Predicting Autocratization: A New NLP-based Approach

Nowcasting and Forecasting with Text as Data

Mayoral, L., Mueller, H., Philipp, M., Rauh, C. and Vassallo, R.

Institute of Economic Analysis (IAE-CSIC) - Barcelona School of Economics

May, 2025

Outline

1. Introduction

2. Model architecture

3. Nowcasting

4. Forecasting

5. More results

Introduction

Wave of De-democratization

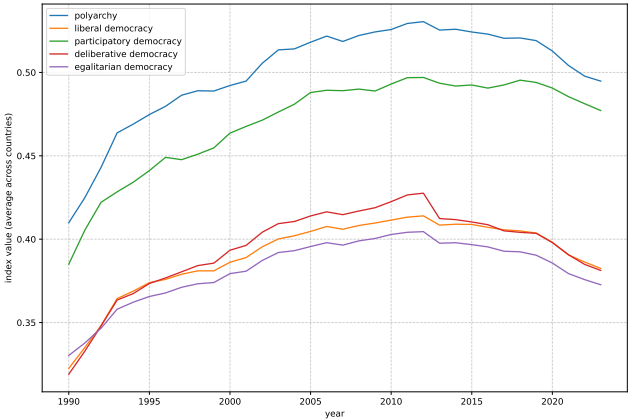


Figure: V-Dem main indicators

De-democratization is costly

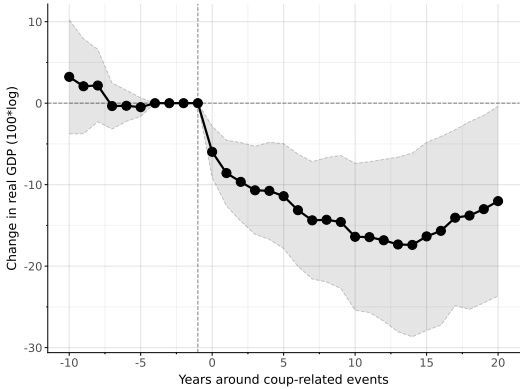


Figure: Dynamic effects of coup-related events on the log of real GDP. The solid line represents the ATE, with shaded areas denoting 95 percent confidence intervals. Estimates are obtained using the local projections difference-in-differences method proposed by Dube, Girardi, Jordà, and Taylor (2024). Own calculations.

Forecasting Challenges

Forecast system in high demand. But there are challenges:

- Sparse (target) data landscape
- Infrequent and delayed updates
- Index based quantification

This Project

- System to forecast de-democratization
- Problems:
  - index based quantification
  - monthly updates and maintenance costs
- Idea: build system from objective **events**
- Forecasting "flows" (events) instead of "stocks" (index level)

Overview

- Monthly forecasts of three event types:
  - coups and coup attempts
  - term limit evasion
  - weakening of the judiciary
- Pragmatic approach working through existing definitions in political science literature and examples.
- Building on our own corpus of ca 6 million articles
- Take existing datasets and expand through:
  - human coding based on literature
  - automated article coding with few-shot learning
  - nowcasting events
  - human coding based on nowcast

Goal

- Goal is a system that can be updated monthly.
  - download articles
  - use title embeddings to nowcast events
  - use human coding to verify
  - update event data
  - forecast forward given the latest dataframe
- We want to go online end of the year
- Critical for application: check media bias

Main Take-aways

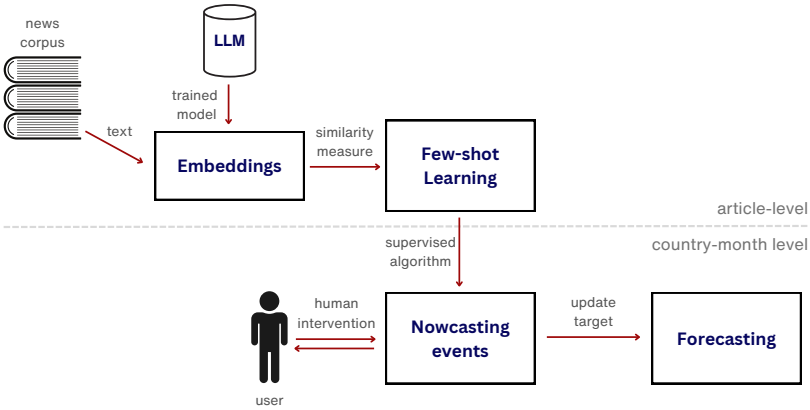
- Very difficult task (needle in the haystack)
- Monthly updates for risk are possible
  - It works for coups! (Cline included 8 of our nowcasts).
  - Good results for term limits
  - But: human coding will remain important
- Some gains from adding more events
- Points towards a measure of institutional *fragility*

Model architecture

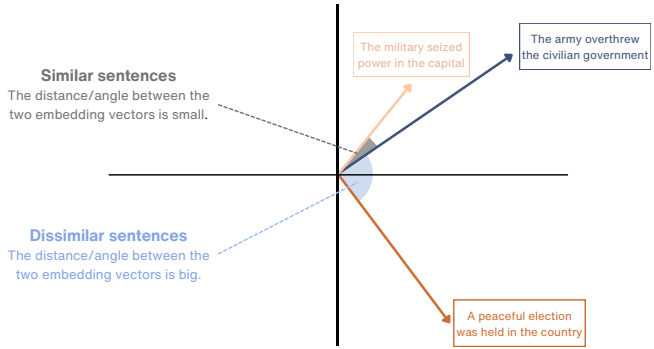
Main Elements of Architecture

1. Ground truth data (very diverse)
2. Event tracking based news corpus
   - Read titles and compare to example sentences
   - Nowcasting checked by human coders
3. Forecasting events

System Architecture



Few shot learning



Dot product as similarity measure (75 embedding dimensions)

Every headline of our ca. 6 million headlines gets 3×75 scores.

Few shot learning

Idea: build on our corpus of ca. 6 million articles.

1) Create 75 prototype example vectors.

2) Transform headlines into embeddings (*all-MiniLM-L6-v2* Wang et al. 2020).

4) Check similarity of headlines and protoype embeddings.

5) Resulting score is used by human coders and nowcasting method.

Ground truth data + nowcasted

Coups

- Cline Centre's Coup d'État Project , Peyton et al. 2023
- merge with Powell and Thyne 2011
- combined dataset has 414 events

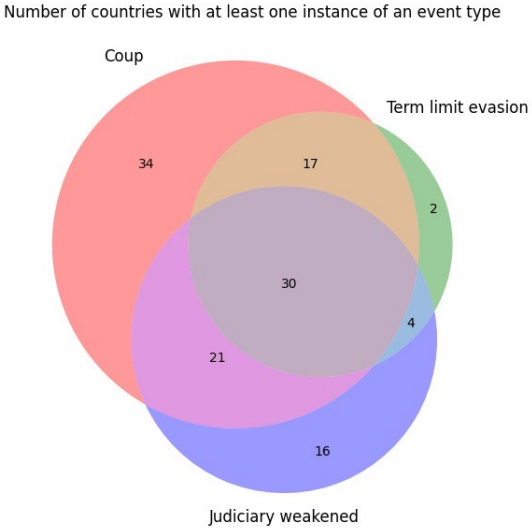
Term-limit evasion

- Ginsburg, Melton, and Elkins 2011
- merge with Versteeg et al. 2020 in Africa and Latin-America
- combined dataset has 117 events

Judiciary weakening

- Democratic Erosion Event Dataset (DEED): 149 events
- merge with Helmke 2017 in Latin-America
- combined dataset has 276 events

Distribution Across Countries



Nowcasting

Model Features

Nowcasting

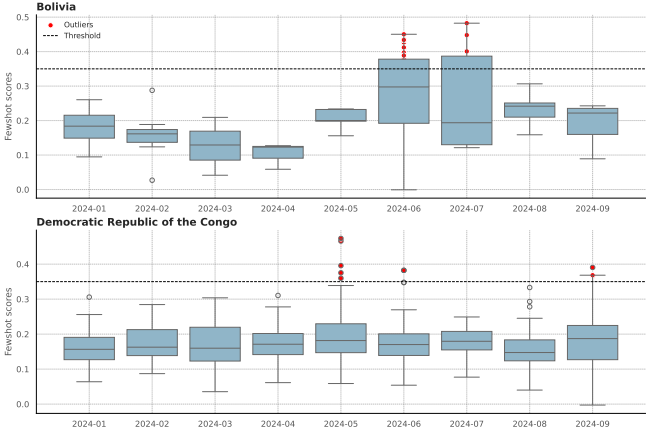
- embedding scores
- feature creation $\rightarrow$ unbalanced panel.

Boosting algorithms perform well in this setting and natively handle missing values.

Let's walk through a simplified implementation.

Nowcast for Target Updates: Coup Examples

Fewshot scores detect autocratic events: coup attempts in Bolivia (June 2024) and Democratic Republic of Congo (May 2024).



Nowcasting: Some Thoughts

Overall we added and verified 15 judiciary events and 22 term limit events. Lot's of positive case studies.

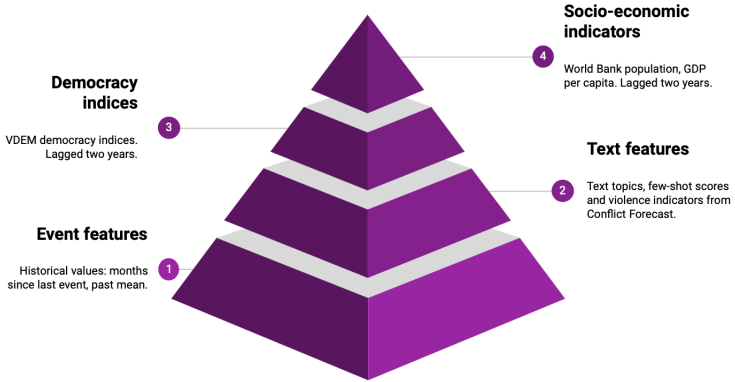
We also picked up Mexico (judicial weakening) in August 2024 and picked up Benin (coup attempt) in September 2024. Timing in Benin was better.

Problem is the system picked up the news reporting on the protests in Mexico not the original policy.

Design choice here: should we do one-country-out cross validation?

Forecasting

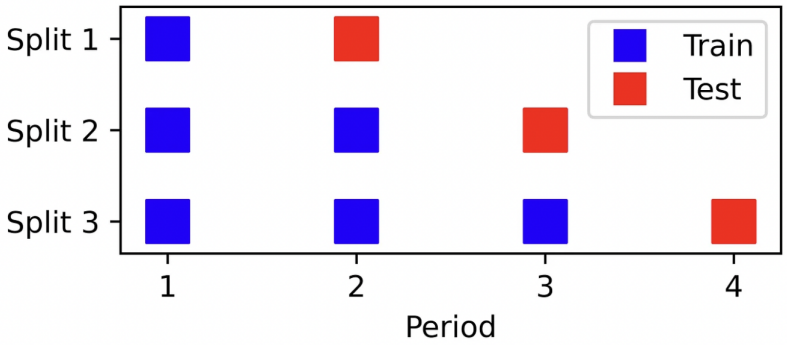
Model Features: forecasting



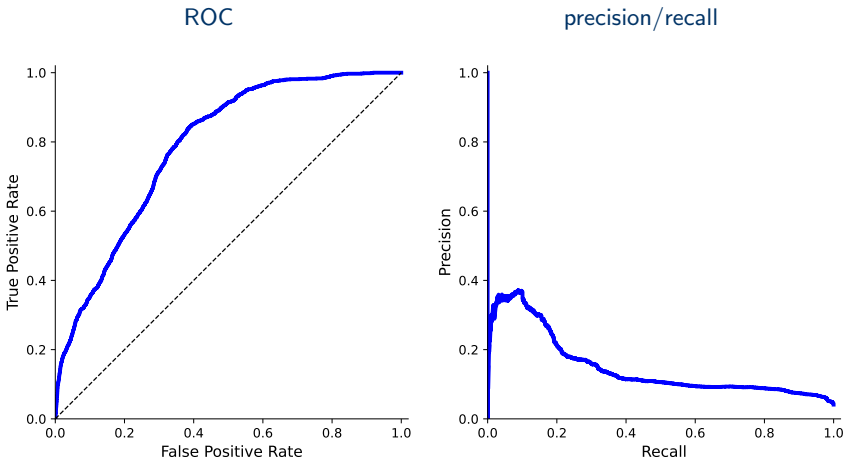
Rolling Forecast

Use data until today to predict tomorrow: pseudo out-of-sample evaluation.

Model: tree-based (Random Forest and XGBoost).



Forecasting Coups: ROC vs. Precision/Recall Curve

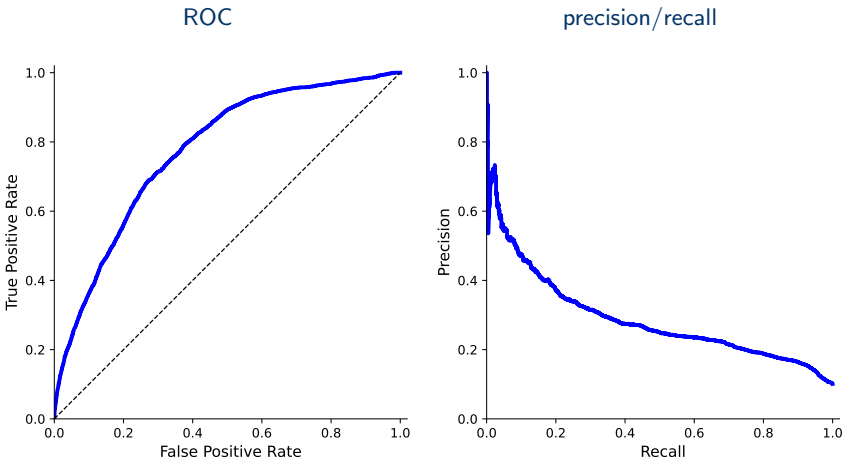


Cumulative Training and Testing Results

Train data	Test data	ROC-AUC	Precision
Coup	Coup	0.786	0.146
Coup + TLE	Coup + TLE	0.789	0.175
Coup + TLE + JW	Coup + TLE + JW	0.774	0.288

Note: Performance metrics for institutional disruption forecast, 12 months ahead. Rolling forecast strategy from Jan 2010 until Jan 2025 using Random Forest Classifier. **Coup**: Coup d'État events, **TLE**: term-limit evasion events, and **JW**: judiciary weakened events. Each row aggregates events to the previous.

Forecasting All: ROC vs. Precision/Recall Curve



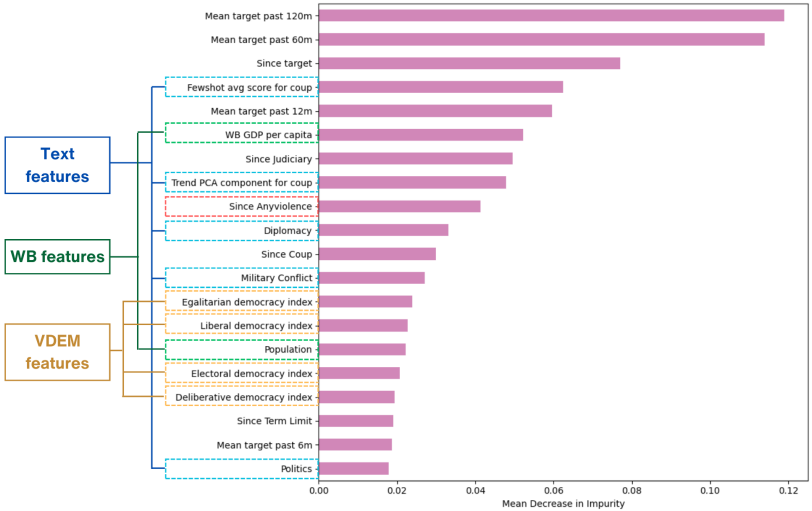
Learning Across Event Types

Train data	Test data (with NaNs)	ROC-AUC		Avg Precision
		Overall	Hard	
Coup	Coup	0.783	0.777	0.147
Coup + TLE	Coup	0.788	0.775	0.158
Coup + TLE + JW	Coup	0.789	0.778	0.141

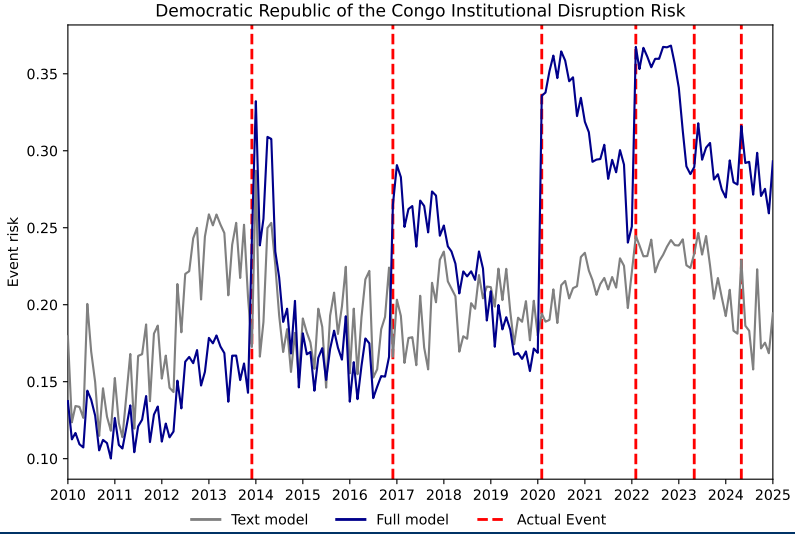
Note: Performance metrics for institutional disruption forecast, 12 months ahead. Rolling forecast strategy from Jan 2010 until Jan 2025 using Random Forest Classifier. **Coup**: Coup d'État events, **TLE**: term-limit evasion events, and **JW**: judiciary weakened events. Testing is performed exclusively on coup events, with TLE and JW events set to NaN. Hard onsets refer to events occurring after more than 60 months with no prior events.

More results

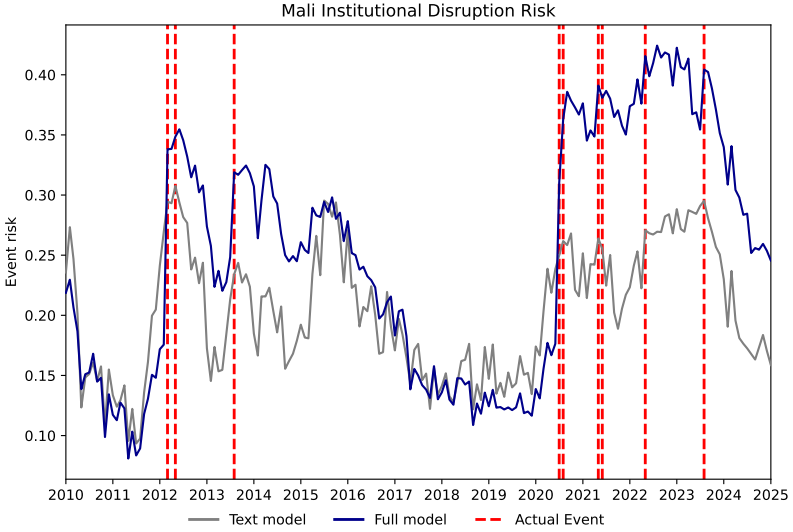
Feature Importance for Full RF model



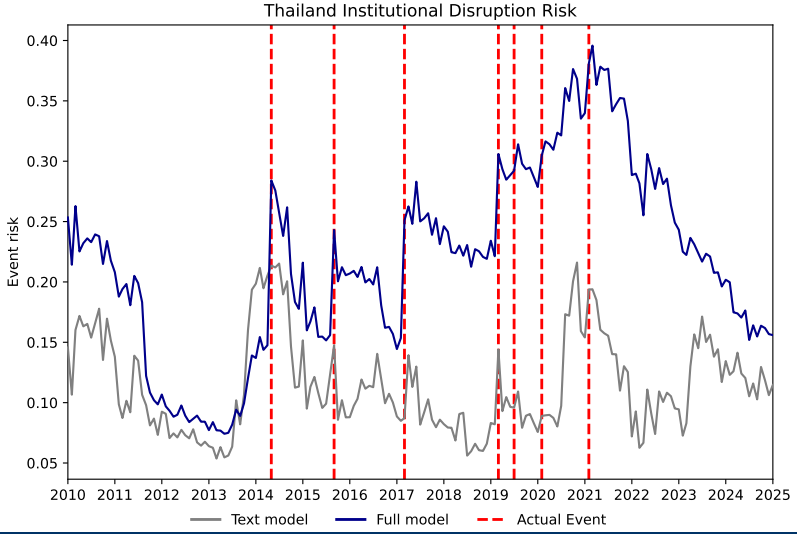
Event risk: DR Congo



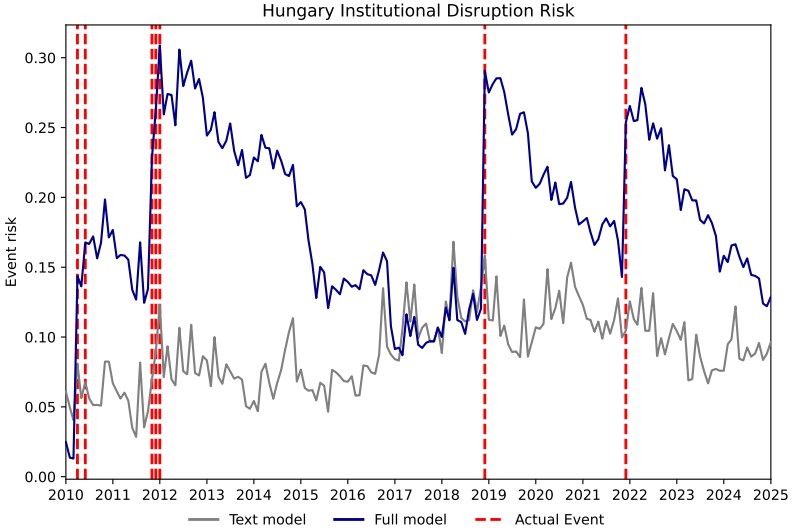
Event risk: Mali



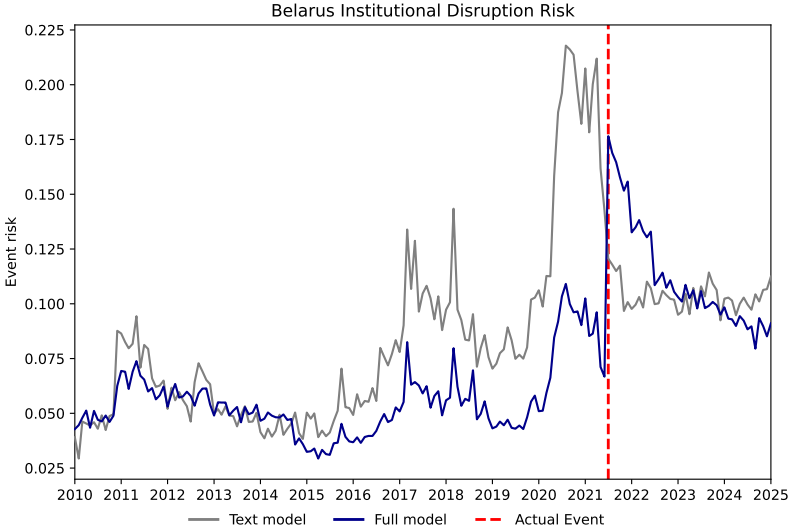
Event risk: Thailand



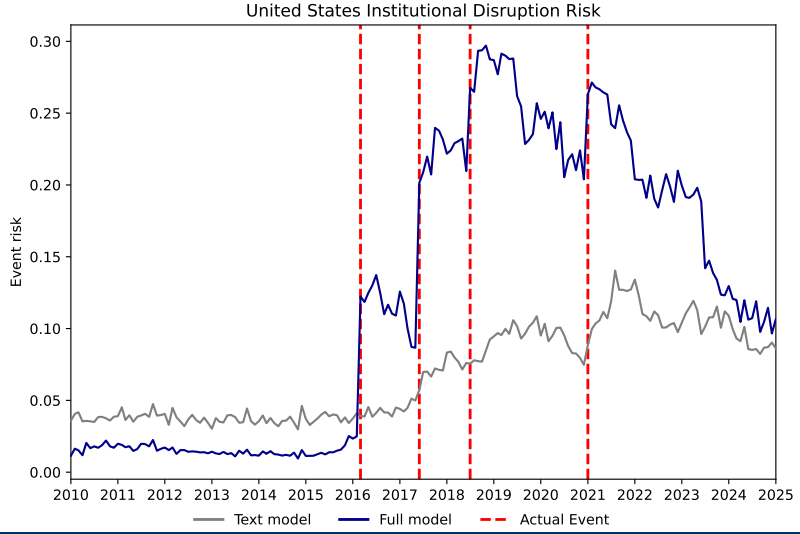
Event risk: Hungary



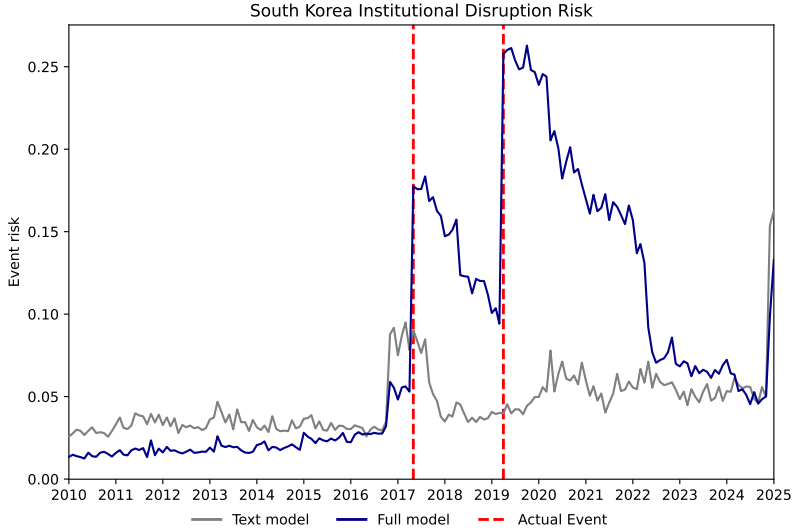
Event risk: Belarus



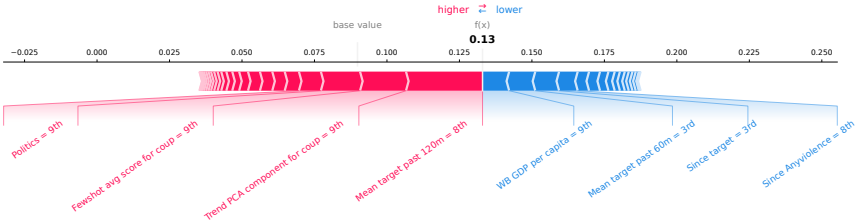
Event risk: United States



Event risk: South Korea



Shapley Korea 2024.12



Main Take-aways

- Very difficult task (needle in the haystack)
- Monthly updates for risk are possible
  - It works for coups!
  - Good results for term limits
  - But: human coding will remain important
- Striking gains from adding more events
- Tree-based algorithms shows its strengths
- Conclusion: system measures *political fragility*

Thank You

Questions?