# Explaining World Economies

David Blankenship

## Introduction

There have been questions in political and economic discussions as to whether the framework of economies as advanced or developing is a measurable or useful model for understanding differences between countries. To answer the question of measurability in a previous project for DSCI 631 we were able to demonstrate that International Monetary Fund (IMF) data could be used to develop a model that predicts this classification. For the question of usefulness, I suggest in this work that XAI can be used with ML models to provide economic development targets showing what is necessary for developing nations to achieve advanced status. In this work I will use SHaply Additive exPlanations (SHAP) to show the most predictive features for economic classification. I will also show how counterfactuals can be used to provide economic development targets based on the necessary changes to achieve advanced status. Finally, I use Mean Absolute value Relative Difference (MARD) as a statistical method for counterfactuals to quantify the largest relative changes required to shift classification. My goal in this report is to provide a proof-of-concept that these techniques can be used on this dataset and be presented in a form where future work with a specialist in economic development or other relevant fields could be done.

## Data/Past Work

This previous work is based on my previous class project for DSCI 63, *Classifying World Economies* [1]. In this work we used the IMF World Economic Outlook [2] which is a biannual dataset released by the IMF that is used to project economic forecasts in the near and medium term. In addition to the forecast, it's also a great collection of past economic data collected by the IMF from 1980 up to 2021 at the time we collected the data. In the original dataset there are 44 features for 196 countries. We consider each country's data for a year as individual data points to increase training data. We impute missing values using a k-nearest neighbor with an N = 3.  After preprocessing we have 975 rows, 40 features, and 1 target column labeling each as either developing or advanced. The reason we dropped some features was because we converted all national currencies into US dollars to allow for similar comparisons which created duplicate columns. We also dropped "employment in persons" because this data was only collected for advanced economies and would introduce a data leakage by its biased presence. We also drop Syria data due to no data being collected since 2011 due to the Syrian Civil War going on at the time. Finally, the column names in the dataset were descriptive and very long so we include a key below in table 1 to pair the simple variable names we use in our visualizations with the full description of what they are.

x0: Gross domestic product, constant prices in US dollars
x1: Gross domestic product, constant prices in Percent change
x2: Gross domestic product, current prices in U.S. dollars
x3: Gross domestic product, current prices in Purchasing power parity; international dollars
x4: Gross domestic product, deflator in Index
x5: Gross domestic product per capita, constant prices in US dollars
x6: Gross domestic product per capita, constant prices in Purchasing power parity; 2017 international dollar
x7: Gross domestic product per capita, current prices in U.S. dollars
x8: Gross domestic product per capita, current prices in Purchasing power parity; international dollars
x9: Gross domestic product based on purchasing-power-parity (PPP) share of world total in Percent
x10: Implied PPP conversion rate in National currency per current international dollar
x11: Total investment in Percent of GDP
x12: Gross national savings in Percent of GDP
x13: Inflation, average consumer prices in Index
x14: Inflation, average consumer prices in Percent change
x15: Inflation, end of period consumer prices in Index
x16: Inflation, end of period consumer prices in Percent change
x17: Volume of imports of goods and services in Percent change
x18: Volume of Imports of goods in Percent change
x19: Volume of exports of goods and services in Percent change
x20: Volume of exports of goods in Percent change
x21: Unemployment rate in Percent of total labor force
x22: Population in Persons
x23: General government revenue in US dollars
x24: General government revenue in Percent of GDP
x25: General government total expenditure in US dollars
x26: General government total expenditure in Percent of GDP
x27: General government net lending/borrowing in US dollars
x28: General government net lending/borrowing in Percent of GDP
x29: General government structural balance in US dollars
x30: General government structural balance in Percent of potential GDP
x31: General government primary net lending/borrowing in US dollars
x32: General government primary net lending/borrowing in Percent of GDP
x33: General government net debt in US dollars
x34: General government net debt in Percent of GDP
x35: General government gross debt in US dollars
x36: General government gross debt in Percent of GDP
x37: Gross domestic product corresponding to fiscal year, current prices in US dollars
x38: Current account balance in U.S. dollars
x39: Current account balance in Percent of GDP
y: target

Table 1: Feature name map

In our previous work we were able to develop a very high-performance model. We used an 80:20 train-test split, recursive feature elimination, and hyperparameter tuning to train a Random Forest model using the scikit-learn library. As we can see in table 2 and the confusion matrix (figure 1) below, we were able to achieve an F1 Score, precision and recall of 0.9756 for each. Of our 195 test data points, only 2 data points were incorrectly classified. Croatia in 2017 was incorrectly labeled as a developing country and Oman in 2021 was incorrectly labeled as an advanced economy. I also note that our train data achieved perfect results which initially worried us about overfitting. However, the high-quality test results suggest that we simply achieved a very high-fidelity model.

| Precision | Recall | F1 |
|---|---|---|
| 0.9756 | 0.9756 | 0.9756 |

Table 2: Random Forest test results

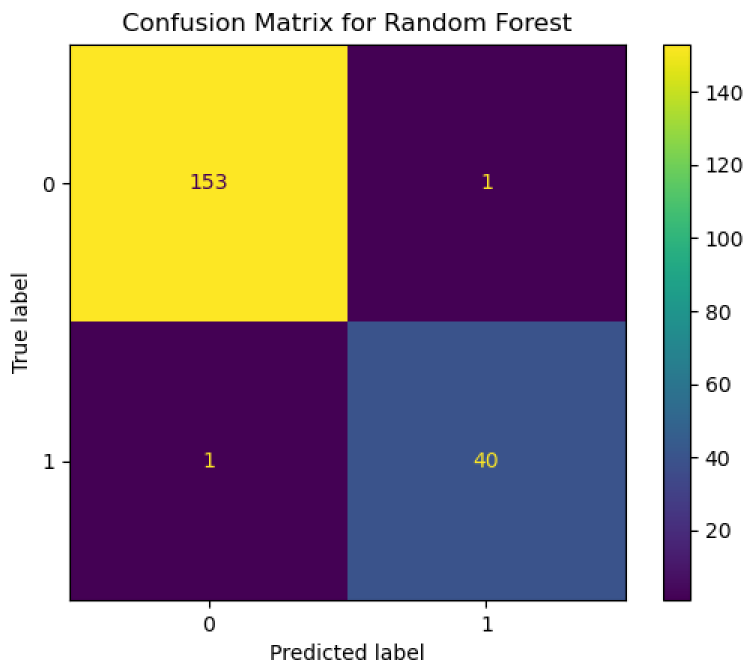


Figure 1: Confusion Matrix for Random Forest

I also looked at the probabilities given in the model's predictions. Due to the size of the dataset, looking at every datapoint isn't feasible. As such I focus on the probabilities already closest to shifting to the other class as well as the incorrectly predicted countries. Table 3 below shows the countries' prediction probability between 0.33 and 0.67. Notably, HRV 2017, which was incorrectly predicted as class 0 had a relatively high probability for class 0 at 0.71.

| | 0 Probability | 1 Probability | y Prediction | y Actual |
|---|---|---|---|---|
| SVK 2017 | 0.36 | 0.64 | 1 | 1 |
| OMN 2021 | 0.45 | 0.55 | 1 | 0 |
| USA 2021 | 0.33 | 0.67 | 1 | 1 |
| HRV 2018 | 0.41 | 0.59 | 1 | 1 |

Table 3: Probabilities closest to different class

## Feature Attribution Method

I looked at the Shapley Additive Explanations (SHAP) values [3] to quantify the feature importance in our model. SHAP uses local explanations of individual predictions to quantify the feature importance values. We display two plots to show this. First, in figure 2 we see the feature importance bar plot showing which features are shown to be the most relevant for prediction. We also look at the beeswarm plot if figure 3 that displays the top features, their values, and where they fall along the SHAP values axis. The beeswarm plot gives us a sense of how the value of a feature pushes the prediction in the model. In our beeswarm plot we look at how it predicts advanced economies or class 1 meaning positive SHAP values predict for class 1. The beeswarm for developing/class 0 would be the exact same but mirrored. SHAP found the top 9 most important features in order of importance are as follows:

1. Population in Persons
2. General Government Revenue in US Dollars
3. Inflation, average consumer prices in Percent change
4. General government net lending/borrowing in Percent of GDP
5. General government net lending/borrowing in US dollars
6. Implied PPP conversion rate in National currency per current international dollar
7. Volume of exports of goods in Percent change
8. Inflation, end of period consumer prices in Percent change
9. Volume of exports of goods and services in Percent change

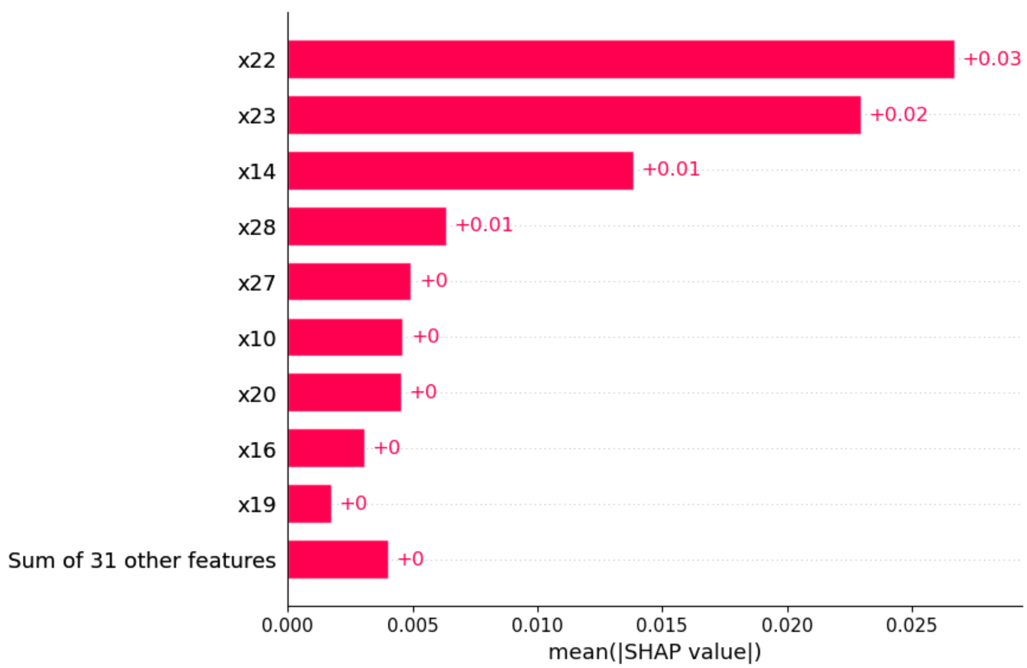


Figure 2: SHAP value feature importance bar plot

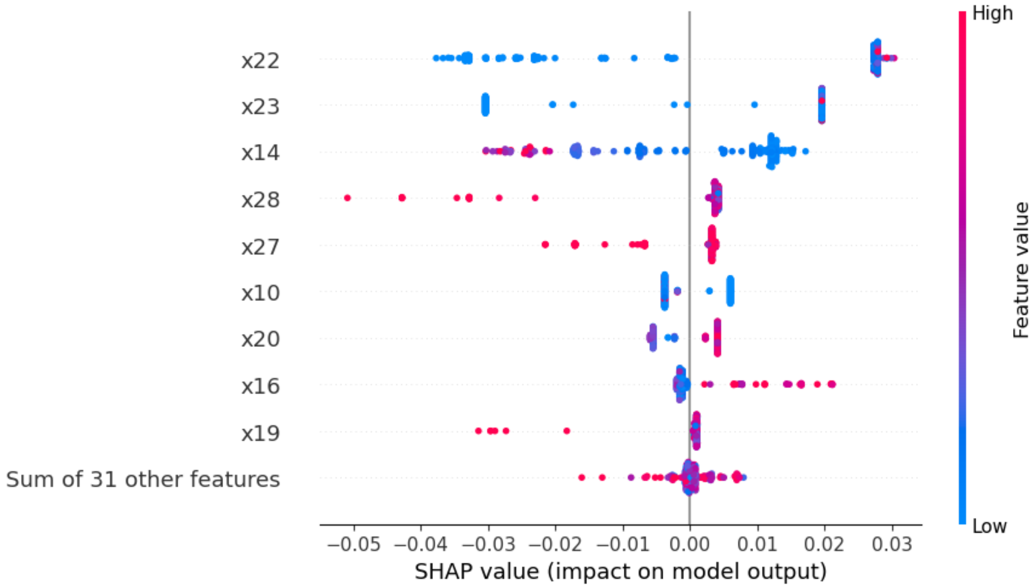


Figure 3: SHAP value beeswarm plot for class 1

# Counterfactual Method

The general concept of a counterfactual is to look at how much the features of a dataset need to change to produce a different outcome [4][5]. In our case we are using them to determine what economic features from the IMF data would need to change to be predicted as the opposite development category. To this end we use the DiCE implementation of counterfactuals as an out-of-the-box Python library available on GitHub and based on the work by Mothilal, Sharma, and Tan [6].

I generated 3 counterfactuals for all data points in the test set leading to 585 counterfactuals total. I then find the aggregate of the Mean Absolute value Relative Difference (MARD) as a measure to quantify how much the features had to change from their original values. This idea is to give a general measure across all countries as an idea for what the most important features are from a counterfactual framework. I also look at individual countries to see what would be necessary to change them. Rather than looking through every country I elected to look at the countries with the prediction probabilities closest to having been selected as the other as well as the two countries that were wrongly chosen earlier. I elected to use the genetic method for the DiCE model rather than random as genetic was significantly quicker and tended to change more features, whereas random took significantly longer and tended to shift fewer features, often only one.

# Analysis

Looking at the top 9 features from the SHAP values we note that most, but not all are included in the selected features in the original model. The Volume of exports of goods in Percent change (x20) is notably not selected in the model. This suggests a disagreement in the model and SHAP method. We can also look at the beeswarm plot and see how the values for each feature influence the class choice. Rather than going through each individual feature we simply note a common pattern in our results. We tend to see that the influence on one direction is either low or high on one side and then a mix in the other direction. For example, in Population in Persons (x22) we see that negative SHAP values have uniformly low values for populations. The positive SHAP values have a mix of low and high values for x22. Only x20 and the sum of features seems to buck this trend.

Since recreating the counterfactual tables in this report would be unwieldy, I point to my notebook on GitHub [1] to display that I was able to generate counterfactuals for the lower probability classifications. We can see that the genetic method provides a broad range of counterfactual values with significant difference between them. One concern is that some counterfactuals may generate values as negative that cannot actually be negative. Some constraints may be necessary to ensure realistic generations. Further analysis of how well these counterfactuals could be used to generate targets would require deeper expertise in the field of economics than I possess. However, we can see that they are in a form that could be presented to an economist for evaluation.

I also show the MARD tables generated by our counterfactual method. Note that in table 4, x35 and x36 are infinite due to Macau 2018 and Macau 2020 having a value of 0 in both of those fields. These are both related to General government gross debt and may be zero due to the unique political situation of Macau with China. As such I removed the two datapoints and recalculated the MARD in table 5. In both tables I highlight the top 9 values requiring the most change to swap the classification. Most notably the General government primary net lending/borrowing in US dollars (x31) with a MARD of 103.24 which corresponds to an average of 10324% change, a major difference. Other major features include General government net debt in US dollars (x33), Current account balance in U.S. dollars (x38), and Gross domestic product per capita, constant prices in US dollars (x5). Comparing the top 9 selected features from SHAP the only overlapping features are x10 and x27.

| Feature | MARD | Feature | MARD | Feature | MARD | Feature | MARD |
|---|---|---|---|---|---|---|---|
| x0 | 7.30 | **x10** | **17.89** | x20 | 2.23 | x30 | 1.04 |
| x1 | 1.61 | x11 | 0.33 | x21 | 0.40 | **x31** | **103.24** |
| x2 | 6.94 | x12 | 0.90 | x22 | 2.40 | x32 | 3.61 |
| **x3** | **8.90** | x13 | 0.37 | x23 | 7.75 | **x33** | **46.44** |
| x4 | 0.57 | x14 | 2.79 | x24 | 0.83 | x34 | 1.10 |
| **x5** | **34.96** | x15 | 0.38 | **x25** | **7.75** | x35 | inf |
| x6 | 6.64 | x16 | 2.65 | x26 | 0.69 | x36 | inf |
| **x7** | **12.03** | x17 | 2.50 | **x27** | **18.44** | x37 | 7.06 |
| x8 | 6.52 | x18 | 4.12 | x28 | 4.53 | **x38** | **42.81** |
| x9 | 2.54 | x19 | 3.86 | x29 | 3.02 | x39 | 3.02 |

Table 4: Mean Absolute Value Relative Difference with top 9 bolded

| Feature | MARD | Feature | MARD | Feature | MARD | Feature | MARD |
|---|---|---|---|---|---|---|---|
| x0 | 7.37 | **x10** | **18.07** | x20 | 2.17 | x30 | 0.99 |
| x1 | 1.62 | x11 | 0.33 | x21 | 0.39 | **x31** | **104.30** |
| x2 | 7.00 | x12 | 0.90 | x22 | 2.41 | x32 | 3.64 |
| **x3** | **8.98** | x13 | 0.37 | x23 | 7.82 | **x33** | **46.90** |
| x4 | 0.57 | x14 | 2.74 | x24 | 0.84 | x34 | 1.11 |
| **x5** | **35.31** | x15 | 0.38 | x25 | 7.82 | **x35** | **11.56** |
| x6 | 6.70 | x16 | 2.59 | x26 | 0.69 | x36 | 1.65 |
| **x7** | **12.15** | x17 | 2.50 | **x27** | **18.62** | x37 | 7.13 |
| x8 | 6.58 | x18 | 4.13 | x28 | 4.57 | **x38** | **42.49** |
| x9 | 2.56 | x19 | 3.76 | x29 | 3.02 | x39 | 3.04 |

Table 5: Mean Absolute Value Relative Difference with top 9 bolded with Macau excluded

# Conclusion

We find that the methods in this paper are capable of being used on our dataset to generate feature importances with different methods and counterfactual examples. I have displayed the proof of concept for these methods as useable for a dataset of this type. Between our two feature importance methods we see some but relatively little agreement regarding what are the most major features. Further investigation by an economist is necessary to determine if the selected features are well suited as these methods suggest. I would also like to investigate further with a specialist or investigate other economic development metrics to compare how the generated counterfactuals compare to real world targets.

# References


[1] Blankenship, David and Vaidya, Jai. (2024). Classifying World Economies. https://github.com/General-Cow/Classifying-World-Economies.

[2] International Monetary Fund. (1980-2021). World Economic Outlook dataset. Accessed 2024.

[3] Lundberg, Scott M., and Su-In Lee. "A unified approach to interpreting model predictions." *Advances in neural information processing systems* 30 (2017).

[4] Byrne, Ruth MJ. "Counterfactuals in explainable artificial intelligence (XAI): Evidence from human reasoning." *IJCAI*. 2019.

[5] Wachter, Sandra, Brent Mittelstadt, and Chris Russell. "Counterfactual explanations without opening the black box: Automated decisions and the GDPR." *Harv. JL & Tech.* 31 (2017): 841.

[6] Mothilal, Ramaravind K., Amit Sharma, and Chenhao Tan. "Explaining machine learning classifiers through diverse counterfactual explanations." *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 2020. https://github.com/interpretml/DiCE