<div align="center">

Yerevan State University
Faculty of Economics and Management
1002/M91, Data Mining
Spring 2020

</div>

Number of Credits: **3**   Instructor: **Lusine Zilfimian**
Web: Github   E-mail: lusinezilfimian@yahoo.com
Office Location: **BB: 310**   Class Hours: **M, 6:20-7:40pm**

Course Description

The course surveys a range of state-of-the-art data mining techniques. The focus of the course is on developing the intuition behind the algorithms and hands-on learning with actual data sets.

Course Web Site

This course takes place on GitHub. You should create your own account (it is totally free) and private repositories for each homework in collaboration with the instructor. You will be able to see announcements, assignments and other information on the above-mentioned page.

Course Goals and Student Learning Objectives

- Be able to collect and process the data, furthermore provide appropriate insights by implementing data mining techniques and interpreting the results.

- Be able to apply predictive data mining techniques including logistic regression, classification trees, and ensemble methods.

- Be able to assess the performance of data mining models.

- Get a sense of the range of applications of these techniques.

- Gain hands-on experience through realistic problems and a course project.

Suggested Texts

1. Pang-Ning Tan, Michael Steinbach, Vipin Kumar "Introduction to Data Mining", 2005

2. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani "An Introduction to Statistical Learning with Applications in R", 2013

   We will be using the R statistical software package. Basics of R and the details of using R will be introduced to perform the specific data mining techniques we will learn. If you would like to dig deeper or gain a broader understanding of R, there are many books about it, including some nice free references. Suggested textbooks for R:

3. Yanchang Zhao "R and Data Mining: Examples and Case Studies", 2015

4. Tilman M. Davies "The book of R", 2016

Course Structure

Class Structure

Two 40-minute class meetings per week (typically, one lesson combines 2 academic hours), with homework in accordance with the schedule. All submissions must be completed before class on the day due.

Assessments and Grading Policy

The course grade will be based on the following components:

		%	Points
1.	Homework assignments	20%	4
2.	Midterm Exam	20%	4
3.	Final project	20%	4
4.	Final Exam	30%	6
5.	Quiz	10%	2

Students are entitled to appeal grades in line with the university's Grades Policies policy which is available online at http://www.ysu.am/education/Archive/Full-Time/2018/Tntes_2018.pdf

Homework Guidelines

Eight home assignments are included in the course to assess the current progress of students during the course. Homework will be due at the end of each topic when assigned and should be turned in through GitHub. Start the homework during the week if your schedule permits. You have a week and a half to fulfill your homework. Homework should be submitted as files like Markdown, HTML/PDF and/or Word format.

Final project

Students will do a group project. They need to use data mining techniques discussed in the class for the project. The topic and the dataset for the project is chosen by the group and confirmed with the instructor. Further description of the final project will be posted on GitHub.

Final project and Exams

Students will do a group project. They need to use data mining techniques discussed in the class for the project. The topic and the dataset for the project is chosen by the group and confirmed with the instructor. Further description of the final project will be posted on GitHub.

The course will include an oral exam covering all topics of the course.

University Policies

Academic integrity and Honesty

Students should know the University's Internal Disciplinary Rules that is available at http://documentation.ysu.am/wp-content/uploads/2015/03/EPH_nerqin_kargapahakan_kanonner.pdf You can find both your rights and responsibilities and the University's Academic Integrity Policy in this file. After entering the exam room during the exams or quizzes the student is forbidden to take with him any telecommunication, internet communication or other technical equipment. In case of detecting such equipment the examiner, who conduct the exam, can stop the student's exam process, remove from the examination room and the examination report can be indicated as failed. Instances of academic dishonesty will not be tolerated. Cheating on exams or plagiarism (presenting the work of another as your own, or the use of another person's ideas without giving proper credit) will result in a failing grade and sanctions by the University. For this class, all assignments are to be completed by the individual student unless otherwise specified.

Course Schedule

N	Topic	Assignment	Materials
1.	Introduction to Data Mining, Getting Started with R, R Shiny, Github		[T1Ch1], [T2Ch2.3] Additional resources
2.	Preprocessing and Exploring Data	Homework #1	[T1CH2], [T1CH3]
3.	Simple & Multiple Linear regression models	Homework #2	[T1AppD], [T2CH3]
4.	GLM: Logistic Regression		[T2CH4.3],[T2CH4.3] Additional resources
5.	GLM: Poisson Regression.	Homework #3	Additional resources
6.	Ridge, Lasso and Elastic Net	Homework #4	[T2CH6.2]
7.	Midterm Exam		
8.	Decision trees		[T1CH4], [T2CH8.1,8.3]
9.	Ensemble methods: Random Forest, Bagging, Boosting	Homework #5	[T1CH5.6], [T2CH8.2,8.3]
10.	Classification	Homework #6	[T1CH5.2, 5.3,5.5]
11.	Cluster Analysis	Homework #7	[T1CH8.2,8.3], [T2CH10.3]
12.	Association Rule Mining		[T1CH6]
13.	Dimensionality Reduction	Homework #8	[T1AppB]
14.	Final Project Presentation		
15.	Final Exam		

The schedule is subject to change with fair notice based on instructor's and students' performance