CS 5683: Big Data Analytics
Fall 2023

Course Information:

Instructor: Arunkumar Bagavathi
Email: abagava@okstate.edu
Office: MSCS 215
Fall 2023 Office hours: Tuesdays 10:00am-1:00pm

Lectures: Monday, Wednesday 1:30pm – 2:45pm
Classroom: Nancy Randolph Davis 4A
Credits: 3

Teaching Assistant: TBA
TA email: TBA
TA Office: TBA
TA Office hours: TBA

Prerequisites:
- *CS 5513 Numerical Computation (with 'C' or better grade) OR permission of instructor* (mandatory)
- Intermediate skills with algorithms and data structures
- Fundamentals of Python programming, and Jupyter notebook
- Familiarity with basic probability theory and linear algebra
- Familiarity with rigorous algorithm analysis and writing proofs

Other information: The class will follow Canvas. All course announcements, assignments, instructions, grades, and course materials will be posted in Canvas

Course Description: This course will discuss several data mining and machine learning algorithms for big data analytics: clustering, recommendation systems, big graph mining, data stream analysis, large-scale machine learning, natural language processing, and data visualizations. The course includes hands-on experience for big data frameworks and algorithms with tutorials and assignments. All programming assignments will be based on Python programming. *Prior experience in machine learning or any big data tools is not required.*

May not be used for degree credit with **MSIS 5683**.

Topics

Following is the tentative syllabus for the course. It is intended to change based on availability of time.

1. Introduction to Big Data – Challenges & Opportunities
2. Dimensionality Reduction
   a. Eigen values and eigenvectors
   b. Singular Value Decomposition
   c. CUR decomposition
   d. Non-Negative Matrix Factorization
   e. t-SNE
3. Recommender systems
   a. Content-based model
   b. Collaborative filtering
   c. Bias and Fairness
4. Machine Learning for Big (text) Data
   a. Clustering
   b. Unsupervised Neural Networks
      i. Auto-encoders
      ii. Attention (Transformers)
   c. Document similarity and retrieval
   d. Other NLP problems like sentiment analysis
5. Intro. to Data Visualization with Tableau
   a. Creating dashboards
   b. Telling a story with big data and visualization
6. Network Mining
   a. Introduction to (social) networks
   b. Link analysis
      i. PageRank
      ii. Personalized PageRank
   c. Machine learning for graphs
      i. Skip-gram models
      ii. Graph neural networks
      iii. Graph attention networks
   d. Community detection (Tentative)
      i. Motif-based clustering
      ii. Louvain algorithm
7. Data Stream Mining (Tentative)
   a. Sampling and Querying
   b. Filtering data streams
   c. Counting elements in data streams
   d. Online learning to integrate ML with streams
8. Introduction to Distributed Computing
   a. Apache Spark + MLlib + GraphX
   b. Distributed deep learning (Tentative)

Grading:

There are only 3 activities in the course throughout the semester: Quizzes, Assignments, and Final project. Quizzes will include programming, multiple-choice, and theory questions. All assignments will be based on Python programming along with problem solving tasks. All submissions will be on Canvas. The final exam will be on the Finals week of the semester. Quizzes and the final exam will be on Canvas. Attendance is not mandatory, but class participation (both in-class and Canvas discussion boards) will be utilized for the Class participation grade. We are planning a total of 5 assignments and 4 quizzes.

Grading split:

    Quizzes: 20%
    Assignments: 50%
    Class participation: 10%
    Final project: 20%

Assignments Policy:

Assignments: The assignments are programming based and are much broader activities which may takes substantial amount of time to complete. We will give two to three weeks of time to complete and turn-in your submissions on Canvas. Assignments can be both independent and team based.

Code of conduct: We follow OSU students code of conduct *extremely seriously*. The standard penalty for students who jeopardize the code of conduct is *suspension from the course or even from the university*. We recommend students to work as team and discuss about assignments and projects. However, each student should write their answers independently for assignment/project submission. Find more details at: More details at: *https://academicaffairs.okstate.edu/site-files/documents/ai-handbook-faculty-personnel.pdf*

Questions: Post your questions on discussion board or email instructor/TA. We will try hard to reply as soon as possible. You can expect our reply anytime within 12 hours of your question. *Please do not post questions in last few hours of the deadline.*

Late submissions: Students have a total of **2 grace periods** for both assignments and projects. Each grace period will extend your deadline by one week.

Regrading: We try our best to make fair and consistent grading for all activities. So, the grades you receive are very unlikely to change significantly. However, if you want to regrade your work, email us your requests within a week after receiving your grades. We do not consider requests coming one week after the grades are posted.

Books:

Books are not mandatory for this course. Following are good books to read for this course:

1. "Mining Massive Datasets", Second Edition by Jure Leskovec, Anand Rajaraman, and Jeffrey David Ullman
2. "Spark: The Definitive Guide", 2018 Edition by Bill Chambers and Matei Zaharia
3. "Introduction to Data Mining", Second Edition by Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, and Vipin Kumar

ALERTS AND RESCHEDULING
If the OSU campus officially closes due to inclement weather or other emergencies, alerts are provided to local news media and posted on the OSU website. Missed exams, classes, or assignments may be rescheduled at times outside the normal meeting schedule. If valid, documented circumstances prohibit students from attending the rescheduled classes, instructors should provide reasonable alternative means for makeup work.

ACADEMIC INTEGRITY
101 Whitehurst/405-744-5627/http://academicintegrity.okstate.edu
OSU is committed to maintaining the highest standards of integrity and ethical conduct. This level of ethical behavior and integrity will be maintained in all OSU courses. Participating in behavior that violates academic integrity (e.g., unauthorized collaboration, plagiarism, multiple submissions, cheating on examinations, fabricating information, helping another person cheat, unauthorized advance access to examinations, altering or destroying the work of others, and altering academic records) will result in an official academic sanction. Violations may subject you to disciplinary action including the following: receiving a failing grade on an assignment, examination, or course, receiving a notation of a violation of academic integrity on your transcript and being suspended from the University. Students have the right to appeal the charge.

COPYRIGHT & FAIR USE POLICY OF COURSE MATERIALS
Course materials may not be published, leased, sold to others, or used for any purpose other than appropriate OSU-related individual or group study without the written permission of the faculty member in charge of the course and other copyright holders. This paragraph grants you a limited license giving you access to materials for this course, including PowerPoint slides, audio/video recordings, written, or other materials, for appropriate OSU-related educational use only. Lectures should not be recorded without permission from the faculty member and must not be further disseminated or shared. Assignments, quizzes, and exams (individual questions or in their entirety) should not be uploaded to websites offering note-sharing, tutoring, or other academic help (free or by paid subscription).

CLASS ATTENDANCE AND PARTICIPATION
Class attendance is a critical component of learning; therefore, you are expected to attend and participate fully in all scheduled class meetings. Many instructors consider attendance so essential your grade may be affected by your absence, and some departments and professors have mandatory attendance policies. If no written attendance policy is provided before the last day to add a class

without instructor permission, no penalty may be assessed for class absences although you may not be permitted to make up certain in-class activities.

If you are ill, you should stay home. Faculty are expected to treat a positive COVID test result as appropriate documentation for an excused absence and are encouraged to remain flexible for students who need to quarantine for other COVID-related issues (e.g., member of household instructed to quarantine). Although instructors may be willing to livestream in-person courses through an online platform, livestreaming is not required or expected. Each instructor will determine how to best assist students with excused absences (including absences due to quarantine requirements) for in-person classes and should include details in their course syllabus.

If you are required to participate in official university-sponsored activities or military training, you should receive an excused absence as discussed in OSU's [Attendance Policy for Students](). If you will be absent from class for sponsored activities, you must provide prior notification of the planned absence to the instructor. You may be required to submit assignments or take examinations before the planned absence.

PRE-FINALS WEEK POLICY
Final examinations are scheduled at the end of each semester and are preceded by pre-finals week, which begins seven days prior to the first day of finals. During pre-finals week, all normal class activities will continue; however, no assignment, test, or examination accounting for more than 5% of the course grade may be given; and no activity or field trip may be scheduled that conflicts with another class. This excludes makeup and laboratory examinations, out-of-class assignments (or projects) made prior to pre-finals week and independent study courses. No student or campus organization may hold meetings, banquets, receptions, or may sponsor or participate in any activity, program, or related function that requires student participation. For additional information, contact the Office of Academic Affairs, 405-744-5627, 101 Whitehurst.

FINAL EXAM OVERLOAD POLICY
In the event you have three or more final exams scheduled for a single day, you are entitled to arrange with the instructor of the highest numbered course (4 digit course number) or two highest, if you have four finals on one day, to re-schedule that examination(s) at a time of mutual convenience during final exam week. If the final exam overload includes a common final exam, the common final exam is excluded from rescheduling unless multiple common exams are scheduled at the same time. You should submit this request in writing, with a copy of your class schedule, at least two weeks prior to the beginning of final exam week. The instructor has one week prior to the beginning of final exam week to arrange a mutually convenient time for administration of the final exam. After one week, if an agreement cannot be reached, take the request to the department head.

EQUAL OPPORTUNITY
409 General Academic Building/405-744-7607
https://1is2many.okstate.edu/
OSU is committed to maintaining a learning environment that is free from discriminatory conduct based on race, color, religion, sex, sexual orientation, gender identity, pregnancy, status as a parent,

national origin, disability (physical or mental), age, family medical history or genetic information, political affiliation, military service, protected veteran status, or other non-merit based factors. OSU does not discriminate on the basis of sex in its educational programs and activities. Examples of sexual misconduct and/or sex discrimination include sexual violence, sexual harassment, sexual assault, domestic and intimate partner violence, stalking, or gender-based discrimination. OSU encourages any student who thinks that they may have been a victim of sexual misconduct or sexual discrimination to immediately report the incident to the Title IX Coordinator (405-744-9153) or Deputy Title IX Coordinator (405-744-5470). If a reporting student would like to keep the details confidential, the student may speak with staff in the Student Counseling Center (405-744-5472) or one of the University's Sexual Assault Victim Advocates (Mon-Fri 8 AM-5 PM, 405-564-2129 or 24 Hour Help Line 405-624-3020).

STUDENT ACCESSIBILITY SERVICES

1202 W. Farm Rd #155/405-744-7116/ https://diversity.okstate.edu/

According to the Americans with Disabilities Act, each student with a disability is responsible for notifying the University of the disability and requesting accommodations. If you think you have a qualifying disability and need accommodations, contact the Office of Student Accessibility Services to start the registration process and to ensure timely implementation of appropriate accommodations. To receive services, you must submit appropriate documentation and complete an intake process to verify the existence of a qualified disability and identify reasonable accommodations. Faculty have an obligation to respond when they receive official notice of accommodations but are under no obligation to provide retroactive accommodations.

STUDENT SUPPORT AND CONDUCT

328 Student Union/405-744-5470/ https://ssc.okstate.edu/

By enrolling at Oklahoma State University, you accept responsibility for complying with all University policies and contracts, and for local, state and federal laws on- or off-campus that relate to the University's mission. **The Student Code of Conduct** educates students about their civic and social responsibilities including policies and procedures involving student misconduct.

In general, the University expects students to aspire to follow and promote the Cowboy Community Standards or integrity, community, social justice, respect, and responsibility.

This course strictly follows Oklahoma State University's academic policies. Get the current set of policies at: https://academicaffairs.okstate.edu/site-files/documents/2022-fall-syllabus-attachment-8-1-22.pdf

USEFUL OSU LINKS:

- CAS Career Services Team:
  https://cas.okstate.edu/department_of_computer_science/students/undergraduate/career_services_syllabus_insert_email.pdf
- CAS Career Services Events (Fall 2022):
  https://cas.okstate.edu/department_of_computer_science/students/undergraduate/cas_fall_2022_events_email.pdf