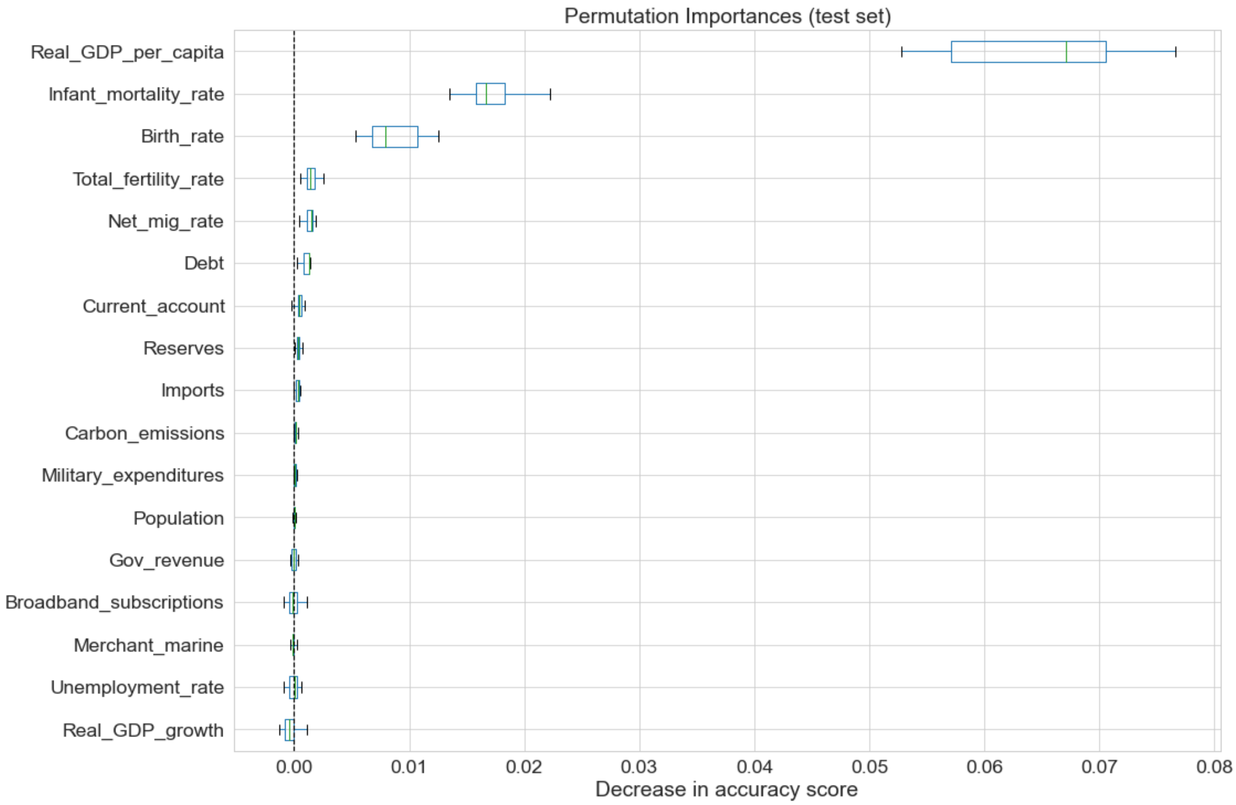

Figure 7. The permutation importance scores of each feature used in the final random forest model.

Discussion

With a model that has an acceptable amount of error in predicting the HDI of each country, analysis can be done on important features and their relation to the HDI. The HDI was built based on three indicators: health, education, and standard of living. The top four important features contributing to the final model's accuracy are similar to the HDI indicators. Real GDP per capita is often used as an indicator of the standard of living in a country. Similarly, the UN uses gross national income (GNI) per capita to calculate its HDI (United Nations, 2020). GNI is very similar to GDP as it measures the economy's overall size. Thus, it makes sense that real GDP per capita has a significant effect on the final model's accuracy.

Additionally, infant mortality rate, birth rate, and total fertility rate all relate to the health aspect of the HDI. The UN uses life expectancy at birth to measure HDI, and all three of the aforementioned features are correlated to life expectancy (see Figure 4). Thus, the health aspect of the HDI is explained through the features present in the final model.

The only part of the HDI that is not explicitly present in the final model is education. While education expenditure is a feature part of the World Factbook that could help explain the education aspect of the HDI, it was removed during VIF feature selection. Thus, education expenditure was multicollinear with the features in the feature set used for the final model. Education expenditure is present implicitly in the feature set due to other features explaining the education aspect of the HDI. This result implies that education and the HDI can be impacted through policies addressing health and standard of living.

Specifically, supply-side policies could see the most impact in impacting the HDI. Supply-side policies are a type of economic policy that increases the total output of an economy. The goal of supply-side policies is to improve productivity and efficiency in an economy, increasing total economic output (Pettinger, 2017). Essentially, through supply-side policies, economic growth is gained so that all economic stakeholders benefit. So, the standard of living is improved. In addition, policies relating to the health industry are associated with supply-side policies (Pettinger, 2019). For example, policies affecting health care insurance, research, and technology are all supply-side policies that could address the health aspect of human development. Supply-side policies offer the option to improve two aspects of the HDI, and as seen through model interpretation, education might also be impacted implicitly.

While this paper focused on the HDI and human development, many other indexes have different focuses. The World Happiness Report ranks happiness in countries based on factors such as perceptions of corruption, freedom to make choices, and generosity (Helliwell et al., 2022). The Happy Planet Index tries to make up for where the HDI lacks in accounting for sustainability by factoring in ecological footprint into its index (*Happy Planet Index – How Happy Is the Planet*, 2021). Accordingly, government policy regarding happiness or sustainability can use the World Happiness Report or the Happy Planet Index with machine learning to guide what features might succeed in improving a country's index score. Machine learning can benefit indexes calculated with factors that do not relate too much to government policy. For example, the World Happiness Report records the generosity levels of countries through surveys. Generosity is not something governments can tackle directly. However, information from the World Factbook could reveal what features could have an impact and allow governments to increase happiness in their countries.

Conclusion

Leveraging machine learning and the HDI, this paper achieved an accurate model that could predict the HDI with an MAE of 0.0284 on the test set. By removing features and most multicollinearity during feature selection, important features could be extracted and analyzed using permutation importance scores. Through these scores, this paper demonstrated that machine learning reveals important features which could guide government policy that attempts to tackle problems addressed by certain indexes. The important features found in this paper show that supply-side policies related to the standard of living and health would have the most success in improving human development. The possibility of applying the machine learning process in this paper to other indexes, such as the Happy Planet Index and the World Happiness Report, also shows promise.

Limitations

There were some limitations to the research that was carried out. For one, the data from the World Factbook had inconsistent dating. Much of the data lagged the present by a few years. In addition, as seen during preprocessing, not all countries had data recorded for every single feature. While imputation made up for the missing data, ideally, all data would have been recorded beforehand. Finally, due to hyperparameter tuning being very computationally expensive and the limited computer resources in this paper, the best hyperparameters may not have been achieved. However, model performance was acceptable, so further hyperparameter tuning may not have had much impact.

Acknowledgments

I would like to extend my gratitude to Mr. Raines for his assistance in writing this paper. In addition, I would like to thank my family for continually supporting me throughout the research process.

References

3.1. Cross-validation: Evaluating estimator performance. (n.d.). [User Guide]. Scikit-Learn. Retrieved August 29, 2022, from https://scikit-learn/stable/modules/cross_validation.html

4.2. Permutation feature importance. (n.d.). [User Guide]. Scikit-Learn. Retrieved August 29, 2022, from https://scikit-learn/stable/modules/permutation_importance.html

Ardila, D., Kiraly, A. P., Bharadwaj, S., Choi, B., Reicher, J. J., Peng, L., Tse, D., Etemadi, M., Ye, W., Corrado, G., Naidich, D. P., & Shetty, S. (2019). End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature Medicine*, *25*(6), 954–961. https://doi.org/10.1038/s41591-019-0447-x

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brownlee, J. (2020a, May 26). How to Scale Data With Outliers for Machine Learning. *Machine Learning Mastery*. https://machinelearningmastery.com/robust-scaler-transforms-for-machine-learning/

Brownlee, J. (2020b, July 30). How to Configure k-Fold Cross-Validation. *Machine Learning Mastery*. https://machinelearningmastery.com/how-to-configure-k-fold-cross-validation/

Central Intelligence Agency. (2022, August 18). *The World Factbook*. https://www.cia.gov/the-world-factbook/

Dhaduk, H. (2021, June 25). *EDA | Exploratory Data Analysis With Python | What is EDA*. https://www.analyticsvidhya.com/blog/2021/06/eda-exploratory-data-analysis-with-python/#h2_3

Elliott, T. (2019, January 24). The State of the Octoverse: Machine learning. *The GitHub Blog*. https://github.blog/2019-01-24-the-state-of-the-octoverse-machine-learning/

Gonfalonieri, A. (2019, May 17). 5 Ways to Deal with the Lack of Data in Machine Learning. *KDnuggets*. https://www.kdnuggets.com/5-ways-to-deal-with-the-lack-of-data-in-machine-learning.html/

Happy Planet Index – How happy Is the planet. (2021). https://happyplanetindex.org/

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Helliwell, J. F., Layard, R., Sachs, J. D., Neve, J.-E. D., Aknin, L. B., & Wang, S. (2022). *World Happiness Report 2022*. https://worldhappiness.report/ed/2022/

Kapoor, A., & Debroy, B. (2019, October 4). GDP Is Not a Measure of Human Well-Being. *Harvard Business Review*. https://hbr.org/2019/10/gdp-is-not-a-measure-of-human-well-being

Khanna, C. (2020, December 5). Multicollinearity—Why is it bad? *Medium*. https://towardsdatascience.com/multicollinearity-why-is-it-bad-5335030651bf

Koehrsen, W. (2018, April 20). Introduction to Bayesian Linear Regression. *Medium*. https://towardsdatascience.com/introduction-to-bayesian-linear-regression-e66e60791ea7

McKinney, W. (2010). Data Structures for Statistical Computing in Python. *Proceedings of the 9th Python in Science Conference*, 56–61. https://doi.org/10.25080/Majora-92bf1922-00a

Ng, A. (2022). *Machine Learning Specialization*. Coursera. https://www.coursera.org/specializations/machine-learning-introduction

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*(85), 2825–2830.

Pettinger, T. (2017, October 8). The importance of supply-side policies. *Economics Help*. https://www.economicshelp.org/blog/31/supply-side/supply-side-policies/

Pettinger, T. (2019, October 30). Supply Side Policies. *Economics Help*. https://www.economicshelp.org/macroeconomics/economic-growth/supply-side-policies/

Radečić, D. (2022, March 21). Data Scaling for Machine Learning—The Essential Guide. *Medium*. https://towardsdatascience.com/data-scaling-for-machine-learning-the-essential-guide-d6cfda3e3d6b

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and Statistical Modeling with Python. *Proceedings of the 9th Python in Science Conference*, 92–96. https://doi.org/10.25080/Majora-92bf1922-011

Singh Chauhan, N. (2022, February 9). Decision Tree Algorithm, Explained. *KDnuggets*. https://www.kdnuggets.com/decision-tree-algorithm-explained.html/

Swalin, A. (2018, July 10). Choosing the Right Metric for Evaluating Machine Learning Models—Part 1. *USF-Data Science*. https://medium.com/usf-msds/choosing-the-right-metric-for-machine-learning-models-part-1-a99d7d7414e4

Tavares, E. (2017, March 8). *Variance Inflation Factor (VIF) Explained—Python*. https://etav.github.io/python/vif_factor_python.html

United Nations. (2020). Human Development Report 2020. In *Human Development Reports*. United Nations. https://hdr.undp.org/content/human-development-report-2020

Vinyals, O., Babuschkin, I., Czarnecki, W. M., Mathieu, M., Dudzik, A., Chung, J., Choi, D. H., Powell, R., Ewalds, T., Georgiev, P., Oh, J., Horgan, D., Kroiss, M., Danihelka, I., Huang, A., Sifre, L., Cai, T., Agapiou, J. P., Jaderberg, M., … Silver, D. (2019). Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature*, *575*(7782), 350–354. https://doi.org/10.1038/s41586-019-1724-z