DS -644 Introduction to Big Data

Syllabus - Fall 2025| Section – 001
Class Time – T,F 4:00 PM – 5:20 PM in KUPF 107
Instructor – Ravneet Kaur rk956@njit.edu or ravneet.kaur@njit.edu
Office Hours – Tuesday/Friday (in-person) at 2.30 PM to 3:10 PM (in GITC 2119)
Teaching Assistant - TBD
TA's office hours – TBD

Required Background:

Programming Skills

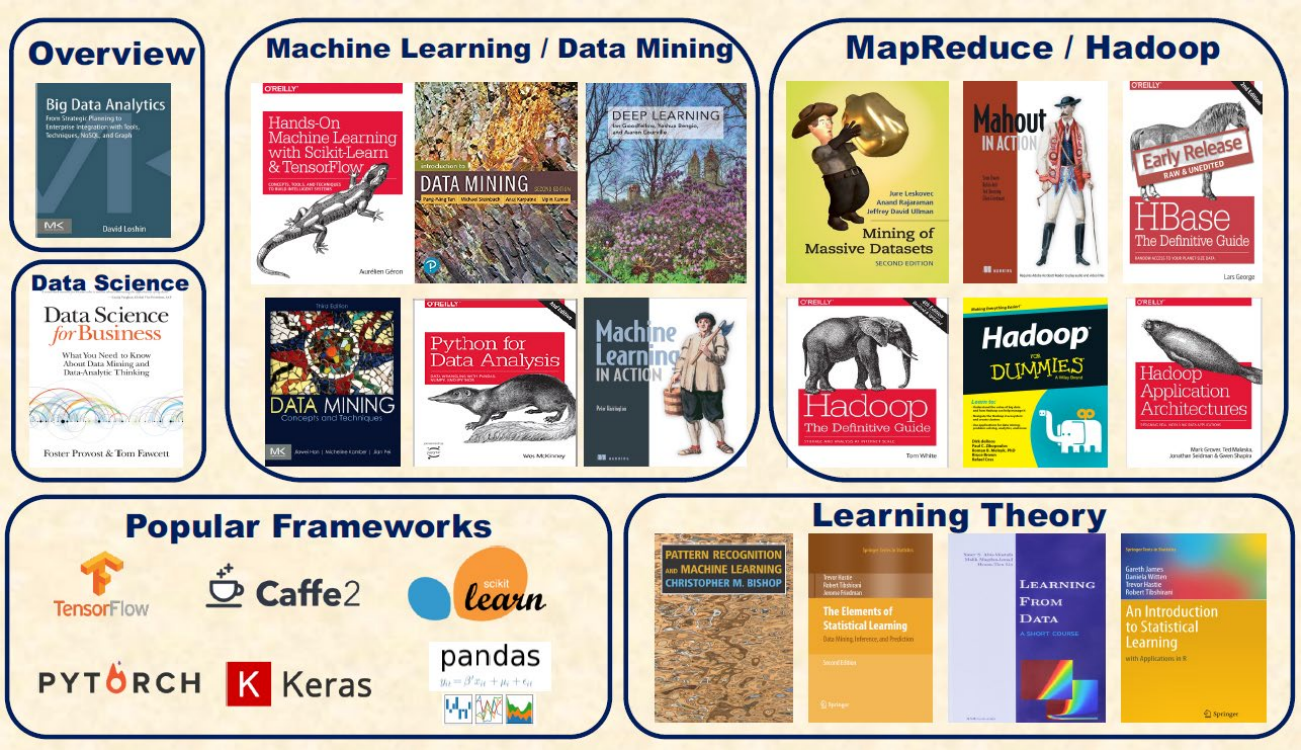
- Java, Python, or C/C++ in Linux

Prerequisite Courses

- CS 610: Data Structures and Algorithms
- Or permission of instructor

Textbooks: Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools, Techniques, NoSQL, and Graph. By David Loshin, Elsevier.

Textbooks and reference books –



Attendance: You are expected to attend all the classes. Participation is highly encouraged to make the class more interactive. In general, students who attend class regularly perform much better than those who come only occasionally. If you miss one class, be sure to consult one of

your classmates about the content of the lecture and use canvas to get notes, exercises, assignments, deadlines, and announcements.

Students who have special/emergency reasons to reschedule exams must apply and get approval of testing accommodation from the Office of Accessibility Resources and Services before the exams. https://www.njit.edu/accessibility/requesting-testing-accommodations

If students have health problems, mental problems, or family problems that will affect their performance negatively, they have to inform the Office of Accessibility Resources and Services (OARS) https://www.njit.edu/accessibility/ as soon as they become aware of the problems.

Course Communication
Canvas (canvas.njit.eduLinks to an external site.) will be the platform for posting lecture notes, submitting assignments and engaging in course discussions. For any questions and additional support, you are welcome to reach out to instructors and classroom assistants via email.

Online Discussions
You are expected to actively engage in weekly discussion forums on Canvas. These forums are designed to foster a collaborative learning environment where all students contribute to discussions. Participation will not only help deepen your understanding of the course material but also enhance your overall success in the class. Your contributions should be thoughtful, demonstrate critical thinking, and reflect on the content covered in lectures and readings.

Policy on AI Usage
Student use of artificial intelligence (AI) is permitted in this course for certain assignments and activities. It is not permitted to be used in the discussion of writing. Additionally, when students use AI in this course, the AI must be cited as is shown within the NJIT Library AI citation page for AI. If you have any questions or concerns about AI technology use in this class, please reach out to the instructor prior to submitting any assignments.

Grade Appeals
If you believe that you deserve more credit than you have been awarded on a particular exam problem, you may request, **at the time the exam is returned**, that it be re-graded. Your entire exam will be regraded, which may result in points being added or subtracted.

If you believe that you deserve more credit than you have been awarded on a particular assignment problem, you may request, **within 48 hours of the grade being posted**, that it be re-graded. Your entire homework will be regraded, which may result in points being added or subtracted.

Course Overview and Learning Outcomes:
This course provides in-depth coverage of various topics in big data from data generation, storage, management, transfer, to analytics, with focus on state-of-the-art technologies, tools,

architectures, and systems that constitute big-data computing solutions in high-performance networks. Real-life big-data applications and workflows in various domains (particularly in the sciences) are introduced as use cases to illustrate the development, deployment, and execution of a wide spectrum of emerging big-data solutions. When you have completed this course, you should be familiar with big data tools, techniques, and systems, and be able to analyze/solve big data problems.

Course Topics *(Subject to changes according to class progress)*

Topics
Introduction to Big Data Analytics
Hadoop Motivation, Ecosystem Mode and AWS Academy Lab instructions
HDFS
Map Reduce and Yarn, Yarn Schedulers, Discussion
HBase
Hive and Oozie
Spark and Pig
When to use Machine Learning, Mahout, Recommendations
Clustering, K-Means, Machine Learning Types
Classification, Decision Tree, Regression
Big Data Visualization

Note – The grades will not be curved unless necessary

Grade	**Marks**
A	100-92
B+	91-84
B	83-74
C+	73-62
C	61-45, anything below 45 is F

3 Homework's	15%
Mid Term Exam 1,2	30% -*(15% each)*
Project	15%
Final Exam	20%
Quiz	10%
Discussions	10%

Extra Credit

There will be **NO** way to get extra credit. Please don't ask.

The department has voted on a stringent curve for all courses. Only the top 25% of the class should expect an A. Furthermore, up to 20% of the class may face a C+, C, or F.

Also note that most students typically get all the points on the Homework assignments. Thus, your position in the curve and your class grade depends almost entirely on the exams.

On the other hand, missing a single homework is highly likely to lower your grade at least by one letter grade.