# MA 540 Data Mining

## Instructor: Prashant Shekhar, PhD

## Tentative Schedule for Spring 2023

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