THE UNIVERSITY OF TEXAS AT TYLER
Soules College of Business
COSC 5371 – Data Mining
Fall 2023

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

Office Hours: (Physical) Tues and Thurs 8:00 am to 9:30 am; 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 data 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; however, for businesses to exploit these large stores of data, these data must 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 so 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 and there is **no** penalty for early submissions. Assignments will be available from 8am on Monday and will be due by 8am on Wednesday. Each assignment will be worth 20 points and there will be six assignments in all – the lowest assignment grade will be dropped. Sum of the five best assignment grades will be used to compute your final letter grade as per the grading policy given below.

Grading Policy:

Points	Grade
≥85	A
≥75, < 85	B
≥65, < 75	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:

Week	Chapter	Topic
8/21 – 8/27	1	Introduction to Data Mining, Introduction to Python Programming
8/28 – 9/3	2,3	Overview, Data Visualization
9/4 – 9/10	4,5	Dimension Reduction, Evaluating Predictive Performance
9/11 – 9/17	6,7	Multiple Linear Regression, k-Nearest Neighbors
9/18 – 9/24	10,11	Logistic Regression, Neural Nets
9/25 – 10/1	14,15	Association Rules, Cluster Analysis
10/2 – 10/7	16	Handling Time Series

University Policies
University policies can be seen at https://www.uttyler.edu/academic-affairs/files/syllabuspolicy.pdf.