

Data Science in Finance and Accounting

Sasan Mansouri

Winter Semester 2021-2

Course description

This course is targeted at students who are in the final semesters of their master's program in economics, management, finance, and accounting, and it delivers the most important aspects of programming that you will face in your prospective career in the financial sector.

Registration

Course examination registration and withdrawal take place electronically via QIS (TBA). Outside this time period, there is no possibility to enroll or withdraw from the course.

Additionally, students must complete the “Entry Application” by completing the Jupyter Notebook [here](#).

Grade determination

There will be no final exam. Grades will be determined according to a weighted average of 4 assignments. The weights increase progressively from the first to the fourth assignment. Assignments will be introduced as an incomplete Jupyter Notebook that you need to complete the missing parts of the codes, and add additional insights/interpretations for the results. Students **must be available** in short notice, upon request, to present their assignments' solutions in front of the classroom.

Required Knowledge

In the first session, I will review the most basics of Python programming. However, I assume you have prior knowledge of basic programming and algorithmic thinking in any other language. If not, provided you still want to take this course, I urge you to take an online crash course in Python programming; there are many nice Python introductory courses that you can find with a simple Google search. My suggestions are something like [this](#) or [this](#) which are free and you can (theoretically) finish in one day or two!

Class rules

Please bring your laptops to class. This will be a very hands-on course, with relatively little in the way of formal theory. Instead, we'll be working through lecture notes together in class and you'll be running code on your own machines.

Software requirements

All of the software requirements for this course are open-source and/or free. All of the lecture notebooks will be accessible in Google Colab. You must have a Google account (which is free of charge, and you can create one by clicking [here](#)) in order to be able to run the codes on your browser.

You can also download the class notebooks and run them locally on your PC. My suggestion is to install Anaconda with Python 3.* accessible [here](#).

Syllabus

The suggested syllabus:

1. Intro to python:

- (a) Programming basics
- (b) Most important libraries: Pandas, Numpy, Matplotlib, ...

→ [Homework 1: cleaning and visualization of a dataset](#)

2. Intro to data acquisition:

- (a) APIs: application on Thomson Reuters
- (b) Website scraping: BeautifulSoup, Selenium (Java-Scripts)

→ [Homework 2: Collect SEC filings & accounting frauds](#)

3. Textual analysis:

- (a) Text cleanings: Regular Expressions, Tf-Idf, Stop-words
- (b) Sentiment Analysis

→ [Homework 3: Banking Sector Riskiness](#)

4. Machine Learning:

- (a) “Traditional Models”: Linear/Logistic Regressions, NB, SVM
- (b) Deep-learning: Keras

→ [Homework 4: Stock market predictions](#)

5. (optional) manage data-sets: SQL

Textbook and other readings

I try to make lecture notebooks as detailed as possible so that you won't need additional books. Fortunately, Python has a great programming community that regardless of your coding issue, you can find a solution with some searches in Stack overflow!

Still, the lectures will closely follow the following resources:

- **“Python for Finance: Mastering Data-Driven Finance [2nd ed.]”** (Yves Hilpisch) - O'Reilly Media
- **“Web Scraping with Python: Collecting Data from the Modern Web”** (Ryan Mitchell) - O'Reilly Media
- **“Applied Text Analysis with Python”** (Benjamin Bengfort, Rebecca Bilbro, Tony Ojeda) - O'Reilly Media
- **“Deep Learning with Python”** (Francois Chollet) - O'Reilly Media
- Etc.

Academic integrity

Plagiarizing would be checked manually and using the software. Students caught cheating or plagiarizing will automatically be assigned a failed grade.