

ROWAN UNIVERSITY
Department of Mathematics

Master Syllabus

DS 01101: In-Depth Introduction to Data Science

CATALOG DESCRIPTION:

This course is a more in-depth version of DS 01.100 Introduction to Data Science. This course provides an introduction to the foundational topics in data science including data manipulation, data analysis with statistics, machine learning, data communication and information visualization. It is intended to introduce the tools and techniques necessary to solve problems involving data. As an introductory course, it will utilize and explain programming and statistical data analyses without requiring prerequisite coursework in these areas. It is designed for the student interested in exploring the data science field, or as a first course in a degree program in data analytics or data science. The course will include additional modules such as ethics of data acquisition, data analysis, and storage mechanisms, and advanced programming functions in command-based languages such as Python and R, as well as more advanced statistical analysis tools.

OBJECTIVES:

By the end of this course,

- Students shall understand the background of data science:
 - Describe the role of a data scientist, including tasks and responsibilities.
 - Identify various types of data.
 - Explain the relationship between correlation and causation.
 - Distinguish between data science, artificial intelligence (AI), and machine learning, and explain their interrelationships.
- Students shall understand the idea of exploratory data analysis:
 - Use mathematical techniques to produce basic summary statistics, such as mean, median, standard deviation, and quartiles.
 - Create and interpret various data visualizations techniques, including scatter plots, histograms, bar plots, and box plots.
- Students shall understand basic programming logic and the application of R programming:
 - Understand basic programming logic.
 - Learn the basics of the R programming language and R Studio interface.
 - Apply R to manage and process data.
 - Apply R to perform data visualization.
 - Use R to perform regression analysis.
 - Use R to perform classification analysis.
 - Use R to perform clustering analysis.
- Students shall understand the machine learning process:
 - Understand and apply the different steps in the machine learning process.
 - Differentiate between types of machine learning, including supervised learning and unsupervised learning.
- Students shall understand the idea and applications of linear regression:
 - Perform linear regression.

- Interpret and apply linear regression models.
- Evaluate the performance of linear regression models.
- Students shall understand the idea and applications of classification:
 - Perform classification.
 - Interpret and apply classification models.
 - Evaluate the performance of classification models.
- Students shall understand the idea and applications of clustering:
 - Perform clustering.
 - Interpret and apply clustering algorithms.
- Students shall understand the idea and applications of advanced modeling techniques:
 - Understand and perform data cleaning.
 - Understand and perform data wrangling.
 - Understand the concepts of feature engineering, feature encoding, and feature selection.
 - Understand the concept of regularization.
 - Understand the concepts of training, validation, and testing.
 - Understand the concept of cross-validation.
- Students shall understand the idea of business analytics and Excel tools:
 - Explain the significance of business analytics to business operations.
 - Explain how businesses use descriptive, predictive, and prescriptive analytics to solve problems.
 - Use Microsoft Excel to manage data, perform basic functions, and visualize data.
- Students shall understand the concept of database queries:
 - Understand the purpose and use of database queries.
 - Learn how to interpret entity-relationship (ER) diagrams in database design.
- Students shall learn the ethical concerns related to AI systems:
 - Learn the moral guidelines that shape how AI should behave and interact in society.
 - Learn the principles of ethical AI.

PREREQUISITES:

None

CONTENT:

- Introduction to data science, AI, and machine learning
- Exploratory data analysis
- R programming and practice
- Data visualization
- Machine learning process
- Linear regression
- Classification
- Clustering
- Advanced modeling topics
- Business analytics and Excel tools
- Database

- Ethical AI

SUGGESTED TEXTS:

- Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. (2013). *An introduction to statistical learning: with applications in R*. New York: Springer