



DS-203

Lecture Schedule : Slot 5, Mon-Wed 9:30 am to 11:30 am.
Venue : LA001
Instructor : [Dr. Vinay Kulkarni](#)

Syllabus and week-wise calendar: [Click here](#)

Course Description

- Programming Basics (Python programming, R, Data Structures), Visualization/Plotting, Data Science Libraries (Pandas, PyPlot, matplotlib) Databases, GPUs/CUDA programming, Parallel/distributed computing for data science (Map/Reduce, Spark/Hadoop), working on the cloud (Amazon Web services, Google Cloud Platform, Azure, etc.).
- The course will be programming heavy, with in-class and take-home programming exercises.
- A project can be optionally included.

Topics Covered

- Introduction to basic probability
- Introduction to basic statistics
- Basic data understanding
- Performance of Python programs
- Exploratory data analysis and data visualization
- Linear and logistic regression
- Supervised machine learning as a Blackbox
- Deep Learning
- Introduction to software engineering
- Graphical User Interface (GUI) programming
- Introduction to databases
- Parallel query processing
- Cloud services

Eligibility

- The course is open to all BTech students.



Prerequisites

- None.

Credit/Audit Requirements

- **Approximate** credit structure
- Performance on assignments will form the majority of the course evaluation plan

Text References

- [Principles and Techniques of Data Science](#), by Sam Lau, Joey Gonzalez, and Deb Nolan, 2019.
- Online tutorials on Python and R
- Learning Python, Mark Lutz, OReilly, 2005
- Python for data analysis, WesMckinney, O Reilly, 2013
- CUDA by Example: An Introduction to General-Purpose GPU Programming, Jason Sanders, Nvidia, 2010
- NORMAN MATLOFF. Parallel Computing for Data Science: With Examples in R, C++, and CUDA. Boca Raton: CRC Press.
- [Neural Networks and Deep Learning](#) by Michael Nilson