

THE UNIVERSITY OF TEXAS AT TYLER
Soules College of Business
Department of Computer Science
COSC 5371 – Data Mining
Summer 2024
Section: 2024-SUMMER7WK2-COSC-5371.460

Instructor: Nary Subramanian, Ph.D.
COB 315.11
Email (*preferred way to contact*): nsubramanian@uttyler.edu
Phone: 430-558-1330

Office Hours: By Zoom and email.

Text (Required): *Data Mining for Business Analytics – Concepts, Techniques, and Applications in Python* by Galit Shmuel, Peter Bruce, Peter Gedeck, and Nitin Patel, First Edition, 2020, Wiley Publishing, ISBN 9781119549840.

Catalog Description: Study of the concepts and techniques of data mining, or knowledge discovery in databases. The automated or convenient extraction of patterns representing knowledge implicitly stored in large databases, data warehouses, and other massive information repositories.

Course Description: It is often said that data is the new oil and just like oil, data needs to be “refined” to extract useful information. This is because the raw data that is being collected at almost all nodes in the modern communication network be they sensors, switches, routers, computers, databases, filesystems, mobile phones, and Internet of Things, needs to be mined for detecting new patterns and insights so that information that will improve competitiveness of the business may be obtained. Modern data mining techniques use computer programs written in Python and other programming languages to mine data. In this course we will learn several techniques to mine data and apply them by writing Python code for real-world situations. This course will introduce Python programming as well so that even if you are new to this language, you will be able to quickly apply Python for solving data mining problems. All course material and videos will be posted on Canvas.

Grading: Grading will be based on homework assignments. Assignments are open book. Assignment submissions should be made electronically to Canvas. Late submissions will not be graded but there is **no** penalty for early submissions. There will be one assignment each week that will be available from 12pm on Sundays and will be due by 8am on Wednesdays; only in the last week you will have a *second* assignment that will appear at 8am that Wednesday due 8am that Friday. Each assignment will be worth 20 points and there will be seven assignments in all. For computing your final letter grade, the lowest assignment grade will be dropped and the sum of the six best assignment grades (maximum possible score = 120) will be used as per the grading policy given below.

Grading Policy:

Points	Grade
≥105	A
≥90, < 105	B
≥80, < 90	C

Course Objectives:

1. Explain data mining and its importance for modern businesses.
2. Compare data mining algorithms.
3. Determine the appropriate algorithm for a given problem.
4. Develop computer programs for data mining using Python.

Make-up Policy

There will be no make-ups for missed homework assignments; missed assignments will get a grade of zero.

Tentative Schedule:

<u>Week</u>	<u>Chapter</u>	<u>Topic</u>
7/1 – 7/7	1	Introduction to Data Mining, Introduction to Python Programming
7/8 – 7/14	2,3	Overview, Data Visualization
7/15 – 7/21	4,5	Dimension Reduction, Evaluating Predictive Performance
7/22 – 7/28	6,7	Multiple Linear Regression, k-Nearest Neighbors
7/29 – 8/4	10,11	Logistic Regression, Neural Nets
8/5 – 8/11	14,15	Association Rules, Cluster Analysis
8/12 – 8/17	16	Handling Time Series

University Policies

University policies can be seen at <https://www.utt Tyler.edu/academic-affairs/files/syllabuspolicy.pdf>.