

Syllabus CS 5293

Text Analytics (Spring 2020)

Class hours: Tuesday/Thursday 9:00 - 10:15am

Location: [Dale Hall Tower, 0104](#)

Instructors

Dr. Christan Grant

- **Email:** cgrant@ou.edu
- **Office Hours:** Online: T/R 2-3p; In-class: T/R 2-2:30pm
- **Location:** Devon Energy Hall 234
- **Online (Google Hangouts):** <https://bit.ly/1ZiLbEH>

Teaching Assistant

Keerti Banweer

- **Email:** keerti.banweer@ou.edu
- **Office Hours:** Monday 4-6pm; Wednesday 10-11am, at DEH 115
- **Location:** Devon Energy Hall 115
- **Online (Google Hangouts):** <https://bit.ly/1ZiLbEH>

Grader

MG Hirsch

- **Email:** mghirsch@ou.edu
- **Office Hours:** TBD
- **Location:** Devon Energy Hall 115
- **Online (Google Hangouts):** <https://bit.ly/1ZiLbEH>

Note: Any email messages to the professors or teaching assistants must include **cs5293 in the subject line**. Any email without this string in the subject line will likely be filtered as junk.

External Tutors

The William Kerber Teaching Scholars will be available for questions during the times listed below. Note that these assistants can provide general help with programming, compiling and editing, but will not know about the class projects. All of their office hours will be held in DEH 115.

Melissa Wilson, Nathan Huffman, Jennifer Pham are available at the following times:

Day	Times
Mondays	9:30a - 3p
Tuesdays	9:30a - 12p 12:30p - 4p
Wednesdays	9:30a - 3p
Thursdays	9:30a - 12p 12:30p - 3:00p
Fridays	12:30p - 2:30p

Prerequisites

Students must have a working understanding of statistics and data structures, in addition to a select set of software skills. The prerequisite courses/skills are listed here: - Statistics: MATH 4743 or MATH 4753 or ISE 3293 or ISE 5013 - Data Structures/Discrete Structures: CS 5005 or CS 2413 and CS 2813 - Software skills: Students should be well versed in Java and/or C++ and also familiar with at least one scripting language such as Python. Students should also be comfortable working with the GNU/Linux command line.

Undergraduate students with a 3.5 GPA or higher may enroll with permission from the instructor.

Course Description

Text Analytics are the methods and techniques used to extract useful knowledge from text to support decision making. This field includes a collection of research from the natural language processing, databases, data mining, and machine learning communities. The aim of this course is to be a primer for text analytics theory and practice. After taking this course, students will have an understanding of how to

independently obtain, parse, and analyze textual information for organizations that want to extract valuable insights.

Topics discussed in the course include: obtaining data sets, understanding data formats, duplicate detection, cleaning data sets, tagging, indexing and search, evaluating algorithms, classification, clustering, topic modeling and entity resolution. Time permitting, we may discuss advanced topic such as relation extraction, slot filling, knowledge graphs, knowledge base construction, the semantic web, question answering or other cutting-edge topics.

Lectures will be a mix of traditional lectures, class discussions, videos and other activities. Participation is required to get the most out of the class.

Learning Management System

We will use the Canvas learning system. This course website can be reached through canvas.ou.edu. Please check this system regularly to keep informed of all announcements, updates, and changes. Important course information will also be distributed through the [course website](#).

Course Materials

Required Textbooks:

- *Text Analytics With Python : A Practical Real-World Approach To Gaining Actionable Insights From Your Data, **Second Edition***, Dipanjan Sarkar, 2019, Apress L. P. (ISBN 978-1-4842-4353-4) (eBook ISBN 978-1-4842-4354-1) [Publisher Website](#)

Computer Accounts and Software

Increasingly, software is developed and executed in “the cloud”. This semester the class will make heavy use of a popular cloud infrastructure. Students will be able to deploy virtual machines with various configurations, on the fly. Credentials for using this infrastructure will be distributed after the first week of class. For questions and issues using this software, students should use the in-class discussion board. All students enrolled in class should also have a CS account and access to a Linux-based systems in the CS department. For most computer science students, an account will be automatically created. All code written for this course MUST run using the compilers or interpreters that will be specified for the assignments. It is your responsibility to ensure that your code runs on these systems. For compatibility reasons, we recommend developing and testing on a Linux-based machine.

Course Policies

- Attendance: You are expected to attend or watch all of the class lectures.
- Readings: For each lecture day, the [course schedule](#) lists a set of readings. You are responsible for this material **before class begins**.
- Laptop Computers: It is the responsibility of each student in this class to have a working laptop computer with ample battery (at least 2 hours of life under moderate usage) and wireless Internet connectivity. You must bring the laptop computer to class. If your computer requires repair during the semester, it is your responsibility to make arrangements to have another computer available and to get the necessary software installed. There exist campus resources (including financial help) to repair broken computers; please see the instructors if you would like information about these programs. Note that temporarily borrowing a computer from a fellow student in the class can present a number of problems, including the potential for academic misconduct.
- Newsgroups and Email: The newsgroup on Canvas should be the primary method of communication (outside of class). This allows everyone in the class to benefit from the answer to your question, and provides students with more timely answers since the TAs and instructors check Canvas at least once a day. Matters of personal interest should be directed to email instead of to the newsgroup, e.g. informing the instructors of an extended personal illness.
- Incompletes: The grade of "I" is intended for the rare circumstance when a student who has been successful in a class has an unexpected event occur shortly before the end of the class. We will not consider giving a student a grade of "I" unless the following three conditions have been met:
 1. It is within two weeks of the end of the semester.
 2. The student has a grade of C or better in the class.
 3. The reason that the student cannot complete the class is properly documented and compelling.
- Religious Holidays: It is the policy of the University to excuse the absences of students that result from religious observances and to provide without penalty for the rescheduling of examinations and additional required classwork that may fall on religious holidays.
- Classroom Conduct: Because cell phones and laptops can distract substantially from the classroom experience, students are asked not to use either during class, except in cases in which they are required as part of a classroom exercise. Disruptions of class will also not be permitted. In the case of disruptive behavior,

we may ask that you leave the classroom and may charge you with a violation of the Student Code of Responsibilities and Conduct. Examples of disruptive behavior include:

- Allowing a cell phone or pager to repeatedly beep audibly.
- Playing music or computer games during class in such a way that they are visible or audible to other class members.
- Exhibiting erratic or irrational behavior.
- Behavior that distracts the class from the subject matter or discussion.
- Making physical or verbal threats to a faculty member, teaching assistant, or class member.
- Refusal to comply with faculty or teaching assistant direction.

Proper Academic Conduct

- Feel free to discuss all assignments with the instructors or the TAs.
- Code (projects and homework): you may **discuss** code solutions with other students. However:
 1. You may not look at or share code with others;
 2. If you discuss a solution with anyone, you must document their names in your assignment;
 3. If you use an external resources (e.g. StackOverflow.com) you must document this in your assignment. Failing to do so properly will result is a breach of the student code of conduct.
- Quizzes, Exams, In-Class Exercises: unless otherwise stated, you may not communicate with others about solutions to these assignments.
- Make sure that your computer account is properly protected. Use an appropriate password, and do not give your friends access to your account or your computer system. Do not leave printouts, computers or thumb drives around a laboratory where others might access them.
- Programming projects will be checked by software designed to detect collaboration. This software is extremely effective and has withstood repeated reviews by the campus judicial processes.
- Upon the first documented occurrence of inappropriate collaborative work or of taking a solution from a network resource, the instructors will report the academic misconduct to the Campus Judicial Coordinator. The procedure to be followed is documented in the University of Oklahoma Academic Misconduct

Code (<http://integrity.ou.edu>). Both the provider of a solution and the receiver of a solution will be treated equally in the misconduct process.

Grading

Points for this class will come from a variety of sources. The different components are weighted as follows:

	Percentage
Quizzes	25%
Assignments	37%
Projects	38%
	100%

Materials will be posted and grades will be posted through the Canvas online platform and or the [course website](#). In-class students will be allowed one unexcused absence. Assignments and quizzes on that days will not be counted towards their final grade.

To perform well, active participation in **In-class assignments** is required. In-class exercises will be often and possibly will be unannounced. These exercises may include group discussion or individual problem-solving.

Other **Assignments** may be assigned weekly. Types include coding assignments, essay questions, online discussions and other similar questions. Online participation will be counted under assignments.

Most homework assignments will due before the start of class in the day indicated in the [class schedule](#). Students can waive one homework assignments without penalty.

Approximatley four **Projects** will be given over the course of the semester. These project will require a substantial amount of planning, programming and debugging. We encourage you to budget your time well for these. The projects will be due at *11:45 pm CST* on the day indicated in the [class schedule](#).

Submission Format

For written student submissions should only be .txt files, portable document format .pdf, or Markdown .md. Files of type .doc, .docx, or .rtf will not be accepted. Compressed files should be of type .gz or .tar.gz. Files of the .rar format will not

be accepted. Other file types, particularly coding files, may be used in the class. The expected file type will be stated. Often, files packaged under non-Unix/Linux flavored operating systems, such as Windows, have a non-negative number of compatibility issues with our grading systems. **If the graders cannot open files for these reasons, the project will not receive credit.**

Late Policy

Projects may be turned in up to 24 hours late for a 10% penalty. After this time window, no late work will be accepted.

Other assignments will not be accepted late.

Final Grade Scale

Grade cut-offs will be at or below the traditional 90, 80, 70, etc. cut-offs.

Grade questions

- Projects/homework/in-class assignments: Grading questions for projects should first be brought to the instructor/TA that originally did the grading. If talking to the TA does not resolve your question, please see one of the instructors. All grading questions must be brought to our attention within one week of them being graded.

Please note that when an exam/assignment is brought with grading questions, we may examine the entire exam/assignment and your final grade may end up lower.

Canvas Grade Summary

Canvas has a grade book that is used to store the data that are used to calculate your course grade. It is the responsibility of each student in this class to check their grades on Canvas after each assignment is returned. If an error is found, bring the graded document to any of the instructors or TAs, and we will correct Canvas.

Miscellaneous

Specific Outcomes of Instruction

By the end of the semester, the students will increase their:

- Ability to apply knowledge of computing and mathematics appropriate to the discipline.
- Ability to analyze a problem and identify and define the computing requirements appropriate to its solution.
- Ability to design, implement, and evaluate a computer-based system, process, component, or program to meet desired needs.
- Understanding of ethical issues and responsibilities.
- Recognition of the need for and an ability to engage in continuing professional development.
- Ability to use current techniques, skills, and tools necessary for computing practice.
- Ability to apply mathematical foundations, algorithmic principles, and computer science theory in the modeling and design of computer-based systems in a way that demonstrates comprehension of the trade-offs involved in design choices.
- Ability to apply design and development principles in the construction of software systems of varying complexity.

Course Evaluations

The College of Engineering utilizes student ratings as one of the bases for evaluating the teaching effectiveness of each of its faculty members. The results of these forms are important data used in the process of awarding tenure, making promotions, and giving salary increases. In addition, the faculty uses these forms to improve their own teaching effectiveness. The original request for the use of these forms came from students, and it is students who eventually benefit most from their use. Please take this task seriously and respond as honestly and precisely as possible, both to the machine-scored items and to the open-ended questions.

Reasonable Accommodation

The University of Oklahoma is committed to providing reasonable accommodation for all students with disabilities. Students with disabilities who require accommodations in this course are requested to speak with the professor as early in the semester as possible. Students with disabilities must be registered with the Office of Disability Services prior to receiving accommodations in this course. The [Office of Disability Services](#) is located in the University Community Center at 730 College Avenue; the phone is 405-325-3852 or TDD only is 403-325-4173.

Adjustment for Pregnancy/Childbirth-Related Issues

Should you need modifications or adjustments to your course requirements because of documented pregnancy-related or childbirth-related issues, please contact one of the instructors as soon as possible to discuss. Generally, modifications will be made where medically necessary and similar in scope to accommodations based temporary disability. Please see <http://www.ou.edu/eoo/faqs/pregnancy-faqs.html> for commonly asked questions.

Title IX Resources

For any concerns regarding gender-based discrimination, sexual harassment, sexual misconduct, stalking, or intimate partner violence, the University offers a variety of resources, including advocates on-call 24.7, counseling services, mutual no-contact orders, scheduling adjustments and disciplinary sanctions against the perpetrator. Please contact the Sexual Misconduct Office 405-325-2215 (8-5) or the Sexual Assault Response Team 405-615-0013 (24.7) to learn more or to report an incident.

Technical Support

For OU IT support, please phone (405) 325-HELP. For help with issues pertaining to any CS department machine (in room DEH 115). There is a OU SharePoint site that you can use for reference <https://sooners.sharepoint.com/sites/OUCSTutorials>.

This syllabus is subject to change. Students are responsible for any changes/additions to this syllabus announced during the semester.

Links

Key Class Resources

- [Syllabus](#)
- [Class Schedule](#)

Dates and details in the syllabus and schedule are subject to frequent change, please check regularly. Major changes will be announced on Canvas.

External Resources

Tools

- [GCP instructions](#)
- [Vim Cheat Sheet](#)

- [Vim reference](#)
- [Emacs reference](#)
- [Perf](#)
- [Valgrind](#)
- [Markdown](#)
- [Tmux Cheat Sheet](#)
- [Screen Cheat Sheet](#)
- [Bash for Windows](#): this gives you an Ubuntu-type environment on your Windows machine.

Tutorials

Others

- [Passwordless login](#)

This page is available online at: <https://oudalab.github.io/cs5293sp20>
