

# DS 540 Data Mining

Instructor: Prashant Shekhar, PhD

Tentative Schedule for Spring 2023

<i>Week Number: Days</i>	<i>Chapter Number</i>	<i>Topic</i>	<i>Homework</i>	<i>Learning Outcome</i>
<b>Data Mining Basics</b>				
1: 12 <sup>th</sup> Jan (Th)	1	Course Introduction		1,2,13
2: 17 <sup>th</sup> Jan/ 19 <sup>th</sup> Jan (Tu,Th)	2	Types of data Data quality		1,3 1,3
3: 24 <sup>th</sup> Jan/ 26 <sup>th</sup> Jan (Tu,Th)	2	Similarity and Distance Data Preprocessing	HW 1 released	1,4 1,4
<b>Classification Basics</b>				
4: 31 <sup>st</sup> Jan/ 2 <sup>nd</sup> Feb 2 (Tu, Th)	3	Decision Trees: I Decision Trees: II		5,6 5,6
5: 7 <sup>th</sup> Feb/ 9 <sup>th</sup> Feb (Tu,Th)	4	Rule Based Naive Bayes	HW 1 due	5,6 5,6
6: 14 <sup>th</sup> Feb/ 16 <sup>th</sup> Feb (Tu,Th)	3	Classifier Evaluation Validation and Overfitting	HW 2 released	5,6,7 5,6,7
<b>Classification Algorithms</b>				
7: 21 <sup>st</sup> Feb/ 23 <sup>rd</sup> Feb (Tu,Th)	4	K-Nearest Neighbor/ Exam review Exam 1		5,6 2,13
8: 28 <sup>th</sup> Feb/ 2 <sup>nd</sup> Mar (Tu,Th)	4	Support Vector Machines: I Support Vector Machines: II		5,6 5,6
9: 7 <sup>th</sup> Mar/ 9 <sup>th</sup> Mar (Tu,Th)	4	Ensemble Methods: I Ensemble Methods: II	HW 2 due HW 3 released	5,6,8 5,6,8
<b>Spring Break</b>				
11: 21 <sup>st</sup> Mar/ 23 <sup>rd</sup> Mar (Tu,Th)	4	Imbalanced Classes: I Imbalanced Classes: II		5,6,9 5,6,9
<b>Association Analysis</b>				
12: 28 <sup>th</sup> Mar/ 30 <sup>th</sup> Mar (Tu,Th)	5	Apriori Algorithm /Exam review Exam 2	HW 3 due HW 4 released	10
<b>Clustering Analysis</b>				
13: 4 <sup>th</sup> Apr/ 6 <sup>th</sup> Apr (Tu,Th)	7	KMeans Algorithm Cluster Evaluation		11 11
<b>Anomaly Detection</b>				
14: 11 <sup>th</sup> Apr/ 13 <sup>th</sup> Apr (Tu,Th)	9	Proximity-based Clustering-based		12 12
<b>Project</b>				
15: 18 <sup>th</sup> Apr/ 20 <sup>th</sup> Apr (Tu,Th)		Course conclusion Project Presentation I	HW 4 due	12 2,13
16: 25 <sup>th</sup> Apr/ 27 <sup>th</sup> Apr (Tu,Th)		Project Presentation II Project Presentation III	Project due	2,13 2,13

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

1. Understand the basics of data mining and its relation to machine learning.
2. Use python as an efficient tool for data mining
3. Understand the types of data and evaluate its quality, distribution etc.
4. Implement foundational data preprocessing techniques for effective data mining.
5. Understand the basics of supervised learning
6. Implement and analyze prominent classification algorithms for data mining.
7. Evaluate and compare various classification algorithms
8. Combining multiple classification models to create better models.
9. Handle unbalanced classes in classification problems.
10. Understand and implement association analysis.
11. Understand and implement clustering analysis.
12. Understand and implement anomaly detection.
13. Apply the concepts learnt in class to problems of practical importance.