

# Forecasting and Nowcasting with Text as Data

Term 3 - 3 ECTS

Elective Course

## Prerequisites to Enroll

Python and STATA, Econometrics, Machine Learning.

## Overview and Objectives

Forecasting and nowcasting are key elements of many policy processes and decisions inside firms and governments. Examples include revenue forecasts, monetary policy, monitoring of human migration, political risk predictions or stock market evaluations. In many of these applications we observe that standard non-textual data is released with a delay which would make a timely response impossible. That is why text mining is becoming increasingly important as a tool as text data is often more readily available. However, the integration of text into forecasting models has many pitfalls.

In this course, students receive an introduction into methods of nowcasting and forecasting in text with an in-depth training for some applications. We will, for example, show how newspaper text can be used to predict the outbreak of armed conflict and how students can build an automated warning system for violence against tourists. We will discuss to what degree international migration, GDP time series or stock market movements can be forecasted or nowcasted with several sources of text data.

Special attention will be given to the case of the Spanish Central Bank (BdE) which used text as data to support the decision-making process in the BdE and ECB and understand the impact of the Corona virus on the economy at a time where no reliable data was available in the first months of the pandemic.

## Prerequisite reading / requirements

To be announced.

# Forecasting and Nowcasting with Text as Data

Term 3 - 3 ECTS

Elective Course

## Course Outline

### Part I. Machine learning for forecasting and nowcasting

- In what decision problems is forecasting and nowcasting important?
- When can machine learning be useful in forecasting and nowcasting?
- Why can unsupervised machine learning be useful?
- How can the forecast and nowcast quality be evaluated?

### Part II. Overview of feature extraction techniques from text used

- Under which circumstances can a text archive help to predict better?
- What are the most common sources of text, how can they be made accessible, what are cheaper alternatives?
- How to use text in “rolling” forecasts? How to streamline the process for faster text processing and prediction?

### Part III. Applications

- nowcasting the economy during the pandemic in Spain
- forecasting political risk worldwide
- nowcasting attacks against tourists
- forecasting stock market returns to patents
- forecasting migration patterns

## Required Activities

Students are expected to attend classes and participate actively. For some of the applications homeworks will be required. Students will also be divided in teams to compete against each other in a real out-of-sample nowcast competition using thousands of texts.

## Evaluation

Homeworks: 40%, Final term paper: 60%

## Competences

Students will understand the usefulness of machine learning for forecasting and nowcasting but also its limitations. They will understand how text can be integrated into the forecasting process and what the best methods of feature extraction are.

## Forecasting and Nowcasting with Text as Data

**Term 3 - 3 ECTS**

**Elective Course**

Students will also be able to conduct all elements of a forecasting task including the analysis of the problem, extracting the correct text features, implementation of the forecast and including it into the decision making problem.

### **Learning Outcomes**

- Understand how to set up forecasting and nowcasting solutions and their limits
- Understand what text features are best extracted in the use of fore- and nowcasting problems.
- Program Machine Learning algorithm for forecasting and nowcasting.
- Implement rolling forecasts over time and standard cross validation for model optimization.
- Be able to produce and interpret the statistics of standard classification and continuous forecast models and their problems.

## Forecasting and Nowcasting with Text as Data

Term 3 - 3 ECTS

Elective Course

### Materials

Besley, Tim, Thiemo Fetzer and Hannes Mueller (2020) Terror and Tourism: The Economic Consequences of Media Coverage.

Böhme, Marcus, Andre Groeger and Tobias Heidland (2020) Searching for a Better Life: Predicting International Migration with Online Search Keywords, *Journal of Development Economics*, vol. 142, 102347, 2020.

Cage, Julia, Nicolas Herve, and Marie-Luce Viaud (2017) The Production of Information in an Online World: Is Copy Right? CEPR Discussion Papers DP12066.

Gentzkow, Matthew, Bryan T. Kelly, and Matt Taddy (2019) Text as Data. *Journal of Economic Literature*. Forthcoming.

Hansen, Stephen, Michael McMahon and Andrea Prat (2018) Transparency and Deliberation within the FOMC: a Computational Linguistics Approach. *Quarterly Journal of Economics*, 133 (2).

Hassan, Tarek A., Stephan Hollander, Laurence van Lent and Ahmed Tahoun (2017) Firm-Level Political Risk: Measurement and Effects. *R&R QJE*.

Mueller, Hannes, and Christopher Rauh (2018) Reading Between the Lines: Prediction of Political Violence Using Newspaper Text. *American Political Science Review*. <https://doi.org/10.1017/S0003055417000570>

Vegard H. Larsen and Leif Anders Thorsrud, (2015) The Value of News. Working Papers No 6/2015. Centre for Applied Macro- and Petroleum economics (CAMP).