

# CSE6250 Syllabus (CSE/OAN)

## Big Data Healthcare

### Instructor Information

Instructor  
Jimeng Sun

Email  
jsun@cc.gatech.edu

Office Hours & Location  
Schedule by Email

Teaching Assistant(s)  
Yuzheng Liu

Email  
liuyz@gatech.edu

Office Hours & Location  
TBD

### General Information

#### Description

Data science plays an important role in many industries. In facing massive amount of heterogeneous data, scalable machine learning and data mining algorithms and systems become extremely important for data scientists. The growth of volume, complexity and speed in data drives the need for scalable data analytic algorithms and systems. In this course, we study such algorithms and systems in the context of healthcare applications.

In healthcare, large amounts of heterogeneous medical data have become available in various healthcare organizations (payers, providers, pharmaceuticals). This data could be an enabling resource for deriving insights for improving care delivery and reducing waste. The enormity and complexity of these data-sets present great challenges in analyses and subsequent applications to a practical clinical environment. In this course, we introduce the characteristics of medical data and associated data mining challenges on dealing with such data. We cover various algorithms and systems for big data analytics. We focus on studying those big data techniques in the context of concrete healthcare analytic applications such as *predictive modeling*, *computational phenotyping* and *patient similarity*. We also study big data analytic technology:

1. Scalable machine learning algorithms such as online learning and fast similarity search;
2. Big data analytic system such as Hadoop family (Hive, Pig, HBase), Spark and Graph DB

#### Pre- &/or Co-Requisites

1. Good machine learning and data mining concepts such as classification and clustering;
2. Proficient programming and system skills in Scala, Python and Java;
3. Proficient knowledge and experience in dealing with data and understand the ETL process (recommended skills include SQL, NoSQL such as MongoDB).
4. Minimum grade of C for MATH 3215 or MATH 3225 or ECE 3077 or ISYE 2027. Two of the following:
  - CX 4240. Introduction to Computing for Data Analysis
  - CS 4400 - Introduction to Database Systems
  - CX 4242. Data and Visual Analytics

## Course Requirements & Grading

### Grading Scheme

- **50% Homework** 4 homework 12.5% each
- **25% Project**
  - Note: Penalty will be reflected due to inactive team project participations
- **20% Final exam**
- **5% Participation**
  - Ed Discussion activities(contributions: viewed/endorsed/hearted answers and threads)

### Grading Scale

Your final grade will be assigned as a letter grade according to the following scale:

A	90-100%
B	80-89%
C	70-79%
D	60-69%
F	0-59%

Prof. Sun will determine if we can curve at the end of semester. Please see <http://registrar.gatech.edu/info/grading-system> for more information about the grading system at Georgia Tech.]

### Course Materials Other Classroom Management Tools

Course website: TBD

Ed Discussion (O01/OAN) for Q&A will be accessible after semester starts.

Gatech Canvas & Gradescope: used for Assignment/Project submission.

### Course Schedule

Check our course website schedule page: TBD

Please email Yuzheng Liu ([liyuz@gatech.edu](mailto:liyuz@gatech.edu)) if you have further question on the schedule.

### Academic Integrity

Georgia Tech aims to cultivate a community based on trust, academic integrity, and honor. Students are expected to act according to the highest ethical standards. For information on Georgia Tech's Academic Honor Code, please visit <http://www.catalog.gatech.edu/policies/honor-code/> or <http://www.catalog.gatech.edu/rules/18/>.

Any student suspected of cheating or plagiarizing on a quiz, exam, or assignment will be reported to the Office of Student Integrity, who will investigate the incident and identify the appropriate penalty for violations.

All course material can only be used by registered students and is not allowed to be shared outside class. Copyright is reserved by the instructional team.

## **Accommodations for Students with Disabilities**

If you are a student with learning needs that require special accommodation, contact the Office of Disability Services at (404)894-2563 or <http://disabilityservices.gatech.edu/>, as soon as possible, to make an appointment to discuss your special needs and to obtain an accommodations letter. Please also e-mail me as soon as possible in order to set up a time to discuss your learning needs.

## **ASSIGNMENT TURN-IN**

Assignments will be released and turned in via the **Canvas** Platform during the week in which they occur.

## **ATTENDANCE AND/OR PARTICIPATION**

Participation is important for this course and learning in general. We use posting on Ed Discussion (Ed Discussion links are listed on the home page) as the proxy to measure the participation level.

## **COLLABORATION & GROUP WORK**

Homework assignments are strictly individual efforts, while final projects must be done by groups (2 people). You can discuss high level concepts regarding to lectures or homework on the course forum but you shouldn't share your own (or others') solution and code with other students.

## **EXTENSIONS, LATE ASSIGNMENTS, & RE-SCHEDULED/MISSED EXAMS**

Each student is allowed 2 days of late submission in total to be used for HOMEWORK only. For each homework, we will count the late day independently (even 1 hour late will be counted as 1 day, and 25 hours counted as 2 days). Once you have used up your late days, late assignments will be penalized at 8 pts per day, assignments more than 3 days late will not be accepted (The deadlines are on Mondays 8:00 AM EST, and no submission will be accepted after the following Thursdays 8:00 AM EST). Late day penalty of each homework will be accumulated and evaluated at the end of the term. We don't consider ANY other homework extension excuses except healthcare emergency incident (doctor's proof is required).

Exam reschedule must be requested at least 24 hours in advance.

## **STUDENT USE OF MOBILE DEVICES IN THE CLASSROOM**

Not applicable

## **STUDENT-FACULTY EXPECTATIONS**

At Georgia Tech we believe that it is important to continually strive for an atmosphere of mutual respect, acknowledgement, and responsibility between faculty members and the student body. See <http://www.catalog.gatech.edu/rules/22/> for an articulation of some basic expectations - that you can have of me, and that I have of you. In the end, simple respect for knowledge, hard work, and cordial interactions will help build the environment we seek. Therefore, I encourage you to remain committed to the ideals of Georgia Tech, while in this class.