# Predicting democratic backsliding with Machine learning
## Model comparison

Paul Elvis Otto
249968
p.otto@students.hertie-school.org

Ujwal Neethipudi
248346
u.neethipudi@students.hertie-school.org

Saurav Jha
249354
s.jha@students.hertie-school.org

**Abstract**

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aeque doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distinguique possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defensa et.

## Contents

1. Motivation and Context ..... 3
2. Additional assumptions ..... 4
3. Subset and EDA ..... 4
   3.1. Subset ..... 4
   3.2. EDA ..... 4
Bibliography ..... 5
A Tables and Data ..... 5
B List of Countries ..... 5

# Appendix

A Tables and Data . . . . . . . . . . 5
B List of Countries . . . . . . . . . . 5

# List of Figures

Figure 1 average of political corruption over time . . . . . . . . . . 3

# 1. Motivation and Context

The Vdem dataset is one of the largest datasets at hand for quantative reserach on states and their political system and the corresponding effects that can be measured. For this final report we have decided to focus on analysising this dataset with state of the art machine learning and evaluate as well as uncover patterns in the data that get missed at glance. As a taget that we have identified we will focus on political corruption as it is defined in the codebook of the dataset as.

"here goes the Vdem corruption definition"

As we assume that corruption it self is inherited in the concept of high trust and low trust societies we subsetted the Vdem dataset to only include a limited number of countiriesg see Appendix B for the full list, that fullfill, at least partialy, those criterias as well as beeing a form of a democratic system. Further we subsetted the available features to only include those with an acceptable concistency across the different countries as well as in the the time series. according to this date we can first focus on the overall corruption index over these countires in perspective to time, Figure 1. The Average of this score does not draw a realy conclusive image of the way that corrupton is taking. To find the underlyin patters we have choosen to:

1. Run multiple models over the compelte subset of our data
2. Evaluate these models for their performance
3. Identify the top predictors of these models
4. Test for different assumptions we have

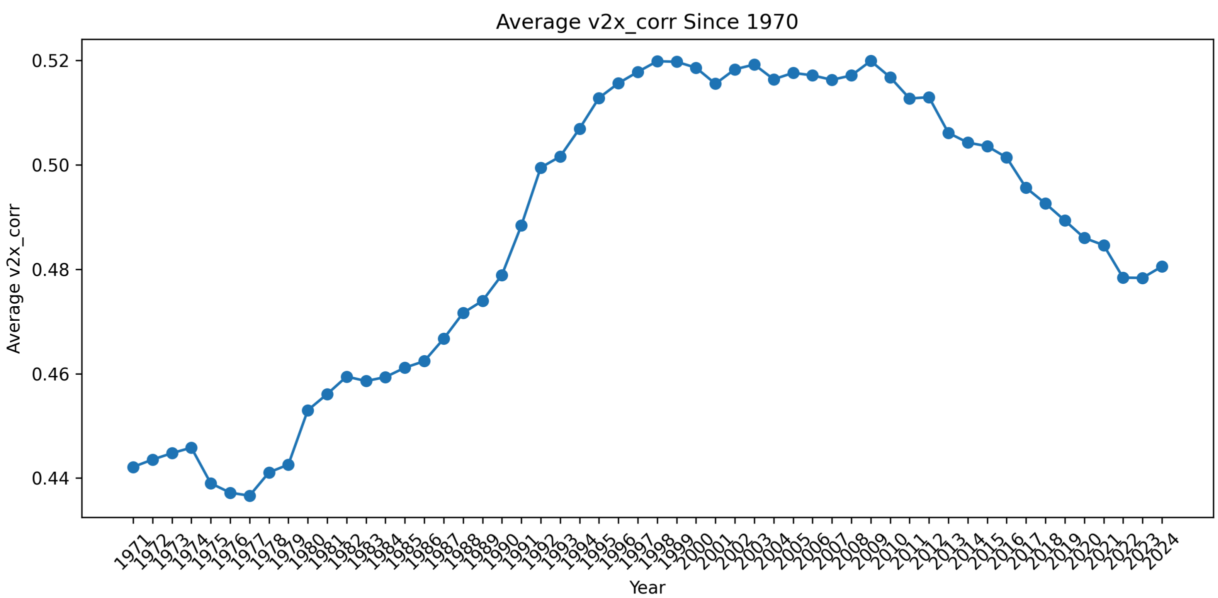


Figure 1: average of political corruption over time

# List of Figures

Figure 1 average of political corruption over time ........ 3

# 2. Additional assumptions

As we dont just want to find the overall best predictors for political corruption we also present the following hypothesys:

1. Allthough not under the top predictors the social equality is a strong predictor for political corruption
2. Gender plays as another strong predictor into `v2x_corr`

# 3. Subset and EDA

## 3.1. Subset

To properly use the model to predict the Target it was necessary to filter the dataset further, for that we implemented a filtering to only keep the bare variables of the dataset without any additional operations as they are included by the authors of vdem.[1] In the next step we filtered for non numeric features and removed as pointed out in Section 1 the missing values.

## 3.2. EDA

To get a first overview of the data and potential cross dependencies that we havent captured yet with our dependencie mapping a correlation matrix was created[2] This correlation matrix directly pointed out more highspots that after a manual correction have been removed. The

[1] To make this process reproduceable we organised the dataset subsetting process in a pipeline that can be run from the root of the project dir on github, manual or with a makefile

[2] Script: `./src/corr_matrix_plot.py`

# Bibliography

# A Tables and Data

… your appendix content …

# B List of Countries

- Austria
- Belgium
- Bulgaria
- Croatia
- Cyprus
- Czech Republic
- Denmark
- Estonia
- Finland
- France
- Germany
- Greece
- Hungary
- Ireland
- Italy
- Latvia
- Lithuania
- Luxembourg
- Malta
- Netherlands
- Poland
- Portugal
- Romania
- Slovakia
- Slovenia
- Spain
- Sweden
- United States of America
- United Kingdom