

SE 611: Software Analytics

Course Syllabus

1. General Information

Instructor: Dr. Preetha Chatterjee

- Email: preetha.chatterjee@drexel.edu

- Website: <https://preethac.github.io/>

Instructor Office Hours: TBD

Mailbox: Dept. of CS Mailbox

Term: Spring 2026

Credits: 3-hour lecture [3 credits]

2. Student Learning Information

2.1 Course Description:

Software repositories archive valuable software engineering data, such as source code, execution traces, historical code changes, mailing lists, bug reports, and chats. This data contains a wealth of information about a project's status and history. By doing data science on software repositories, researchers can gain an understanding of software development practices, and practitioners can better manage, maintain, and evolve complex software projects.

In recent years, the advances in Machine Learning (ML) and Natural Language Processing (NLP) technologies did not go unnoticed in the field of Software Engineering (SE). Researchers have applied software analytics techniques to various tasks such as code summarization, code comment generation, question-answer extraction, sentiment analysis, etc.

SE 611 aims to give students a deep understanding of and a hands-on approach to how ML and NLP techniques (including large language models) are used to represent knowledge and solve existing SE problems in novel ways.

2.2 Prerequisites:

- Students should be able to code in Python.
- Students should be familiar with using [GitHub](#).
- Familiarity with basic machine learning, and natural language processing techniques is preferred, but not required.

2.3. Statement of Expected Learning

Course Objective: The overall goal of this course is to allow you to participate collaboratively in a research project that is scoped to make significant progress through the research process within the allotted time of this term. The goal is to explore and problem-solve creatively with one or more software engineering-related data sets that would be made available to you.

Learning Outcomes: This course will enable students to

- Analyze related work in the area of Software Analytics.
- Apply techniques from machine learning, and natural language processing on software engineering-related datasets.
- Analyze data from software repositories and extract new insights (i.e., mining software repositories).
- Evaluate the applicability of results in the software analytics literature on practical problems.

2.4. Course Purpose

This course counts towards the depth requirements of a CS Ph.D. degree and will count towards CS, AI, and SE electives. Similarly, this course counts towards the requirements of an M.S. degree and will count as an elective for the MS-CS, MS-SE, or MS-AIML degrees. This course will be particularly useful for students with research interests in software engineering, machine learning, and natural language processing.

2.5 Tools/Programming Language

This course uses Python3. You are allowed to use any IDE of your choice.

3. Course Materials

Required Readings

- No textbooks
- Research papers as assigned during the weekly assignments and related to the course project.

References (for programming and data science) This is optional

- Python Data Science Handbook (PDSH). ISBN: 978-1491912058, O'Reilly Media, 2016.
- Data Science from Scratch (DSFS): First Principles with Python. ISBN: 978-1492041139, O'Reilly Media, 2019
- The Data Science Handbook (TDSH). ISBN:9781119092940, John Wiley Sons, 2017

Python Data Science Handbook (PDSH) is available free of charge on Github:

- <https://jakevdp.github.io/PythonDataScienceHandbook/>

The Data Science Handbook (TDSH) and Data Science from Scratch (DSFS) are available to Drexel students through the University Libraries.

- https://drexel.primo.exlibrisgroup.com/permalink/01DRXU_INST/1pvv3q/alma991013112932104721
- https://drexel.primo.exlibrisgroup.com/permalink/01DRXU_INST/1pvv3q/alma991008049759704721

4. Assignments and Final Project

The course grading is focused on assignments (e.g., responses to readings) and a final project.

Assignment submission requirements:

Each week (with a few exceptions) you will need to respond to one or more research papers, no later than 11.59 PM the day before class. The idea is to read the paper, write a summary (more instructions on this later), and **most importantly come up with 2-3 interesting questions to be discussed in the class.**

- All submissions must be made through Blackboard Learn. You are allowed multiple resubmissions. However, we will grade the latest submission you sent before the due date. In other words, if you make a **resubmission** after the due date, that submission will not be graded.

Project submission requirements:

You will have to work on a research problem, which would typically consist of investigating one or more software engineering-related data sets provided to you.

- This course requires Python3. Submissions using other versions of Python will not be graded.
- The final project should be uploaded on **a private** GitHub repository under the organization [Software Analytics \(github.com\)](https://github.com/Software-Analytics). You will be invited to this GitHub organization by the instructor. You will be required to submit the link to your private repository on Blackboard.
- The course project is an **individual project**, this is NOT a group project.

Late policy:

- Assignments submitted up to 1 week late will receive a 20% penalty.
- Assignments submitted more than 1 week late will not be accepted.
- Late submissions for the final project **are not allowed**.

5. Grading

Grading Matrix:

- Assignments: 50%
- Final Project: 50%

There will be no exams in this course.

The instructor reserves the right to make modest adjustments (5% or 10% for a category) in the weighting.

The following scale will be used to convert points to letter grades:

Points	Grade
97-100	A+
92-96.99	A
90-91.99	A-
87-89.99	B+
82-86.99	B
80-81.99	B-
77-79.99	C+
72-76.99	C
70-71.99	C-
67-69.99	D+
60-66.99	D
0-59.99	F

Note that the instructor may revise this conversion if/when necessary.

Attendance

In-person section: Drexel's stated policy is that course attendance is mandatory for students in the in-person section. I will not take attendance explicitly in every class, because it's tedious and takes away time from actual material. But I do expect you to come, and if you are absent on a regular basis it will negatively affect your term grade, beyond the grading percentages above. That said, I understand things happen (you might get sick, your car might break down). If you're in the in-person section and will miss class, please let me know.

Online section: You are welcome to email me with questions related to the materials. I hope to interact with each one of you via Zoom, so please feel free to reach out for fixing appointments for office hours.

6. Course Schedule

[This schedule is tentative and may change during the course.]

Week by week:

1. Introductions, Syllabus, Overview of Software Analytics, How to read a research paper

2. Paper discussion and in-class activities

Topic: Understanding Datasets in Software Engineering

(projects could be based on you selecting to work on one or more of these datasets)

- a. Paper title and link: TBD

3. Initial project description and research plan (Student presentation)

In-person section: Students need to present in class

Online section: Students need to record their presentations and upload them on BBLearn.

4. Paper discussion and in-class activities
Topic: Qualitative and Empirical Studies
 - a. Paper title and link: TBD
5. Paper discussion and in-class activities
Topic: Large Language Models for Software Maintenance
 - a. Paper title and link: TBD
6. Project Update and Q&A (Student presentation)
In-person section: Students need to present in class
Online section: Students need to record their presentations and upload them on BBLearn.
7. Paper discussion and in-class activities
Topic: Mining real-world GitHub software repositories
 - a. Paper title and link: TBD
8. Paper discussion and in-class activities
Topic: Agentic AI workflows in industry
 - a. Paper title and link: TBD
9. Paper discussion and in-class activities
Topic: Software Developer Productivity and Tools
 - a. Paper title and link: TBD
10. Final Project Presentation (Student presentation)
In-person section: Students need to present in class
Online section: Students need to record their presentations and upload them on BBLearn.
11. Final Project Report Submission

7. Academic Policies

Be careful about using public code available on the internet or ChatGPT. If you are keen on using some resources that you think would be useful, you need to get the instructor's permission prior to using them. You would also need to cite the resource you have used in your project.

This course follows university, college, and department policies, including but not limited to

- Academic Honesty: http://www.drexel.edu/provost/policies/academic_dishonesty.asp
- Student Life Honesty Policy from Judicial Affairs: https://drexel.edu/studentlife/community_standards/overview/
- Students with Disability Statement: <https://drexel.edu/oed/disabilityResources/students/>

- Course Drop Policy: http://www.drexel.edu/provost/policies/course_drop.asp
- Course Withdrawal Policy: <http://drexel.edu/provost/policies/course-withdrawal>
- Department Academic Integrity Policy: <http://drexel.edu/cci/resources/current-students/undergraduate/policies/cs-academic-integrity/>
- Drexel Student Learning Priorities: <http://drexel.edu/provost/assessment/outcomes/dslp/>
- Office of Disability Resources: http://www.drexel.edu/ods/student_reg.html

The instructor(s) may, at his/her/their discretion, change any part of the course before or during the term, including assignments, grade breakdowns, due dates, and schedule. Such changes will be communicated to students via the course website. This website should be checked regularly and frequently for such changes and announcements.

Students [requesting accommodations](#) due to a disability at Drexel University need to request a current Accommodations Verification Letter (AVL) in the [ClockWork database](#) before accommodations can be made. These requests are received by Disability Resources (DR), who then issues the AVL to the appropriate contacts. For additional information, visit the DR website at drexel.edu/oed/disabilityResources/overview/, or contact DR for more information by phone at 215.895.1401, or by email at disability@drexel.edu.