Topics in Fintech
(Introduction to Textual Analysis Using Python)

Spring, 2021
Syllabus (Session 3)

INSTRUCTORS

Dr. Zhigang QIU (Session 1)
Email: zhigang.qiu@ruc.edu.cn

Dr. Ke WU (Session 2)
Email: ke.wu@ruc.edu.cn

Dr. Dun JIA (Session 3)
Email: dun.jia@ruc.edu.cn

LECTURES

2:00 PM - 5:15 PM on Fridays

DAYS OF MEETINGS

May 7, May 14, May 21, May 28

COURSE DESCRIPTION

This course is a master-level course that introduces basic concepts in Fintech and important analytic tools in mid of the advancement of IT technology and big data analytics. We will be covering the theoretical rationales behind the machine learning and textual analysis along with their applications in asset pricing, financial analysis, macroeconomic analysis, and portfolio management.

Here is the 3rd session of the course, which focuses on textual analysis and aims to equip students with a framework and the toolkit to examine important questions in finance using text-based data, given limitations on the usages of more traditional form of data and measurements. We will be primarily using Python for understanding mechanics in textual analysis. Through going through various applications, students will be introduced to frontier themes including basics of Natural Language Processing (NLP), textual extraction and representation, thematic clustering, tone and sentiment analysis, web scraping and ultimately deep learning for textual analysis. Students are expected to be at least better informed, if not

becoming hands-on practitioners, of these enhanced tech tools for enriched understandings via lens of Fintech.

PREREQUISITE

Students should have taken introductory statistics and possess very basic programming intuition and experience (any computer language is fine). Advanced courses on financial markets and investment, intermediate micro and macroeconomics, accounting, and corporate finance are highly recommended before taking this class.

REFERENCE BOOKS

This course will heavily draw on materials from a range of reference books. Note that any newer edition of this book should suffice:

- Bhargav Srinivasa-Desikan. *Natural Language Processing and Computational Linguistics. A practical Guide to Text Analysis with Python, Gensim, spaCy and Keras*, Packt, 2018
- Benjamin Bengfort, Tony Ojeda, and Rebecca Bilbro. *Applied Text Analysis with Python: Enabling Language-Aware Data Products with Machine Learning*, O'Reilly Media, 2018
- Dirk Hovy. *Text Analysis in Python for Social Scientists, Discovery and Exploration*, Cambridge University Press, 2020
- Seppe vanden Broucke and Bart Baesens. *Practical Web Scraping for Data Science, Best Practices and Examples with Python*, APress, 2018

GRADING

My part of Session 3 of this course constitutes 1/3 of your final grade, which can be further broken down into three parts with relative weights:

- 5 % Class participation
- 35% ONE Pop-Quiz (45 Minutes)
- 60% Final Exam.

The grade of my session is an *weighted average* of scores of categories above. Your overall course grade will be simple averages of grades obtained across the three sessions. If at any point during the semester, you face circumstances which prevent you from attending the lecture, handing in the assignment on time, and/or participating in exams, please contact me as *early* as possible to manage the situation. There is little that can be done after an unsatisfactory grade has been assigned.

STUDY PREPARATION AND PARTICIPATION

1. You are responsible for ALL the materials delivered in the class to better prepare yourself for the successful completion of this course.

2. Do *read* and *study* the relevant PPT slides and related book chapters

3. Carefully take *notes* from the class.

POP QUIZ

- There will be ONE pop quiz that will be given out to students for completion in class. This pop quiz is not pre-scheduled, which means that it is mostly likely taking place on the 3rd or the 4th lecture day. There will be 20 questions related to concepts for textual analysis and syntax related to Python. The questions in the quiz will also help you work through some parts of the materials covered in this class and will be useful for the preparation of the final exam.

EXAMS

There will be **ONE** final exam. The exam covers ALL we have covered in class. You are expected to do simple calculations, derivations and to provide intuitions and explanations in short answers. If you fail to take the exam, you will get a clear zero for the grading part of final exam, unless your absence is due to one of the legitimate causes and you provide documentation timely.

ACADEMIC INTEGRITY

As a student, you are responsible for upholding the academic integrity with full commitment to all the ethics, codes, and standards of the Renmin University of China. It is very important for you to be aware of the consequences of cheating, fabrication, facilitation, and plagiarism. If there is any breach and violation of the academic integrity detected, as instructor, I will be filing a complaint for further actions against the violation, and this carries very serious consequences.

COURSE EVALUATION

At any time of our semester, please let me know what you think about this class and what can be improved. I appreciate all your feedbacks. Incorporating your suggestions will greatly help me in updating the course for future generations of students.

COURSE PLAN

- Disclaimer: the course plan is adjustable and subject to due changes

Lecture Day 1 (May 7, 2021)

- introduction to Python
- basic concepts in Natural Language Processing
- text parsing
- understanding the Chinese texts

Lecture Day 2 (May 14, 2021)

- clustering and representation
- sentiment analysis

Lecture Day 3 (May 21, 2021)

- web scrapping
- data visualization

Lecture Day 4 (May 28, 2021)

- topic modelling
- textual analysis based on deep learning
- Extending beyond textual analysis using Python: application to portfolio analysis