CS7311 Fall 2020 Syllabus
updated Aug 28 2020

Meets: Thu 6:30 p.m.

Instructor: Jelena Tešić

Contact

Instructor: Jelena Tešić, jtesic@txstate.edu

Instructions Thu 6:30 p.m. – 7:50 p.m.

Discussions Thu 8:00 pm – 9:20 p.m. (can change)

- ❖ Start with Zoom, move to Teams -> see calendar
- ❖ Videos available on MediaFlo – link in Canvas**, Modules Tab**

Q&A Assignment presentations and Discussion: after 8pm break or other times

- ❖ Team chat to all your availability Mon – Fri

Contact Best way to reach me is to send me a message in Teams

Office hours (subject to change, check canvas):

- ❖ Tue 8:30 – 10 a.m. Thu 3 – 5 p.m. on Microsoft Teams
- ❖ By apt – I expect at least 3 appt with team this semester
  - Teams will have private project channels in class Teams

About the Course

- This is a **hybrid course**:
  - lectures will be online Thu 6:30 p.m.
  - All lectures will be recorded and posted on Canvas via MediaFlo link.
- Students will have the opportunity to attend the F2F class **Oct 1 Oct 22 and Dec 3.**
  - Class meets in Comal 212 conference room on those days.

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.

The objective is for students to 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 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

Assessment

Participation 5 pt

- Attending and participating when other teams are presenting
- Participating in Discussions (Student Hours)
- Surveys, quizzes

Homework 40 pt

- 8 HW, 5 pt each
- Submitted to git.txstate.edu/CS7311/NetID.git

Class Project 70 pt

❖**Total: 115 pt**

Max 100 pt

Homework

- HW 1 - review of the tutorial
- HW 2 - critical assessment of data used in COVID 19 articles
- HW 3 - tool tutorial and code presentation
- HW 4 - paper presentation
- HW 5 - *ipynb submission*
- HW 6 - Amazon Dataset EDA and visualization, ipynb submission and presentation
- HW 7 - data driven architecture paper presentation
- HW 8 - *ipynb submission*

Class Project

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

Class Project

Project teams should form by Sep 7 (Monday).

- Project proposals: teams must contact instructor w ideas before Sep 10th class

Benchmark Challenge provides:

- A clear problem statement
- An accessible dataset
- In this semester project, you will develop:
- Modeling plan
- Research and choose appropriate tools
- Research and commit to a framework
- Develop, experiment
- Visualize (EDA and Results)

Class Project

- Report: Midterm, and Final report
  - Exploratory Data Analysis and Implications
  - Template on overleaf and/or git.txstate.edu/CS7311/2020Fall
    - Submit latex or markdown – report will be a documentation for the package
  - Interpret analysis results and identify their implications for business concerns or scientific interest
- Code: deliver the code using **git.txstate.edu** class repo
  - Create and run data processing and analysis pipelines on modern high-performance computing infrastructure, and demo
  - Clean, package, share
- Present: Exploratory Data Analysis, Data Science Pipeline Proposal, Visualization and justification of results, and demo of the code

Class Project Assessment

- **10 pt Delivery:** git.txstate.edu/CS7311/ID.git package
  - 5 pt Includes appropriate documentation in markdown format
    - Introductory README.md is a tutorial for novice users to use your code and what is input what is output example
    - Separate document from report, shorter
  - 5 pt Clean, package, structure
- **20 pt** Reports: Proposal (5pt), Midterm (5pt), and Final report (10pt)
  - Exploratory Data Analysis and Implications
  - Template on overleaf and/or git.txstate.edu/CS7311/2020Fall
    - Submit latex or markdown – report will be a documentation for the package
  - Interpret analysis results and identify their implications for business concerns or scientific interest
- **20 pt** Code:
  - 10 data processing and analysis pipelines with all elements
  - 5 Code demo (ipynb)
  - 5 Quality, extension, focus, innovation
- **20 pt** Present: Project Proposal (5pt), Midterm (5pt), Final (10 pt)
  - Exploratory Data Analysis, Data Science Pipeline Proposal, Visualization and justification of results, and demo of the code

Remote Learning Setup

Download the phone apps

- Mobile app Canvas student
- Microsoft teams (free for TXST students)

Sign-in with TXST credentials

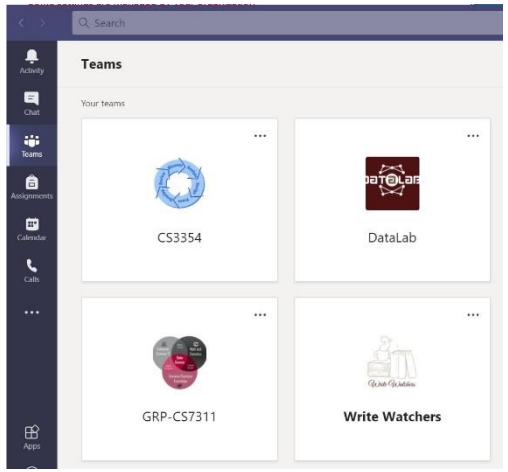
- NetID and pass (do not use email alias)
- Enable notifications for announcements (canvas) and chat (MS Teams)

Off-campus

- Follow the software instructions: https://itac.txstate.edu/remote/students.html

Tools for the Class

- Office hours on MS Teams



- All announcements posted on Canvas

  canvas.txstate.edu

  - **Make sure to enable announcements**

https://itac.txstate.edu/remote/students.html

Tools for the Class

- Classes on Teams
  - Synchronous Thu 6:30 p.m.
  - Recordings available EOD
  - Possibility of moving to Teams if Zoom issues persist
- **Video Links in Canvas**
  - LinkedIn Learning Collection
  - Lecture Asynchronous recordings on MediaFlo
- All recordings (lectures and tutorials) will be available on

https://mediaflo.txstate.edu/Playlist/CS7311 pass: 2453436

https://itac.txstate.edu/remote/students.html

Tools for the Class

git.txstate.edu

- Login with NetID and password to activate your account
- If activated, you are added to: https://git.txstate.edu/CS7311 org for the semester
  - https://git.txstate.edu/CS7311/NetID - class central
  - https://git.txstate.edu/NetID - your txstate github
- git.txstate.edu is an enterprise version of GitHub

LinkedIn learning:

- https://doit.txstate.edu/services/online-training
- git tutorial videos
- **GitHub tutorial:** https://guides.github.com/activities/hello-world/

Course Material and Resources

Learning Path

- Become a Data Scientist https://www.linkedin.com/learning/paths/become-a-data-scientist?u=51086649

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

- https://canvas.txstate.edu/courses/1434270
  - -> Modules tab
- https://git.txstate.edu/orgs/CS7311/teams/2020fall
  - -> material folder

Resources

❖ ssh NetID@zeus.cs.txstate.edu (or use bitvise)

- LEAP: ssh to leap.txstate.edu – contact instructor for the account
- ssh to GPU server – contact instructor for the account

❖ **Google Colab (with your gmail account)**

Communication

- Best way to contact the instructor is
  - Microsoft Teams
  - Email
- Communicate all issues and hurdles early
- All announcements, resources, and updated will be posted on Canvas.
- When in doubt:
  - Step 1: check Syllabus
  - Step 2: check canvas.txstate.edu
  - Step 3: check git.txstate.edu/CS7311/2020Fall repo
  - Step 4: post message on canvas forum
  - If Step 1-4 do not work, contact instructor

Schedule (subject to change)

	Date	Lecture	Discussion (Time TBD)
1	27-Aug-20	Introduction, Syllabus	Q&A
2	3-Sep-20	Data Preparation	HW #1 Presentation
3	10-Sep-20	Data Cleaning	HW #2 Presentation
4	17-Sep-20	Project Proposals and Discussion	Project Proposals and Discussion
5	24-Sep-20	Data Analysis	HW #3 Presentation
6	**1-Oct-20**	**Machine Learning Overview 1**	**HW #4 due, Project Discussion**
7	8-Oct-20	Machine Learning Overview 2	HW #5 Presentation
8	15-Oct-20	Machine Learning and Visualization	HW #6 Presentation
9	***22-Oct-20***	***Midterm Project Presentation***	***Midterm Project Presentation***
10	29-Oct-20	Visualization 2	HW #7 due, Discussion
11	5-Nov-20	Graph Analysis	HW #8 Presentation
12	12-Nov-20	Spark and GraphX	HW #9 Presentation
13	19-Nov-20	Bias and Fairness in Data Driven Methods	HW #10 due, Discussion
14	***3-Dec-20***	***Demos and Final Project Presentation***	***Demos and Final Project Presentation***
15	10-Dec-20	Demo and Final Project Presentation	We will start early not 8 pm

Students are expected to

PARTICIPATION

- **Attend** instructional meetings or watch videos
  - Ask questions on Microsoft Teams CS7311 General Channel synchronously
  - Live discussion only after instructor announces she has stopped recording
- Weekly check the updates posted on
  - https://canvas.txstate.edu/courses/1434270 – lecture slides, deadlines and general class info
  - https://git.txstate.edu/CS7311/2020Fall - new information, material and code
- **Read announcements** on Canvas course site
- Check out material from
  - on weekly basis
- Be informed and prepared for the class, complete participation and homework assignments.
- Show up for presentation (in person or online) and discussion sessions with camera ON
  - To participate in the discussion
- Participate in Synchronous conversation is on Microsoft Teams CS7311 General channel
- **Schedule project meetings and presentations on time.**
  - Clearly communicate with the instructor regarding and issues, delays or unforeseen circumstances in timely manner.
  - Teams chat is fastest way to reach the instructor.
- Student Hours are for students
  - Time to ask questions, get feedback on the project, guideline, and mentoring

Students are expected to

HOMEWORK

- Submit homework assignments on time
  - Each homework assignment is ~5% of the final grade point
  - Each HW is due date Thursday 6:30 p.m. no extension
- Submit homework assignments on TXST git class repo
  - Students will submit programming assignments through https://git.txstate.edu. Your login is your NetID.
  - If you have questions about using git, please contact cs_helpdesk@txstate.edu.
  - Students repository for the class is https://git.txstate.edu/CS7311/NetID.git.
  - This is different than your central repository https://git.txstate.edu/NetID.git.
  - Group coding project means the entire group gets an identical grade score.
  - Each student in the group MUST check in the code to its individual to its git.txstate.edu repository to receive that grade.
- Check grading, and grade grievance policy

Students are expected to

PROJECT

- Select a project partner and propose project work
- **Initiate communication** with the instructor on stages of the project
- Put in several hours a week in the project development
- Master the use of tools needed for the class
  - source versioning, command line, Teams, Canvas, and overleaf
  - Python notebooks, ML framework for the project
  - At least one tool is new and previous unused
    - This class is for you to fail in building a data science pipeline and try again
- Actively address instructor's feedback on the project progress and approach in the subsequent presentation
- Weekly check-ins of the project progress code and report to https://git.txstate.edu/CS7311/NetID repository
  - Asynchronous comments from the instructor in Teams project channel
- Check grading, and grade grievance policy

Learning in Time of COVID 19

- Student Roadmap: https://www.txstate.edu/coronavirus/road-map/student-roadmap.html
- Attendance policy under the current guidelines (subject to change)
  - Students are expected to attend at least 50% of the classes when scheduled
  - Everyone is present during team midterm and final presentation (Participation)
- Reporting tool for notification of positive COVID-19 tests/illnesses to the university: https://webapps.sa.txstate.edu/bobcattrace/
- In-person meets (F2F) under current guidelines (subject to change):
  - F2F meets will be scheduled during regular class time in Derrick 0235, see schedule (F2F)
  - We will adhere to Texas State social distancing measures as listed: https://www.txstate.edu/coronavirus/road-map/social-distancing-measures.html
  - We all fit in Comal 212 and follow 50% occupancy rule. Plan to bring your laptops.
  - Rest of the class can attend the meetings via Teams.

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.

Free Resources for Students

- cs.txstate.edu -> News, Job Announcements, Lab and Tutoring Announcements (@txstCS)
- LinkedIn Learning: https://doit.txstate.edu/services/online-training.html
- Food Insecurity: Bobcat Bounty https://bobcatbounty.txstate.edu
- Texas State Counseling Center offers free, and confidential services to currently enrolled Texas State students while classes are in session. https://www.counseling.txstate.edu/
- Texas State Sextual Misconduct Policy and Reporting:
  - https://cm.maxient.com/reportingform.php?TexasStateUniv&layout_id=10
- Discrimination Complaints: Texas State prohibits discrimination and harassment based on race, color, national origin, age, sex, religion, disability, veterans' status, sexual orientation, gender identity, or gender expression:
  - https://www.txstate.edu/oei/policies/complaints.html
- 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

Policies

- Civility Policy: Civility in the classroom is very important for the educational process and it is everyone's responsibility.
  - If you have questions about appropriate behavior in a class, please address them with your instructor first.
- Disciplinary procedures may be implemented for refusing to follow an instructor's directive, refusing to leave the classroom, not following the university's requirement to wear a cloth face covering, not complying with physical distancing or sneeze and cough etiquette, and refusing to implement other health and safety measures as required by the university.
- Additionally, the instructor, in consultation with the department chair/school director, may refer the student to the Dean of Students Office for further disciplinary review.
- Such reviews may result in consequences ranging from warnings to sanctions from the university. For more information regarding conduct in the classroom, please review the following policies at
  - https://policies.txstate.edu/division-policies/academic-affairs/02-03-02.html, Section 03: Courteous and Civil Learning Environment, and
  - https://studenthandbook.txstate.edu/rules-and-policies/code-of-student-conduct.html, number II, Responsibilities of Students, Section 02.02: Conduct Prohibited.

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 student's 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.

Policies

- Incomplete Policy Computer Science department has a strict policy regarding 'Incomplete grade’.
  - It must 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.
- Makeup Policy
  - No make up. 20% extra credit is built in the grade to account for strenuous circumstances.
- 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.
  - Do not copy “as is” code from the open source in your coding project: use of open source is encouraged, but personal stamp on comments classes and flow needs to be visible – penalty is 0 points for coding project.

Policies

- **Best way to contact the instructor is via Microsoft Teams chat**
- **E-mail Policy**:
  - If you need to reach me by email, please use the subject line: Your Name, Course Name/Number, Topic.
  - If the information can be found on canvas (Sylalbus, Modules, Announcement) or on git.txstate.edu/CS7311/2020Fall, instructor will not reply
  - During the work week, instructor will respond to personal emails during working hours, within 24 hour window.
    - 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.

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.