TEXTUAL ANALYSIS AND MACHINE LEARNING WITH APPLICATIONS TO ECONOMICS AND FINANCE

June 19th – June 23rd, 2023.
Venue: National Bank of Slovakia

Course Content

The objective of this course is to study how we can use the millions of textual contents published on the Internet and social media every day to improve our understanding of various economic and financial phenomena. After an introduction to the Python programming language, we will start by seeing how it is possible to extract online content via the use of existing APIs or the implementation of web scraping tools. We will create an application to collect articles from a major media site and we will use an API to extract tweets from a social network dedicated to finance.

Next, we will see how to analyse a text using Natural Language Processing (NLP) methods and create a full NLP pipeline (cleaning, stop words, Part-of-Speech tagging, Named Entity Recognition, Stemming/Lemmatization) relevant to a given research project. We will apply this to the press conferences made by the European Central Bank to show how it is possible to give structure to unstructured data. The next session will be dedicated to sentiment analysis and will present the different methods (dictionary approach and machine learning) with an application on a database of media articles. The fourth session will be devoted to machine learning using text as data with an application on StockTwits data (asset pricing). In the last session, we will introduce methods of textual analysis on unsupervised data (topic modelling and transformers). We will perform an application of a Latent Dirichlet Allocation on a large corpus of Glassdoor reviews.

For the different sessions, we will first present both the related theories and methods - in a language accessible to non-mathematicians - and their latest applications in the economic and financial literature. We will then study and share with the participants all scripts and codes to realize different tasks in Python. We will also offer participants the opportunity to present their research and/or projects, and if possible, we will assist them with their projects - both on the data collection side and on the data analysis side.

Pre-requisite

Participants should have a basic understanding of computer programming. It is possible to follow the tutorial available at
https://www.learnpython.org/ or https://www.datacamp.com/
to learn or review the basics of programming in Python.

Participants must install Anaconda (https://www.anaconda.com/products/individual) to have a functional programming environment before the beginning of the course.

About the instructors

Matthieu Picault is an Assistant Professor at the University of Orléans (France) in the Laboratoire d'Economie d'Orléans (LEO). He received his Ph.D. diploma from the University of Aix-Marseilles (AMSE) in 2017. His research focuses primarily on central bank communications and their impact on financial and macroeconomic variables. It includes textual analysis of both official documents and media. He teaches in the Applied Econometrics Master courses of Introduction to Python and Natural Language Processing with Python. His research has been published in journals including the Journal of International Money and Finance, Finance Research Letters, and in the International Journal of Finance & Economics.

Thomas Renault is an Assistant Professor at the University Paris 1 Panthéon-Sorbonne (France) and a scientific advisor at the French Council of Economic Analysis. He has received his Ph.D. diploma in 2017 from Paris 1 Panthéon-Sorbonne. His work focuses on the use of textual data – mostly from social media – to forecast financial markets and construct novel indicators to track economic conditions. He is teaching the following classes at the University Paris 1 Panthéon-Sorbonne: Applied Data Science in Finance, Applied Big Data in Finance, Introduction to Python, Digital Data, and Network Analysis. His research has been published in journals including the Journal of Public Economy, the Journal of Banking and Finance, the Economic Journal, and the Journal of International Money and Finance.

Academic papers:

- Altig, D., Baker, S., Barrero, J. M., Bloom, N., Bunn, P., Chen, S., ... & Thwaites, G. (2020). Economic uncertainty before and during the COVID-19 pandemic. Journal of Public Economics, 191, 104274.
- Kearney, C., & Liu, S. (2014). Textual sentiment in finance: A survey of methods and models. International Review of Financial Analysis, 33, 171-185.
- Picault, M., Pinter, J., & Renault, T. (2022). Media sentiment on monetary policy: determinants and relevance for inflation expectations. Journal of International Money and Finance.
- Picault, M., & Renault, T. (2017). Words are not all created equal: A new measure of ECB communication. Journal of International Money and Finance, 79, 136-156.
- Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. Journal of Accounting Research, 54(4), 1187-1230.
- Renault, T. (2020). Sentiment analysis and machine learning in finance: a comparison of methods and models on one million messages. Digital Finance, 2(1), 1-13.
- Renault, T. (2017). Intraday online investor sentiment and return patterns in the US stock market. Journal of Banking & Finance, 84, 25-40.
- Thorsrud, L. A. (2020). Words are the new numbers: A newsy coincident index of the business cycle. Journal of Business & Economic Statistics, 38(2), 393-409.

Books:

- Mitchell, R. (2018). Web scraping with Python: Collecting more data from the modern web. " O'Reilly Media, Inc.".
- Bengfort, B., Bilbro, R., & Ojeda, T. (2018). Applied text analysis with python: Enabling language-aware data products with machine learning. " O'Reilly Media, Inc.".
- Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python: analyzing text with the natural language toolkit. " O'Reilly Media, Inc.".

About the keynote lecturer

Michael McMahon is Professor of Economics at University of Oxford and Senior Research Fellow at St Hugh's College. He worked at the Bank of England for many years. Since April 2019, he serves as a member of the Irish Fiscal Advisory Council. He is a research fellow of the CEPR and Director of the Research Policy Network on Central Bank Communication. His interests lie in macroeconomics of fiscal policy, business cycles, monetary economics, inventories and applied econometrics. A key feature of his recent research is the use of interdisciplinary, data science techniques to understand communication and deliberation in central banks. His research has been published in journals including the Quarterly Journal of Economics, Review of Economic Studies, Journal of Monetary Economics, Review of Economics and Statistics, Journal of International Economics, and numerous others.

NÁRODNÁ BANKA SLOVENSKA
EUROSYSTÉM

Schedule

Welcome drink: Sunday evening (June 18th)
Social event: Wednesday evening (June 21st)

	Monday
09.00 – 12.00	**Introduction to Python**
13:00 – 14:00	**Keynote lecture by Michael McMahon**
14.30 – 16.30	**Application: Web scraping of the Wall Street Journal**
	Tuesday
09.00 – 12.00	**Natural Language Processing (Creating an NLP pipeline)**
13.00 – 15.30	**Application: Using NLP to analyse central bank communication**
	Wednesday
09.00 – 12.00	**Sentiment Analysis**
13.00 – 15.30	**Application: Media sentiment and CB communication**
	Thursday
09.00 – 12.00	**Machine learning using text as data**
13.00 – 15.30	**Application: Predicting Asset Prices using StockTwits**
	Friday
09.00 – 12.00	**Advanced methods in text mining (LDA, Transformers, Bert)**
13.00 – 15.30	**Application: Latent Dirichlet Allocation on Glassdoor**

The program is subject to change.