

<https://www.andrewoarnold.com/nyu7871/> (Last year's website, for reference)

The fast-growing field of news analytics requires large databases, fast computation, and robust statistics. This course introduces the tools and techniques of analyzing news, how to quantify textual items based on, for example, positive or negative sentiment, relevance to stocks or other indicators, and the amount of novelty in the content. Applications to trading strategies are discussed, including both absolute and relative return strategies, and risk management strategies. Students will be exposed to leading software, tools and datasets in this area.

Students will benefit from some familiarity with basic probability, statistics and programming (python), and an interest in natural language processing (NLP) or computational linguistics. While the course will introduce a few trading strategies, it will also focus on NLP as a tool in its own right, applicable to domains outside of quantitative trading strategies.

There will be weekly reading, homework, a midterm exam and a final project.

Diversity, equity and inclusion: 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.

Content:

1. Introduction to NLP:
 - 1.1. Literature
 - 1.2. Vocabulary
 - 1.3. Methods
 - 1.4. Tools
 - 1.5. Datasets
2. Dealing with data:
 - 2.1. Text
 - 2.2. Speech
 - 2.3. Cleaning up text
 - 2.3.1. Stop words
 - 2.3.2. Stemming
 - 2.3.3. Lemmatization
 - 2.3.4. Tokenization
 - 2.3.5. Feature engineering
 - 2.3.6. Challenges
3. Models
 - 3.1. Dictionaries
 - 3.2. N-grams
 - 3.3. Supervised vs. Unsupervised learning
 - 3.4. Similarity metrics
 - 3.5. Sentiment analysis
 - 3.6. Topic modeling
 - 3.7. Classification

- 3.8. Collaborative filtering
- 3.9. Ensemble methods
- 3.10 Embedding models
- 3.11 Attention models
- 4. Applications of NLP to Quantitative Trading:
 - 4.1. Fundamentals of markets
 - 4.2. Trading strategies