

Final Output Machine Learning

Erik Schulte

30.6.2022

Introduction

References are to be made as follows: Fama and French (1997, 33) and Grinold and Kahn (2000). Such authors could also be referenced in brackets (Grinold and Kahn 2000) and together Grinold and Kahn (2000). Source the reference code f

Data

First of all, I present all relevant data sources and explain essential features to get a proper overview. Afterwards, I explain the process of loading the data into R and how we process the raw text corpus, to obtain meaningful results.

Data Sources

For our empirical analysis we use three main data sources. The first two data sources are used for the text analysis and the building of the subgroups. Data on the speeches of the UNSC were obtained from Schönfeld et al. (2019). It contains 82165 speeches over the years 1995 until 2017.

```
#load required packages
library(tidyverse)
library(readtext)
library(quanteda)
library(quanteda.textplots)
library(quanteda.textstats)
library(ggplot2)
library(magrittr)
library(gdata)
library(kableExtra)

#load the UN Security Council Debates dataset
load("data/dataset.RData")
load("data/docs.RData")

#set a seed for reproducibility
set.seed(2333)

# View data
#head(meta_meetings)
#head(meta_speeches)
```

- Fama, Eugene F, and Kenneth R French. 1997. “Industry Costs of Equity.” *Journal of Financial Economics* 43 (2): 153–93.
- Grinold, Richard C, and Ronald N Kahn. 2000. “Active Portfolio Management.”
- Schönfeld, Mirco, Steffen Eckhard, Ronny Patz, and Hilde van Meegdenburg. 2019. “The UN Security Council Debates 1995-2017.” *arXiv Preprint arXiv:1906.10969*.