Special Topics Course in NLP and the Investment Process – Syllabus (Fall 2021)
Prof. Dan Rodriguez
Meeting Time: Saturdays from 10am – 12:40pm in Rogers Hall Rm 214
Sept. 4th – Oct 26th, 2021

The purpose of this course is to introduce the student to the process of natural language processing (NLP) and how this process can be employed for the purpose of supporting the investment process or portfolio management at an asset management firm

The reading materials for the course include two reference textbooks that will help you with the processing of unstructured text data: 1) *Applied Text Analysis with Python* (2018) by Benjamin Bengfort, Rebecca Bilbro & Tony Ojeda, and 2) *Natural Language Processing with Python* (2009) by Steven Bird, Ewan Klein and Edward Loper, which is also available at the following link: https://www.nltk.org/book/ . In addition to these references, I have included four applications of NLP in investment management.

Topic 1 – **Textual Analysis, Dictionaries, and 10-Ks**: In this topic, we replicate the first well-known NLP application for financial analysis, the evaluation of annual reports from Loughran and McDonald(2011), "When is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks", The Journal of Finance Vol. 66, No. 1, pp. 35-65.

Topic 2 – Stock Prediction Using **Twitter Sentiment Analysis**: In this topic, we will learn how to collect and analyze Twitter data to determine if we can leverage this data to help inform investments on both the long and short side for specific stocks, both around earnings announcements and outside of earnings announcements. Some of the primary references supporting this topic include: 1) "Stock Prediction Using Twitter Sentiment Analysis", (2011) by Anshul Mittal and Arpit Goel, 2) "Sentiment Analaysi of Twitter Data for Predicting Stock Market Movements," (2016) by Pagolu, et al. and 3) Nisar and Yeung(2018) "Twitter as a tool for forecasting stock market movements: A short-window event study," *The Journal of Finance and Data Science*, Vol. 4, pp. 101-119.

Topic 3 - Web Scraping for Alternative Data: In this topic, we learn how to apply web scraping techniques in order to collect financial data from a variety of websites that can help inform the investment process that can be supported by complementary sentiment analysis. See "Applications of web scraping in financial services industry," by Preetish, Nov. 29, 2017 (promptcloud.com/blog/applications-of-web-scraping-in-financial-services-industry/.

Topic 4 – Applying **customized NLP** to the analysis of recurring announcements such as Earnings Announcements and the decisions of the FOMC. Materials to be distributed in class.

Since each of these topics requires multiple steps of data collection, cleaning and analysis, we will allocate two to three sessions per topic, in order to ensure comprehension of the necessary methods to effectively implement each of the analyses required for each topic.

Course evaluation will be based upon project submissions for each major topic at 15 points per submission for a total of 60 points from submitted projects. In addition, there will be a final exam addressing each of the topics discussed above worth 40 points.

Assignment	Points	% of Total Grade
Textual Analysis w/Dictionaries	15	15
Twitter Sentiment Analysis	15	15
Web Scraping	15	15
Customized NLP Application	15	15
Final Exam/Final Course Project	40	40
Total	100	100

Please go to GitHub to set-up an account so that you can submit your homework assignments through this channel. My GitHub username is rangerrod1. You will share your completed assignments using this username.

Class Schedule

Wk 1	Course Overview and Introduction to Textual Analysis and Dictionaries
Wk 2	2nd Session for Textual Analysis and Dictionaries
Wk 3	Introduction of Twitter Sentiment Analysis
Wk 4	Twitter Sentiment Analysis Applications
Wk 5	Twitter Sentiment Analysis During Earnings and Tracking Macro Events
Wk 6	Introduction to Webscraping and Applications
Wk 7	Customized NLP Applications in Earnings and FOMC
Wk 8	Course Summary and Final Exam

The course instructor can be reached at dr1902@nyu.edu or rangerrod1@gmail.com or by cell phone via text message at 512-676-1246. Please text me since I am usually unable to take calls during the trading day.

Best of luck,
Prof. Rodriguez

Brief Bio

Currently the CRO for Light Sky Macro, a discretionary global macro fund investing in all asset classes across all global regions.
Dr. Dan Rodriguez previously worked as a Senior Risk Officer at Point 72 Asset Management with responsibilty for overseeing the long-short portfolio holdings for generalists and for

overseeing the risk management of the Global Macro Group of the fund. Dan was previously a Managing Director in the Credit-Suisse Global Equities Division, and served as the Chief Risk Officer for the Systematic Market-Making Group of the Investment Bank. His responsibilities included the front-office risk management for a global cross-asset trading portfolio, which includes the development of portfolio risk limits, daily and intra-day monitoring of portfolio risk profiles.

Dan joined Credit Suisse in June 2007 and moved to Point 72 in October of 2014. He has fifteen years of risk management experience, including previous positions at Morgan Stanley, reporting to the CRO of the firm, with responsibility for market risk methodology for the Institutional Securities Group. He began his career monitoring the risk of the Commodities Division at Morgan Stanley.
Dan holds a Ph.D. in Economics from M.I.T. and a Bachelor's of Science degree from the United States Military Academy and received a National Science Foundation Fellowship upon completion of his undergraduate studies. He received the PRMIA Higher Standard Award in 2011 for outstanding service in risk management and has published in leading academic journals such as the Journal of Finance, International Corporate Governance, and the Industrial Labor Relations Review. Dan served as the co-Regional Director of the New York City Chapter of PRMIA, with a membership of over 17,000 from 2011 – 2014.

Resources

- **Access your course materials**: NYU Classes (nyu.edu/its/classes)
- **Databases, journal articles, and more**: Bern Dibner Library (library.nyu.edu)
  NYU Virtual Business Library (guides.nyu.edu/vbl)
- **Obtain 24/7 technology assistance:** Tandon IT Help Desk (soehelpdesk@nyu.edu, 646.997.3123)
  NYU IT Service Desk (AskIT@nyu.edu, 212-998-3333)

Policies

Academic Misconduct

A. Introduction: The School of Engineering encourages academic excellence in an environment that promotes honesty, integrity, and fairness, and students at the School of Engineering are expected to exhibit those qualities in their academic work. It is through the process of submitting their own work and receiving honest feedback on that work that students may progress academically. Any act of academic dishonesty is seen as an attack upon the School and will not be tolerated. Furthermore, those who breach the School's rules on academic integrity will be sanctioned under this Policy. Students are responsible for familiarizing themselves with the School's Policy on Academic Misconduct.
B. Definition: Academic dishonesty may include misrepresentation, deception, dishonesty, or any act of falsification committed by a student to influence a grade or other academic

evaluation. Academic dishonesty also includes intentionally damaging the academic work of others or assisting other students in acts of dishonesty. Common examples of academically dishonest behavior include, but are not limited to, the following:

1. Cheating: intentionally using or attempting to use unauthorized notes, books, electronic media, or electronic communications in an exam; talking with fellow students or looking at another person's work during an exam; submitting work prepared in advance for an in-class examination; having someone take an exam for you or taking an exam for someone else; violating other rules governing the administration of examinations.
2. Fabrication: including but not limited to, falsifying experimental data and/or citations.
3. Plagiarism: Intentionally or knowingly representing the words or ideas of another as one's own in any academic exercise; failure to attribute direct quotations, paraphrases, or borrowed facts or information.
4. Unauthorized collaboration: working together on work that was meant to be done individually.
5. Duplicating work: presenting for grading the same work for more than one project or in more than one class, unless express and prior permission have been received from the course instructor(s) or research adviser involved.
6. Forgery: altering any academic document, including, but not limited to, academic records, admissions materials, or medical excuses.

Disability Disclosure Statement

Academic accommodations are available for students with disabilities. Please contact the **Moses Center for Students with Disabilities** (212-998-4980 or mosescsd@nyu.edu) for further information. Students who are requesting academic accommodations are advised to reach out to the Moses Center as early as possible in the semester for assistance.

Inclusion Statement

The NYU Tandon School values an inclusive and equitable environment for all our students. I hope to foster a sense of community in this class and consider it a place where individuals of all backgrounds, beliefs, ethnicities, national origins, gender identities, sexual orientations, religious and political affiliations, and abilities will be treated with respect. It is my intent that all students' learning needs be addressed both in and out of class, and that the diversity that students bring to this class be viewed as a resource, strength and benefit. If this standard is not being upheld, please feel free to speak with me.