

# CS 7311: Data-Driven Computational Methods and Infrastructure

Fall 2019

Syllabus updated Aug 26 2019

<b>Instructor</b>	Dr. Jelena Tešić (pronounced as Yeh-LE-nah TE-shich)
<b>Contact</b>	<a href="mailto:jtesic@txstate.edu">jtesic@txstate.edu</a> (office x5-3436)
<b>Faculty Profile</b>	<a href="https://cs.txstate.edu/accounts/profiles/j_t463/">https://cs.txstate.edu/accounts/profiles/j_t463/</a>
<b>Office Hours</b>	Tue 11:00 a.m. – 1:30 p.m. or Thu 3:30 – 6 p.m. by appt
<b>Class Meets</b>	Thursday 6:30 p.m. – 9:20 p.m. @ Comal 212
<b>Open Labs</b>	DERR 231 (Linux Lab) MCS 590 (Windows Lab)
<b>TA</b>	Hanie Samimi, <a href="mailto:h_s163@txstate.edu">h_s163@txstate.edu</a>
<b>TA Office Hours</b>	Mon 2:00 p.m. – 4:00 p.m. and Fri 10:00 am – noon Comal 209

## Course Description

This course covers computational and statistical methods for using large-scale data sets ('big data') to answer scientific and business questions. It focuses on framing research questions, understanding how data can answer them, and using modern software tools such for scalable data storage, processing, and analysis.

## Course Objectives

The students will be able to:

- Formulate concrete research questions to address business or scientific objectives
- Identify or collect data to answer research questions
- Design tools to process, clean, and organize data for subsequent analysis
- Create and run data processing and analysis pipelines to compute statistical results over large-scale data sets using modern high-performance computing infrastructure
- Present results clearly using data visualizations and written prose
- Interpret analysis results and identify their implications for business concerns or scientific interest
- Determine appropriate data processing technology to support a desired analysis method

Objective of this class is for teams of two or three students to

1. Formulate the business or scientific objective, identify or collect the data to answer the objective, and research and propose data-driven choice of computational or programming methods
2. Design tools to process, clean, and organize data for subsequent analysis
  - a. Determine appropriate data processing technology to support a desired analysis method
3. Create and run data processing and analysis pipelines on modern high-performance computing infrastructure
4. Interpret analysis results and identify their implications for business concerns or scientific interest
5. Clean, package, share, and deliver the code using github.txstate.edu repository

## Course Material

There is no official textbook

Instructor will provide set of online references and tutorials for specific data science problems

## What is expected of student registered for CS 7311?

- (1) Attend instructional meetings, in person or via Zoom.
- (2) **Read announcements** from the instructor posted on TRACs course site.
- (3) Be informed and prepared for the class, complete pre-class assignments.
- (4) **Schedule presentations on time and submit assignments on time.**
- (5) Clearly communicate with the instructor regarding and issues, delays or unforeseen circumstances in timely manner. Emailing is the fastest way to reach the instructor.
- (6) Attend office hours and ask questions in timely manner.

## Assessment

### Participation 30 pt

- Pre-Class Assignments 10 pt
- Critique and presentation/summary of papers 10 pt
- **Walk-through tutorial, web series, demo w git code on**
  - data driven computation methods **OR**
  - data driven infrastructure and programming methods
  - FCFS schedule on TRACS 10 pt

### Research Project 50 pt - Final Project Report and Code due Dec 12 6:30 PM

1. Exploratory Data Analysis and Implications 10 pt, part of final report due Dec 12 6:30 PM
2. Create and run data processing and analysis pipelines on modern high-performance computing infrastructure 10 pt, demo due Dec 12 6:30 PM
3. Interpret analysis results and identify their implications for business concerns or scientific interest 10 pt, due Dec 12 6:30 PM
4. Clean, package, share, and deliver the code using github.txstate.edu repository 10 pt, due Dec 12 6:30

### Progress Presentation 40 pt

- Project Proposal Presentation: research idea, related work, data source, and what tools and framework you plan to use 10 pt Sep 19 and Sep 26
- Midterm Checkpoint Presentations, Oct 17 and Oct 24 10 pt
- Midterm Report Draft (on overleaf) 10 pt, pdf due Oct 17th 6:30 PM
- Final Project Presentation, Dec 5 and Dec 12

Extra Credit: 100 pt range, you can earn up to 120 pt

## Communication

Best way to contact the instructor is to send her an email. All announcements will be posted on TRACS, and all resources on git.

## Course Schedule

Following is a tentative schedule for the class. **Exact topics and dates may be updated as the course progresses.**

Class	Date	Lecture	Assignments/In Class Presentations
1	Aug 29	Introduction	Introduction and Objectives
2	Sep 5	Data Preparation	Present Research/Tool/Paper
3	Sep 12	Data Cleaning	Present Research/Tool/Paper
4	Sep 19	Data Analysis	Project Proposals
5	Sep 26	Machine Learning Overview 1	Project Proposals
6	Oct 3	Machine Learning Overview 2	Present Research/Tool/Paper
7	Oct 10	Machine Learning and Visualization	Present Research/Tool/Paper
8	<b>Oct 17</b>		<b>Project Checkpoint Presentation</b>
9	<b>Oct 24</b>		<b>Project Checkpoint Presentation</b>
10	Oct 31	Visualization 2	Analysis of proposed pipelines
11	Nov 7	Graph Analysis	Analysis of proposed pipelines
12	Nov 14	Spark and GraphX	Analysis of proposed pipelines
13	Nov 21	Bias and Fairness in Data Driven Methods	
14	<b>Dec 5</b>	<b>Demos and Final Project Presentation</b>	
15	<b>Dec 12</b>	<b>5- 7:30 PM Demo and Final Project Presentation (FINAL EXAM DATE)</b>	

## Policies

**Grade Grievance Policy:** If a student believes a mistake has been made in grading an assignment, the student has one week after an assignment is returned to resubmit an assignment for re-grading if they believe there is an error.

**Drop Policy** Students will not be automatically dropped for non-attendance: if you are planning to drop the class or withdraw from the class, follow the instructions listed on registrar's web site:

<http://www.registrar.txstate.edu/registration/drop-a-class.html>

It is your responsibility to be familiar with the University Policy on dropping classes as described in the catalog and the TXSTATE website (see), to observe relevant deadlines, and to follow proper procedures for dropping classes.

**Incomplete Policy** CS department has a strict policy regarding 'Incomplete grade'. It has to be approved by the chairman and thus an 'Incomplete grade' will only be granted under unexpected and truly severe situations, which must be supported by some official documents.

**E-mail Policy:** During the work week, instructor will respond to personal emails within 24 hours. Instructor will review communication over the weekend but will respond on Monday to most situations. If you need to reach me by email, please use the subject line: Your Name, Course Name/Number, Topic. Please allow a full 24 hours before emailing me again about the same question or issue, and on Monday for inquiries sent over the weekend.

**Plagiarism Policy:** Except where explicitly and specially allowed (such as group project), all work submitted in class is expected to be your individual work. Plagiarism will not be tolerated and if detected will result in an automatic 'F' grade. Please refer to <http://www.txstate.edu/effective/upps/upps-07-10-01.html> for Texas State's Honor Code.

**Extra Credit Policy:** See Assessment

## Accommodations for students with disability

Any student requiring special accommodations, should inform me during the first two weeks of classes. The student should also contact the office of disability services at the LBJ student center. Students who qualify for extra time for exams must take their test with ATSD and must schedule their test at the same time the test is given in class.

## Academic Honor Code and Conduct

You are expected to adhere to

- the University's Academic Honor Code <http://www.txstate.edu/honorcodecouncil/Academic-Integrity.html>
- Code of Student Conduct - <http://www.dos.txstate.edu/handbook/rules/cosc.html>
- Texas State Mission and Shared Values: <http://universityplan2023.avpie.txstate.edu/overview/Texas-State-Mission-and-Goals.html>.

## Relevant Campus Resources

**cs.txstate.edu -> News, Job Announcements, Lab and Tutoring Announcements (@txstCS)**

**Free Online Courses through LinkedIn Learning Hub:** <https://doit.txstate.edu/services/online-training.html>

**Food Insecurity: Bobcat Bounty** is the first student-run, on-campus food pantry at Texas State University. It is run by students under the direction of faculty from the Food Security Learning Community in the Nutrition & Foods Program. Our goal is to decrease food insecurity by providing healthy food to the students, faculty, and administration at Texas State University. The majority of our food comes from the Hays County Food Bank, a partnership to create a grocery store style experience for clients.

<https://bobcatbounty.txstate.edu/Eligibility.html>

**Texas State Counseling Center:** Counseling Center services are free, confidential, and provided by trained professionals to currently enrolled Texas State students while classes are in session. Counseling Center Services include brief individual, group, and couples counseling, consultation and crisis response, and workshops about coping with stress and other mental health topics.

<https://www.counseling.txstate.edu/>

**Texas State Sexual Misconduct Policy and Reporting:** If you need to make a report, please contact Ameerah McBride, Title IX Coordinator at 512.245.2539 or you can file a report for someone else, anonymously, or using a pseudonym here:

[https://cm.maxient.com/reportingform.php?TexasStateUniv&layout\\_id=10](https://cm.maxient.com/reportingform.php?TexasStateUniv&layout_id=10)

Title IX details: <https://www.txstate.edu/oei/title-IX.html>

**Discrimination Complaints:** Texas State prohibits discrimination and harassment on the basis of race, color, national origin, age, sex, religion, disability, veterans' status, sexual orientation, gender identity, or gender expression. Additionally, Texas State prohibits retaliation against a person who files a harassment or discrimination complaint, or who assists or participates in the investigation of a report.

<https://www.txstate.edu/oei/policies/complaints.html>

Download the discrimination complaint form:

<https://gato-docs.its.txstate.edu/jcr:1376ecad-b820-4bb5-bd4a-508ef63844f0/Discrimination%20Complaint%20Form.v3.fillable.pdf>

**Office of Student Diversity and Inclusion:** <https://www.sdi.txstate.edu/>

Texas State Alliance: <https://twitter.com/TxstAlliance>

<https://www.sdi.txstate.edu/Support-and-Empowerment/LGBTQIA-and-Allies.html>

**Full List of On-Campus Resources:**

<https://www.studentsuccess.txstate.edu/programs/faces/On-Campus-Resources.html>

**Search Student Organizations:**

<https://www.lbjsc.txstate.edu/soc/join/search-orgs.html>