

Name of the Programme: M.Sc. in Artificial Intelligence

Course Code: CSI-506

Title of the Course: Data Science Fundamentals

Number of Credits: 2(2L+ 0T+ 0P)

Effective from AY: 2023-24

<u>Prerequisites for the course</u>	Statistics and Probability theory and Python Programming	
<u>Objectives</u>	To get started with basics of Data Science and learn all aspects of Data Science in its entirety	
<u>Content</u>	Introduction: Typology of problems - Data science in a big data world: Benefits and uses of data science and big data-Facets of data-The data science process-The big data ecosystem and data science-The data science process: Overview of the data science process- Defining research goals and creating a project charter-Retrieving data-Cleansing, integrating, and transforming data-Exploratory data analysis-Build the models- Presenting findings and building applications on top of them.	4 hours
	Mathematics for Data science - Importance of linear algebra, statistics and optimization from a data science perspective; Structured thinking for solving data science problems. Linear Algebra: Matrices and their properties (determinants, traces, rank, nullity, etc.); Eigenvalues and eigenvectors; Matrix factorizations; Inner products; Distance measures; Projections; Notion of hyperplanes; half-planes.	8 hours
	Probability, Statistics and Random Processes: Probability theory and axioms; Random variables; Probability distributions and density functions (univariate and multivariate); Expectations and moments; Covariance and correlation; Statistics and sampling distributions; Hypothesis testing of means, proportions, variances and correlations; Confidence (statistical) intervals; Correlation functions; White-noise process. Data clearing (EDA)	3 hours
	Introduction to Data Science Methods: Linear regression as an exemplar function approximation problem; Linear classification problems-PCA	4 hours
	Handling large data on a single computer The problems you face when handling large data-General techniques for handling large volumes of data-General programming tips for dealing with large data sets - Case study 1: Predicting malicious URLs - First steps in big data-Distributing data storage and processing with frameworks	4 hours
	Introduction to NoSQL The rise of graph databases Introducing connected data and graph databases Introducing Neo4j: a graph database Data visualization to the end user Data visualization options Crossfilter, the JavaScript MapReduce library Creating an interactive dashboard with dc.js Dashboard development tools	3 hours 4 hours
<u>Pedagogy</u>	Lectures/ Tutorials/Hands-on assignments/Self-study	
<u>References / Readings</u>	1. Practical statistics for data science by peter bruce and andrew bruce 2. Naked statistics by charles wheelon	

	<ul style="list-style-type: none"> 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
<u>Course Outcomes</u>	<ul style="list-style-type: none"> 1. Understanding of data science principles. 2. Proficiency in data manipulation and preprocessing. 3. Ability to visualize and communicate data insights. 4. Knowledge of statistical analysis and predictive modeling techniques.