CMPE428- Data Science

Department: Computer Engineering		
Program Name: Computer Engineering	**Program Code:** 25	
Course Number: CMPE428	**Credits:** 4 Cr	**Year/Semester:** 2020-2021 FALL

☒ Required Course ☐ Elective Course

Prerequisite(s):
MATH322 Probability and Statistical Methods

Catalogue Description:

Introduction to data science process and its lifecycle. The role of data scientist, problem definition, data preparation, model planning and building, delivery of the results. Introduction to R and Rstudio. Graphical user interfaces, data import from different sources such as csv, xls, JSON, SPSS, SAS, ARFF and online sources (URLs). Attributes and their types. Vectors, matrices, lists and classes in R. Data frames and operations on data frames. Data Exploration and wrangling using R. Cleaning data. Data Visualization using ggplot2. Supervised versus unsupervised learning from data. Clustering for unsupervised learning. Supervised learning for regression and evaluation of the models in terms of degree of fit. Logistic regression models. Classification models. Decision trees and naïve Bayes classifier. Implementation of the classifiers and their evaluation. Performance metrics. Extraction and selection of attributes. Dimensionality reduction using principal component analysis and exploratory factor analysis. Selecting most discriminative attributes using forward and backward selection methods. Visualization of high-dimensional data using principal components.

Course Web Page: https://staff.emu.edu.tr/hakanaltincay/en/teaching/cmpe428

Textbook(s):
Data Science and Big Data Analytics – D. Dietrich, B. Heller and B. Yang, WILEY, 2015

Python for Data Analysis – Wes McKinney, O'REILLY, 2013

A learning Guide To R, R. Duursma, J. Powell and G. Stone, W. Sydney Univ press

R Graphics Cookbok, W. Chang, O'Reilly, 2012

R in Action, R. I. Kabacoff, Manning press, 2011

Topics Covered and Class Schedule: (4 hours of lectures per week)

Week 1	Introduction to data science process and its lifecycle.
Week 2-3	Introduction to R and Rstudio, Python and Spider.
Week 4	Data Exploration using R and Python
Week 5	Data Visualization using ggplot2 (R) matplotlib (Python).
Week 6	Linear regression models
Week 7	Logistic regression models and performance metrics.
Week 10	Nearest Neighbour and feature line classifiers
Week 11	Decision trees
Week 12	Data clustering
Week 13	Dimensionality reduction
Week 14-15	Term project presentations/discussions

Week 16-17	Final Exams

Assignment Schedule:

The preliminary schedule is as follows:

#1– TBA
#2– TBA
#3– TBA
#4– TBA
#5– TBA

Course Learning Outcomes:

On successful completion of this course, all students are expected to be able to:

(1) Learn Programming in R and Python
(2) Analysis of data
(3) Visualization of data
(4) Develop regression models
(5) Design classifier models
(6) Performance evaluation of classifiers
(7) Implement data clustering
(8) Select/extract discriminative attributes

	Method	No	Percentage
Assessment	Midterm Exam	1	25%
	Assignments	5	25%
	Term Project	1	25%
	Final Examination	1	25%

Contribution of Course to Criterion 5

Credit Hours for:

Mathematics & Basic Science : 0

Engineering Sciences and Design : 4

General Education : 0

Relationship of Course to Program Outcomes

The course has been designed to contribute to the following program outcomes:

1. an ability to identify, formulate, and solve complex engineering problems by applying principles of engineering, science, and mathematics
2. an ability to apply engineering design to produce solutions that meet specified needs with consideration of public health, safety, and welfare, as well as global, cultural, social, environmental, and economic factors

Exams:

- You have re-sit exam chance at the end of semester if you fail. Note that; if your letter grade is "D" or above and you have no warning, you will not be able to enter re-sit exam. Yet, be aware that if you attend the re-sit exam, grade you get will be replace your final exam grade even if your grade is decreased.
- If you miss the midterm or the final exam, you must **submit a written report to the course instructor, stating your excuse within 3 days of that examination. The report will be evaluated and, if approved, you will be able to take the make-up exam**.

- If you miss both midterm and final exams and do not submit any written report, you will get an "NG" grade. In the same case, if you submit report for both missed exams and both are accepted, you will be able to enter make-up for one of them only.

Plagiarism:

- Plagiarism (which also includes any kind of cheating in exams, assignments, and lab works) is a disciplinary offence and will be dealt with accordingly. Furthermore, the penalty of plagiarism is to get grade zero for the corresponding exam, assignment, or lab work.

Important Remarks:

- You should have regular attendance to the lectures for being successful in the course. Course related materials, exercises, laboratory experiments, old exam questions and announcements will be published on the course web site and you will be responsible from all. Note that the course web site can update during the semester. Therefore, please check it regularly.

| **Prepared by** Hakan Altınçay | 12/10/2020 |