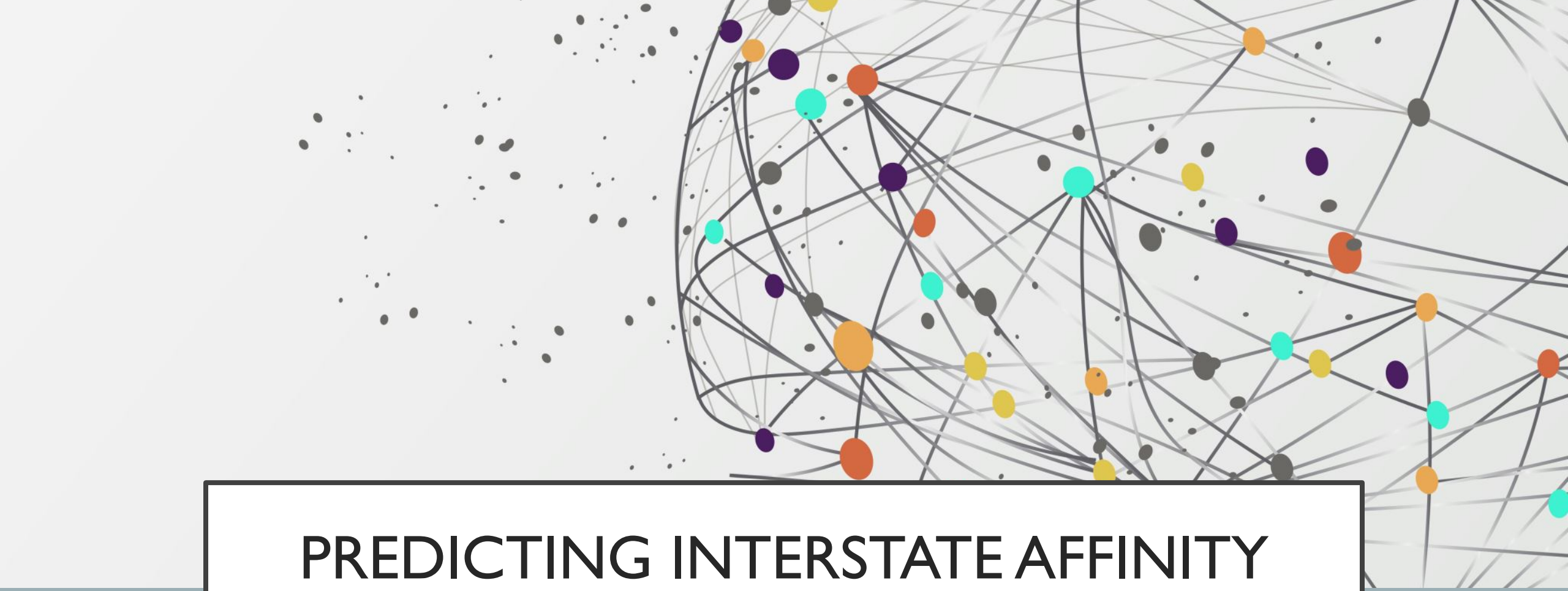


PREDICTING INTERSTATE AFFINITY USING MACHINE LEARNING

Alden Felix, Wei-chen Huang, Jim Pan, Samuel Adelusi

AREA/TOPIC

- Inspired by Dr. Vito D'Orazio – predicted violent conflict in Africa using machine learning models
- Wanted to generalize globally
- Predict inter-state affinity with machine learning using political, social, economic. and demographic data

Event/Affinity Data

- Integrated Crisis Early Warning System
  - ICEWS: Integrated Crisis Early Warning System from Harvard Dataverse from January 1, 1995, to April 11, 2023
  - Affinity score computed from intensity found in event data

Coded Event Data

Source.Country	Event.Text	CAMEO.Code	Intensity	Target.Country
Ukraine	Use conventional military force	190	-10.0	Ukraine
Ukraine	Use unconventional violence	180	-9.0	Ukraine
Pakistan	Make statement	10	0.0	Pakistan
United Kingdom	Consult	40	1.0	Pakistan
Pakistan	Consult	40	1.0	United Kingdom
Taiwan	Grant diplomatic recognition	54	6.0	Taiwan
Taiwan	Conduct hunger strike	142	-6.5	*NA*
Thailand	Arrest, detain, or charge with legal action	173	-5.0	Thailand
China	Praise or endorse	51	3.4	Occupied Palestinian Territory
North Korea	Express intent to meet or negotiate	36	4.0	South Korea

Socioeconomic Data

- World Bank
  - World Development Indicators
  - Education Statistics
  - Gender Statistics
  - Health Nutrition and Population Statistics
  - Doing Business (Objective measures of business regulations and their enforcement)
- International Monetary Fund
- Freedom House
  - Trade data
  - Freedom scores

METHODS

Dependent Variable

- Continuous – affinity score from coded event data
- Categorical – affinity scored converted into categories from -1 to 3

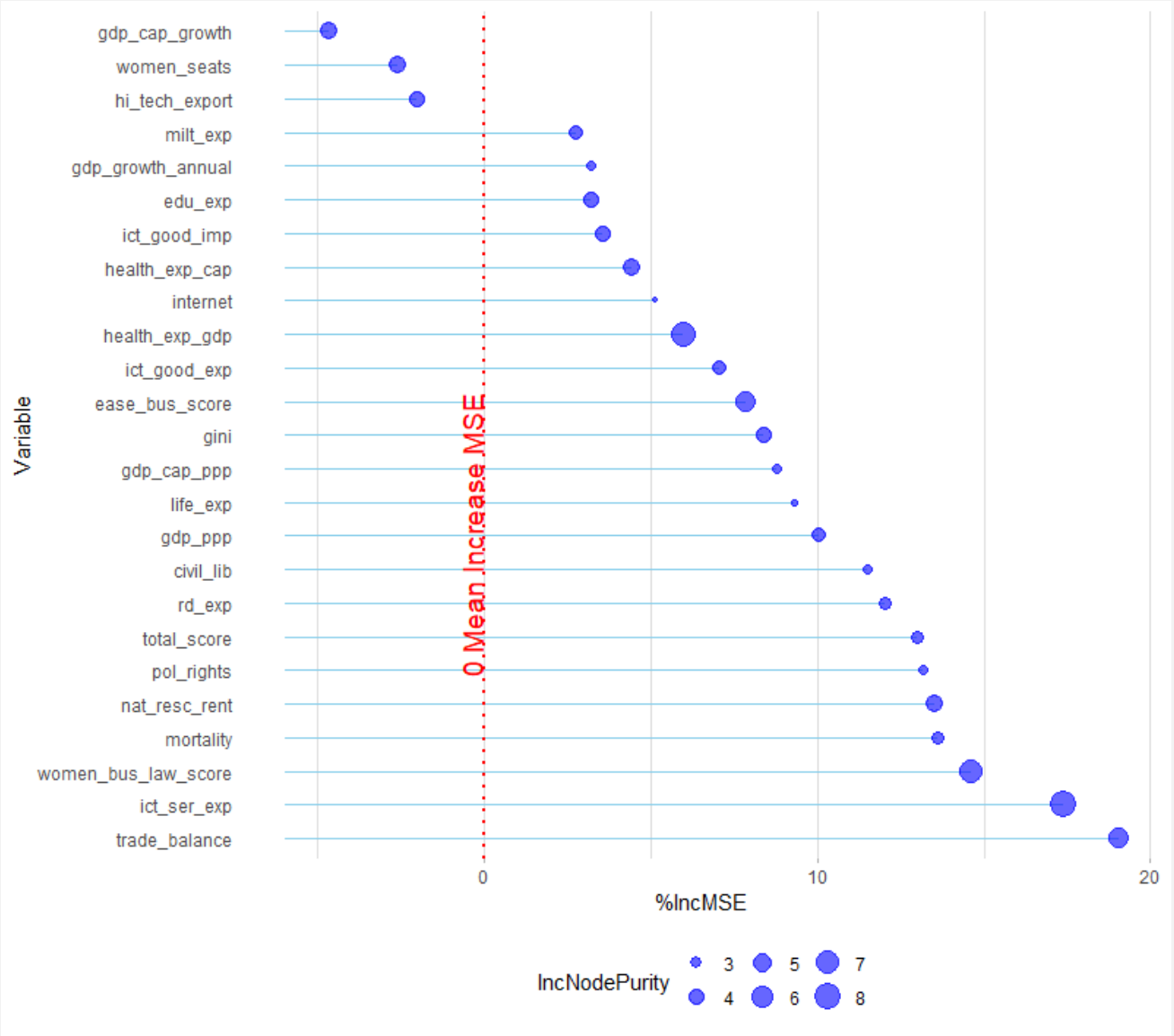
Random Forests

- Continuous – Regression
- Categorical – Classification

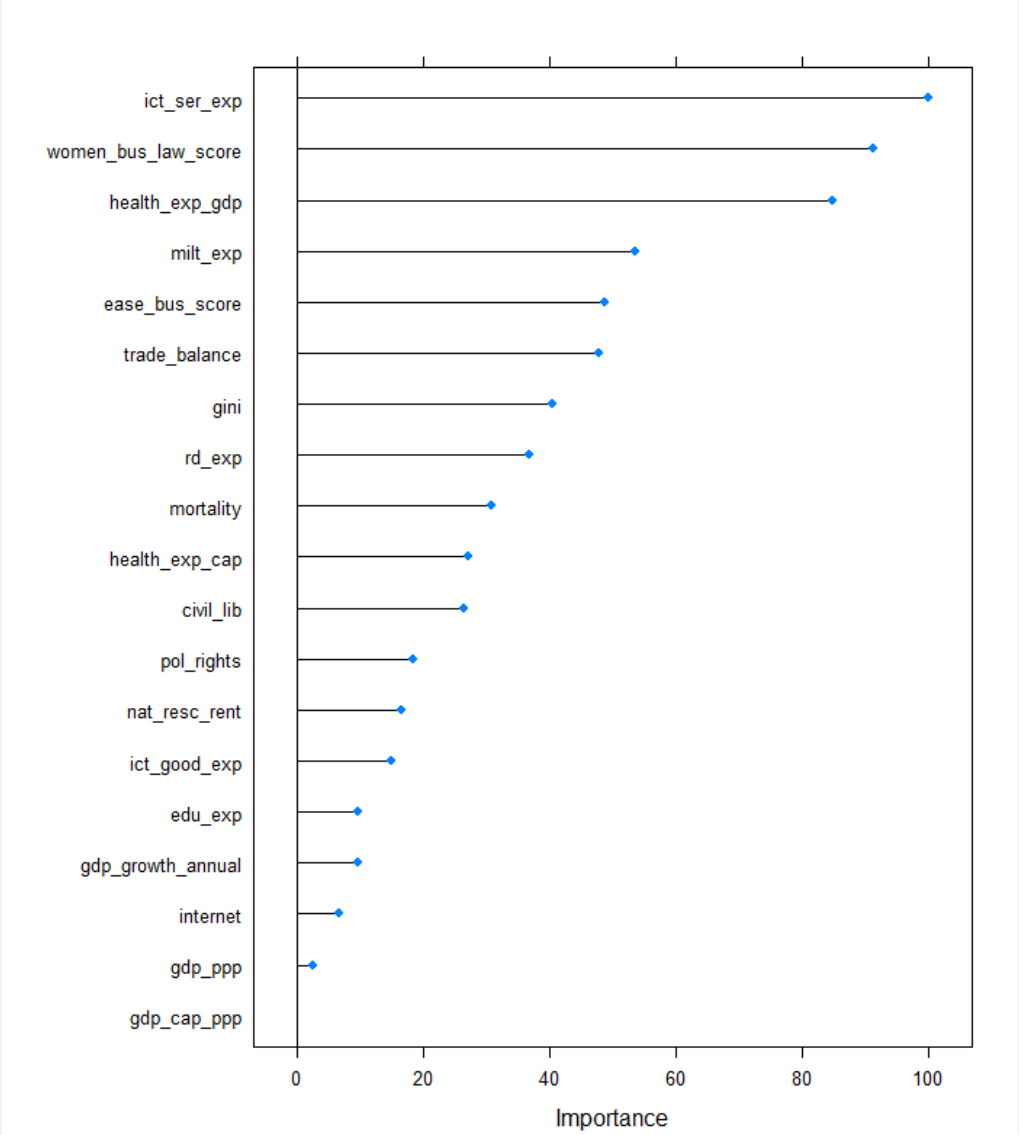
Deep Learning

- Categorical – TensorFlow classification
- Possible overfitting and little insight using shap library

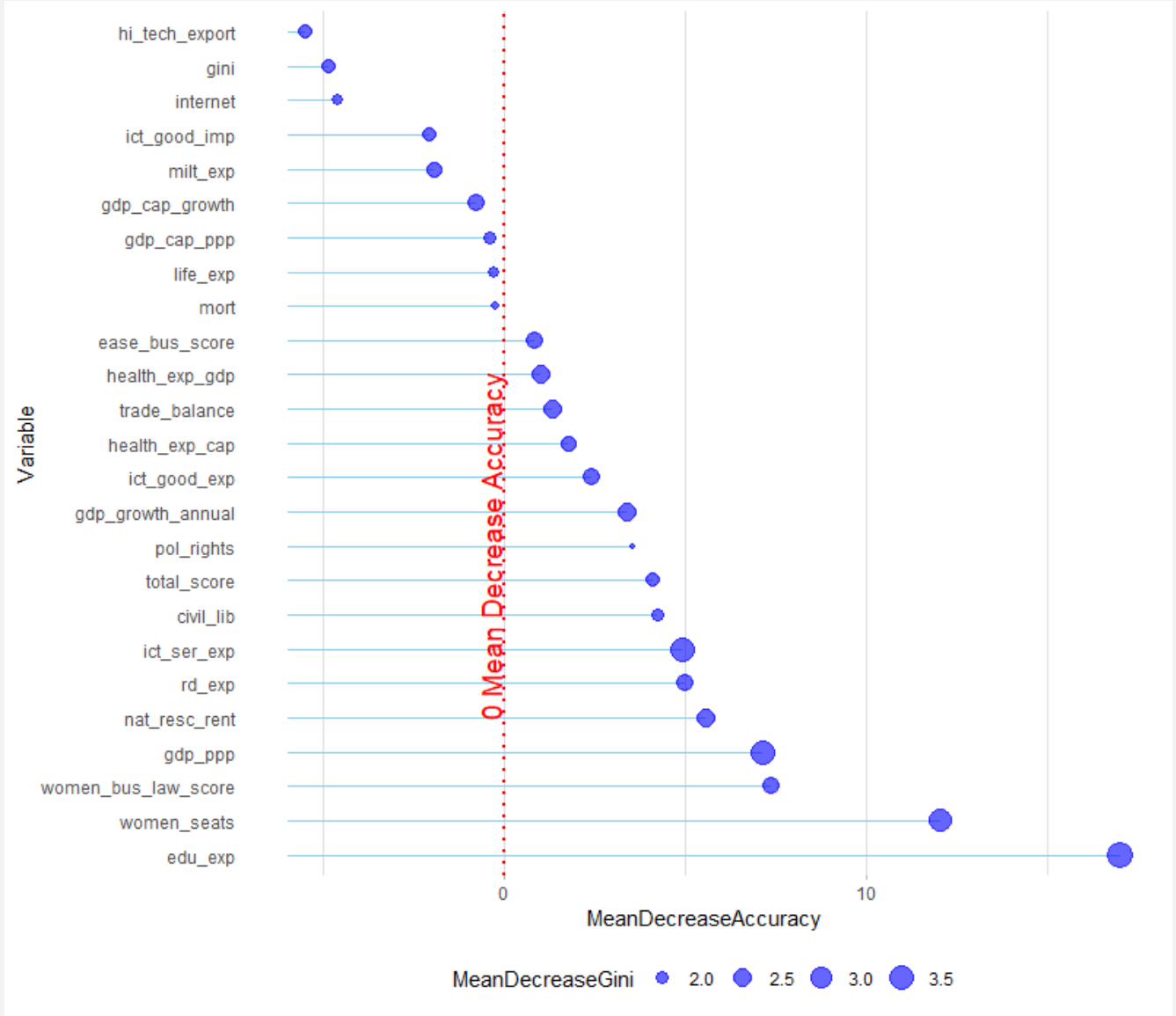
Random Forest Regression



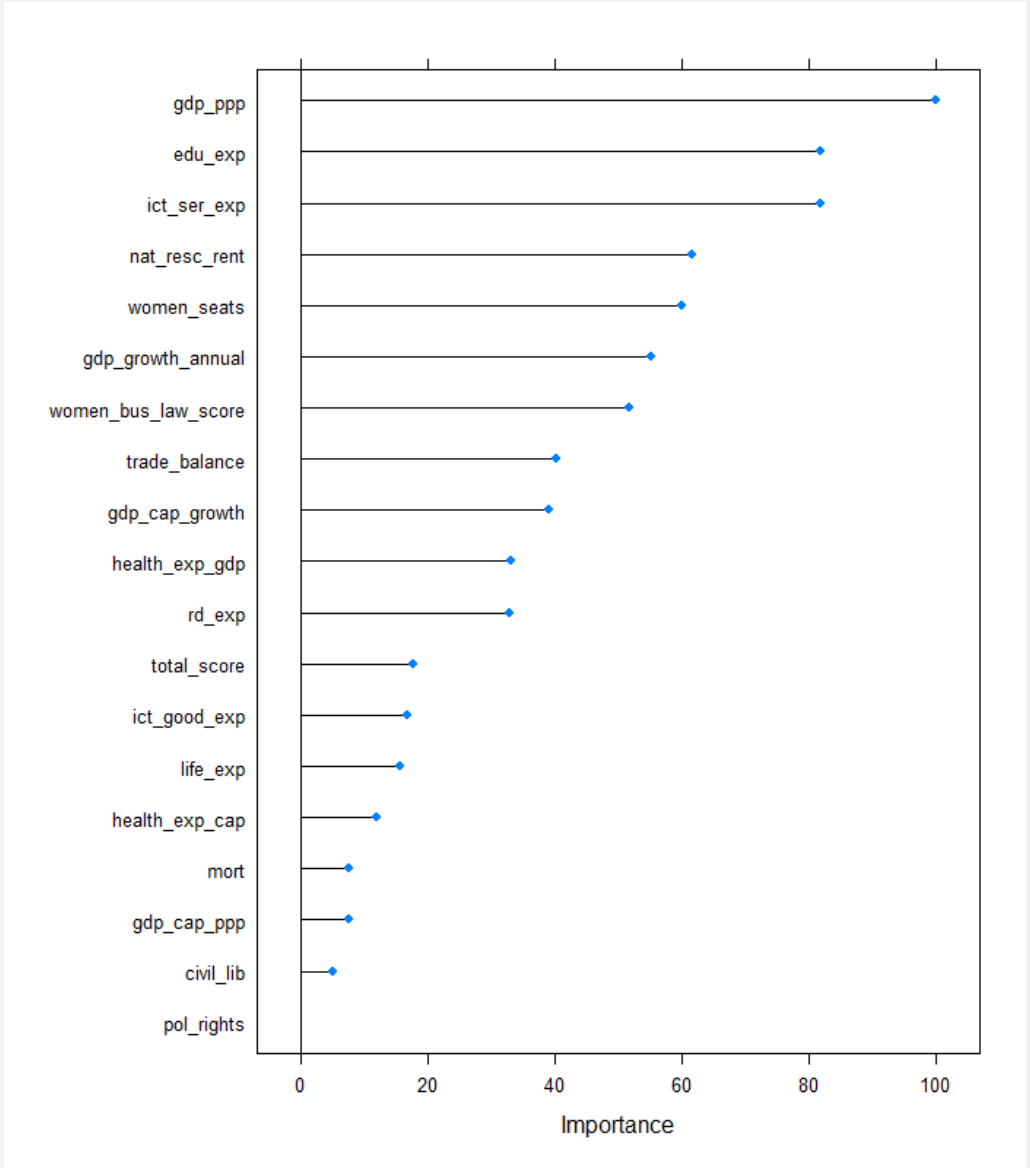
Random Forest Regression: $R^2 = 0.27$



Random Forest Classification

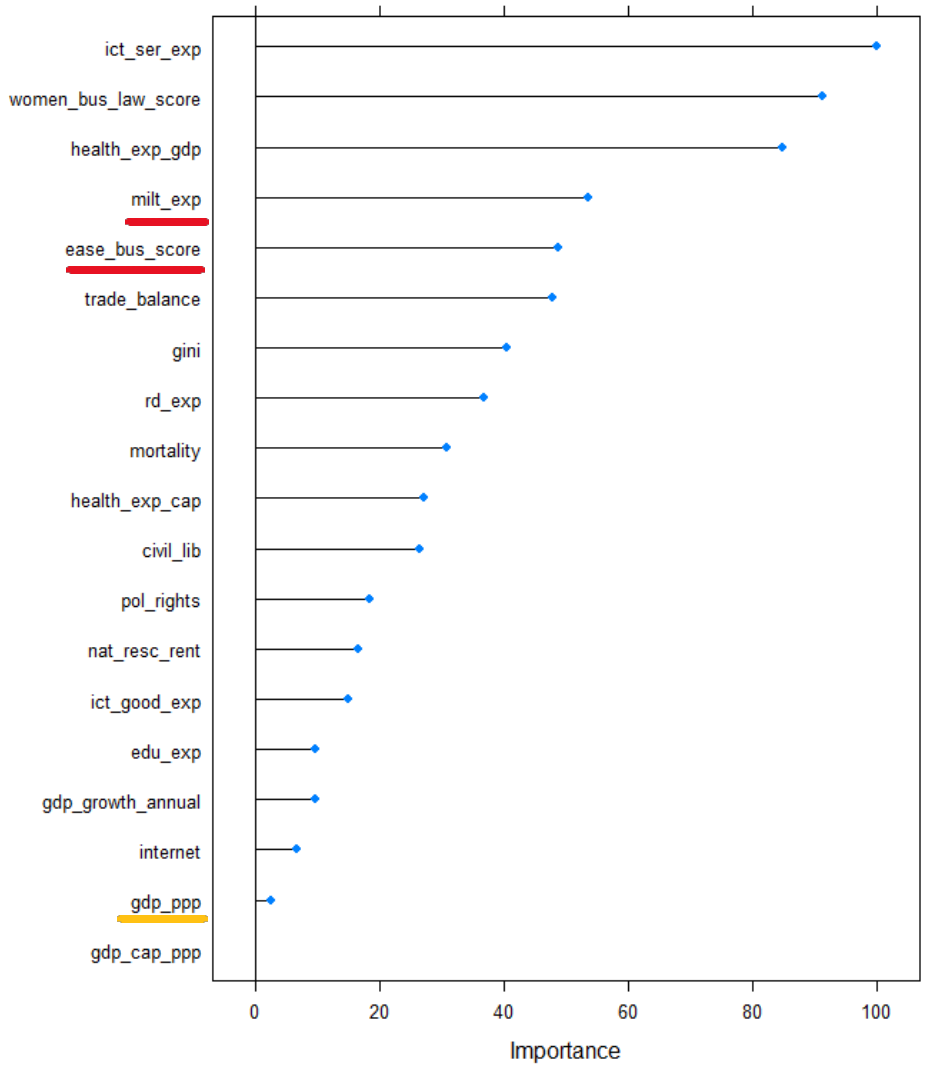


Random Forest Classification: $R^2 = 0.43$

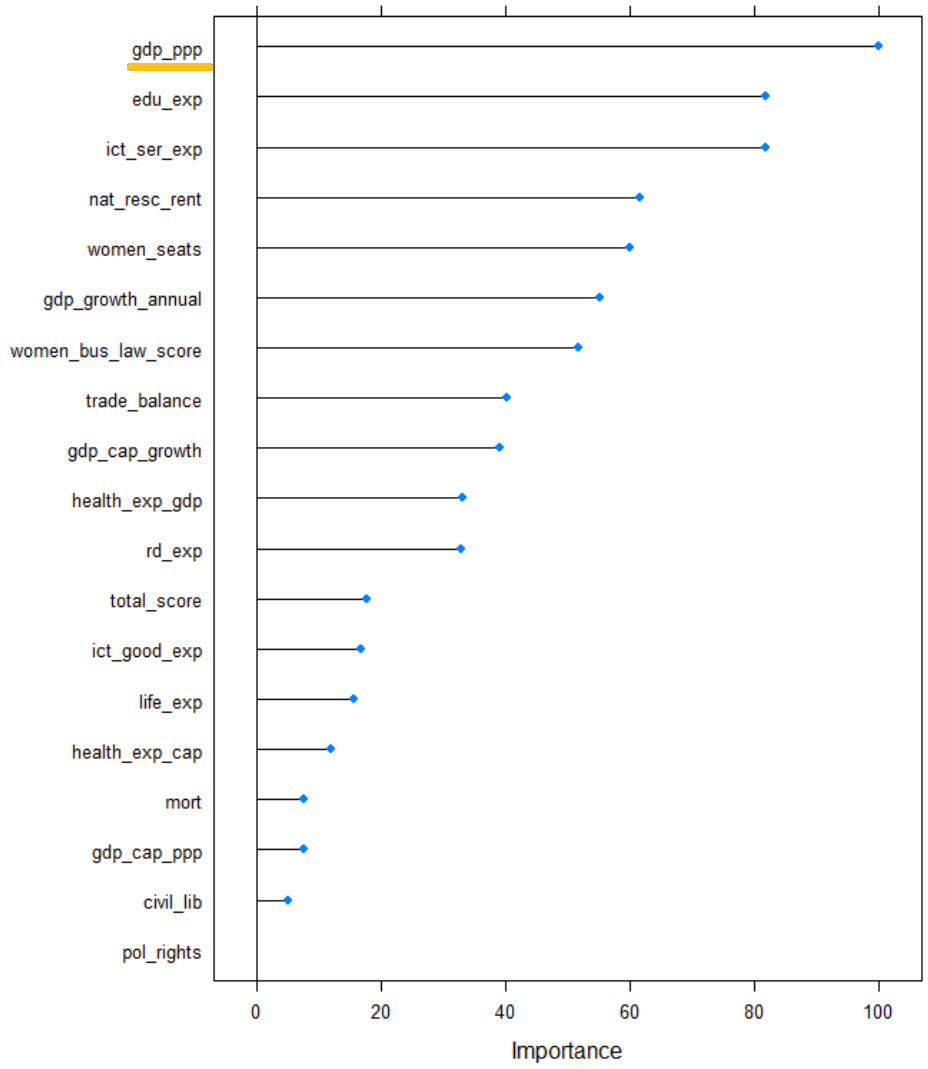


Variable Importance between Classification & Regression

Regression



Classification



Random Forest Classification: $R^2 = 0.76$

```
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
```

```
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 1 3 4 2 3 3 3 4 4 3 2 3 0
```

Future Improvements

Current Model		Improved Model
• Simple affinity calculation	→	• Clustering/Network methods to derive affinity communities (Pauls et al., 2017)
• ICEWS Coded Event Database	→	• POLECAT Database
• Imputed independent variables and scope of variables	→	• New sources of socioeconomic and demographic data