



Predicting Civil Wars from Socioeconomic Indicators

By: Morgan Chen, Theobard Habineza, Chloe Jo, Junyuan Li, Alexander Mekovsky, Ziheng Zeng

Introduction

While the eruption of civil wars in nations undoubtedly afflict their citizens, their occurrences can have significant consequences outside of domestic borders. These include economic disruptions in regional and global markets, the emergence of international terrorism and uncontrollable large-scale refugee crises.¹ Therefore, a better understanding of civil war dynamics and predictive capabilities to prevent their unfolding is valuable to the global community.

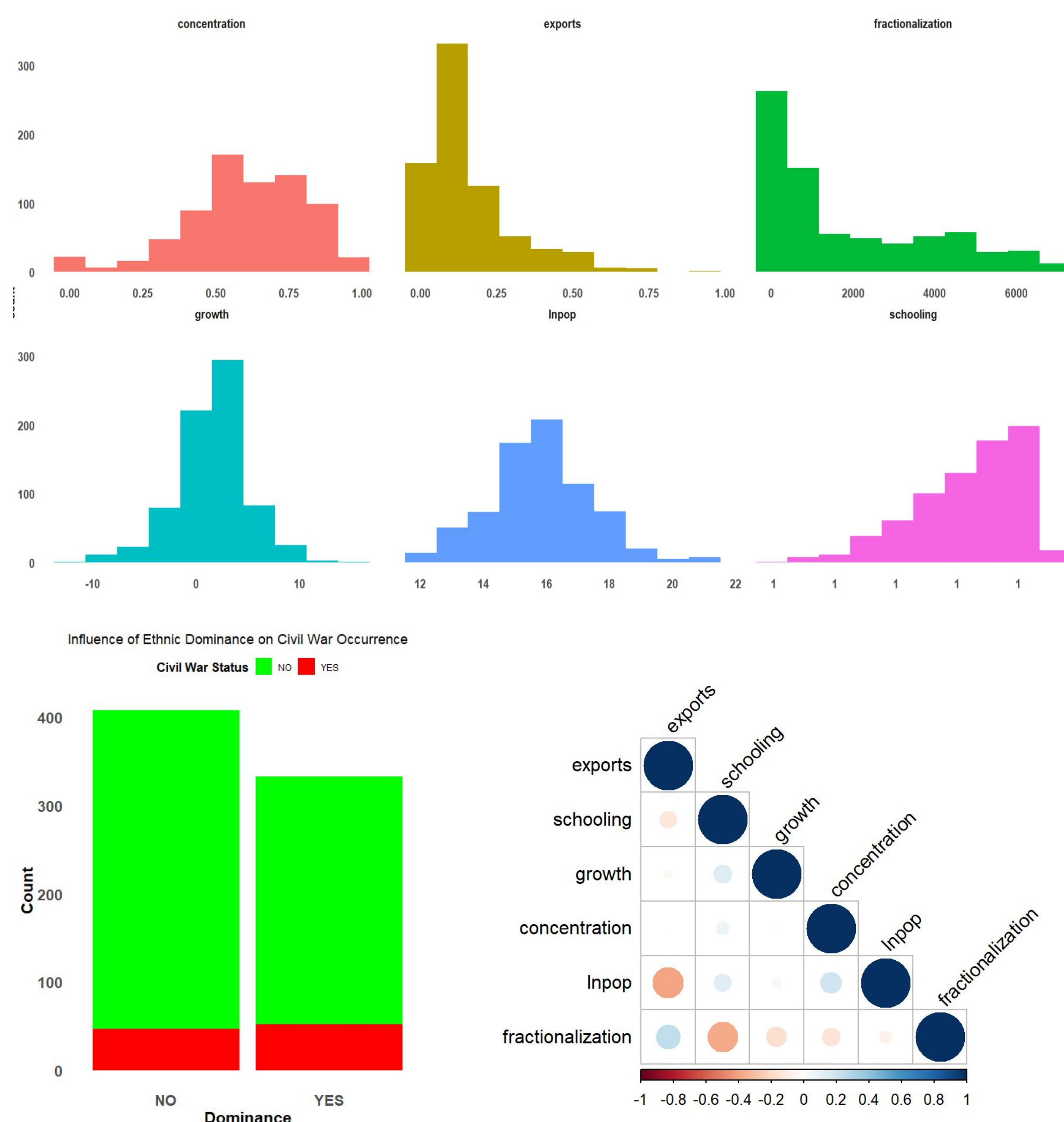
❖ The goal of this study is to create a model to best predict the possibility of civil war given present socioeconomic factors.

Methods

- ❖ We divided the data into 80% training data and 20% testing data.
- ❖ We use binary classification models to determine which one works best with the data by judging its misclassification rate and area under the curve.
 - Models used: Decision Trees, K-Nearest-Neighbors, Logistic Regression, Random Forest, SVM.

Analysis

Data Pre-Processing



Predictors like *schooling* and *fractionalization* exhibit skewed distributions. The correlation matrix reveals low multicollinearity among predictors, ensuring their independent contribution to the model. Notably, civil wars are more common in countries with ethnic dominance, emphasizing its potential importance as a predictor.

Model Comparison

Model	AUC	MCR
Logistic Regression	0.706	0.480
Random Forest	0.814	0.074
SVM	0.602	0.412
Decision Trees	0.647	0.203
KNN	0.729	0.345

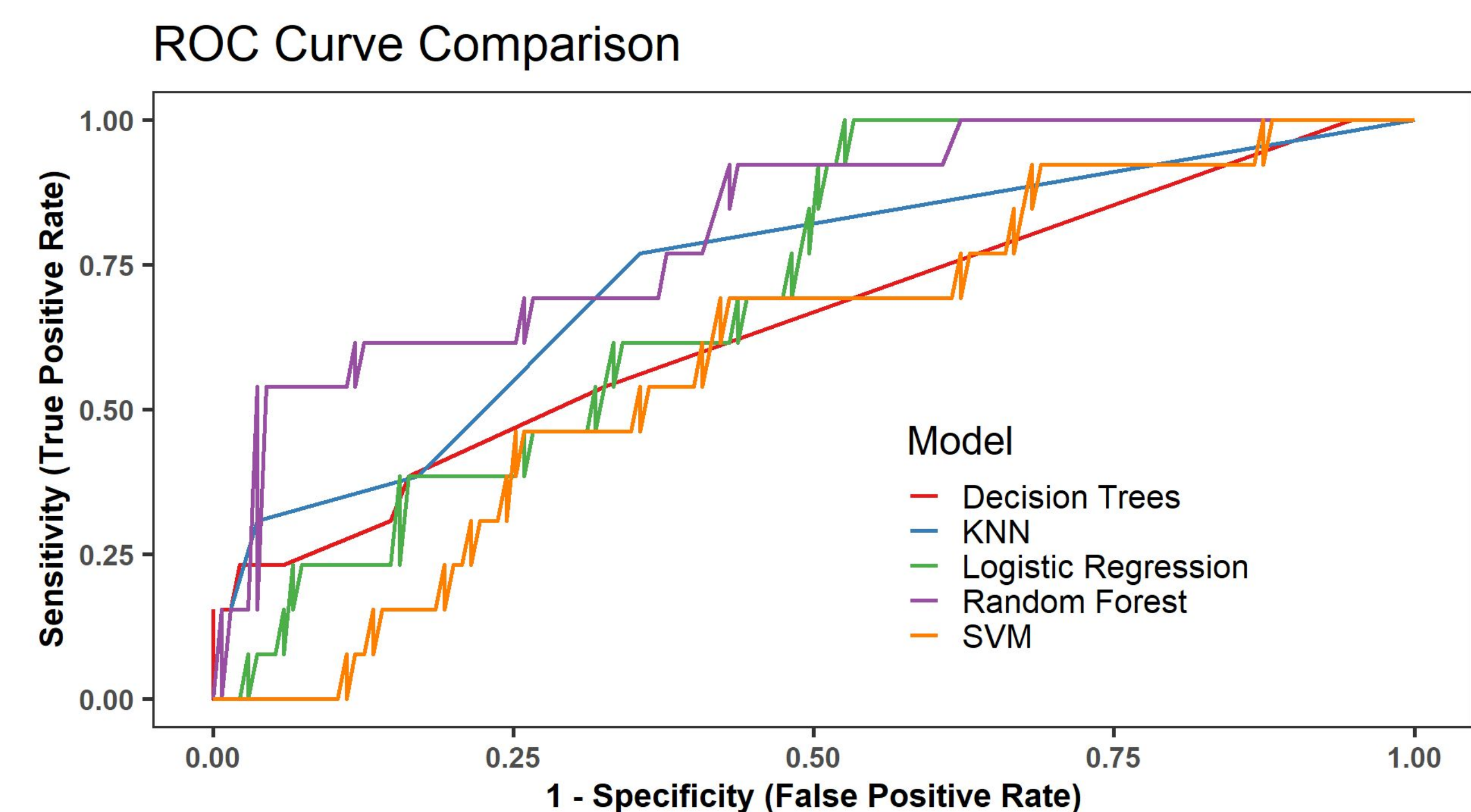
Best model is Random Forest

Optimal Probability Threshold	0.38
MCR	0.074
AUC	0.814

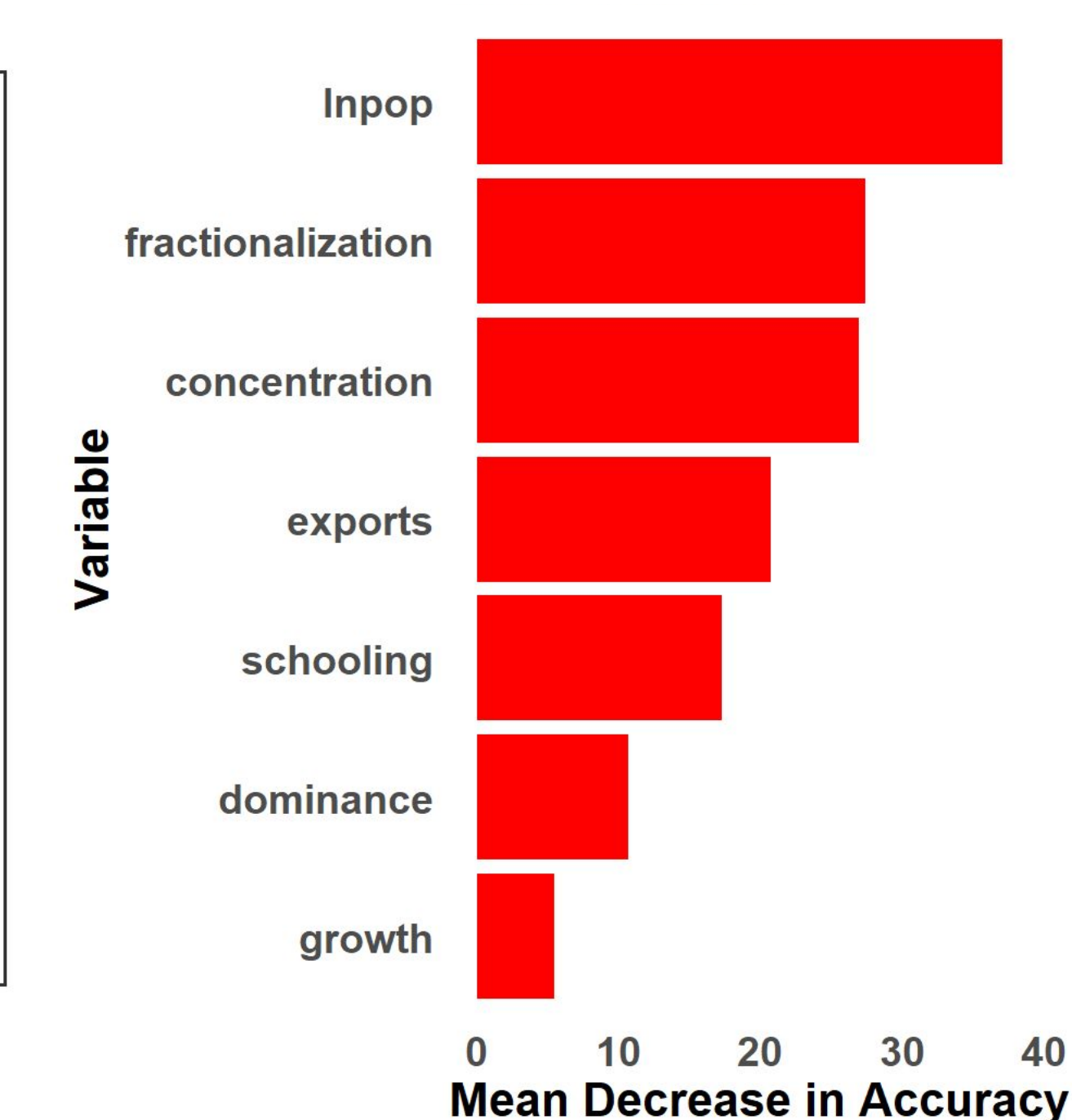
Confusion Matrix of Random Forest Model

	Actual = NO	Actual = YES
Predict = NO	130	6
Predict = YES	5	7

ROC Curve of All Models



Variable Importance Plot of Random Forest Model



Model Comparison: Random Forest achieves the best performance with an AUC of 0.814 and MCR of 0.074, outperforming other models in predicting civil war occurrence.

Confusion Matrix: Random Forest accurately predicts "NO Civil War" cases (130) but struggles with "YES Civil War" cases.

Variable Importance: Key predictors like *Inpop* and *fractionalization* drive model performance, while *dominance* and *growth* contribute less significantly.

Conclusions

The analysis demonstrates that the Random Forest model outperforms other classifiers with the highest AUC and lowest MCR, making it the best model for predicting civil war occurrence. Despite this success, we believe that we have an insufficient amount of data needed to properly train the models needed to predict civil wars.

References

1. Eikenberry, Karl W., and Stephen D. Krasner, eds. *Civil Wars & Global Disorder: Threats & Opportunities*. MIT Press, 2017.