

TEXT ANALYTICS

Course Instructor Kang Pyo Lee Campus Address: W383 PBB Phone: (319) 384-1205 Email: kangpyo-lee@uiowa.edu Office Hours: Wednesdays, 10:00 AM – 11:00 AM via Zoom by appointment	Class Meeting Times Thursdays, Jan 28, 2021 – Apr 8, 2021 6:00 PM - 9:40 PM This course is organized within a structure of scheduled course work. Students will progress through the course as a cohort and will complete and/or submit course work online. Recorded lectures will be viewable at choice of time within week. Required (online) class sessions are a part of this course.
Academic Course Home Tippie College of Business DEO: Prof. Ann Campbell Phone: (319) 335-0918 Email: ann-campbell@uiowa.edu	Course Site To access the course site, log into Iowa Courses Online (ICON) using your Hawk ID and password. Prerequisites (MSCI/BAIS:6060 Data Programming in R or MSCI/BAIS:9060 Data Programming in R or MSCI/BAIS:6040 Data Programming in Python) and (MSCI/BAIS:6070 Data Science or MSCI/BAIS:9110 Advanced Analytics)

Program Goals

The Tippie MSBA has learning goals that drive decisions about curriculum and assignments within courses:

Program Goal 1: Graduates will exhibit knowledge and skills relevant to data and its application in business

Program Goal 2: Graduates will create and communicate solutions to data-related business problems that impact their organizations and communities

Program Goal 3: Graduates will understand and contemplate ethical and privacy issues arising in their own work

Program Goal 4: Graduates will demonstrate the ability to be effective team members in a diverse and complex world.

Course Description and Goals

Unstructured text data in business is being generated at an unprecedented speed from various sources such as customer reviews, emails, electronic publications, blogs, tweets, websites, and so on. Given this large volume of text data, businesses face the challenge of finding valuable information and insights within such data to support decision making. This course will introduce the concepts and techniques of text analytics using the Python programming language. Topics covered will include Python basics for text processing, Natural Language Processing (NLP) techniques, keyword analysis and visualization, text data acquisition, term-document matrix representation, text classification, text clustering and topic modeling, sentiment analysis, text similarity, and full-text search. This course will combine lectures, assignments, and projects to reinforce the learning of these concepts and techniques and introduce students to the real-world practice of large-scale text

mining.

By the end of the course, students will be expected to demonstrate that they have met the following learning objectives:

- Use Python to perform text processing.
- Use Python to apply Natural Language Processing techniques.
- Use Python to perform keyword analysis and visualization.
- Use Python to perform text classification, clustering, and topic modeling.
- Use Python to collect text data using web scraping and APIs.
- Be able to explain the main challenges in text data analysis and their solutions.

Course activities will include formal and active-learning lectures, individual in-class exercises, individual homework assignments, individual tests, and a group project.

Media/System Requirements

Technical requirements for completing this online course include:

- Student-provided personal computer.
- Computer with reliable Internet access. A wired Ethernet connection to the internet is very strongly suggested. Wireless and cellphone data connections may experience connection problems. Android and iOS operating systems are not fully supported at this time. See specific requirements on the [Distance and Online Education Technical Requirements/Download page](#).
- USB headset with microphone and a web camera capable of conferencing. Most laptops have a built in camera which typically works fine. However, a USB headset with microphone is recommended to minimize audio feedback and room noise and maximize sound quality.
- While tablets, smartphones and other mobile devices may allow for some completion of coursework, they are not guaranteed to work in all areas. Please ensure you have a Windows or Mac based computer available to complete coursework in the event your selected mobile device does not meet the needs of the course.

Students who need assistive technologies will have different computer and technology requirements. Please check with your [Student Disability Services](#) to determine the requirements for the specific technologies needed to support your online classes.

For questions regarding virtual classrooms (i.e. Zoom) or UICapture (Panopto) please contact [Continuing Education Technical Support](#) (319-335-3925).

Need help with ICON or your HawkId? Please contact the [ITS Helpdesk](#) (319-384-HELP).

Required Textbook/Media

There are no required textbooks for the course.

Grading Criteria

Final grades will be assigned according to the following planned curve (however, the instructor reserves the right to deviate from this curve if deemed necessary):

- A range: \approx 60% of students
- B range: \approx 40% of students
- C, D, F: as needed

In addition, the A range and B range will be divided into +/- designations.

Course Structure

This course is being offered over the World Wide Web. Students will **login to the course site** on ICON to access the course materials. For details of the course assignments and activities, see the **“Course Work”** section of this syllabus.

Students are expected to visit the course site regularly to:

- **Access assigned course materials (posted on the “Modules” page)** such as pre-recorded lectures and journal articles.
- **Review the course homepage regularly** for any updates related to the course **“Announcements”** and/or **“Calendar.”**
- **Submit** assignments to the course instructor via the **ICON “Assignments.”**
- **Participate** in the **“Discussion”** forums.

Course Work

The coursework breakdown is as follows:

- 35% homework assignments (five assignments, see calendar)
- 50% tests (two in-class exams, equally weighted, see calendar)
- 15% group project (presented during the last week of semester)

Course-Specific Policies and Guidelines

As a registered student you are responsible for the course policies posted below.

Communications: Students can expect to receive weekly communications from the instructor (via course “Announcements”). Students are also responsible for all official correspondence sent through their University of Iowa email address. Privacy considerations, such as federal law, may apply when using an address other than the standard University e-mail address. You can expect to receive responses to your inquiries within 24-48 hours.

Exams: Students are expected to take the exams at the regularly scheduled times, unless permission has been granted by the instructor at least 1 week in advance in writing. All exams are **INDIVIDUAL WORK** and collaboration on exams is not permitted under any circumstance. All exams are open book and open Internet but closed communication. Use of any means for communication during the exam is strictly prohibited and will be grounds for a grade of zero on the exam.

Attendance: Attendance will not be taken. However, students are expected to check all the announcements made during each class that they need to be absent for via class recordings.

Late Work: All assignments are expected on time. You may turn in an assignment late, but you will receive a 20% deduction for each day that it is late, including the first/same day. For example, if Assignment X is due at 6:00 PM on Thursday, and you submit the assignment at 6:01 PM on Thursday, then you will lose 20%; if you submit the assignment anytime on Friday, then you will lose 40%; etc.

Collaboration: I encourage discussion and collaboration, but all assignments must be written by the student (or students in the case of group project) alone. Students must not share code for any homework assignment and test under any circumstance.

Netiquette: The term “netiquette” refers to the do’s and don’ts of online communication. As it applies to this online course, it is my expectation that students will communicate effectively and respectfully with each other,

the instructor, and our guest speakers (if applicable). [Follow this link to learn more about The Core Rules of Netiquette.](http://www.albion.com/netiquette/corerules.html) <http://www.albion.com/netiquette/corerules.html>

Inclement Weather/Class Cancellation: Although it is our intent to offer every class at its assigned time, on rare occasion there are weather or other emergency events that require that alternative arrangements are made for class delivery. Because these sessions are online, even if Tippie College or University in-person classes are cancelled, I anticipate holding the regularly scheduled Zoom session of the class. If a situation occurs such that Zoom cannot be used, I will cancel the class completely – all reading and work that was due is expected to be completed and turned in on time through ICON.

Collegiate and University Policies

The administrative home of this course is the Tippie College of Business, which governs academic matters relating to the course such as the add/drop deadlines, issues concerning academic misconduct, and how credits are applied for various graduation requirements. Different colleges might have different policies.

As a registered student in a course in the Tippie College of Business, you are responsible for the collegiate policies posted at <https://tippie.uiowa.edu/collegiate-policies-mba>.

As a registered student in a course through The University of Iowa, you are responsible for the University policies posted at https://provost.uiowa.edu/sites/provost.uiowa.edu/files/wysiwyg_uploads/Syllabi%20updates%20for%20spring%202021.pdf.

Course Calendar (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 23)
6	Mar 4	Group Project Announcement Modules 5 & 6. Text Data Acquisition Using Twitter APIs and Web Scraping	
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 30)

11	Apr 8	Group Project Presentations and Course Wrap-Up	Project Deliverables
----	-------	--	----------------------