King Fahd University of Petroleum and Minerals
Department of Mathematics & Statistics
MATH 506 Syllabus, Term 221

Code: MATH 506
Title: Fundamentals of Data Science
Credit Hours: 3-0-3
Instructor: Dr. Ali N. Duman
E-mail: aliduman@kfupm.edu.sa (Use your KFUPM e-mail to communicate)

Office Hours:
Time: Sunday and Tuesday 3:30pm – 5:00pm and 8:00pm - 8:30pm
or by appointment (contact me via MS Teams)
Venue: 5-325 or 59-Cafeteria (contact me via MS Teams to inquire my location)

Lectures:
- Section 1
  Sunday-Tuesday 5:20pm – 6:35pm
  Room: See registrar

Objective: The main objective of the course is to
- Introduce to 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:
- Describe the learning process.
- Distinguish data science tasks.
- Prepare data for analysis.
- Build a model in a computer environment.

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

Supplementary Material:
1. [S1] Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications by Igual, Laura, Seguí, Santi, Springer
2. [S2] A Hands-On Introduction to Data Science, by Chirag Shah, Cambridge University Press

3. [S3] An Introduction to Statistical Learning with Applications in R, by G. James, Springer

Grading Policy: Projects (40%), Quizzes (15%), CLW (10%), Exam (15%), Final Exam (20%)

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

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

Schedule:

Weeks	Topics	Reference
1	**Introduction to Data Science** What is Data Science? Data Science Methodology, Data Science Tasks	Ch 1 [TB], [S1],[S2]
2-3	**Tools for Data Scientists** Python, SQL, Libraries	Ch 2.1, 2.2 [TB] Ch 2.1-2.6 [S1] Ch 5.1-5.3 [S2]
4-5	**Data** Data Types, Data Sources, Data Formats, Data Pre-Processing	Ch 3 [TB] Ch 2 [S2]
6-7	**Data Analysis Techniques** Descriptive, Multivariate Analysis, Feature Engineering	Ch 4 [TB] Ch 3 [S1] Ch 3 [S2]
8	**Visualization, Reporting**	Instructor Notes
9-10-11	**Introduction to Modeling** What is a dataset?, What is learning?, What is a model?, Regression, Classification, Bias vs Variance, Training-Testing-Validation	Ch 5, Ch 11[TB] Ch 6.1 [S1] Ch 8.1 – 8.3, Ch 9.4 [S2]
12-13	**Evaluating Models** Metrics, Cross-Validation, Hyperparameters	Ch 7 [TB] Ch 12.4 [S2]
14	**Automating Models** Building Pipelines, Joining Pipelines, Saving Models	Instructor Notes
15	**Catch-Up & Project Presentations**	

Important Dates:
Exam 1: 7th week
Final Exam: During final exam week. Check registrar's website.