Grace Saunders, Ravza Aykan, Mariam Seshan
DS 3001: Foundations of Machine Learning
May 07, 2025

Final Paper

Abstract/Executive Summary

This study investigates the determinants of Official Development Assistance (ODA) allocations by exploring which economic and societal indicators best predict the amount of aid a country receives. Using data from the World Bank's World Development Indicators, we construct and compare two Random Forest regression models to forecast lagged Net ODA received (in constant 2021 USD).

In the first model, we use a broad set of economic and societal indicators, limited to the most relevant topics in key domains of economic health, governance, and human development. This model achieved strong performance, with an R^2 of 0.76 on training data and 0.53 on test data. However, upon closer analysis, we found that redundancy and multicollinearity likely contributed to overfitting.

To improve interpretability, we construct a second model using a curated subset of 29 non-redundant indicators. This refined model achieved slightly lower predictive performance (train $R^2 = 0.73$; test $R^2 = 0.49$), but offered clearer insights. The most important predictors included domestic credit to the private sector by banks (as a percentage of GDP), infant mortality, and the use of IMF credit. Both models produced residuals generally centered around zero, though with long tails indicating rare deviations, possibly driven by geopolitical or humanitarian crises not captured in the economic data.

Our results suggest that while economic indicators can reasonably predict aid allocation patterns, careful feature selection is essential for generalizability and interpretability. The presence of outliers and the model's modest predictive power on unseen data also highlight limitations of econometric approaches to capturing complex aid dynamics. This analysis offers a foundation for future research to enhance the predictive modeling of ODA flows and analysis of policy decisions related to ODA provision.

Introduction

Foreign aid remains a critical component of international development and diplomacy, yet the factors behind aid allocation are often unclear when compared to stated motivations. Although various donor nations and international organizations claim to base aid on developmental need or economic hardship, empirical patterns often reflect a blend of humanitarian, economic, and geopolitical considerations. This paper investigates the economic and societal indicators most predictive of Official Development Assistance (ODA) allocation using machine learning techniques applied to cross-national panel data.

Using World Bank data, specifically the World Development Indicators Dataset, we built two Random Forest models. Both predict as their target variable a one-year lagged measure of net ODA received (in constant 2021 USD), but differ in the features used as predictors. The first model uses a broad set of indicators filtered for availability but not redundancy. While this model performs well, further analysis reveals that multicollinearity and duplication may inflate its predictive accuracy. To address this, we constructed a second model using a curated set of 29 distinct and relevant variables specifically chosen to avoid overlap and improve interpretability.

Our findings show that while both models capture meaningful variation in ODA receipt, preprocessing decisions such as feature selection substantially influence model performance and clarity. Across models, indicators related to private sector credit, infant mortality, and reliance on IMF loans emerge as particularly important, hinting at how donors may respond to signs of economic stress, underdevelopment, or governance challenges.

This paper proceeds by detailing our data sources and methodology, presenting results from both models, and discussing the implications for research on foreign aid allocation. We conclude by reflecting on the limitations of our approach and suggesting avenues for future work that might integrate additional political or crisis-related variables.

Data

This study uses data from the World Bank's World Development Indicators (WDI), a global database, which compiles data from multiple international sources like the UN, IMF, and the OECD, to create over 1,000 time-series indicators covering topics such as economic development, health outcomes, and financial flows. The WDI provides annual data for most countries beginning in 1960, though coverage varies by indicator, year, and region. We limited our study to only 2010-2020, since many indicators were missing data for earlier years and we expected that the COVID-19 pandemic would have potential effects on aid flows.

Data Selection

To model the determinants of Official Development Assistance (ODA), we began by downloading the full WDI dataset and extracting country-year level observations. The target variable for prediction was Net ODA received (constant 2021 USD), which was obtained from the WDI under the "Economic Policy & Debt: Official development assistance" topic. This variable was lagged ahead by one year, so that the model would predict the amount of aid that a country would receive based on indicators from the previous year.

Data Processing

After filtering the entire dataset in its original long format to exclude indicators missing over 50% of their observations, we filtered to exclude variables with keywords directly related to international aid flows to avoid circular reasoning in the model. We also removed the sections of the dataset focused on regions, including only country-level observations. After filtering, we reshaped to a wide format with one row per country-year combination. We then dropped any rows for which the target indicator on net ODA received was missing, and created the time lagged version of the ODA variable. Because of its skew, we used an inverse hyperbolic sine transformation on the target variable. We addressed missing values for other columns using zero imputation, accompanied by binary indicators to flag missingness. Finally, we used the standard scaler to normalize the variables.

Methods

Overview

To explore which country-level indicators best predict foreign aid allocations, we used a supervised machine learning approach using Random Forest regression, as we are using a high-dimensional dataset with likely non-linear relationships and interactions between predictors. We developed two models with different feature sets to evaluate the impact of indicator selection on predictive performance and interpretability.

Feature Engineering

Model 1 (Broad Feature Set):
We kept all indicators from the WDI dataset that had at least 50% non-missing observations across all country-year pairs. While this increased coverage and flexibility, it also introduced redundancy, such as indicators measured in both current and constant terms or multiple versions of similar economic metrics.

Model 2 (Curated Feature Set):
To improve interpretability and reduce collinearity, we manually selected 29 non-redundant indicators based on relevance to aid allocation. These included measures related to macroeconomic conditions (e.g., GDP per capita, inflation), human development (e.g., infant mortality, literacy), governance (e.g., corruption perception), and debt exposure (e.g., IMF credit usage).

Model Training

Both models used the RandomForestRegressor from the scikit-learn library. We tuned hyperparameters such as the number of trees, maximum depth, and minimum samples per split to improve model fit.

To reflect real-world forecasting, we used a time based split: data from years up to and including 2017 were used for training, and 2018 onward was reserved for testing. This ensured temporal separation (i.e., the preventing the model learning patterns from future data that would not be available at the time of prediction).

Evaluation Metrics

We assessed model performance using the Root Mean Squared Error (RMSE) to measure absolute prediction error and the R-squared (R^2) to assess how much variation in the target variable is explained by the indicator set. We used Kernel Density Estimates (KDE) of residuals to visualize prediction error distributions and diagnose potential bias or skewness. We also extracted and plotted feature importances to determine which indicators were most influential in determining predicted aid amounts. These measures of model performance allow us to compare both the accuracy and interpretability of each model.

Results

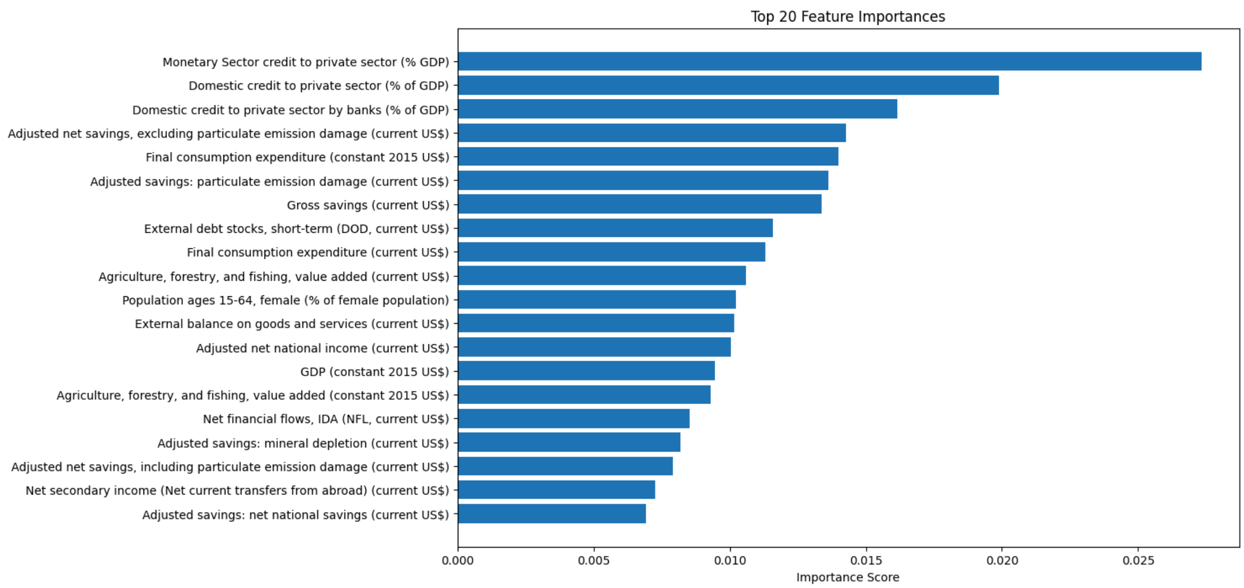
Model 1

Our first model used a broad array of indicators from the dataset. After filtering for missingness (retaining indicators with at least 50% coverage), we trained a Random Forest Regressor to predict one-year lagged Net ODA received. The model was trained on data up to 2017 and tested on data from 2018 onward.

Model 1 achieved strong in-sample performance, with an RMSE of 2.59 and an R^2 of 0.76 on the training set. On the test set, the model yielded an RMSE of 4.04 and an R^2 of 0.53, indicating a reasonable ability to generalize, though with some performance drop.

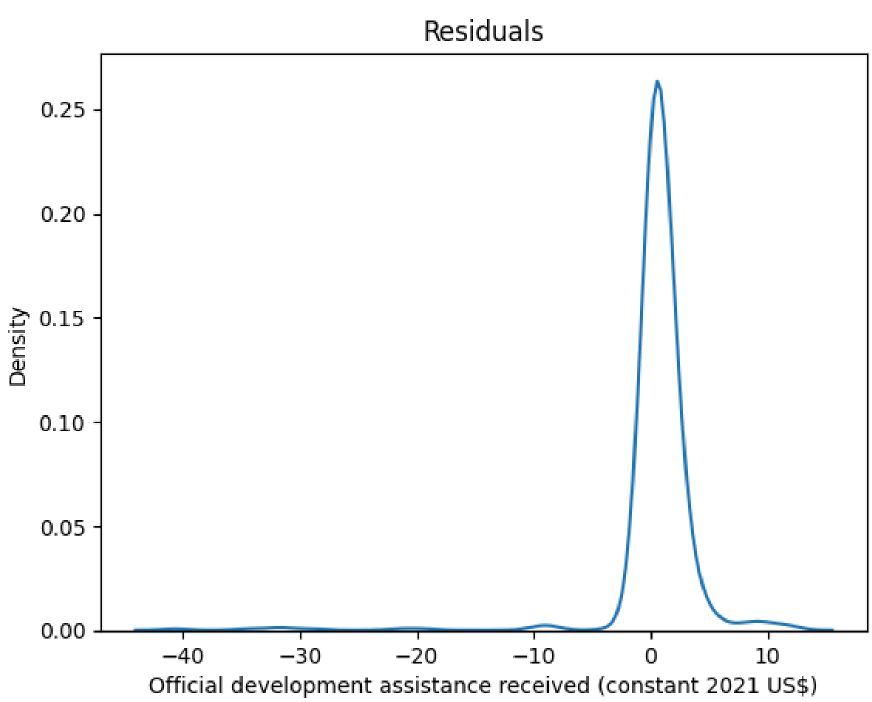
To understand what was driving the model's predictions, we visualized the top 20 most important features, which were heavily concentrated in GDP-related metrics and savings measures. This suggests that economic size and savings behavior are closely correlated with aid receipt.

Model 1, Top 20 Feature Importances (Github Link)



To further evaluate model performance, we created the residual KDE plot below, which showed that residuals were generally centered around zero, indicating that the model does not systematically over- or under-predict ODA. However, the distribution displayed long, asymmetric tails, with extremely low-density values extending to approximately -40 on the left and 10 on the right.

Model 1, KDE Residuals Plot (Github Link)



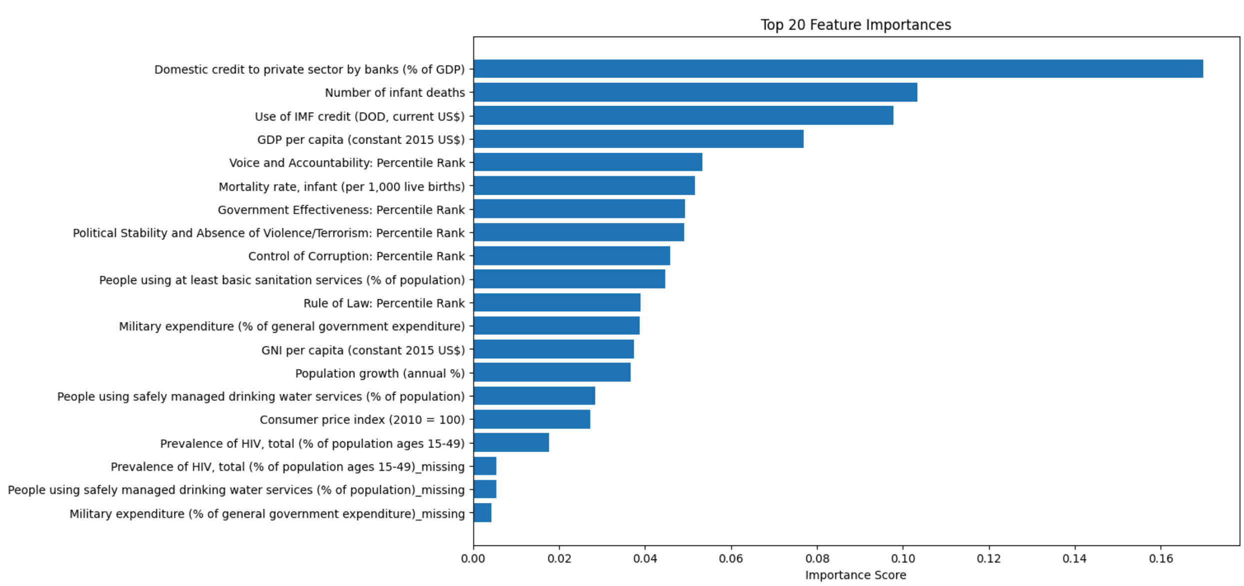
While these extreme residuals were rare, their presence suggests that there are outliers in aid receipt for a few country-year cases, which could point to atypical geopolitical or humanitarian events that are not captured by the economic indicators used in the model. Despite these few extreme values, the overall distribution of residuals is centered and consistent, suggesting that the model is not systematically over- or under-predicting for most countries.

However, upon further inspection of the feature importance plot, we found that this first model included redundant features, such as indicators measured in both constant and current USD or overlapping metrics for similar concepts, which introduced noise and multicollinearity and may have artificially inflated our models performance. Almost all of these indicators were related to, or components of, a country's GDP. To address this, we refined our approach by selecting a targeted subset of 29 distinct and relevant indicators that avoided duplication.

Model 2

We constructed a second model using the curated set of indicators. These included variables representing macroeconomic conditions, population health, governance, and debt burden. Model 2 achieved slightly weaker performance than Model 1, with a training RMSE of 2.73 and R^2 of 0.73, and a test RMSE of 4.22 with R^2 of 0.49. This decline likely reflects a reduction in overfitting and offers a clearer view into which distinct indicators actually matter for predicting aid.

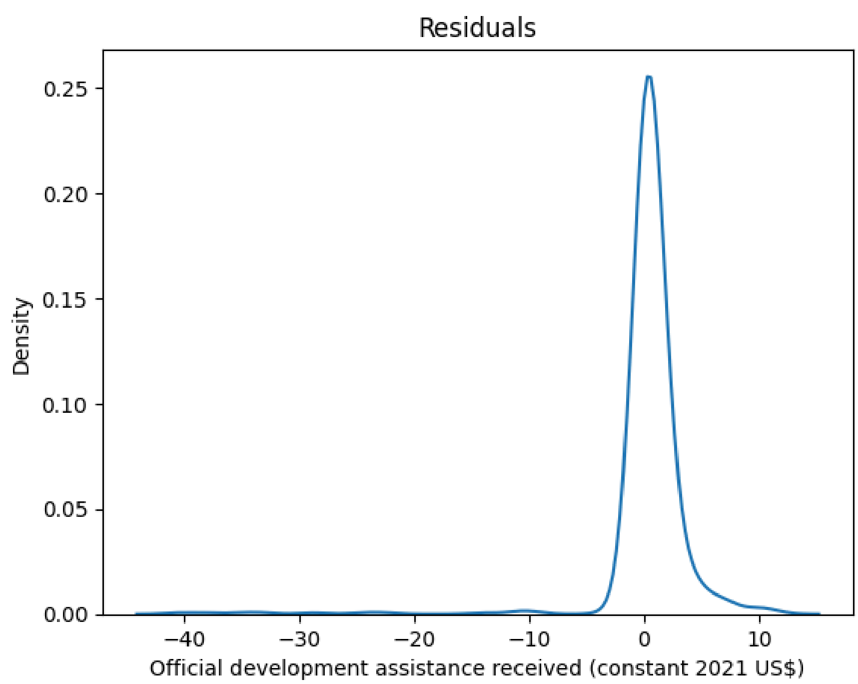
Model 2, Top 20 Feature Importances (Github Link)



We found that, after removing redundant and irrelevant indicators, the top three most important features were domestic credit to the private sector by banks (as a percentage of GDP), the number of infant deaths, and the use of IMF credit (debt outstanding, current USD). The rest of the top 20 features are related to government stability and corruption, and the health of the population.

To evaluate model bias and error behavior, we again analyzed the distribution of residuals. The KDE plot showed a similar peak centered near zero, with slightly thinner tails compared to Model 1. This suggests that the curated feature set may have helped reduce some of the prediction errors seen earlier.

Model 2, KDE Residuals Plot (Github Link)



Summary

Model	Feature Set	Train R^2	Test R^2	Top Feature Domains
1	Broad, partially redundant	0.76	0.53	1. Monetary sector credit to private sector (3 measures) 2. Adjusted net savings (2 measures) 3. Final consumption expenditure
2	Curated, distinct variables	0.73	0.49	1. Domestic credit to private sector by banks 2. Number of infant deaths 3. Use of IMF credit

While both models successfully captured variation in ODA allocation, they reflect different trade-offs between predictive accuracy and interpretability.

Model 1, with a broader and partially redundant feature set, achieved slightly higher performance (Train $R^2 = 0.76$, Test $R^2 = 0.53$). However, many of the top features were overlapping economic indicators, and this redundancy may have artificially boosted performance by capturing similar signals multiple times.

Model 2, in contrast, performed slightly worse (Train $R^2 = 0.73$, Test $R^2 = 0.49$) but relied on a more curated and conceptually distinct feature set. Its top predictors, domestic credit to the private sector, number of infant deaths, and use of IMF credit, offer clearer interpretations and align with theoretical expectations around how donors may respond to signs of financial instability, public health challenges, or debt vulnerability.

Both models produced residuals centered near zero but with long tails, reflecting errors tied to unpredictable aid spikes, likely from crises or political events. While development indicators capture general patterns, donor decisions are also influenced by factors beyond standard economic data. Our findings also show that feature selection plays a major role in shaping both model performance and interpretability.

Conclusion

This study set out to identify which measures of economic and societal well-being are most useful in predicting the amount of Official Development Assistance (ODA) a country receives. By applying machine learning techniques, specifically Random Forest regression models, to the World Bank's World Development Indicators dataset, we assessed how different economic and social indicators contribute to patterns of foreign aid allocation.

In our first model, we employ a broad set of indicators filtered only for data coverage. While this model achieved relatively strong performance (test $R^2 = 0.53$), further investigation revealed that the inclusion of redundant or overlapping variables likely introduced multicollinearity and inflated performance metrics. In response, we constructed a second model using 29 curated and distinct indicators. This model performed slightly worse on unseen data (test $R^2 = 0.49$), but offered much clearer insights into the drivers of aid.

Across both models, several predictors appeared consistently. Indicators related to financial sector development (especially domestic credit to the private sector), human development (e.g., infant mortality), and financial distress (e.g., IMF credit usage) ranked among the most predictive features. Interestingly, governance indicators such as corruption perception and political stability also appeared among the top features, implying that institutional quality may play a role in aid allocation decisions, either directly or indirectly through its effects on economic conditions.

While the models do not provide causal explanations, they offer several valuable takeaways. First, preprocessing decisions, particularly around feature selection, play a major role in shaping model outcomes. The shift from a broad to a curated feature set revealed trade-offs between predictive power and interpretability. Second, the residuals of both models displayed long tails, indicating the presence of extreme outliers likely corresponding to years in which a country received unusually high or low aid due to crises, conflict, or sudden political shifts. These outliers point to the limitations in predicting aid: political, humanitarian, or strategic factors may significantly alter aid patterns in ways that standard development metrics cannot fully capture.

There are several potential extensions and future directions for this work. One would be to incorporate political or crisis-oriented variables, such as flags for armed conflict, refugee flows, or diplomatic relationships, which are likely to influence foreign aid but are absent from standard economic datasets. Another extension could involve disaggregating aid by donor or sector, enabling more targeted analysis of thematic priorities (e.g., health, education, infrastructure).

Finally, while this paper focuses on predictive performance, future research could combine machine learning with causal inference techniques to better understand not just which, but how certain indicators are associated with aid receipt.

References

World Bank Group. (2024, April 21). *World Development Indicators*. Data Catalog. https://datacatalog.worldbank.org/search/dataset/0037712/World-Development-Indicators