



Web Mining (BIA 660B)
School of Business
Spring 2020

Meeting Times: Thursdays 12:30 – 3:00 p.m.

Classroom Location: Babbio Center 319

Instructor: Jordan W. Suchow

Contact Info: jws@stevens.edu

TA: Sujith Kochupurackamuriyil Sam

TA Contact Info: skochupu@stevens.edu

Office Hours: TBD based on poll

Course Web Address: <https://sit.instructure.com/courses/35559>

Prerequisite(s): Students must have some programming experience. It is also highly recommended for the students to have taken Multivariate Data Analytics (BIA 652), Data & Knowledge Management (MIS 630), and Knowledge Discovery in Databases (MIS 637).

COURSE DESCRIPTION

In this course, students will learn through hands-on experience how to extract data from the web and analyze web-scale data using distributed computing. Students will learn different analysis methods that are widely used across the range of internet companies, from startups to online giants like Amazon and Google. At the end of the course, students will apply these methods to answer a real scientific question or to create a useful web application.

STUDENT LEARNING OUTCOMES

Data collection and preprocessing skills: students will learn how to identify and profile candidate sources of valuable data, as well as how to automatically collect and manage the information they need for their analytics tasks.

Diverse Analytic Skills: students will be exposed to a wide range of both quantitative and qualitative analytics techniques with applications across multiple business domains.

Team Skills: the students will be organized in teams and collaborate on projects for the duration of the course.

FORMAT AND STRUCTURE

Each class will be a mixture of lecture, discussion, and hands-on applications.

TENTATIVE COURSE SCHEDULE

Class date	Topic(s)	Readings	Assignment
January 23	An introduction to the web	See Canvas.	Assignment 1
January 30	On scraping the web		Assignment 2
February 6	On the structure of the web		Assignment 3
February 13	On crawling the web		Assignment 4
February 20	On natural language processing		Assignment 5
February 27	On topic modeling		Assignment 6
March 5	On sentiments and opinions		Assignment 7
March 12	A midterm exam		
March 26	Project discussions		
April 2	On indexing and searching the web		Assignment 8
April 9	On making recommendations		Assignment 9
April 16	On analyzing web usage I		Assignment 10
April 23	On analyzing web usage II		
April 30	Project presentations		

COURSE MATERIALS

Textbook(s): None.

Other Readings: Readings provided on Canvas.

Materials: Access to a computer with Python 3 and internet access.

COURSE REQUIREMENTS

Attendance — Attendance is mandatory and is worth 1 point per class; it's all-or-none for showing up to that class prepared, unless excused. If you don't show up to class without having been excused, you won't receive any credit for attendance. Of course, you are excused if there is an emergency (medical, family, or other), and you will receive credit, just let me know.

Please note that there are two classes where an unexcused absence will be considered a flagrant absence, much worse than a usual unexcused absence and therefore worth -5 points: the day of

the midterm exam and the group project presentations. The reason why these are considered flagrant absences are that they negatively affect other people in the class.

Assignments — There are 10 assignments, each worth 3 points for submitting it on time.

Project — There is one group project.

Exams — There is one midterm exam.

GRADING PROCEDURES

Grades will be based on:

Class Attendance	15 points (1 point each, plus 1 for good luck)
Assignments	30 points (3 points each)
Midterm	20 points
Team Project	35 points
<i>Total</i>	<i>100 points</i>

GRADE SCALE

(Minimum points needed.)

A	94
A-	90
B+	87
B	84
B-	80
C+	77
C	74
C-	70
D+	67
D	64
F	0

ACADEMIC INTEGRITY

Undergraduate Honor System

Enrollment into the undergraduate class of Stevens Institute of Technology signifies a student's commitment to the Honor System. Accordingly, the provisions of the Stevens Honor System apply to all undergraduate students in coursework and Honor Board proceedings. It is the responsibility of each student to become acquainted with and to uphold the ideals set forth in the [Honor System Constitution](#). More information about the Honor System including the constitution, bylaws, investigative procedures, and the penalty matrix can be found online at <http://web.stevens.edu/honor/>

The following pledge shall be written in full and signed by every student on all submitted work (including, but not limited to, homework, projects, lab reports, code, quizzes and exams) that is assigned by the course instructor. No work shall be graded unless the pledge is written in full and signed.

"I pledge my honor that I have abided by the Stevens Honor System."

Reporting Honor System Violations

Students who believe a violation of the Honor System has been committed should report it within ten business days of the suspected violation. Students have the option to remain anonymous and can report violations online at www.stevens.edu/honor.

Graduate Student Code of Academic Integrity

All Stevens graduate students promise to be fully truthful and avoid dishonesty, fraud, misrepresentation, and deceit of any type in relation to their academic work. A student's submission of work for academic credit indicates that the work is the student's own. All outside assistance must be acknowledged. Any student who violates this code or who knowingly assists another student in violating this code shall be subject to discipline.

All graduate students are bound by the Graduate Student Code of Academic Integrity by enrollment in graduate coursework at Stevens. It is the responsibility of each graduate student to understand and adhere to the Graduate Student Code of Academic Integrity. More information including types of violations, the process for handling perceived violations, and types of sanctions can be found at www.stevens.edu/provost/graduate-academics.

Special Provisions for Undergraduate Students in 500-level Courses

The general provisions of the Stevens Honor System do not apply fully to graduate courses, 500 level or otherwise. Any student who wishes to report an undergraduate for a violation in a 500-level course shall submit the report to the Honor Board following the protocol for undergraduate courses, and an investigation will be conducted following the same process for an appeal on false accusation described in Section 8.04 of the Bylaws of the Honor System. Any student who wishes to report a graduate student may submit the report to the Dean of Graduate Academics or to the Honor Board, who will refer the report to the Dean. The Honor Board Chairman will give the Dean of Graduate Academics weekly updates on the progress of any casework relating to 500-level courses. For more information about the scope, penalties, and procedures pertaining to undergraduate students in 500-level courses, see Section 9 of the [Bylaws of the Honor System](#) document, located on the Honor Board website.

EXAM ROOM CONDITIONS

The following procedures apply to quizzes and exams for this course. As the instructor, I reserve the right to modify any conditions set forth below by printing revised Exam Room Conditions on the quiz or exam.

1. Students may use the following devices during quizzes and/or exams. Any electronic devices that are not mentioned in the list below are not permitted.

Device	Permitted?	
	Yes	No
Laptops		X
Cell Phones		X
Tablets		X
Smart Watches		X
Google Glass		X
Other		X

2. Students may use the following materials during quizzes and/or exams. Any materials that are not mentioned in the list below are not permitted.

Material	Permitted?	
	Yes	No
Handwritten Notes <i>Conditions:</i>		X

Typed Notes <i>Conditions:</i>		X
Textbooks <i>Conditions:</i>		X
Readings <i>Conditions:</i>		X
Other (specify)		X

3. Students are not allowed to work with or talk to other students during quizzes and/or exams.
Specific Parameters: None.

LEARNING ACCOMODATIONS

Stevens Institute of Technology is dedicated to providing appropriate accommodations to students with documented disabilities. The Office of Disability Services (ODS) works with undergraduate and graduate students with learning disabilities, attention deficit-hyperactivity disorders, physical disabilities, sensory impairments, psychiatric disorders, and other such disabilities in order to help students achieve their academic and personal potential. They facilitate equal access to the educational programs and opportunities offered at Stevens and coordinate reasonable accommodations for eligible students. These services are designed to encourage independence and self-advocacy with support from the ODS staff. The ODS staff will facilitate the provision of accommodations on a case-by-case basis.

Disability Services Confidentiality Policy

Student Disability Files are kept separate from academic files and are stored in a secure location within the Office of Disability Services. The Family Educational Rights Privacy Act (FERPA, 20 U.S.C. 1232g; 34CFR, Part 99) regulates disclosure of disability documentation and records maintained by Stevens Disability Services. According to this act, prior written consent by the student is required before our Disability Services office may release disability documentation or records to anyone. An exception is made in unusual circumstances, such as the case of health and safety emergencies.

For more information about Disability Services and the process to receive accommodations, visit <https://www.stevens.edu/office-disability-services>. If you have any questions please contact: Phillip Gehman, the Director of Disability Services Coordinator at Stevens Institute of Technology at pgehman@stevens.edu or by phone (201) 216-3748.

INCLUSIVITY

Name and Pronoun Usage

As this course includes group work and in-class discussion, it is vitally important for us to create an educational environment of inclusion and mutual respect. This includes the ability for all students to have their chosen gender pronoun(s) and chosen name affirmed. If the class roster does not align with your name and/or pronouns, please inform the instructor of the necessary changes.

Inclusion Statement

Stevens Institute of Technology believes that diversity and inclusiveness are essential to excellence in academic discourse and innovation. In this class, the perspective of people of all races, ethnicities, gender expressions and gender identities, religions, sexual orientations, disabilities, socioeconomic backgrounds, and nationalities will be respected and viewed as a resource and benefit throughout the semester. Suggestions to further diversify class materials and assignments are encouraged. If any course meetings conflict with your religious events, please do not hesitate to reach out to your instructor to make alternative arrangements.

You are expected to treat your instructor and all other participants in the course with courtesy and respect. Disrespectful conduct and harassing statements will not be tolerated and may result in disciplinary actions.

MENTAL HEALTH RESOURCES

Part of being successful in the classroom involves a focus on your whole self, including your mental health. While you are at Stevens, there are many resources to promote and support mental health. The Office of Counseling and Psychological Services (CAPS) offers free and confidential services to all enrolled students who are struggling to cope with personal issues (e.g., difficulty adjusting to college or trouble managing stress) or psychological difficulties (e.g., anxiety and depression). Appointments are strongly encouraged and can be made by phone (201-216-5177) or in-person (on the 7th floor of the Howe Center). CAPS is open from 9:00 am – 5:00 pm Mondays, Wednesdays, Thursdays and Fridays and from 9:00 am – 7:00 pm on Tuesdays during the Fall and Spring semesters.

EMERGENCY INFORMATION

In the event of an urgent or emergent concern about the safety of yourself or someone else in the Stevens community, please immediately call the Stevens Campus Police at 201-216-5105 or on their emergency line at 201-216-3911. These phone lines are staffed 24/7, year round. Other 24/7 resources for students dealing with mental health crises include the National Suicide Prevention Lifeline (1-800-273-8255) and the Crisis Text Line (text “Home” to 741-741). If you are concerned about the wellbeing of another Stevens student, and the matter is *not* urgent or time sensitive, please email the CARE Team at care@stevens.edu. A member of the CARE Team will respond to your concern as soon as possible.