Predicting and explaining Illicit Financial Flows in developing countries: A machine learning approach

Akshith Putta

Abstract:

Cross-border corruption and the illicit movement of financial assets, referred to as illicit financial flows (IFFs), have a strongly deleterious effect on the economies of developing nations. Over the past 20 years, there has been a concerted international effort to mitigate cross-border corruption, however, the most important economic and political factors leading to IFFs are unclear. In this work, I use multiple machine learning (ML) approaches - including linear regression, logistic regression, support vector machines, random forests, and neural networks - to predict the levels of corruption using various economic and political measures from the years 2009 to 2018. Furthermore, to make clear the relative importance of these factors, I use several ML model interpretation tools. Out of the various regression ML Models, the Artificial Neural Network (ANN) had the most success in predicting the IFFs, with a Pearson correlation coefficient of 0.803. The most important features, as quantified using Shapley values, were Aid Percent of Gross National Income (GNI), control of corruption, and population. Taken together, these models and their interpretation provide a method for predicting the IFFs as well as the features that drive them, enabling policy makers to focus on these factors to decrease corruption.

Introduction:

Illicit financial flows (IFFs) are defined as any transfer of money that has been earned, moved, or stored illegally (Global Financial Integrity, 2020a), posing a major challenge to developing countries. These funds often pass through the control of corrupt officials or corporations, diverting resources away from benefiting the country's citizens and ending up in tax havens. It has been estimated that financial assets equal to 10% of global GDP are held by individuals in tax havens (Brandt, 2022). Since hiding financial assets through IFFs is practiced mostly by wealthy and powerful members of society in developing countries (Alstadsæter et al., 2019), this practice may lead to widening financial inequalities that arise from such corruption, in addition to other social issues. Additionally, knowledge of others evading taxes may lead to less tax compliance among the general populace (Alm et al., 2017).

In large supranational organizations such as the United Nations (UN) and Organization for Economic Co-operation and Development (OECD), efforts have been made to address the issue of IFFs (Olken & Pande, 2012). 44 countries have signed the OECD Anti-Bribery Convention, which came into effect in February 1999. Signatories included the 38 OECD countries and 6 other countries. The convention targets bribery of foreign public officials in order to foster equal grounds for international bribery prevention efforts. Similar efforts have been undertaken by the UN, with the adoption of the United Nations Convention Against Corruption (UNCAC), which came into effect in December 2005.

At the national level, analyzing and discovering the indicators of corruption can enable leaders to address the problem of illegal international money flows through policy. As an example, if a country has a high GDP per capita, it can be linked to a high amount of corruption in an area with high urbanization, policymakers can address this issue. Also, estimating levels of corruption in a region allows policy makers to focus there, and it also allows businesses to understand the difficulties they will face when operating in an area.

To identify factors leading to high IFFs, this paper uses multiple types of machine learning models - including regression, random forests, and neural networks - to predict IFFs given an initial set of features. I will provide an explanation of the models' predictions using various machine learning interpretation tools. For the data set features, I obtain political indicators defined by the Worldwide Governance Indicators (WGI), such as Voice and Accountability, Political Stability and Absence of Violence/Terrorism, Government Effectiveness, and others (Kaufmann et al., 2010). Global Financial Integrity (GFI) is a think-tank which keeps track of IFFs and various other financial crimes, and this is where I acquire my data about the levels of IFFs (Global Financial Integrity, 2020b), measured in millions of USD. Other social and economic factors in this dataset come from the study, "Transparency and corruption: Measuring real transparency by a new index" (Mungiu-Pippidi, 2022). For the purpose of standardization, I limit my analysis to the years from 2008 to 2018. These years were selected because this is the range in which data from all three datasets overlap.

Prior works have also used machine learning to predict corruption before, for both individual countries, such as Brazil and Spain, as well as globally (Colonnelli et al., 2020; Lima & Delen, 2020; López-Iturriaga & Sanz, 2018). However, these papers focus primarily on predicting local corruption, as opposed to cross-border flows of financial assets. Furthermore, these studies use area-specific features rather than the broad economic and political data which this paper uses.

Currently, there is not enough information about how political and economic metrics, such as those used in this study, affect IFFs. This study addresses this gap in knowledge by providing clarity about the relationship between these metrics and the levels of IFFs in order to give insight into which factors facilitate cross border illegal transactions. I use a combination of easily interpretable models, such as regression and random forests, as well as neural networks supplemented with Shapley values, to determine the most important factors driving each model's predictions, the features in the model.

Related Works-

The most similar works to this have been papers which have used artificial intelligence to predict corruption. Lima & Delen (2020) has been the most similar paper to this, since it tries to predict corruption globally. Another paper has been published analyzing corruption in Brazil at the municipal level (Colonnelli et al., 2020), and a paper also developed maps to predict corruption in Spain (López-Iturriaga & Sanz, 2018). However, none of these studies address the issue of predicting IFFs. In the field of analyzing IFFs, Brandt (2022) is very well written about the topic of IFFs in developing countries. It has definitions of IFFs and describes the causes of such illegal financial transactions. For additional information on IFFs, Collin (2020) is useful for understanding measurements of IFFs.

Methods:

Experimental Setup-

There are 134 countries in the dataset, with each country appearing 10 times in the dataset (one column for each year), resulting in the dataset having 1340 distinct data lines. I have set the year as a feature, to analyze if that might also contribute to predicting the level of IFFs. I replaced null data with averages of the features, so that we could run analysis and train ML models on the data. The outlier countries with an extremely high and low IFF rate have been excluded to make the functionality of the models much better, since including the extremely large outliers, such as China, made the models inaccurate. This has been done by excluding outliers as determined by the 1.5*IQR rule. This leaves the dataset with 1190 lines of data. As is commonly used for machine learning regression and classification tasks, 75% of the data went towards training and 25% went toward testing.

Machine Learning Models-

To carry out the analysis, I used several machine learning models, which will be explained below. I approached the problem by framing it as both a classification task as well as a regression task.

For the classification task, I first split the data into classes by using either the median or quartile values. The dataset split on the median was trained on the logistic regression model. Next, I split the dataset into 4 quartiles. The Random Forest Classifier was used on this dataset.

For the regression task, I first used Linear Regression, which is the most simple regression model. This was done as a preliminary experiment, to gauge the feasibility of the project. This model simply tries to fit the variables onto a line of best fit and predicts the values from that line of best fit. However, since it just fits values onto a line of best fit, it has limited ability to comprehend the features. Next, I tried to use a support vector machine (SVM) for regression on the dataset. The SVM model was first introduced by Boser, Guyon, & Vapnik (1992). The model is based on a combination of factors that were learned from training data. During training, a theoretical hyperplane is generated from the data. All the training instances, which are called support vectors, influence the format and position of the hyperplane, which is a linear line (Zhang et al., 2015). The SVM has better results than the linear regression model because the SVM is more effective in high dimensional spaces. Essentially, the training instances reduce the dimensionality of the features on the hyperplane and then use these simplified features to draw a linear line, which is used to predict the data.

After this, I decided to use the Random Forest model to perform regression, because it had very promising results in the quartile classification. Due to having similar promising results the same as the Random Forest Classification, the feature importances were extracted for the results section below. One of the most important attributes of the Random Forest Regression is its capability to show which features were most important. This feature was especially useful to pinpoint which statistics contributed to predicting the IFF values. The model's good performance is likely due to its ability to store much information between the individual trees, and the lack of collinearity between the separate decision trees. After this, an Artificial Neural Network (ANN) was developed for the regression task. It is the most complex model that was developed for predicting the IFFs. To visualize the ANN model, SHapley Additive exPlanations (SHAP) were used, since it is an effective way to understand ANNs (Rozemberczki et al., 2022). The SHAP values come from game theory and are used to assign contributions to the features in the ML Model. These will be displayed below in the results section.

Results:

Features-

I constructed Violin Plots (Figures 1, 2, 3 and 4) to show the distribution of some important features. The ones being displayed show the IFFs and most important features in the ML models. As I will discuss shortly, the most important features for the performance of the models are Aid Percent of Gross National Income (GNI), Population, and Gross Domestic Product Per Capita (GDPPC). Note that this is being done with the same dataset that has been used for training the ML models.

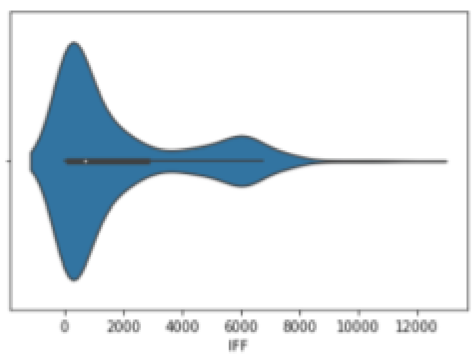

Figure 1

This graph displays the distribution of the IFFs. I hypothesize that this distribution has occurred because of patterns of corruption in developing countries.

Figure 1 displays the distribution of the IFFs. Some interesting trends to note are that there is a large concentration of values around 350 Million USD, and a smaller concentration of countries around 6,000 Million USD. There also seems to be a significant outlier in the dataset, near the 13500 Million USD mark, even though only values in the IQR Range are included. I hypothesize that this has occurred because of two cases with developing countries. The larger concentration is the vast majority of developing countries, which have low IFF levels because the economy of these countries is too small, or the government has cracked down effectively upon these activities. Another possibility might be that political officials, through public companies, might transport money overseas. This doesn't count in IFF statistics, since IFF statistics take into account 'the spirit of the law' when calculating these statistics (Brandt, 2022).

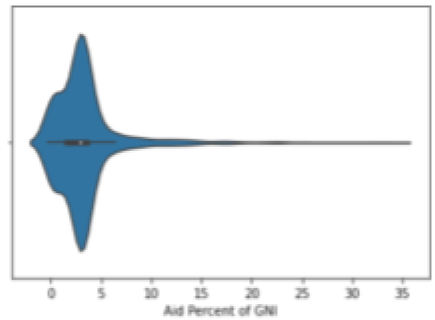

Figure 2

The Aid Percent of GNI (Gross National Income), has played a significant role in the model results. In the Random Forest Regressor and the ANN, this was the most important feature.

This factor, the Aid Percent of GNI (Gross National Income), has played a significant role in the model results (Figure 2). While most of the countries lie in the 0-5% range, some outliers are present, which have extremely high levels of Aid percent of GNI. This might be due to the foreign aid being embezzled by the people in charge and might be a reason for the ineffectiveness of foreign aid (In'airat, 2014). When aid money gets embezzled, it is usually sent or invested overseas, which is an IFF. This is very prevalent in countries undergoing civil unrest or countries which do not have strong governance (In'airat, 2014).

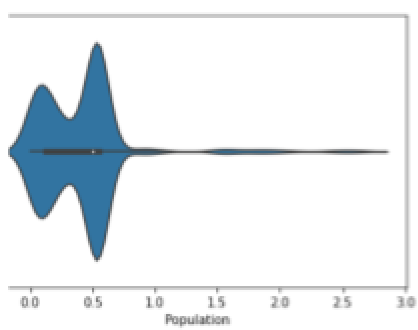

Figure 3

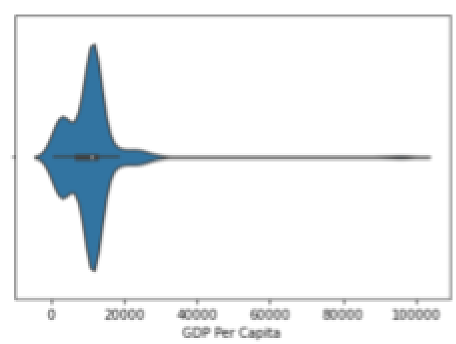

Figure 4

Population has also been one of the more important factors in both of the best models; it is the 3rd most important factor in the ANN. Intuitively, it makes sense that a relatively high GDPPC would lead to lower IFFs in developing countries. However, it is not clear why that is not the case in this model.

Population has also been one of the more important factors in both of the best models; it is the 3rd most important factor in the ANN, as can be observed in Figure 3. Population was the factor which was most surprising, since there is no intuitive way to point out the connection between population and a high rate of IFFs. I have chosen to not include the Control of Corruption feature in this, due to the feature being quite explanatory. GDPPC (Gross Domestic Product Per Capita) was not a strongly predictive feature in the Random Forest Regressor and the ANN. Intuitively, it makes sense that a relatively high GDPPC would lead to lower IFFs in developing countries (Figure 4). However, it is not clear why that is not the case in this model. Most countries have a GDPPC from 5000 to 15000. GDP Per Capita has a somewhat similar spread of distribution to 3 features: Voice and Accountability, Government Effectiveness, and Regulatory Quality. This is consistent with the conclusions of (In'airat, 2014). This could be an interesting case for further study.

Model Results-

Linear Regression had a Pearson correlation coefficient of 0.359. Contrary to what was expected, The SVM only performed marginally better than the linear regression, with a Pearson correlation coefficient of 0.366 (Table 1). This was surprising because one advantage of the SVM over the ANN is that the SVM is less prone to overfitting because it is based on structural risk minimization (Lima & Delen, 2020). The vector coefficient model will not be discussed in this section due to poor results, Pearson correlation coefficient of 0.366. The Random Forest Regression model performed very well (Table 1), with a Pearson correlation coefficient of 0.756. With the Artificial Neural Network regressor model, the Pearson Correlation Coefficient of 0.803 was achieved (Table 1). The mean absolute error was 913.11, and the root mean squared error was 1482.35. Table 1 is visualized in Figure 5.

Model Type (Regression Models)	Performance (Pearson Correlation Coefficient)
Linear Regression	0.359
Support Vector Machine (SVM)	0.366
Random Forest Regressor	0.756
Artificial Neural Network Regressor	0.803

Table 1

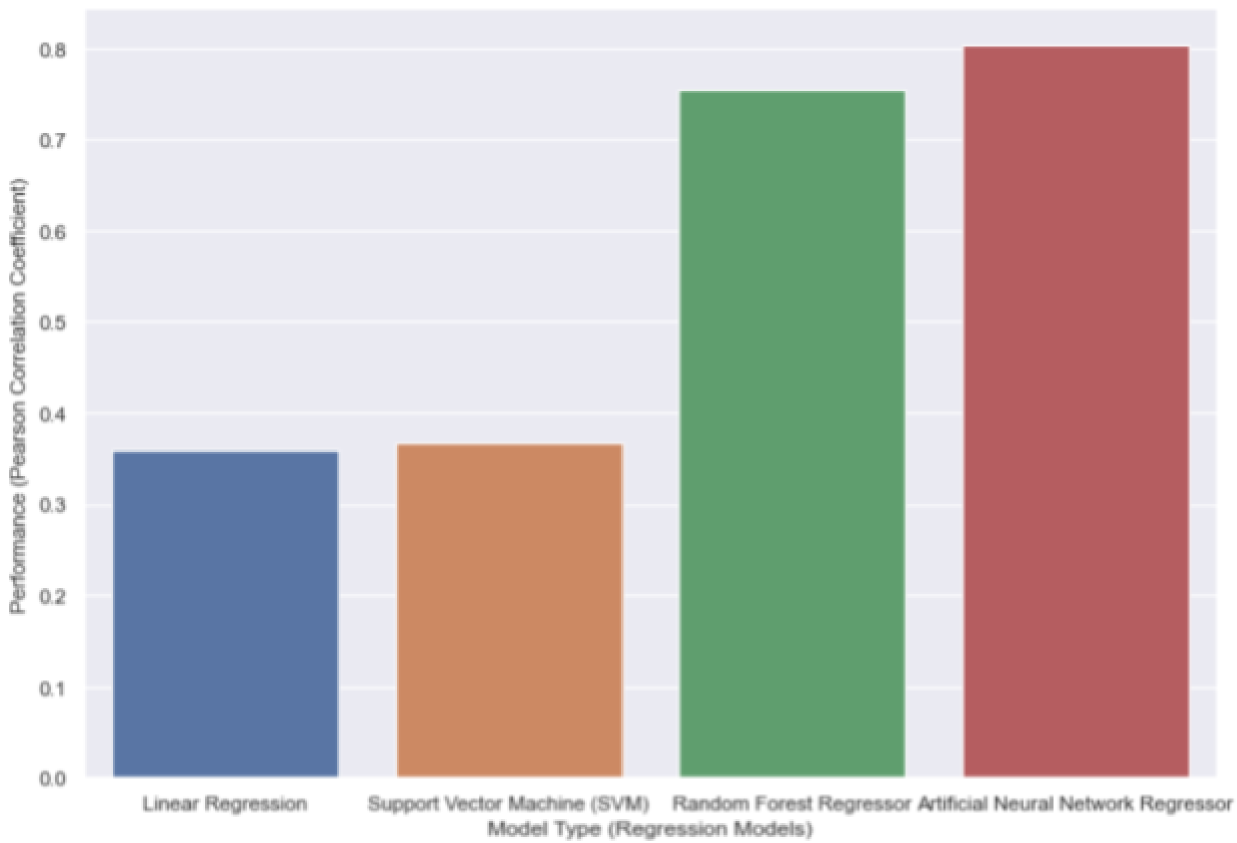

Figure 5

In the performance of our Machine Learning models when predicting the numeric quantity of IFFs, the ANN had the best results, with a Pearson Correlation Coefficient as a metric for evaluation. This was followed closely by the Random Forest Regressor. The Support Vector Machine and the Linear Regression were the least accurate models.

The logistic regression performed at the same level as the linear regression, with a weighted average F-1 score of 0.56 (Table 2). When doing the logistic regression by binarizing along the mean, the model just predicted everything as belonging to one class. Therefore, it can be concluded that logistic regression does not do very well when trying to predict values binarized by the mean, due to outliers greatly skewing the model. This did much better, even though it had more classes to predict, having a weighted average F-1 score of 0.75 (Table 2). Random Forest had a good accuracy on the dataset, with the first and third quartiles having the best individual F-1 scores of 0.78. Quartiles 2 and 4 were predicted less accurately, with F-1 scores of 0.73 and 0.74, respectively. Due to the good F-1 scores, I visualized the feature importances in the model. The last classification model was the Artificial Neural Network classifier, which had a weighted average F-1 score of 0.85 (Table 2). In later regression models, a more accurate model was developed with the ANN, which made the classification tasks less explanatory, because the numeric value of the IFF could explain more than simply the quartile to which the country belongs. Table 2 is visualized in Figure 6.

Model Type (Classification Models)	Performance (Weighted Average F-1 Score)
Logistic Regression (Split along median)	0.56
Random Forest Classifier (Split into Quartiles)	0.75
Artificial Neural Network Classifier (Split into Quartiles)	0.85

Table 2

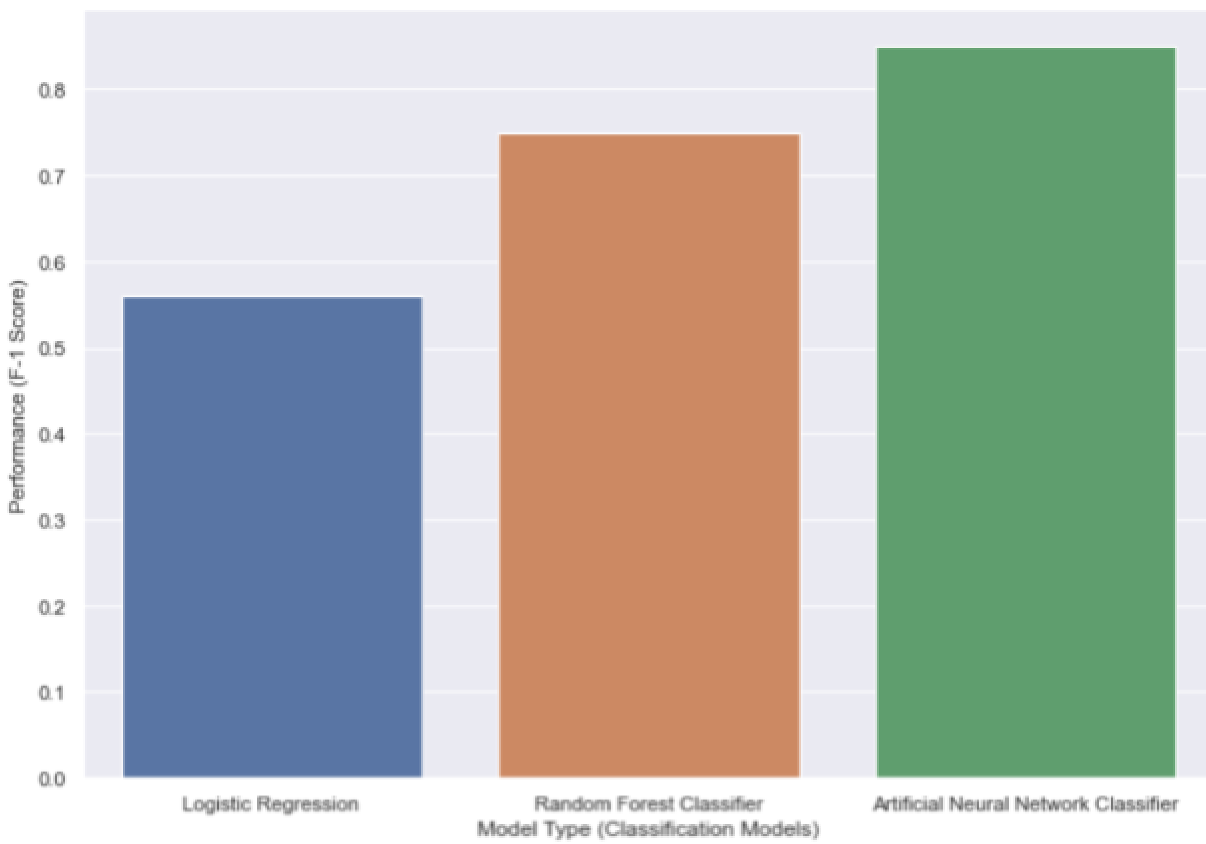

Figure 6

In the performance of our Machine Learning models when predicting IFFs split along the median or into quartiles, the ANN had the best results, with a F-1 score as a metric for evaluation. This was followed closely by the Random Forest Classifier. The Logistic Regression was the least accurate model.

Discussion-

Model Feature Loadings-

Model feature loadings were acquired from three separate models: the Random Forest Classifier (Figure 7), Random Forest Regressor (Figure 8) and Artificial Neural Network (Figure 9).

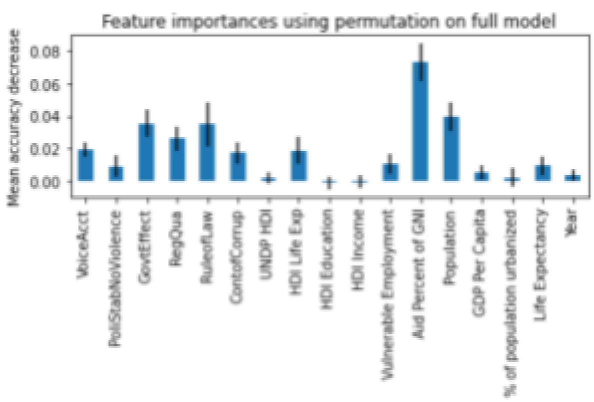

Figure 7

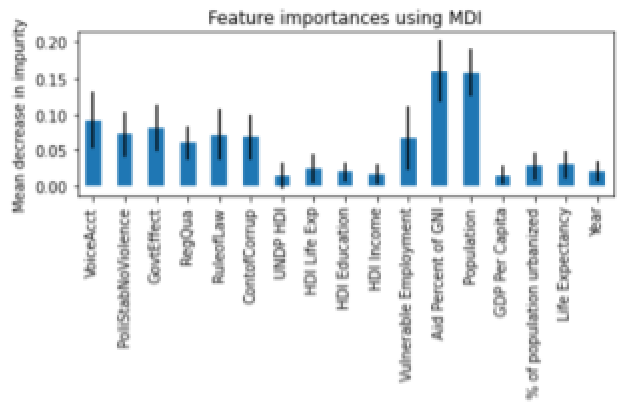

Figure 8

The figure on the left shows Feature Importances using permutation on full model for the Random Forest Classifier. The figure on the right shows Feature Importances using MDI for the Random Forest Regressor. It can be observed that the most important factors (Aid Percent of GNI, Population, Control of Corruption, Government Effectiveness, and Rule of Law) in the Random Forest Regressor are also important features in the ANN.

For the Random Forest models, only the feature importances in terms of how much they contributed to the Mean Decrease in Impurity (MDI) was displayed. The important features of the Random Forest Regressor and Classifier were also important in the ANN. Therefore, it is logical to conclude that the features which led to an increase in the rate of IFFs in the ANN model would do the same in the Random Forest Regressor.

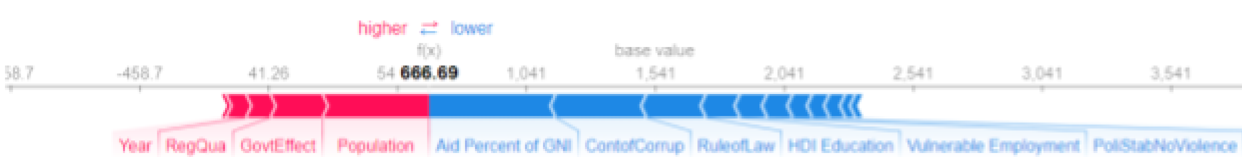

Figure 9

Above is the feature importances for the significant features in the Artificial Neural Network. As noted before, the most important features which predict a higher rate of IFFs are Aid Percent of GNI, Control of Corruption, and Rule of Law. The most important features that predicted a lower rate of IFFs were the Population, Government Effectiveness, Regulatory Quality, and the year in which the IFF was being predicted.

Above in Figure 9 are the feature importances for the significant features in the Artificial Neural Network. As noted before, the most important features which predict a higher rate of IFFs are Aid Percent of GNI, Control of Corruption, and Rule of Law. There are other factors which play a smaller role, such as the Human Development Index (HDI) Education ratings for each country, the percent of population which has Vulnerable Employment (not having consistent wages, etc), and the Political Stability and Absence of Violence. The most important features that predicted a lower rate of IFFs were the Population, Government Effectiveness, Regulatory Quality, and the year in which the IFF was being predicted.

As shown above in Figure 7, Figure 8, and Figure 9, the most important feature was the amount of Aid Percent of GNI. The country with the highest rate of this is Yemen, followed by similar war-stricken low-income countries, mostly in Africa and the Middle East. In the Random Forest Regressor (Figure 9) and the ANN (Figure 7), this was one of the most important features. This might be due to the foreign aid being embezzled by the people in charge and might be a reason for the ineffectiveness of foreign aid (In'airat, 2014). When aid money gets embezzled, it is usually sent or invested overseas, which is an IFF (Brandt, 2022). This is very prevalent in countries undergoing civil unrest or countries which do not have strong governance (In'airat, 2014). For example, data from Madden (2020), shows that for the $2 Trillion that comes into Africa as foreign direct investment or official development assistance, $1 Trillion left the country in the form of IFFs. Also, it should be noted that financial assets or money can change multiple mediums of exchange before it becomes an IFF. Therefore, it is likely that there is a connection between the levels of IFFs and how much aid consists of the GNI. This highlights some issues with the way FDI (Foreign Direct Investment) and EDA (Economic Development Aid) are provided to lower income countries. Both FDI and EDA are counted as a part of the aid in the feature above.

For population, there is no intuitive way to point out the connection between population and a high rate of IFFs, in Figures 7, 8, and 9. However, we can notice that there are two large concentrations of population, around 50 million and 10 Million.The relationship between population and IFFs might be simply because there are more opportunities to take money overseas with a higher population. In the ANN (Figure 7), a high population indicated high levels of IFFs, while a high GDP (Gross Domestic Product) Per Capita indicated a low levels of IFFs.

The Control of Corruption variable was very important in the ANN (Figure 7), which indicated that it led to higher amounts of IFFs. This is intuitive, since corruption is very often the cause of money being illegally transferred across national borders. This feature has a central spread around -0.4, with a sharp drop off after the center of distribution, before the center of distribution, it starts increasing from -2.0. Higher levels of Government Effectiveness lead to lower rates of IFFs (Figures 7,8 and 9), since efficient government prevents money being used for non-state purposes. Most of the IFFs happen due to improper use of government funds and the inability of governments to properly track its expenditures. From this, we can assume that the study which formulated this feature (Mungiu-Pippidi, 2022) does quite a good job in capturing the levels of government control on corruption. In the study of IFFs, trade misinvoicing has largely not been used to track IFFs (Brandt, 2022). Yet, a connection can be made between government effectiveness and trade misinvoicing, since the lack of government controls of trade can lead to trade misinvoicing and even IFFs being noted as legitimate trade, especially by multinationals (Forstater, 2018). Further evidence can be found in the example Côte d'Ivoire coffee exports (Forstater, 2018). Côte d'Ivoire has a low Government Effectiveness rating. The distribution of the Government Effectiveness feature is similar to that of the control of corruption, with a center of distribution around -0.3. The rule of law variable is the 3rd most predictive feature of high IFFs. The rule of law has a clear connection to corruption, since enforcement of existing laws is the first deterrent to IFFs. As noted in the introduction, less respect for the rule of law promotes more IFFs and domestic tax avoidance. This feature has a lot of overlap with the Government Effectiveness feature, but they each predict opposite levels of IFFs. This has a similar center of distribution and overall distribution to the government effectiveness feature. The rest of the features had a relatively minimal impact on the rate of IFFs.

We can also observe that effects have been made in reducing IFFs, since as time goes on the predicted number of IFFs go down in the model (Figure 7). This is a relatively small factor in the model, which is to be expected, since these anti-IFF efforts take quite a bit of time to start showing observable results in the model. IFFs have been identified as a major issue by the international community during the early 2000s, and this dataset encompasses the years 2009-2018. Therefore, the results of this dataset show that efforts such as the United Nations Convention Against Corruption (UNCAC), and OECD (Organization for Economic Co-operation and Development) Anti-Bribery Convention have had some success in combating IFFs. This might be a point for further research, to gauge the effectiveness of these anti-IFF efforts more accurately.

Conclusion:

With the success in predicting IFF levels, it is now possible to successfully estimate the level of IFFs in a region or country using machine learning. This will help in addressing corruption, since it is necessary to know where corruption is present in order to mitigate it (Colonnelli et al., 2020). I have found that the amount of foreign aid, population, and measures of effectiveness of government functionality were very important in predicting levels of IFFs. Rather than basing international efforts to mitigate IFFs on traditional means, using ML models to predict IFFs is a more reliable method to curb this issue. Attempts to curb IFFs have promoted many reforms in government bureaucracy, spending, and other areas. However, this paper shows that there might be some other areas that should be addressed, such as the mismanagement of aid that is sent to developing countries. For future work, I plan to incorporate other datasets, which could enable us to predict IFFs more accurately. This could be further expanded upon through the development of more sophisticated neural networks, which would help produce more accurate results.

References:

[1] Alm, J., Bloomquist, K. M., & McKee, M. (2017). When You Know Your Neighbour Pays Taxes: Information, Peer Effects and Tax Compliance. *Fiscal Studies*, *38*(4), 587–613. https://doi.org/10.1111/1475-5890.12111
[2] Alstadsæter, A., Johannesen, N., & Zucman, G. (2019). Tax Evasion and Inequality. *American Economic Review*, *109*(6), 2073–2103. https://doi.org/10.1257/aer.20172043
[3] Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, 144–152. https://doi.org/10.1145/130385.130401
[4] Brandt, K. (2022). Illicit financial flows and developing countries: A review of methods and evidence. *Journal of Economic Surveys*, *37*(3), 789–820. https://doi.org/10.1111/joes.12518
[5] Collin, M. (2020). Illicit Financial Flows: Concepts, Measurement, and Evidence. *World Bank Research Observer*, *35*(1), 44–86. https://doi.org/10.1093/wbro/lkz007
[6] Colonnelli, E., Gallego, J. A., & Prem, M. (2020). *What Predicts Corruption?* (SSRN Scholarly Paper 3330651). https://doi.org/10.2139/ssrn.3330651
[7] Forstater, M. (2018). *Illicit Financial Flows, Trade Misinvoicing, and Multinational Tax Avoidance: The Same or Different?*
[8] Global Financial Integrity. (2020a). *Illicit Financial Flows*. https://gfintegrity.org/issue/illicit-financial-flows/
[9] Global Financial Integrity. (2020b). *Trade-Related Illicit Financial Flows: Data by Country* [dataset]. https://gfintegrity.org/iif-data-by-country/
[10] In'airat, M. (2014). Aid allocation, selectivity, and the quality of governance. *Journal of Economics Finance and Administrative Science*, *19*(36), 63–68. https://doi.org/10.1016/j.jefas.2014.03.002
[11] Kaufmann, D., Kraay, A., & Mastruzzi, M. (2010). The Worldwide Governance Indicators. *World Bank Policy Research*, *Working Paper No. 5430*. https://ssrn.com/abstract=1682130
[12] Lima, M. S. M., & Delen, D. (2020). Predicting and explaining corruption across countries: A machine learning approach. *Government Information Quarterly*, *37*(1), 101407. https://doi.org/10.1016/j.giq.2019.101407
[13] López-Iturriaga, F. J., & Sanz, I. P. (2018). Predicting Public Corruption with Neural Networks: An Analysis of Spanish Provinces. *Social Indicators Research*, *140*(3), 975–998. https://doi.org/10.1007/s11205-017-1802-2
[14] Madden, P. (2020, March 2). *New trends in illicit financial flows from Africa*. Brookings. https://www.brookings.edu/articles/new-trends-in-illicit-financial-flows-from-africa/
[15] Mungiu-Pippidi, A. (2022). Transparency and corruption: Measuring real transparency by a new index. *Regulation & Governance*. https://doi.org/10.1111/rego.12502
[16] Olken, B. A., & Pande, R. (2012). Corruption in Developing Countries | Annual Review of Economics. *Annual Review of Economics*, *4*, 479–509.

[17] Rozemberczki, B., Watson, L., Bayer, P., Yang, H.-T., Kiss, O., Nilsson, S., & Sarkar, R. (2022). *The Shapley Value in Machine Learning* (arXiv:2202.05594). arXiv. https://doi.org/10.48550/arXiv.2202.05594