

26:010:685:01: Decoding of Textual Corp Com

Fall 2018

NWK: 1WP: 358

MON 1:00-3:50PM

Dr. Mauricio Codesso

NWK: 1WP: 919C

mauricio.codesso@rutgers.edu

Office Hours: MON 04:00-05:00PM

Course Description

In Decoding of Textual Corporate Communications, you will learn the literature and techniques needed to incorporate unstructured text into your research.

Academic Integrity

Academic dishonesty will not be tolerated in class and could result in a failing grade. Please read the complete Rutgers policy on academic integrity at:

<http://academicintegrity.rutgers.edu/>

Harassment and Discrimination

Harassment and discrimination of any sort will not be tolerated in this class. If you feel you are a victim of harassment or discrimination, or if you are a witness to such behavior, please contact the instructor immediately. Please see the complete Rutgers policy on harassment and discrimination at:

<http://policies.rutgers.edu/PDF/Section60/60.1.12-current.pdf>

Point Allocation

Attendance (140)

Participation (260)

Final Research Proposal (200)

Attendance and Participation Assignments

Attendance will be taken every day. You will not earn these points if you are absent. You will not lose points for University excused absences that are also reported through the University systems at <https://sims.rutgers.edu/ssra/>. If you do miss a day of class due to an excused absence you will still be responsible for turning in the required assignments by the following Monday at noon.

Students are required to read all papers assigned for each class and prepare a 40-minute PowerPoint presentation on one of the papers for each class (these will be assigned in class). PowerPoint files must be turned in before class starts. From time to time coding assignment will also be given and will count toward the participation grade.

Final Research Proposal

A minimum 2000-word research proposal (excluding the required references) will be turned in at the end of the semester. More details will be given during the sixth week of class.

Required Book

Make your own Neural Network E-Book (Rashid, 2016)

Course Schedule

The course schedule is subject to change according to our progress in class.

Week	Topic	Papers	Techniques
1	Overview of Text Mining Research		Install Anaconda environment. Create Python 3.x instance with NLTK Python basics Program control flow Read files Write to .txt files Write to .csv files Accessing Data sources
2	Dictionaries	When is a Liability not a Liability. (Loughran and McDonald, 2011) MD&A Disclosure and the Firm's Ability to Continue as a Going Concern (Mayew, Sethuraman, Venkatachalam, 2014)	Dictionary building Words Phrases Lower case Stemming and lemmatization Removing stop words
3	Readability	Annual report readability, current earnings, and earnings persistence (Feng Li, 2008) Readability and understandability: Different measures of the textual complexity of accounting narrative. (Smith and Taffler, 1992).	Sentence length Word length Syllable counting FOG, Flesch-Kincaid, ARI
4	Regular Expressions	Automated contract analysis in auditing. (Yan, Moffitt, Titera, 2017...On Blackboard) Kravet, Todd, and Volkan Muslu. Textual risk disclosures and investors' risk perceptions. (Kravet, Todd, Muslu, 2013)	Regular Expressions Part 1 RegEx Basics RegEx in Python Regular expression exercise Use regular expressions to count and write to CSV Use regular expressions to parse a document Use regular expressions to count phrases, permutations of words, etc... RegEx Parsing an XBRL document
5	Regular Expressions	Making words work: Using financial text as a predictor of	Regular Expressions Part 2 RegEx Advanced Topics

		<p>financial events. (Cecchini et al., 2010)</p> <p>Textual Analysis in Accounting and Finance: A Survey. (Loughran and McDonald, 2016)</p>	<p>RegEx Parsing HTML 10-K</p> <p>RegEx Implementing Dictionary</p>
6	POS Tagging	<p>Holton, Carolyn. "Identifying disgruntled employee systems fraud risk through text mining: A simple solution for a multi-billion-dollar problem. (Holton, 2009)</p> <p>Which spoken language markers identify deception in high-stakes settings? Evidence from earnings conference calls. (<i>Burgoon et al., 2015</i>)</p>	<p>POS Tagging Techniques</p> <p>Penn Treebank</p> <p>Expressivity</p> <p>Parsing Sentences by Tense</p> <p>Passive Voice</p> <p>Noun Phrase Identification</p>
7	Document Similarity	<p>Jegadeesh, Narasimhan, and Di Wu. "Word power: A new approach for content analysis. (Jegadeesh et al. 2013)</p> <p>Large-sample evidence on firms' year-over-year MD&A modifications. (Brown and Tucker, 2011)</p>	<p>TF-IDF</p> <p>Cosine Similarity</p>
8	MYSQL	<p>Natural language processing in accounting, auditing and finance: a synthesis of the literature with a roadmap for future research. (Fisher et al. 2016)</p> <p>Larcker, David F., and Anastasia A. Zakolyukina. "Detecting deceptive discussions in conference calls. (Larcker and Zakolyukina, 2012)</p>	<p>Writing results to MySQL database</p> <p>Install MySQL</p> <p>Connect to MySQL</p> <p>Build conceptual model</p> <p>Query using MySQL</p> <p>Query using Python</p>
9	Topic Modeling	<p>Simultaneously Discovering and Quantifying Risk Types</p>	<p>Topic modeling</p> <p>LDA</p>

		<p>from Textual Risk Disclosures (Bao and Datta, 2014)</p> <p>The information content of mandatory risk factor disclosures in corporate filings. (Campbell et al., 2014)</p>	<p>Dictionary-based</p> <p>Search engine based</p>
10	Naïve Bayes	<p>The information content of forward-looking statements in corporate filings—A naïve Bayesian machine learning approach. (Feng Li, 2010)</p> <p>Accounting variables, deception, and a bag of words: assessing the tools of fraud detection. (Purda and Skillicorn, 2015)</p>	<p>Bayesian Probability</p> <p>Naïve Bayes Algorithm</p>
11	Neural Network	<p>Make your own Neural Network E-Book (Rashid, 2016)</p>	<p>Neural networks Part 1</p> <p>Neural network basics</p>
12	Neural Network	<p>Make your own Neural Network E-Book (Rashid, 2016)</p>	<p>Neural networks Part 2</p> <p>Convolutional networks</p> <p>Recursive networks</p> <p>Sentence classification</p> <p>Tensor Flow</p>
13	Sentiment Analysis	<p>Evaluating sentiment in financial news articles. (Schumaker et al., 2012)</p> <p>Processing fluency and investors' reactions to disclosure readability. (Rennekamp, 2012).</p>	<p>Word2Vec</p> <p>Doc2Vec</p> <p>Dictionaries</p>
14	Student Research Presentations		