

King Fahd University of Petroleum and Minerals
Department of Mathematics & Statistics
MATH 506 Syllabus, Term 231

Code: MATH 506

Title: Fundamentals of Data Science

Credit Hours: 3-0-3

Prerequisite: Graduate Standing

Instructor: Dr. Jamal Al-Smail

E-mail: jamalhas@kfupm.edu.sa (use your KFUPM e-mail for communications)

Office Hours: Sundays & Tuesdays, 3:00 pm – 4:50 pm
Building 5-407

Objective: The main objective of the course is to

- Introduce the mechanism of the learning process,
- Implement solutions using data scientific software, toolboxes, and libraries.

Description: All aspects of the data science pipeline using the software, toolboxes, and libraries like NumPy, SciPy, Pandas, SymPy, Matplotlib, and Seaborn: Data acquisition, cleaning, handling missing data, EDA, visualization, feature engineering, modeling, model evaluation, bias-variance tradeoff, sampling, training, testing, experimenting with a classical model.

Learning Outcomes: Upon completion of the course, students should be able to:

- Distinguish data science tasks.
- Prepare data for analysis.
- Describe the learning process.
- Build a model in a computer environment.

Textbook [TB]: Data Science using Python and R by C. Larose and D. Larose, Wiley, 2019.

Supplementary Material:

1. [R1] A Hands-On Introduction to Data Science, by Chirag Shah, Cambridge University Press, 2020.
2. [R2] Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications by Igual, Laura, Seguí, Santi, Springer, 2017.

Grading Policy:

Group Assignments (15%), Data Science Project (15%)
IBM SPSS Modeler Completion (5%) & Certification(5%)
Exam1 (15%), Exam2 (15%), Final Exam (30%)

Attendance: Attendance is a University Requirement. A DN grade will be awarded to any student accumulating 6 unexcused absences.

Academic Integrity: All KFUPM policies regarding ethics apply to this course.

Course Outline:

Weeks	Topics	Reference
1	Introduction to Data Science Data Science Methodologies and Tasks Implementation: A Comprehensive Example in Data Science	Ch 1 [TB], [R1],[R2]
2-3	Toolboxes for Data Scientists Python, IBM SPSS Statistics, Libraries Implementation: Basics on using Python and Libraries, Basics on using IBM SPSS Statistics and Libraries	Ch 2.1, 2.2 [TB] Ch 5.1-5.3 [R1] Ch 2.1-2.6 [R1]
4-5	Data Preparation Types, Sources, Formats, Pre-Processing Implementation: Data Preparation using IBM SPSS Statistics, Data Preparation using IBM SPSS Modeler	Ch 2 [R1] Ch 3 [TB]
6-7	Data Analysis Techniques Descriptive, Multivariate Analysis, Feature Engineering Implementation: Feature Engineering using IBM SPSS Statistics	Ch 4 [TB] Ch 3 [R1] Ch 3 [R2]
8	Data Visualization Implementation: Data Visualization using Python, Data Visualization using IBM SPSS Statistics	External Notes
9-10-11	Introduction to Modeling Datasets, Machine Learning, Modeling, Training-Testing-Validation, Regression, Classification Implementation: Data Science Applications using IBM SPSS Statistics, Data Science Applications using IBM SPSS Modeler	Ch 5, Ch 11[TB] Ch 8.1 – 8.3, Ch 9.4 [R1] Ch 6.1 [R2]
12-13	Evaluating Models Metrics, Cross-Validation, Hyperparameters Implementation: Model Validation using Python and IBM SPSS Statistics	Ch 7 [TB] Ch 12.4 [R1]
14	Automating Models Building Pipelines, Joining Pipelines, Saving Models Implementation: Building Pipelines using IBM SPSS Statistics	External Notes
15	Project Presentations IBM SPSS Modeler Certifications	

Important Dates:

Exam1: Week 5(Tuesday); **Exam2:** Week 10(Tuesday)

Final Exam: During final exam week. Check registrar’s website.

Data Science Project Proposal: Week 7 (Sunday)

Project Report Submission: Week 14 (Tuesday)

Project Presentations: Week 15 (Sunday and Tuesday)

IBM SPSS Modeler Completion: Week 14 (Thursday)

IBM SPSS Modeler Certification: Week 15 (Thursday)