

DS 440 Data Mining

Instructor: Prashant Shekhar, PhD

Tentative Schedule for Fall 2022

<i>Week Number: Starting Date (days)</i>	<i>Topic</i>	<i>Homework</i>	<i>Learning Outcome</i>
Unit I: Data Mining Basics			
1: 29 th Aug (M,W,F)	Course introduction Python basics Computations in python: numpy		1,2 5 5
2: 5 th Sept (W,F)	Computations in python: scipy Data visualization in python: matplotlib		5 5
3: 12 th Sept (M,W,F)	Data characteristics Data quality and preprocessing Machine Learning in python: sklearn	HW1 released	1,2 1,2,5 1,2,5
Unit II: Supervised Learning			
4: 19 th Sept (M,W,F)	Introduction to regression Linear regression Ridge regression		4,5,7 4,5,7 4,5,7
5: 26 th Sept (M,W,F)	Lasso regression Other regression models Overfitting & model selection in regression	HW1 due HW2 released	4,5,7 3 4,5,7
6: 3 rd Oct (M,W,F)	Introduction to classification Logistic regression Decision trees		4,5,7 4,5,7 4,5,7
7: 10 th Oct (M,W,F)	Random forest Classifier evaluation Overfitting and classifier model selection	Project details due HW2 due	4,5,7 6 3
8: 17 th Oct (M,W)	Ensemble methods: bagging Ensemble methods: boosting	HW3 released	3,4,7 3,4,7
9: 24 th Oct (M,W,F)	K-nearest neighbor classification Support vector machines I Support vector machines II		4,5,7 4,5,7 4,5,7
Unit III: Unsupervised Learning			
10: 31 st Oct (M,W,F)	Association analysis: apriori Clustering: K-means Cluster evaluation	HW3 due	4,5,7 4,5,7 4,5,7
11: 7 th Nov (M,W)	Anomaly detection I Anomaly detection II	HW4 released	4,5,7 4,5,7
12: 14 th Nov (M,W,F)	Test review Test Project day 1		3, 10 3, 10 6
13: 21 th Nov (M)	Project day 2		6
Thanksgiving Break			
Course Conclusion			
14: 28 th Nov (M,W,F)	Course review Project presentation I Project presentation II	HW4 due	6 6
15: 5 th Dec (M,W)	Project presentation III Project presentation IV	Project due	6 6

Learning outcome: After successful completion of this course, you will acquire knowledge to:

1. Understand the main goals and types of data mining.
2. Identify a broad variety of real-world applications of data mining.
3. Identify the strengths and limitations of popular data mining techniques.
4. Explain the mathematics concepts behind several data mining methods such as decision trees, k-nearest neighborhood, Bayesian method, support vector machine, neural network, etc.
5. Gain hands-on experience in the use of machine learning software tools in Python.
6. Gain teamwork experience to handle real-world data-mining projects and expand their expertise beyond traditional book learning exercises.
7. Demonstrate the ability to solve problems beyond the scope of textbook exercises.