

DS-301 Introduction to Data Science

Credit Hours: 2-0

Prerequisite: None

Course Objectives: At the conclusion of the course, students should learn the skill sets required to be a data scientist. Basic statistical concepts such as probability distributions, statistical inference etc. will be covered during the course. Python language will be utilized to carry out basic statistical modeling and analysis. Significance of exploratory data analysis (EDA) in data science will be explored together with basic tools (plots, graphs, summary statistics).

Course Contents: Introduction: What is data science? Big data and data science hype - and getting past the hype, skill sets needed, statistical inference, populations and samples, statistical modelling, probability distributions, fitting a model, exploratory data Analysis and the data science process, basic tools of EDA and introductory concepts involved in machine learning.

Course Outcomes: Upon completion of this course, the student should be able to:

- Describe data science
- Explain statistical inference in basic terms
- Explain the significance of exploratory data analysis (EDA) in data science

Text Book: Cathy O'Neil and Rachel Schutt. Doing Data Science, Straight Talk From The Frontline. O'Reilly, 2014.

Reference Books:

1. Van Der Aalst, Wil. Process mining: data science in action. Vol. 2. Heidelberg: Springer, 2016.
2. De Brouwer, Philippe JS. The Big R-Book: From Data Science to Learning Machines and Big Data. John Wiley & Sons, 2020.

Weekly Breakdown		
Week	Section	Topics
1	Chap 1	Introduction: What is Data Science? Big Data and Data Science hype - and getting past the hype Skill sets needed

2	2.1	Statistical Inference Populations and samples Statistical modelling, probability distributions, fitting a model
3	2.2	Exploratory Data Analysis and the Data Science Process Basic tools (plots, graphs and summary statistics) of EDA
4	2.3	Philosophy of EDA, The Data Science Process
5	2.4	SQL and Enterprise Data Management
6	3.1	Three Basic Machine Learning Algorithms, Linear Regression
7	3.2-3.3	k-Nearest Neighbours (k-NN) and k-means
8	4.1,4.6-4.7	Data Wrangling: APIs and other tools for scrapping the Web
9	Mid Semester Exam	
10	7.4	Feature Generation and Feature Selection (Extracting Meaning From Data)
11	7.5-7.8	Mining Social-Network Graphs
12	7.9-7.10	Social networks as graph
13	9.1-9.3	Data Visualization, Basic principles, ideas and tools for data visualization
14	9.4-9.6	Data Science and Ethical Issues,
15	16.1-16.3	Next-generation data scientists
16	16.4-16.5	Next-generation data scientists
17		Review
18	End Semester Exam	