Name of the Programme: M.Sc. in Artificial Intelligence
Course code: CSI-507
Title of course: Data Science Fundamentals Lab
Number of credits: 2(0L+0T+2P)
Effective from AY: 2023-24

Prerequisites for the course	Basic programming skills, Statistics	
Objectives	To introduce Basic process of data science, Python and Jupyter notebooks. To understanding how to manipulate and analyze uncurated datasets To learn basic statistical analysis and machine learning methods and effectively visualize results	
Content	Jupyter and Numpy: Jupyter notebooks are one of the most commonly used tools in data science as they allow you to combine your research notes with the code for the analysis. After getting started in Jupyter, we'll learn how to use numpy for data analysis. numpy offers many useful functions for processing data as well as data structures which are time and space efficient.	10 hours
	Pandas: Pandas, built on top of numpy, adds data frames which offer critical data analysis functionality and features.	10 hours
	Visualization: When working with large data sets you often need to visualize your data to gain a better understanding of it. Also, when you reach conclusions about the data, you'll often wish to use visualizations to present your results.	10 hours
	Mini Project: With the tools of Jupyter notebooks, numpy, pandas, and Visualization, you're ready to do sophisticated analysis on your own. You'll pick a dataset we've worked with already and perform an analysis for this first project.	10 hours
	Machine Learning: To take your data analysis skills one step further, we'll introduce you to the basics of machine learning and how to use sci-kit learn - a powerful library for machine learning.	10 hours
	Working with Text and Databases: You'll find yourself often working with text data or data from databases. This week will give you the skills to access that data. For text data, we'll also give you a preview of how to analyze text data using ideas from the field of Natural Language Processing and how to apply those ideas using the Natural Language Processing Toolkit (NLTK) library.	5 hours
	Mini-Project	5 hours
Pedagogy	Tutorials/ Lab assignments/ Project work	
References/ Readings	1. Practical statistics for data science by Peter Bruce and Andrew Bruce 2. Naked statistics by Charles Wheelon 3. Business data science by Matt Taddy 4. Elements of statistical learning by Trevor Hastie, Robert and jerome 5. Python for data analysis 6. Data science and big data analytics -EMC2	
Course Outcomes	1. Application of data science techniques to real-world problems. 2. Proficiency in data acquisition and preprocessing. 3. Ability to perform exploratory data analysis. 4. Building and evaluating predictive models.	

(Back to top)