BAIS:6100 Text Analytics

Text Clustering
Topic Modeling

Kang-Pyo Lee

Course Schedule (Subject to Change)

Week	Date	Topics	Due
1	Jan 28	Introduction to Text Analytics Introduction to Python, Jupyter Notebook, and UI Interactive Data Analytics Service (IDAS)	
2	Feb 4	Module 1. Python Basics for Text Processing, Part 1 : Strings, Collections, Built-in Functions, Flow Control, and User-Defined Functions	
3	Feb 11	Module 2. Python Basics for Text Processing, Part 2 : Files, Dataframes, and Pattern Matching Using Regular Expressions	HW 1
4	Feb 18	Module 3. Basic Natural Language Processing (NLP) Techniques : Tokenization, Part-of-Speech Tagging, Stemming, Lemmatization, N-grams, Noun Phrase Extraction, Language Detection and Translation, and Gender Prediction Module 4. Keyword Analysis and Visualization	HW 2
5	Feb 25	Test 1	HW 3 (Feb 24)
6	Mar 4	Modules 5 & 6. Text Data Collection Using Twitter APIs and Web Scraping Group Project Announcement	
7	Mar 11	Module 7. Document-Term Representation Module 8. Text Classification	HW 4
8	Mar 18	Module 9. Text Clustering and Topic Modeling	Project Proposal
9	Mar 25	Module 10. Text Similarity Module 11. Keyword Network Analysis	
10	Apr 1	Test 2	HW 5 (Mar 31)
11	Apr 8	Group Project Presentations and Course Wrap-Up	Project Deliverables

Group Project Proposals

Proposals from 10 groups

- **BKP - www.poandpo.com (Post Online Media)**
- **CryptoPredictors - news.bitcoin.com (Bitcoin news)**
- **DuchessOfData - www.slate.com/technology (daily magazine)**
- **Fish_Thanks - www.rottentomatoes.com/ (movie/show reviews)**
- **JD - theconversation.com/us/topics/climate-change-27 (news for public good)**
- **K-means - www.eater.com/ (food blog)**
- **LaughTrack - www.rogerebert.com/ (movie reviews)**
- **TBD - pitchfork.com/reviews/albums/ (music review)**
- **TextMiners - www.cnet.com/topics/tech-industry (tech news)**
- **The_Onion - www.sciencenews.org/all-stories (science news)**