

CS 484 – Spring 2018

Data Mining – CRN: 19891 – CS 484 - 002

Prerequisites: Grade of C or better in CS 310 and STAT 344.

Instructor: Prof. Harry Wechsler wechsler@gmu.edu

TA: Priya, Mani ~ **Office Hours** ~ M & W 3 pm – 4 pm ~ ENGR 4456

Email correspondence using GMU accounts with subject CS 484-002

Course Description – Basic principles and methods for data analysis and knowledge discovery. Course emphasizes developing basic skills for modeling and prediction and performance evaluation. Topics include system design; data quality, preprocessing, and association; event classification; clustering; biometrics; business intelligence; and mining complex types of data.

Course (ABET) Outcomes:

1. The ability to apply computing principles, probability and statistics relevant to the data mining discipline to analyze data.
2. A thorough understanding of model programming with data mining tools, algorithms for estimation, prediction, and pattern discovery.
3. The ability to analyze a problem, identifying and defining the computational requirements appropriate to its solution: data collection and preparation, functional requirements, selection of models and prediction algorithms, software, and performance evaluation.
4. The ability to understand performance metrics used in the data mining field to interpret the results of applying an algorithm or model, to compare methods and to reach conclusions about data.
5. The ability to communicate effectively to an audience the steps and results followed in solving a data mining problem (through a term project).

Time, Day, and Venue: January 22, 2018 – May 16, 2018, TR – 1:30 – 2:45 pm – Art and Design Building L008.

Office Hours: T: 3:15 – 4:00 pm, R 12:15 – 1:00, or by appointment: ENGR 4448.

<https://registrar.gmu.edu/calendars/spring-2018/>

Spring Break: March 12 – March 18 (no class on March 13 and 15)

Last day of classes: May 3

<https://registrar.gmu.edu/calendars/spring-2018/final-exam/>

Final Exam: Tuesday, May 15, 1:30 – 4:15 pm

Required Textbook (including online slides): T <http://www-users.cs.umn.edu/~kumar/dmbook/index.php>

Topics covered: Data, Classification, Association Analysis, Cluster Analysis, and Anomaly Detection.

Reference Textbook (including online slides): I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques* (4th ed.), Morgan Kaufmann, 2016. <http://www.cs.waikato.ac.nz/ml/weka/book.html> and *The WEKA Workbench* [http://www.cs.waikato.ac.nz/ml/weka/Witten et al 2016 appendix.pdf](http://www.cs.waikato.ac.nz/ml/weka/Witten_et_al_2016_appendix.pdf)

CLOSED BOOK & 2 – 3 TYPED FORMULAE PAGES & SCIENTIFIC CALCULATOR EXAMINATIONS

Grading Composition (100 points)

- 3 (three) Programming Homework (Classification, Association, and Clustering) and 1 (one) Problem Solving Homework – 30 %
- Midterm – March 22 – 10 %
- TEAM TERM PROJECT PRESENTATION – 25 %, May 1 and May 3
- FINAL – May 15 – 35 %

Academic Integrity

You are expected to abide by the GMU honor code. Homework assignments and exams are individual efforts. Information on the university honor code can be found at

<https://oai.gmu.edu/mason-honor-code/>

Additional departmental CS information:

<https://cs.gmu.edu/resources/honor-code/>

Learning Disability Accommodation

If you have a documented learning disability or other condition which may affect academic performance, make sure this documentation is on file with the Office of Disability Services (ODS) and then discuss it with the professor in his office regarding accommodations.