Using Machine Learning to Predict Inter-State Affinity

Alden Felix

Wei-chen Huang

Jim Pan

Samuel Adelusi

May 3, 2023

Contents

Abstract

Machine learning has been able to accurately predict conflict at the local and state level. This study examines if machine learning can be used to make accurate predictions of generalized relationships between states, with the United States as the base of reference. International relations were measured using an aggregate affinity score derived from event data, and various socioeconomic and demographic factors were used as predictive variables. Multiple models, including a regression and classification random forest and a deep learning model, were built and evaluated in regards to their accuracy in predicting interstate affinity. The strongest predictors of interstate affinity varied significantly depending on the model.

Introduction

Studies on why states engage in conflict and who they engage in conflict with, who they trade with, how and why multinational organizations such as the United Nations come about, and numerous other subfields that seek to explain the relationship between international actors are not new. However, advancements in data science have led conflict prediction and network studies using machine learning to become an extremely prevalent area of focus within international relations studies.

While a party's motivation for studying conflict can change depending on the interests of the involved groups, some of the reasons include improving state defense or making policies, academic interest in the field, or international organizations with any number of social and economic interests that are affected by conflict in an area. This might explain why predicting conflict has become such a large and heterogeneous field.

One of the motivations behind this study is the Violence Early-Warning System (ViEWS) that aims to predict the probabilities of armed conflict in the next 36 months using machine learning (Helle et al., 2018). We wanted to generalize this study to instead focus on the general relationships between states, in a sense combining the previously mentioned subfields of international relations. Therefore this study uses a multitude of social and economic variables based on previous literature and the team's own experience to predict affinity between a pair of states.

Literature Review

The existing literature on using machine learning to predict conflict is expansive and new methodologies and sources of data are continuously being explored. One of the most important ideas found in the existing literature and more generally in machine learning studies is that “the selection of appropriate machine learning methodologies can offer substantial improvements in accuracy and performance” (Perry 2013). Therefore, understanding the benefits and drawbacks of a method and what would best fit a dataset is crucial in building good models. In addition, machine learning models offer the ability to include a substantial amount of variables without concern of violating assumptions seen in classical methods such as OLS. Of the variables used in predicting conflict, the existence of violence in the previous time period is often the most significant predictor of violence in the next period. However, studies find that socioeconomic and demographic variables alone can offer as good of predictive performance as conflict history (Bazzi et al., 2022).

Network studies are another important aspect to understanding the causal forces behind conflict and international relations. Although not used in this study, a viable method of discovering networks of similar states, or affinity communities, are clustering methods applied to situations such as United Nations voting, where state preferences are revealed (Pauls et al., 2017). These clusters can then be used as independent or dependent variables. Forming the dependent variable in this way allows for more specific aspects of state relationships to be explored.

Methods and Data

The starting point for our study was finding a reliable source of coded event data that we could use to compute a value for affinity between a pair of states. The two viable sources we initially identified were the Integrated Crisis Early Warning System (ICEWS) Coded Event Data database, and the University of Texas at Dallas Event Data database. However, the UT Dallas event database was ultimately not used over stability concerns stability in terms of retrieving data with the API. These databases automatically scrape events from online sources and detail interactions between individuals, groups, and states, sorted by year. The events are coded by the parties involved, a brief descriptor of the event, location of the event, the source that the event was published in, and an intensity value. The intensity value ranges from -10 to 10, describing the effect of that event on the relationship between two actors, with any use of violent force being automatically categorized as -10. With this dataset, we created an aggregate statistic called affinity score, which is the mean of all intensity scores in events that happened between two countries in any given year.

Using this computed affinity score shaped the study moving forward in a significant way, as this value became the dependent variable which we predicted in our machine learning model. The simplicity in computing affinity in this way meant we could focus on other aspects of the study we felt were important such as choosing the independent variables in our model, as literature stated the significance of this step in well built machine learning algorithms. However, there were drawbacks in using such a simple computational method which will be discussed later.

We obtained our socioeconomic and demographic data from various sources, including the World Bank and their World Development Indicators, the UN, and the International

Monetary Fund for trade data. In gathering our independent variables we were focused on including a wide variety of different socioeconomic and demographic variables that were included in previous literature and others that we believed could be useful based on our own domain knowledge. Our original list of variables included a few variations of the same or related values, such as GDP per capita, GPD, GDP PPP, etc., as we knew that we could prune variables later on based on their performance in the model. A significant problem we encountered in this stage of data collection was the amount of missing data in the databases we came across, particularly the WDI indicators from which the majority of our independent variables came from. Removing any countries that had missing values significantly diminished the number of observations in our dataset to the point that we believed it was better to impute the missing values, although we did also build a machine learning model on the dataset with missing data removed.

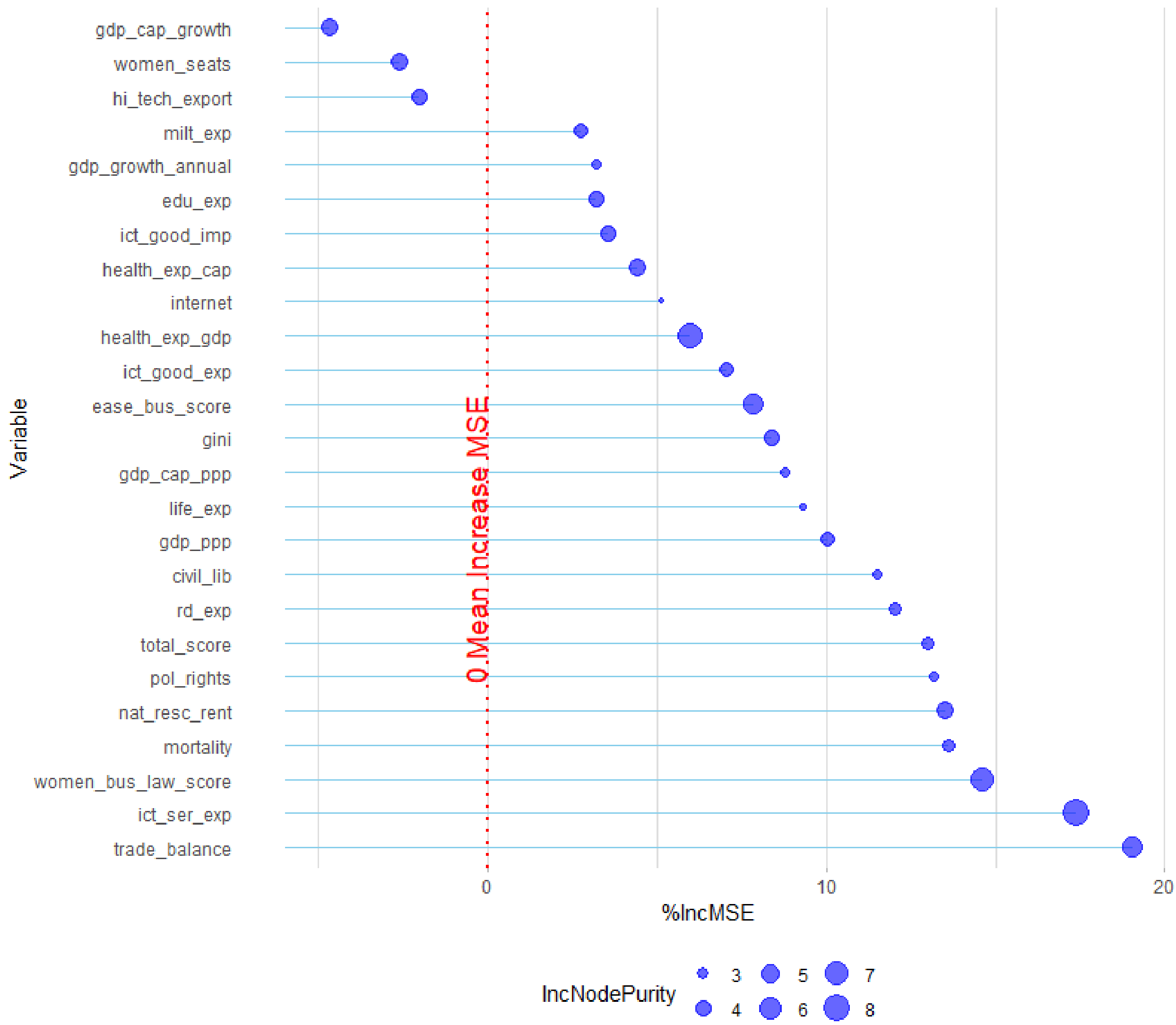
The method we used for imputing missing values was k-nearest neighbor (KNN) imputation from the 'VIM' library. After research on different imputation methods we believed that KNN imputation would be the most robust while still maintaining enough simplicity to not significantly hinder progress. KNN imputes a missing value by "finding the samples in the training set 'closest' to it and averages these nearby points to fill in the value" (Kuhn et al., 2013). We also tested soft imputation for a matrix using nuclear-norm regularization from the 'softImpute' library, of which the finer details won't be discussed here, but the results were not satisfactory enough to be considered. Overall, KNN imputation provided results that we felt were normal enough to build a machine learning model on.

The machine learning model we started with was a random forest built using the 'caret' library. The decision to use a random forest came from past experience with the method, and a literature review on machine learning for conflict prediction which showed robust results from random forests. Using such a well studied and applied method served as a good starting point so that we could be confident that our results were more due to variable specification and the inclusion or absence of observations rather than specification of hyperparameters such as the mtry value and number of trees. This is not to say that hyperparameters were not thoroughly tested, as the ‘caret’ library has functions to test various mtry values.

Due to the low initial R^2 of the random forest regression model, we decided to try a deep learning model to improve prediction.

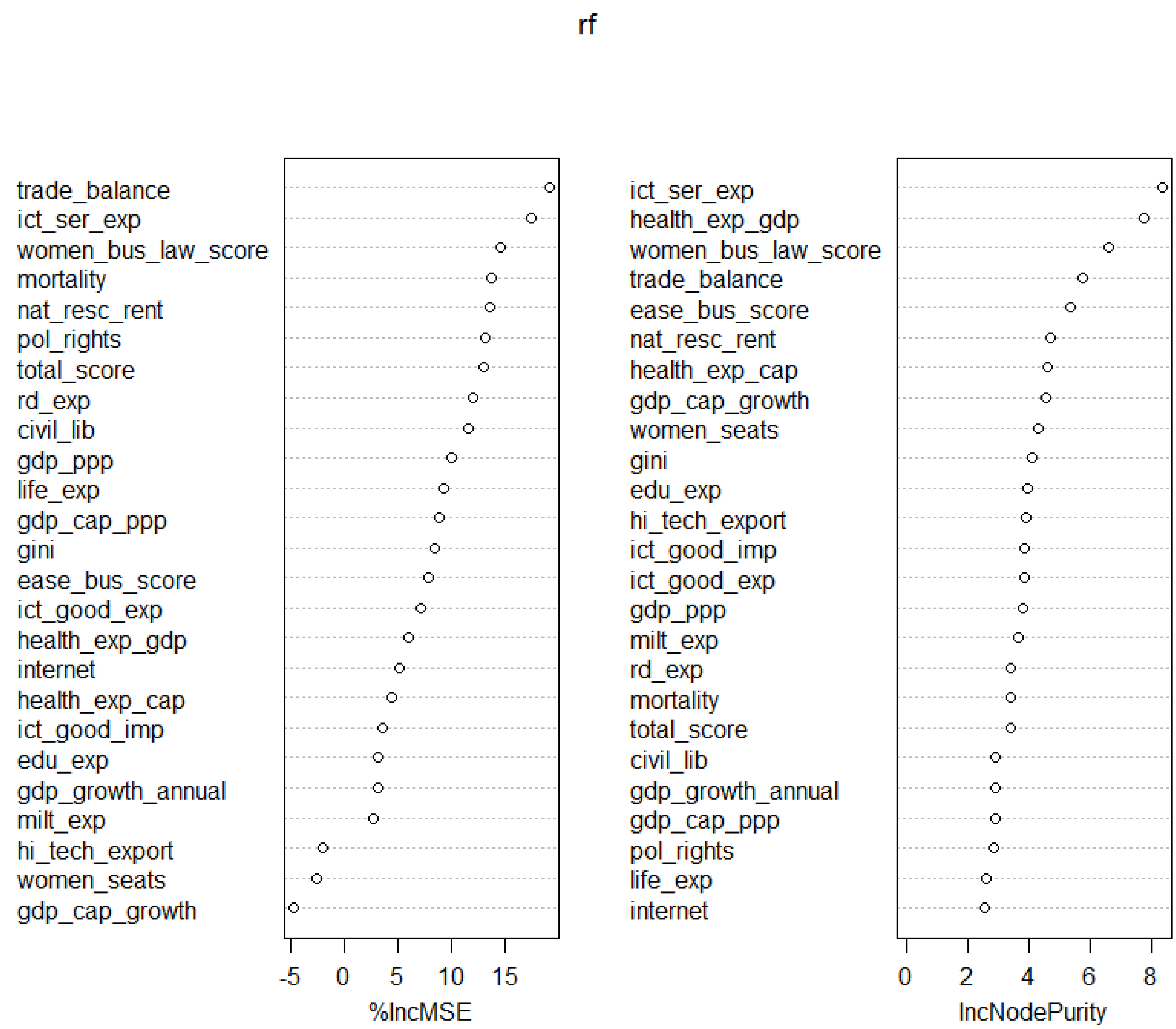
Analysis

Our first model was a random forest regression. After building the model with all variables and generating a variable importance plot, any variables with a negative value for increase in mean square error were removed and a new model was built. Variables that were highly correlated with each other that did not have proper justification for remaining in the model were also removed. The percent increase in mean square error term indicates how much the mean square error would increase by if that variable were to be removed, so any variables with a negative value indicate that the model would actually benefit from their removal. In addition, the model was set up to use 10 fold cross validation resampling and pick an mtry value that gave the smallest root mean square error. The initial increase in mean square error and node purity plot can be seen below.

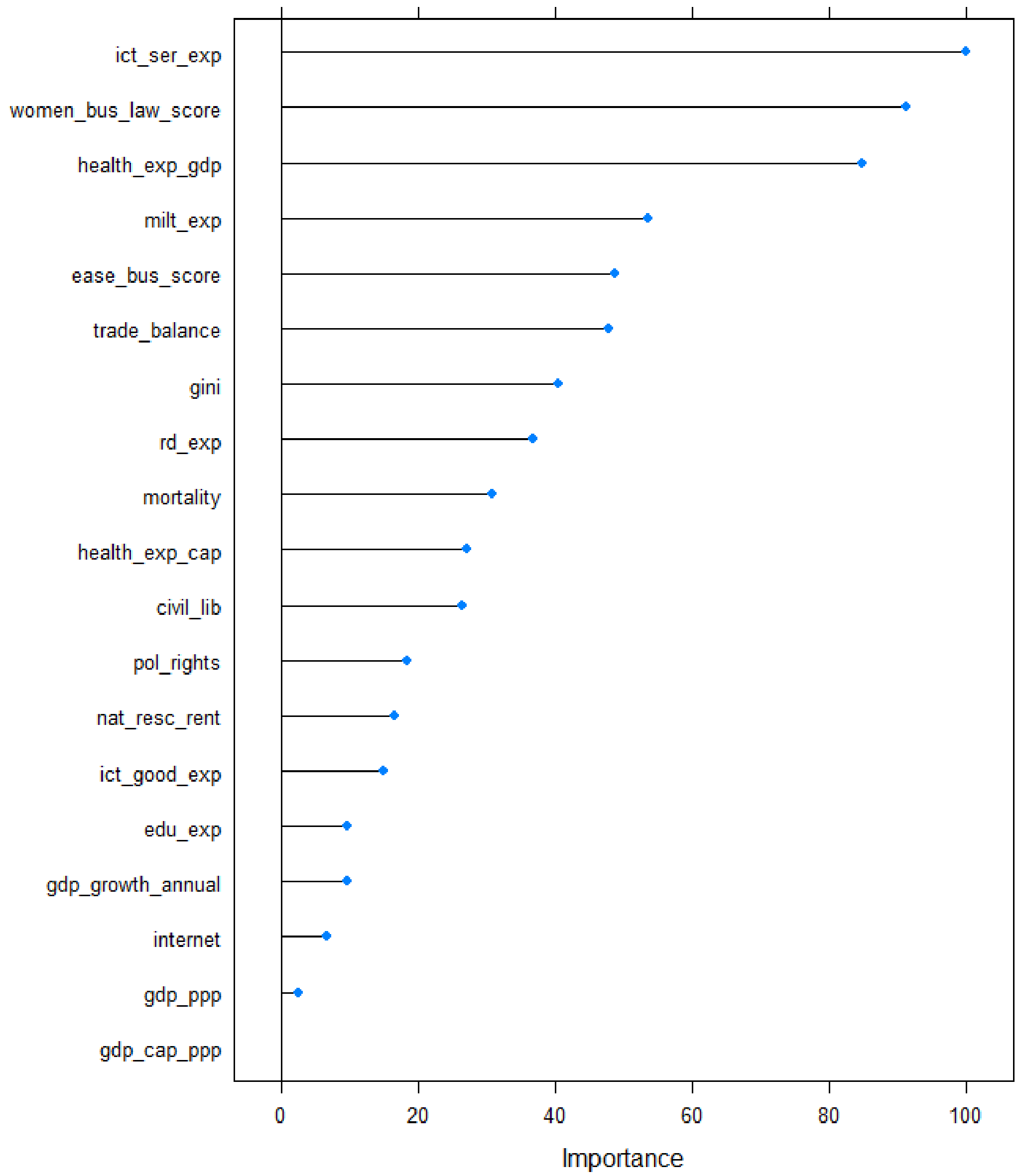


gdp_cap_growth
women_seats
hi_tech_export
milt_exp
gdp_growth_annual
edu_exp
ict_good_imp
health_exp_cap
internet
health_exp_gdp
ict_good_exp
ease_bus_score
gini
gdp_cap_ppp
life_exp
gdp_ppp
civil_lib
rd_exp
total_score
pol_rights
nat_resc_rent
mortality
women_bus_law_score
ict_ser_exp
trade_balance
Variable
0 Mean Increase MSE
0
10
20
%IncMSE
IncNodePurity
3
4
5
6
7
8

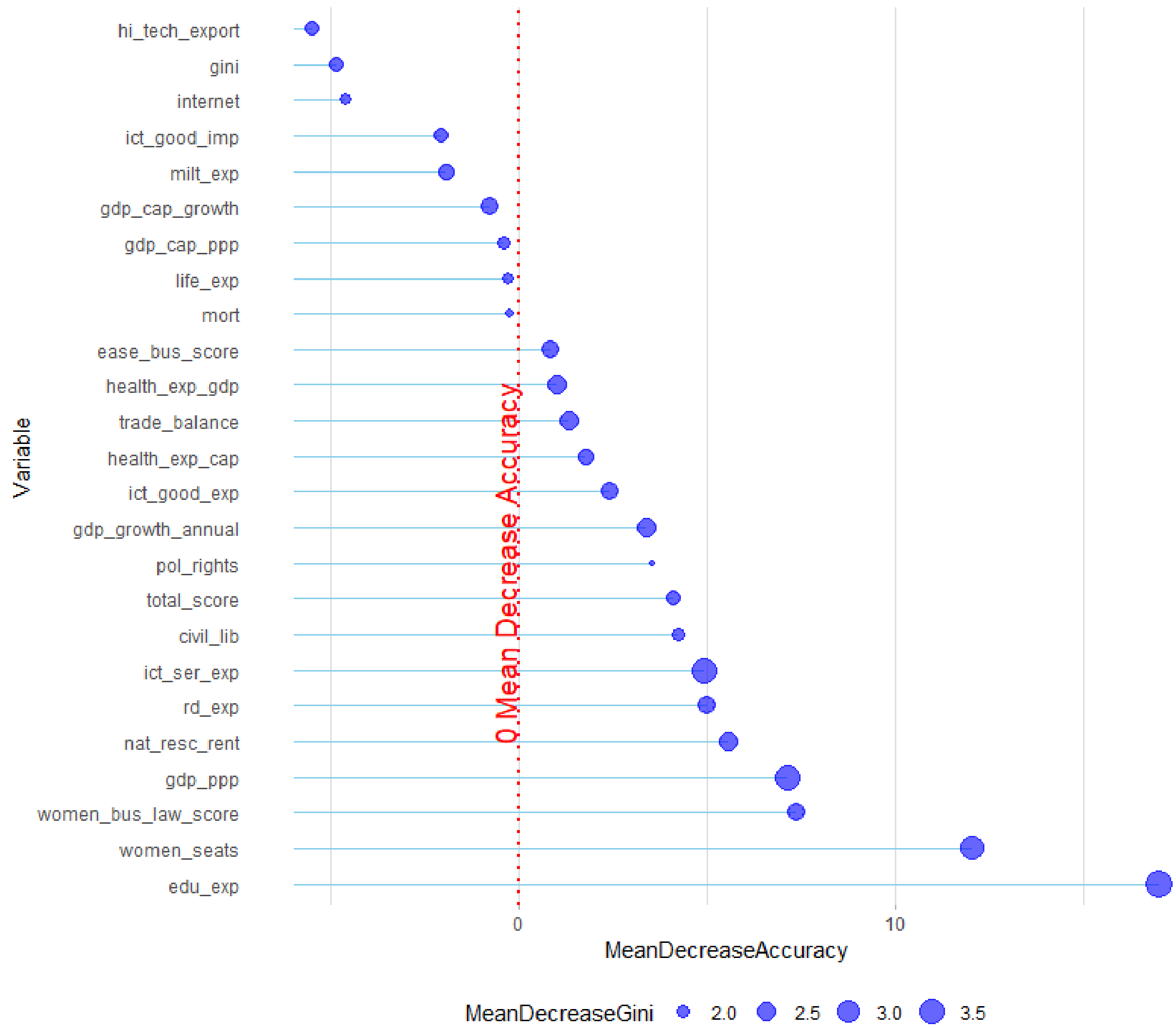
After removing the variables with negative increase in mean square error and those that were highly correlated with each other, the variables that were left in the model and the respective mean square error and node purity plot can be seen below.



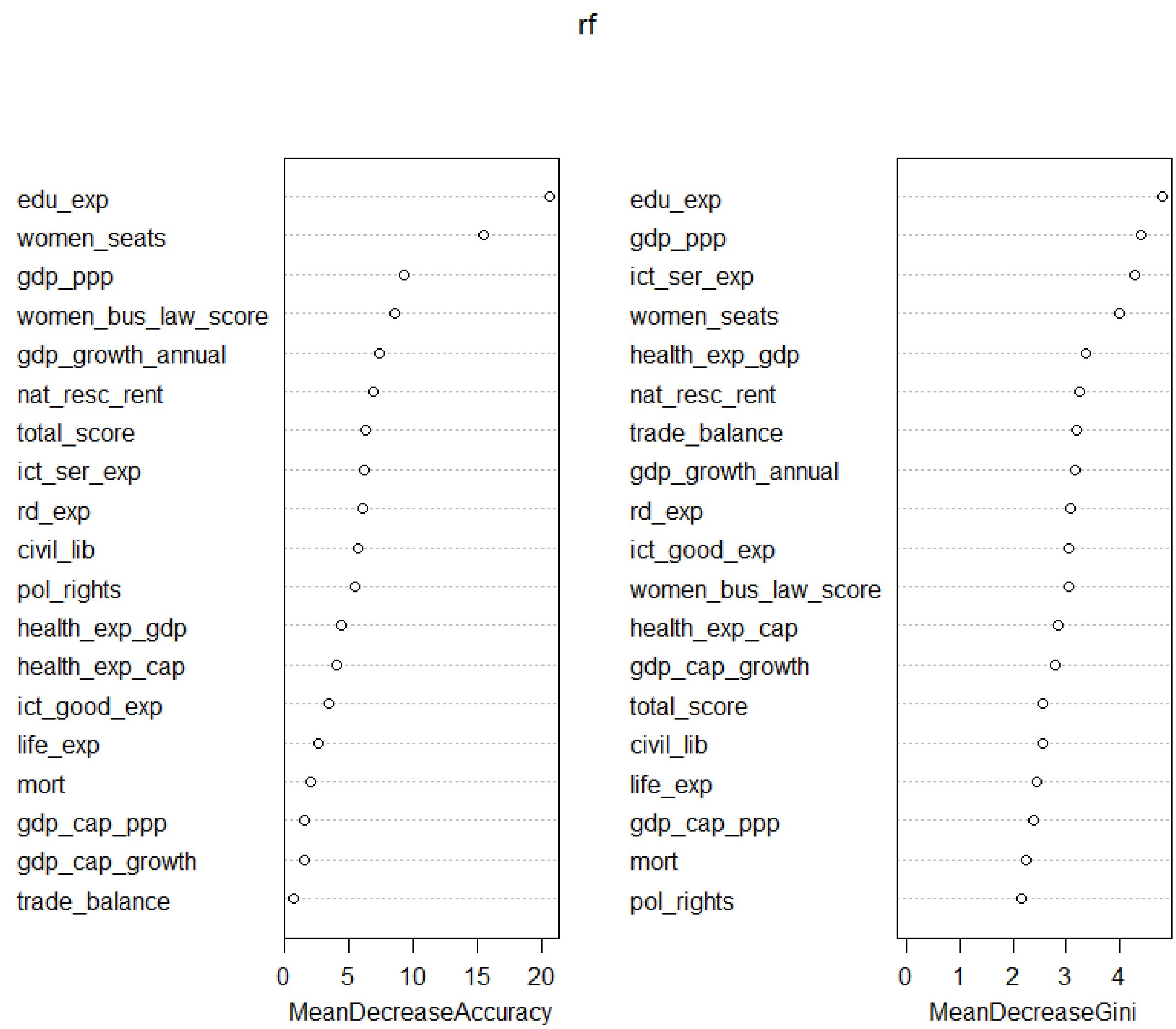
The final regression model's R^2 value was about 0.27. The low R^2 indicates that the model is not good at predicting affinity between states. The most likely problems with this model are the variable specification and low variance of the small dataset which it was trained on. However, as the first model it served as a good baseline for further development to explore what changes could be implemented to improve the predictive accuracy. A variable importance plot for the model with values scaled between 0 and 100 can be seen below.



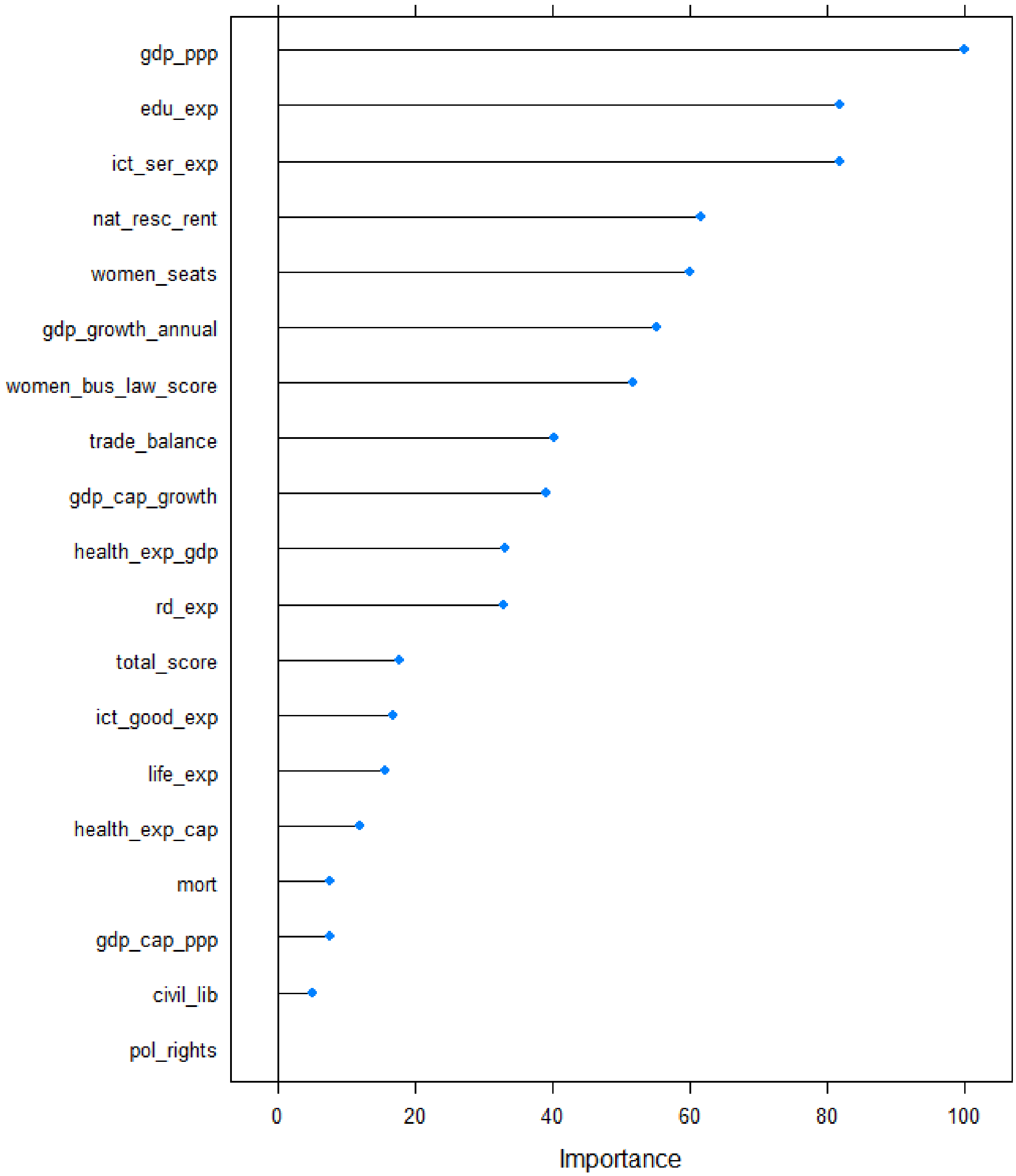
The next step we took towards improving model accuracy was turning the numerical dependent variable, affinity, into a categorical variable. This would reduce the amount of information available in the variable, but with enough categories the variable would still provide a useful interpretation on the relationship between two states. The new affinity variable now had integer categories from -1 to 3. The initial classification random forest's mean decrease accuracy and mean decrease gini plot can be seen below. Compared to the random forest regression model there are notable differences in the significance of some variables, while others seem to stay constant. There also seem to be more variables with a negative mean decrease in accuracy value.



The remaining variables in the model after removing those with negative mean decrease in accuracy and the respective mean decrease accuracy and mean decrease gini plots can be seen below.



The final random forest classification model had an R^2 value of about 0.43. Although this was decently higher than the regression model, it can not be said that it is very good at predicting affinity between states either. A variable importance plot for the model with values scaled between 0 and 100 can be seen below. Once again, there are significant differences in the variable importance between the regression and classification trees.



Another model we made was a deep learning model built using the Keras and tensorFlow libraries in python. It has four hidden layers, with 256, 128, 128, and 256 nodes respectively, and classified countries into three groups: negative (affinity less than -0.75), neutral (affinity between -0.75 and 0.75), and positive (affinity greater than 0.75). The first build we created had an accuracy of 0.57, which then increased to 0.76 after hyperparameter tuning.

```
Epoch 199/200
11/11 [==============================] - 0s 4ms/step - loss: 0.5551 - accuracy: 0.7560
Epoch 200/200
11/11 [==============================] - 0s 4ms/step - loss: 0.5550 - accuracy: 0.7560
3/3 [==============================] - 0s 5ms/step - loss: 0.6478 - accuracy: 0.7619
Test accuracy: 0.761904776096344
```

Despite the large increase in R^2, we were hesitant to use this deep learning model for reliable predictions due to the small sample size of the training dataset, as we feared that the model would be prone to overfitting. We attempted to analyze the strength of each variable using the SHAP library, but was unable to as the SHAP library only supported image classification for deep learning models. Another method we tried was the IntegratedGradient function in the Keras library. We ran into the same problem, as for deep learning models, the function was meant to plot out how much each pixel of the image contributed to the final output, and used only integers from 0-4. This does give us some small insight regarding the strength of variables but is very minimal.

```
Received a label value of 4 which is outside the valid range of [0, 3).  Label values: 2 3 3 1
 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 1 3 4 2 3 3 3 4 4 3 2 3 0
```

Conclusion and Implications

In the scope of this study the random forest regression and classification models did not have very good predictive accuracy, and the deep learning model was not able to be appropriately trained due to a lack of a large enough dataset. However, even these initial models showed enough promise to justify further research on using machine learning to predict relationships between states. This is especially true when considering that the dependent variable used in this study, a very simple computation of affinity, can be vastly improved upon to more accurately represent interstate relations. As mentioned in the literature review Pauls et al., 2017 use clustering methods on UN voting to find affinity communities and then go on to use these communities as an independent variable. In the same way a clustering or network study method could be used to compute a significantly more informative and accurate dependent variable.

There are also recent developments in the social sciences that can contribute greatly to this study and similar studies. First of these developments is the new POLECAT dataset, which is the successor to the Integrated Conflict Early Warning System database used in this study. The most significant improvements of this new database is its use of machine learning in the form of a “combination of NLP tools, transformer-based neural networks, and actor information sourced from Wikipedia”, that “offer substantial improvements in the scope and accuracy of political event data in terms of the what, how, why, where, and when of domestic and international interactions” (Halterman et al., 2023). Having the context behind an event is especially crucial in transforming the affinity score used in this study into a more valuable metric for studying international relations. The inclusion of more news sources in languages other than English also contributes to forming a more accurate representation of international events.

Similar to the expansion of the event database, the independent variables dataset can be grown to include more states and variables leading to more variance of the data and increasing the model's robustness. With other sources of data, future studies can also decrease the amount of missing values and imputation needed to be done.

At a worldwide scale, an accurate predictor tool for state affinity and conflict would hopefully prove to be useful for many parties. Government officials could use predictions to brace a country for war, and anti-war parties could use a prediction of future war to coax government officials into peace talks. Affinity and conflict predictions could also be utilized by third parties in a meaningful manner. Businesses and commoners could change their spending patterns to brace for economic hardships from international disputes or war. In addition, humanitarian organizations such as Medecins Sans Frontieres could prepare aid responses ahead of time.

Synergy Report

Alden Felix - Team Coordinator

Helped with initial brainstorming, literature review, and data gathering and wrangling. Contributed to project proposal and proposal presentation. Built and tuned random forest regression and classification models. Also contributed to the final paper and presentation.

Wei-chen Huang

Helped with the initial brainstorming, data gathering, including scraping data from various sources, and variable inclusion and pruning. Contributed to project proposal and presentation. Built, trained, and fine-tuned various classification models in TensorFlow and interpreted results. Contributed to the final paper and presentation as well.

Jim Pan

Assisted with the initial brainstorming, data gathering, and data wrangling for the project. Additionally, contributed to both the project proposal and the proposal presentation. Improved the data frame and assisted with imputing. Finally, also contributed to the final presentation.

Samuel Adelusi

References

Kuhn, & Johnson, K. (2013). *Applied predictive modeling*. Springer.

Halterman, A., Bagozzi, B. E., Beger, A., Schrodt, P., & Scraborough, G. (2023, April 3). PLOVER and POLECAT: A New Political Event Ontology and Dataset. https://doi.org/10.31235/osf.io/rm5dw

Helle, V., Negus, A.S., & Nyberg, J. (2018). Improving armed conflict prediction using machine learning : ViEWS+.

Pauls, & Cranmer, S. J. (2017). Affinity communities in United Nations voting: Implications for democracy, cooperation, and conflict. *Physica A*, *484*, 428–439. https://doi.org/10.1016/j.physa.2017.04.177

Perry. (2013). Machine Learning and Conflict Prediction: A Use Case. *Stability (Norfolk, VA)*, *2*(3), 56–. https://doi.org/10.5334/sta.cr

Bazzi, Blair, R. A., Blattman, C., Dube, O., Gudgeon, M., & Peck, R. (2022). The Promise and Pitfalls of Conflict Prediction: Evidence from Colombia and Indonesia. *The Review of Economics and Statistics*, *104*(4), 764–779. https://doi.org/10.1162/rest_a_01016