1. **Course Number and Course Title:**

   MLR 503 – Data Mining and Knowledge Discovery

2. **Credit Hours:**

   3 – 0 – 3

3. **Prerequisites and/or Co-Requisites:**

   Prerequisites: Approval of the CSE Head of Department.
   Co-requisites: None
   Competencies: Undergraduate-level knowledge of statistics and programming

4. **Name and Contact Information of Instructor:**

   Dr. Alex Aklson
   Office: Engineering and Science Building, ESB 2172
   Email: aaklson@aus.edu
   Phone: (06) 515-4893
   Office Hours: Monday and Thursday: 11:00AM – 12:00PM
   Tuesday: 11:00AM – 12:00PM and 1:30PM – 3:30PM
   Wednesday: 11:00AM – 12:00PM and 3:30PM – 4:30PM
   By appointment

5. **Course Description (Catalog Description):**

   Presents the principles of data mining and knowledge discovery. Covers key topics including data transformation and visualization, pattern recognition, experimental validation, and model interpretation. Examines various techniques from decision trees and rule induction to probabilistic methods and association mining, lazy learning, and clustering. Studies examples of data mining applications using state-of-the-art software such as RapidMiner or Weka.

6. **Textbook and other Supplemental Material:**

   Textbook:

   - Witten I.H., Frank E., and Hall M.A., Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann, 4th ed., 2017.

   Other supplemental material:

   - Han J., Pei J., and Tong H., Data Mining: Concepts and Techniques, Morgan Kaufmann, 4th ed., 2022.
   - Alpaydin E., Introduction to Machine Learning, MIT Press, 4th ed., 2020.
   - Hastie T., Tibshirani R., and Friedman J., The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer, 2nd ed., 12th print, 2017.
   - Machine learning and data mining software (free, open source); articles and data sets.

7. **Course Learning Outcomes:**

   Upon completion of the course, students will be able to:
   1. Appraise the importance of and issues with data and pattern mining for society.
   2. Apply appropriate data mining techniques and machine learning models for various tasks.

3. Compare pattern mining algorithms such as rules, trees, probabilistic, clustering.
4. Evaluate the performance of many data mining techniques to solve real world problems.
5. Use modern software to experiment with machine learning and data mining techniques.
6. Critically review current research work in data mining and knowledge discovery.
7. Conduct independent research in the field of data mining and machine learning.

8. Teaching and Learning Methodologies:

Methods include lectures, problem and project-based learning methods (assignments, exams, survey paper, research project, presentation), and class discussions.

9. Course Topics and Schedule:

Topic/Activity	**Weeks**
Introduction: Machine Learning and Data Mining Principles – CRISP-DM	Week #1
Python Primer – Virtual Environments, Git and Github	Week #2
Exploratory Data Analysis with Python – Data Analysis and Visualization	Week #3
Logistic Regression, Support Vector Machines, Naïve Bayes	Week #4
Ensemble Methods and XGBoost	Week #5
Experimental Validation, Model Evaluation, Imbalanced Classes, Model Interpretation	Week #6
Data Transformation and Dimensionality Reduction, PCA, and t-SNE	Week #7
Anomaly Detection – Gaussian Distribution and Multivariate Gaussian Distribution	Week #8
Midterm Exam	Week #9
Gaussian Mixtures and Mixture of Experts Survey Paper	Week #10
Lazy Learning and Active Learning	Week #11
Association Mining	Week #12
Model Deployment and Monitoring	Week #13
Case Studies in Data Mining Applications	Week #14
Project Presentations	Week #15
Final Exam	Week #16

10. Schedule of Laboratory and other Non-Lecture Sessions:

The survey paper is done individually and is due in week 10. Students must conduct a review on a specific topic related to machine learning and data mining then write a short paper to summarize their findings.

The research project is conducted by teams of 2 students and is due at the end of the semester. Students are required to select one of three problems provided by the instructor. The submission will be in the form of a documented report submitted as Github repository as well as a final presentation in class.

11. Out-of-Class Assignments with Due Dates:

Assignment	Due Date (tentative)
Homework 1	Week #4
Homework 2	Week #6
Homework 3	Week #8
Homework 4	Week #12
Homework 5	Week #14
Survey Paper	Week # 11
Research Project and Presentation	Week # 15

12. Student Evaluation:

Assessment	Weight	Due Date (tentative)
Homework assignments	15%	Cf. Section 11
Survey	10%	Cf. Section 11
Research project and presentation	20%	Cf. Section 11
Midterm exam	25%	Week #9
Final Exam	30%	Week #16

13. Assessment Instruments:

Assessment	Course Learning Outcomes
Homework	O2–O5
Class presentation	O1, O6, O7
Review paper	O1, O6, O7
Research project	O2–O5, O7
Midterm exam	O3, O4
Final Exam	O1–O4

14. Contribution of Course to Program Outcomes:

MSMLR Program Outcomes	Emphasis in this course	Course Learning Outcomes
1. Perform research emphasizing creativity, independent learning, and scientific methods in the field of Machine Learning.	◑	O1–O7
2. Apply advanced mathematics, computer science knowledge, and software tools in identifying, formulating, and solving real world problems.	●	O2–O4
3. Demonstrate an in-depth understanding of modern Machine Learning approaches, algorithms, and tools.	●	O1–O7
4. Select and use techniques, skills, and modern tools necessary for research or professional practice.	◑	O2, O4, O5

5. Communicate effectively through technical presentations and reports.	○	O6, O7
6. Recognize the need for, and engage in, lifelong learning in professional areas.		
7. Attend to professional and ethical responsibilities within global and societal contexts.	○	O1, O6, O7

Emphasis: ● High; ◑ Medium; ○ Low; Blank – Nothing Specific Expected

15. Letter Grade Policy:

Total (T)	Letter Grade
90 ≤ T	A
85 ≤ T < 90	A-
80 ≤ T < 85	B+
75 ≤ T < 80	B
70 ≤ T < 75	B-
65 ≤ T < 70	C+
60 ≤ T < 65	C
T < 60	F