## The University of Georgia
### Department of Computer Science

## *Course Information Sheet*
# CSCI 4380
### Data Mining

| | |
|---|---|
| **Brief Course Description**<br>(50-words or less) | The course aims to provide students with a broad introduction to the field of Data Mining and related areas and to teach students how to apply these methods to solve problems in complex domains |
| **Extended Course Description / Comments**<br><br>Use this section to put additional information that's relevant to whom this course is targeting | The course is appropriate both for students preparing for research in Data Mining and Machine Learning, as well as Bioinformatics, Science and Engineering students who want to apply Data Mining techniques to solve problems in their fields of study. |
| **Pre-Requisites and/or Co-Requisites** | Prerequisite: CSCI 2720 (Data Structures) OR CSCI 2725 (Data Structures for Data Science) |
| **Approved Textbooks**<br>(If more than one, course text used during a semester is at the discretion of the instructor) | Author(s): Ian Witten, Eibe Frank & Mark Hall<br>Title: Data Mining: Practical Machine Learning Tools and Techniques<br>Edition: Third Edition, 2011<br>ISBN-13: 978-0123748560 |

**Specific Learning Outcomes (Performance Indicators)**

These are a (non-exhaustive) list of specific, measurable outcomes, as they relate to the course & program objectives.

These learning outcomes should avoid using ambiguous language such as "understand" or "familiar".

Performance indicators must include an action verb (indentifying the depth to which students should demonstrate performance), and the content referent that is the focus of the instruction (from ABET)

Target number 5 - 10

This course presents a survey of topics in data mining. At the end of the semester, all students will be able to do the following:

1. Analyze a real-world data set and identify appropriate data mining techniques to apply thereto.
2. Write a program or use a package to implement a data mining algorithm.
3. Conduct data mining experiments and properly report and discuss the results.
4. Effectively present a data mining article to an audience.
5. Review and critique data mining articles.

**Relationship Between Course Outcomes and Learning Outcomes**

| | | Program Outcomes | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | A | b | c | d | e | f | g | h | I | j | k |
| Learning Outcomes | 1 | • | • | | | | | | | • | | |
| | 2 | • | | • | | | | | | • | | |
| | 3 | • | | • | | | | | | • | | |
| | 4 | • | | | | | • | | | | | |
| | 5 | • | | | | | | | | • | | |

**Program Outcomes**

(These are ABET-specified and should not be changed)

a. An ability to apply knowledge of computing and mathematics appropriate to the discipline.
b. An ability to analyze a problem, and identify and define the computing requirements appropriate to its solution.
c. An ability to design, implement, and evaluate a computer-based system, process, component, or program to meet desired needs.
d. An ability to function effectively on teams to accomplish a common goal.
e. An understanding of professional, ethical, legal, security and social issues and responsibilities.
f. An ability to communicate effectively with a range of audiences.
g. An ability to analyze the local and global impact of computing on individuals, organizations, and society.
h. Recognition of the need for and an ability to engage in continuing professional development.
i. An ability to use current techniques, skills, and tools necessary for computing practice.
j. An ability to apply mathematical foundations, algorithmic principles, and computer science theory in the modeling and design of computer-based systems in a way that demonstrates comprehension of the tradeoffs involved in design choices.
k. An ability to apply design and development principles in the construction of software systems of varying complexity.

**Major Topics Covered**
(Approximate Course Hours)

3 credit hours = 37.5 contact hours
4 credit hours = 50 contact hours

Note: Exams count as a major topic covered

Introduction (5-hours)
Input: Concepts, instances, attributes (5-hours)
Output: Knowledge representation (5-hours)
Algorithms: The basic methods (7.5-hours)
Credibility: Evaluating what's been learned (5-hours)
Implementations: Real machine learning schemes (7.5-hours)
Data Transformations  (2.55-hours)
Ensemble Learning  (2.5-hours)
Paper Presentations (10-hours)

**Assessment Plan for this Course**

Each time this course is offered, the class is initially informed of the Course Outcomes listed in this document, and they are included in the syllabus. At the end of the semester, an anonymous survey is administered to the class where each student is asked to rate how well the outcome was achieved.  The choices provided use a 5-point Likert scale containing the following options: Strongly agree, Agree, Neither agree or disagree, disagree, and strongly disagree.  The results of the anonymous survey are tabulated and results returned to the instructor of the course.

The course instructor takes the results of the survey, combined with sample student responses to homework and final exam questions corresponding to course outcomes, and reports these results to the ABET committee.  If necessary, the instructor also writes a

recommendation to the ABET committee for better achieving the course outcomes the next time the course is offered.

Each course Learning Outcome, listed above, directly supports one or more of the Program Outcomes, as is listed in "Relationships between Learning Outcomes and Program Outcomes". For CSCI 4380, Program Outcomes (a), (b), (c), (f) and (i) are supported.

Dr. Khaled Rasheed