

Special Topics Course in NLP and the Investment Process – Syllabus (Fall 2019)

Prof. Dan Rodriguez

Meeting Time: Saturdays from 10am – 12:40pm

Oct. 26th – Dec. 21st

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. In addition to these references, I have included four applications related to NLP applications in finance.

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 Analysis 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. 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 – 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.

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	2 nd 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 dr190@nyu.edu or drodr87@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 responsibility 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 in risk management, 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.