OSTBAYERISCHE TECHNISCHE HOCHSCHULE REGENSBURG

RCAI REGENSBURG CENTER FOR ARTIFICIAL INTELLIGENCE

www.oth-regensburg.de

Strategic Forecasting with Machine Learning for Geopolitical Analysis

Introduction

- Strategic forecasting for conflict regions is of increasing relevance to governments and NGOs alike
- Machine learning (ML) tools can assist in overcoming two fundamental problems with currently available data:
  - It can include qualitative (in addition to quantitative) data of social phenomena
  - It requires a structure within the data (i.e. categories that can be classified)
- Our analysis focuses on parts of Northern and Western Africa, which share sufficiently similar background conditions and are of current interest to policymakers

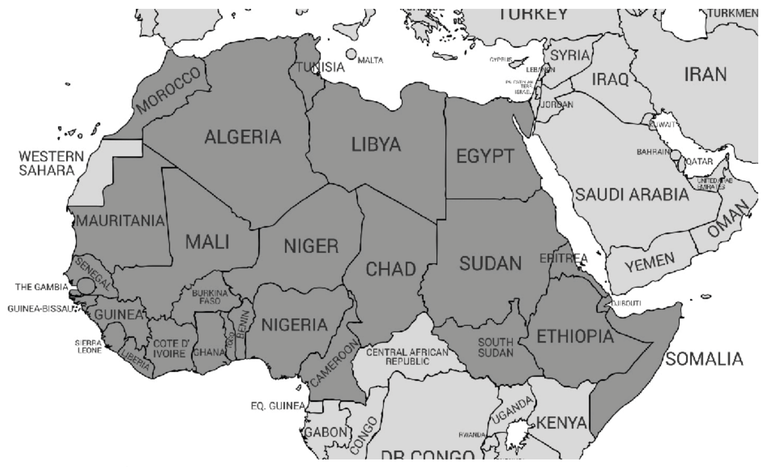

Figure 1. Map of countries included in the analysis.

Methodology

- We used a supervised ML model to identify patterns in the dataset
- This was done through the Waikato Environment for Knowledge Analysis (WEKA)
- The decision tree was generated using a C4.5 algorithm (referred to as J48 in WEKA)

Data

- The data was taken from the 2020 Bertelsmann Transformation Index (BTI), which covers over 52 variables for a time period between (2006-2020)
- „Conflict intensity“ was selected and transformed into a predictive class
- The minimum number of instances per leaf was set to 5

Class	BTI Values	Instances	%
No violent incidents	1	1	0.5%
Few violent incidents	2; 3; 4	67	33.8%
Violent incidents	5; 6; 7	90	45.5%
Widespread violent conflict	8; 9; 10	40	20.2%

Table 1. Generated classes for conflict intensity.

Findings and Discussion

- Analysing the modified BTI dataset using the J48 algorithm in WEKA resulted in a decision tree drawing on the following 11variables

Name in WEKA	Name in BTI
Reli_dogma	No interference of religious dogmas
force_mnply	Monopoly on the use of force
State_idntty	State identity
Judiciary_indpnt	Independent judiciary
Implementation	Implementation
Conflict_mngmnt	Cleavage/conflict management
Groups_intrst	Interest groups
Freedom_expr	Freedom of expression
Participation_civsoc	Civil society participation
Consensus_gls	Consensus on goals
Admin	Basic administration

Table 2. Name of Variables used in the decision tree.

- For a detailed overview of these variables, please refer to the 2020 BTI Codebook
- 153 out of 198 instances were identified correctly, (resulting in an accuracy of 73.23% (the baseline accuracy for the most common class is 45.5%, see table 1)
- All variables present in the decision tree can be logically explained and are sufficiently backed by the corresponding academic literature
- There are some inconsistencies towards the bottom of the decision tree, which are mainly a result of the shortage of data
- The main variables for predicting the level of conflict intensity are reli_dogma, force_mnply, and state_idntty
- Where force_mnply is rated above 7, the highest category of conflict intensity (i.e. widespread violent conflict) was never observed
- This indicates that the existence of a strong force monopoly over the use of force is related to the observation of violent conflict in a state in the first place
- The category „no violent incidents“ is omitted from the decision tree because it applied to only one instance from the countries under examination
- Only in two cases a relatively high error rate was observed (implementation and freedom_expr)

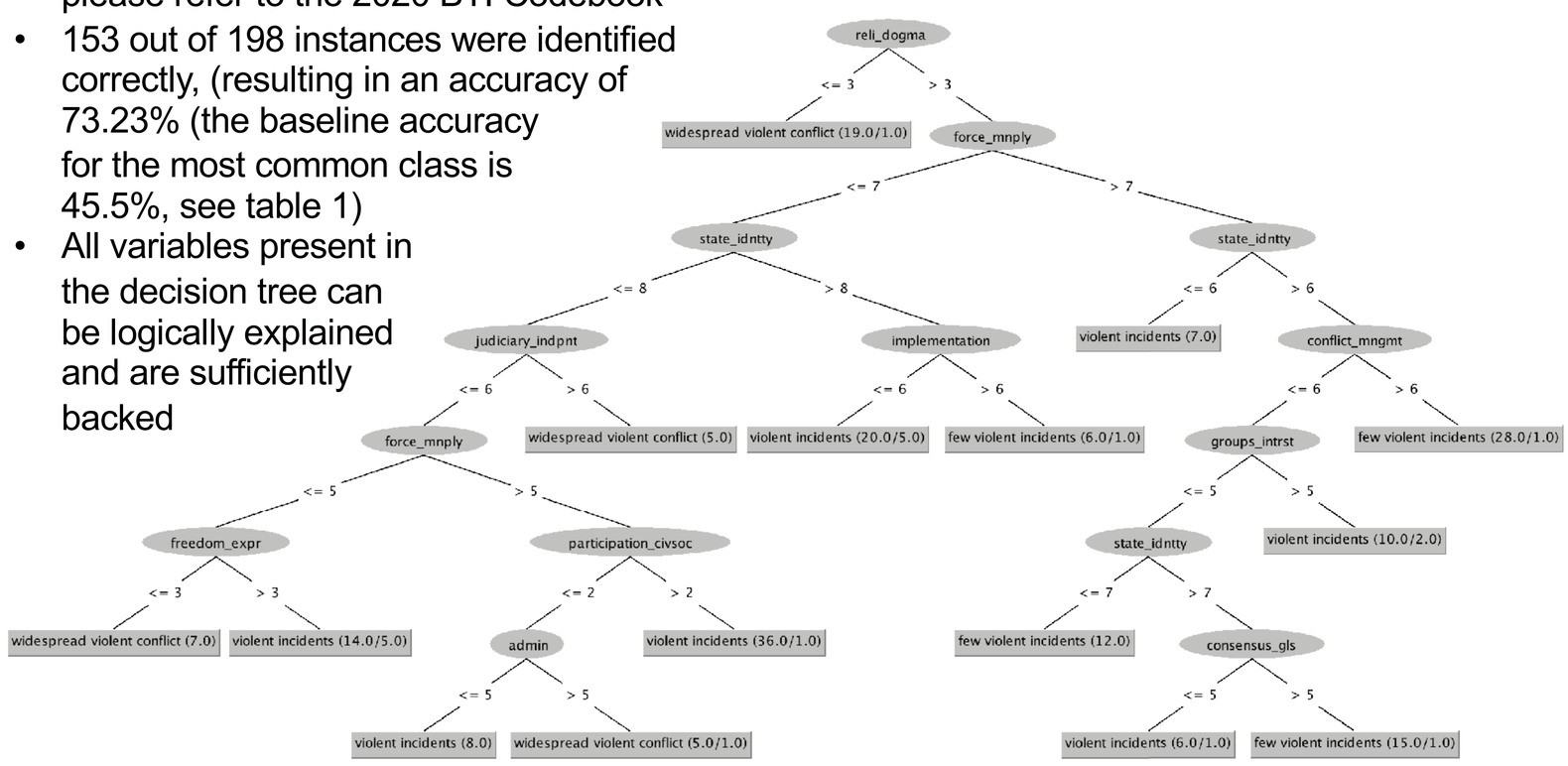

Figure 2. Decision tree calculated in WEKA using J48.

Limitations

- The BTI dataset focuses on transformation and therefore includes variables which are not entirely relevant to predicting conflict intensity
- Other important variables, such as environmental factors or the existence of an illicit economy are entirely absent in the BTI
- The findings are inevitably subjective; the BTI relies on expert judgement in their rating and the categories for conflict intensity were defined by the researchers

Conclusions

- Using publicly available ML platforms for strategic forecasting can provide valuable insights
- ML offers an alternative to regression analysis through which both quantitative and qualitative variables can be analysed
- Nevertheless, the accuracy of the decision tree is only as good as the data provided
- More data is required for more accurate predictions
- Datasets which measure the relevant variables and categories are desirable

Project manager: Prof. Dr. Markus Bresinsky, Assistance: Johanna Schroeder, Valerio Viscardi (Faculty for Applied Natural Sciences and Cultural Studies)

References:
Bertelsmann Stiftung (2020). „Codebook for Country Assessments“. BTI Project. Accessed 02.07.2020, https://www.bti-project.org/content/en/downloads/codebooks/BTI_2020_Codebook.pdf.