**Summary**

This course aims at taking the student to the frontier of text-analysis applications in finance as well as teaching some traditional estimation techniques.

**Content**

In recent years, Finance has seen a revolution due to the arrival of new statistical techniques, large datasets (think about Facebook) and powerful computational tools. The objective of this course is to introduce the participant to both natural language processing and the relevant finance literature related to this technique. Since 80% of information is supposedly unstructured text, whereas only 20% is in actual data, such text analysis appears most promising and opens many doors.

From an economic point of view, the emphasis is on how to use modern techniques to better predict stock markets.

In the later part of the course, we will also spend some lectures on more traditional techniques such as GMM estimation and Bayesian learning.

From a programming point of view, the participant should be able to program both in Python and R after attending this course. For efficient data handling, we will discuss both Pandas (Python) and data.table (R). Using RStudio’s markdown, we will learn how to simplify the research pipeline between data-analysis and paper.

**Grading**

20% of grade: There will be 2 referee reports and some general exercises.

80% of grade: Final 3h written, closed book examination.

Retake: 3h written, closed book for 100%.

**Note**

The subject of the course remains the same but the tools have been updated.

**Keywords**

- Return predictability, Natural language processing, Network characteristics
- GMM / Bayesian estimation, Python / R and tidy data logic

**Learning Prerequisites**

- Programming in a formal language

**Resources**

- Websites