Code: MATH 506  
Title: Fundamentals of Data Science  
Credit Hours: 3-0-3  
Instructor: Dr. Ahmet Emin Tatar  
E-mail: atatar@kfupm.edu.sa (Use your KFUPM e-mail to communicate)

Office Hours:  
MTW 8:30pm – 9:30pm on Zoom  
Meeting Id: 95670508793

Lectures:  
- Section 1  
  MW 5:40pm – 6:55pm on Zoom  
  Meeting Id: 92342356230

- Section 2  
  MW 7:05pm – 8:20pm on Zoom  
  Meeting Id: 99766280403

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

**Grading Policy:** Homework (10%), Quizzes (15%), CLW (10%), Exam (15%), Project (15%), Project Presentation (10%), Final Exam (25%)

**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:**

<table>
<thead>
<tr>
<th>Weeks</th>
<th>Topics</th>
<th>Reference</th>
</tr>
</thead>
</table>
| 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, VM, Docker, 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: 10/22/2020  
Project Announcement By 09/23/2020 (4th week)  
Project Part 1 due by 11/04/2020 (10th week)  
Project Part 2 due by 12/09/2020 (15th week)  
Final Exam: During final exam week. Check registrar’s website.