

CS 484 – Summer 2017

Data Mining – CRN: 41928 – CS 484 - C01

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

Instructor: Prof. Harry Wechsler wechsler@gmu.edu

Email correspondence using GMU accounts with subject CS 484

Course Description – Course covers 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: Session C (5 weeks), June 26, 2017 – July 28, 2017, MWF – 1:20 – 4:20 pm – Innovation Hall 136.

Office Hours: MWF 12:00 – 1:00 pm or by appointment: ENGR 4448.

<http://registrar.gmu.edu/calendars/summer-2017/>

First day of classes: Monday, June 26

No class on Monday, July 3, (Independence Day Observance)

Last day of classes: Wednesday, July 26

<http://registrar.gmu.edu/calendars/fall-2016-semester/final-exams/>

Final Exam: Friday, July 28, 1:30 – 4:15 pm

Required Textbook (including online slides): P. N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*, Addison Wesley, 2006. <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

OPEN BOOK and NOTES EXAMINATIONS

Grading Composition (100 points)

- 3 (three) Homework (on Classification, Association, and Clustering) – 30 %
- Midterm – W, July 12 – 20 %
- TERM PROJECT – 30%, July 24 and 26
- Final – July 28 – 20 %

Honor Code

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

<http://oai.gmu.edu/the-mason-honor-code-2/>

Additional departmental CS information:

<http://cs.gmu.edu/wiki/pmwiki.php/HonorCode/CSHonorCodePolicies>

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.