Chair: Tracy Christofero | **GC#6: Course Addition**

Request for Graduate Course Addition

1. Prepare one paper copy with all signatures and supporting material and forward to the Graduate Council Chair.
2. E-mail one identical PDF copy to the Graduate Council Chair. If attachments included, please merge into a single file.
3. ***The Graduate Council cannot process this application until it has received both the PDF copy and the signed hard copy.***

College: CITE | Dept/Division: Computer Science | Alpha Designator/Number: CS 515 | (●) Graded ( ) CR/NC

Contact Person: Hyoil Han | Phone: 304-696-6204

NEW COURSE DATA:

New Course Title: Data Mining

Alpha Designator/Number: C S 5 1 5

Title Abbreviation: D a t a M i n i n g

(Limit of 25 characters and spaces)

Course Catalog Description: (Limit of 30 words)

Covers (1) the process of knowledge discovery, (2) algorithms (association rules, classification, and clustering), and (3) real-world applications. Focuses on efficient data mining algorithms and scaling up data mining methods.

Co-requisite(s): None.

First Term to be Offered: Fall 2015

Prerequisite(s): None.

Credit Hours: 3

Course(s) being deleted in place of this addition (*must submit course deletion form*): None.

Signatures: if disapproved at any level, do not sign. Return to previous signer with recommendation attached.

Dept. Chair/Division Head	Date 20 Oct 2014
Registrar 110101	Date 10/22/14
College Curriculum Chair	Date
Graduate Council Chair	Date

Request for Graduate Course Addition - Page 2

College: CITE Department/Division: Computer Science Alpha Designator/Number: CS 515

Provide complete information regarding the new course addition for each topic listed below. Before routing this form, a complete syllabus also must be attached addressing the items listed on the first page of this form.

1. FACULTY: Identify by name the faculty in your department/division who may teach this course.

Drs. Hyoil Han and Venkat N Gudivada

2. DUPLICATION: If a question of possible duplication occurs, attach a copy of the correspondence sent to the appropriate department(s) describing the proposal. Enter "***Not Applicable***" if not applicable.

Not Applicable.

3. REQUIRED COURSE: If this course will be required by another deparment(s), identify it/them by name. Enter "***Not Applicable***" if not applicable.

Not Applicable.

4. AGREEMENTS: If there are any agreements required to provide clinical experiences, attach the details and the signed agreement. Enter "***Not Applicable***" if not applicable.

Not Applicable.

5. ADDITIONAL RESOURCE REQUIREMENTS: If your department requires additional faculty, equipment, or specialized materials to teach this course, attach an estimate of the time and money required to secure these items. (Note: Approval of this form does not imply approval for additional resources.) Enter "***Not Applicable***" if not applicable.

Not Applicable.

6. COURSE OBJECTIVES: (May be submitted as a separate document)

See attached syllabus.

7. COURSE OUTLINE (May be submitted as a separate document)

See attached syllabus.

8. SAMPLE TEXT(S) WITH AUTHOR(S) AND PUBLICATION DATES (May be submitted as a separate document)

See attached syllabus.

9. EXAMPLE OF INSTRUCTIONAL METHODS (Lecture, lab, internship)

Lecture, lab., group discussion, and paper critiques.

Request for Graduate Course Addition - Page 4

10. EXAMPLE EVALUATION METHODS (CHAPTER, MIDTERM, FINAL, PROJECTS, ETC.)

Quiz, midterm exam, final exam, project, paper presentation, and assignment,

11. ADDITIONAL GRADUATE REQUIREMENTS IF LISTED AS AN UNDERGRADUATE/GRADUATE COURSE

Graduate students will be required to do a research-oriented project and term paper, augmented assignments and augmented exam.

12. PROVIDE COMPLETE BIBLIOGRAPHY (May be submitted as a separate document)

See attached syllabus.

Request for Graduate Course Addition - Page 5

Please insert in the text box below your course summary information for the Graduate Council agenda. Please enter the information exactly in this way (including headings):

Department:
Course Number and Title:
Catalog Description:
Prerequisites:
First Term Offered:
Credit Hours:

Department: Computer Science

Course Number and Title: CS 515 Data Mining

Catalog Description:
Covers (1) the process of knowledge discovery, (2) algorithms (association rules, classification, and clustering), and (3) real-world applications. Focuses on efficient data mining algorithms and scaling up data mining methods.

Prerequisites: None.

First Term Offered: Fall 2015

Credit Hours: 3

Marshall University
Syllabus

Course Title/Number	Data Mining / CS 515
Semester/Year	Fall / 2015
Days/Time	Tu, Th 9:30 – 10:45 a.m.
Location	GH 211
Instructor	Hyoil Han
Office	Gullickson Hall 205B
Phone	(304)696-6204
E-Mail	hanh@marshall.edu
Office/Hours	Tu, Th 11 – noon, 3:30 – 5:30 p.m. (or by appointment)
University Policies	By enrolling in this course, you agree to the University Policies listed below. Please read the full text of each policy by going to www.marshall.edu/academic-affairs and clicking on "Marshall University Policies." Or, you can access the policies directly by going to http://www.marshall.edu/academic-affairs/policies/ . Academic Dishonesty/ Excused Absence Policy for Undergraduates/ Computing Services Acceptable Use/ Inclement Weather/ Dead Week/ Students with Disabilities/ Academic Forgiveness/ Academic Probation and Suspension/ Academic Rights and Responsibilities of Students/ Affirmative Action/ Sexual Harassment: For more information: http://muwww-new.marshall.edu/academic-affairs/policies/

Course Description: From Catalog

Covers (1) the process of knowledge discovery, (2) algorithms (association rules, classification, and clustering), and (3) real-world applications. Focuses on efficient data mining algorithms and scaling up data mining methods.

The table below shows the following relationships: How each student learning outcomes will be practiced and assessed in the course.

Course Student Learning Outcomes	**How students will practice each outcome in this Course**	**How student achievement of each outcome will be assessed in this Course**
Students identify what data mining is, the components of knowledge discovery process, and data mining applications.	**Reading assignment** **Discussion** **In-class exercise**	**Assignment** **Exam** **Quiz**
Students identify the data preparation tasks and its implications.	**Reading assignment** **Discussion** **In-class exercise**	**Assignment** **Quiz**
Students demonstrate an understanding of the alternative knowledge representations such as rules and decision trees.	**In-class exercise** **Discussion**	**Assignment** **Exam** **Quiz**
Students examine the design and analysis of data mining algorithms and identify	**Reading assignment** **In-class exercise**	**Assignment** **Project**

approaches for scaling up data mining methods.	**Discussion**	**Exam**
Students compare alternative data mining algorithms and identify what might be most appropriate for a given data mining task.	**Paper review** **Reading assignment** **Discussion** **In-class exercise**	**Assignment** **Exam** **Paper** **Project**
Students will employ publicly available latest data mining tool(s).	**In-class exercise**	**Assignment** **Project**

Required Texts, Additional Reading, and Other Materials

1. Jiawei Han, Micheline Kamber, and Jian Pei, *Data Mining: Concepts and Techniques, 3rd Edition,* Morgan Kaufmann., 2012, ISBN: 978-0-12-381479-1
2. Technical papers will be distributed in class

References (Additional references appear at the end syllabus):

1. Ian H. Witten, Eibe Frank, and Mark A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques, 3rd Ed.,* Morgan Kaufmann, 2011.
2. Tom Mitchell, *Machine Learning*, McGraw-Hill, 1997.
3. D. Hand, H. Mannila, and P. Smyth, *Principles of Data Mining*, MIT Press, 2001.

Course Requirements / Due Dates

1. **In-class exercise / Quiz**
2. **Assignment**
3. **Project**
4. **Paper presentation**
5. **Midterm exam**
6. **Final exam**
7. **Discussion**
8. **Graduate level requirements: Refer to items under Course Requirements below.**

Grading Policy

Midterm Exam	20%
Final Exam	20%
Project	20%
Homework	10%
Paper reviews and presentation	10%
Discussion / in-class exercises & quizzes	10%
In-class contributions and attendance	10%

Total Score	Letter Grade
>= 90	A
>= 80 & <90	B
>= 70 & < 80	C
>= 60 & < 70	D

Course Requirements

Reading Assignments

You are expected to review the sections of the textbook that compose the in-class topics before the class in which they are covered.

In-class Exercises & Quizzes

There will be a series of exercises or quizzes covering topics discussed in class. There will be a series of hands-on exercises that reinforce the practical utility of the concepts presented in the text and the lecture slides.

Graduate Level Requirements: Graduate students should solve additional questions marked graduate-only.

Homework Assignments

There will be a number of assignments during the semester that emphasize the quantitative aspects of data mining.

Graduate Level Requirements: Graduate students should solve additional questions marked graduate-only.

Project Report

Each student will have two options for his/her project: (i) explore a current trend in data mining by performing independent research on a focused technical topic, and (ii) implement Data Mining procedure using real-world data and WEKA.

A written poster report of this research must be prepared and submitted in electronic form (in either Microsoft Word or Adobe PDF format). In addition, each student will give a 15-minute presentation to the class describing the results of this research.

Graduate Level Requirements: Graduate students should submit both (1) a poster from an implementation-oriented real-world data mining project and (2) a research-oriented term paper using ACM format.

Paper Presentation

Each student will present a paper related to real-world applications of data mining algorithms.

Exams

There will be a mid-term and a final exam.

Graduate Level Requirements: Graduate students should solve additional questions marked

graduate-only.

Class Attendance, Participation, and Decorum

Students are expected to attend all class sessions and participate in class activities. Students are also expected to maintain a certain level of decorum that includes turning off (or silencing) cell phones, arriving to class on time, not sleeping during class, and keeping side conversations to a minimum.

Attendance Policy

Students are expected to attend and participate in every class. After **four** unexcused absences, your grade will be decreased by one letter grade. Students who are absent more than 6 classes will get "F". Coming late to class or leaving the class early without permission is considered **half** absent.

Course Schedule

This is the tentative list of topics and due dates. This will be adjusted as the semester progresses: please see the course entry on MUOnline for the current information of course topics.
Lecture slides and other handouts of each class will be posted to MUOnline on the same day after class.

Weeks	Topics & Readings
1	Introduction
2	Getting to Know Your Data
3	Data Preprocessing
4	Introduction to Association Rule Mining
5	The Apriori Algorithm
6	Introduction to Classification (Chapter 8)
7	ID3 and C4.5 algorithms and beyond
8	Midterm review & exam
9	Introduction to Clustering (Chapter 10)
10	K-means Clustering algorithm and beyond
11	Ensemble Learning
12	K-nearest neighbor and Naïve Bayes
13	Paper Presentation
14	Thanksgiving week
15	Project presentation

Exam Attendance

Students are required to take exams at the scheduled class period. Students may take an exam at a different time under one of the following conditions:

- They present a University Excused Absence
- They present a valid medical excuse
- Other extraordinary circumstance as determined by the instructor

Communication

Class handouts (slides, labs, and homework) are all available on MUOnline. Course announcements will be sent to your myMU e-mail account.

Academic Conduct:

You are allowed and encouraged to work with other students on the completion of these assignments,

subject to the following constraints:

- copying someone else's work and submitting it as your own is plagiarism and will not be tolerated
- you may work with others to develop a solution to a problem but the material you submit must be your own work and you must acknowledge your collaborators
- unless designated as a team exercise, you may not sub-divide the tasks of an assignment; each student is expected to complete the whole assignment

It is your responsibility to satisfy the spirit of this conduct. If you have any questions, please ask the instructor for clarification. Depending on the severity of the offense, the instructor may:

- Take no action
- Penalize the student on the assignment in question
- Assign the student a failing grade in the course

Web Resources:

WEKA	http://www.cs.waikato.ac.nz/ml/weka/
R	http://www.r-project.org/
KDnuggets	http://www.kdnuggets.com/

Affirmative Action Policy:

This course will follow Marshall University's policy on Affirmative Action, which can be found on page 67 of the 2014-2015 undergraduate catalog. (http://www.marshall.edu/ucomm/files/web/UG_14-15_published.pdf) Specifically, all students will be afforded equal opportunity without regard to race, color, sex, religion, age, disability, national origin, or sexual orientation.

Inclement Weather Policy

Students can find information concerning Marshall's policy regarding inclement weather on p. 68 of the 2014-2015 undergraduate catalog.

Policy for Students with Disabilities:

Marshall University is committed to equal opportunity in education for all students, including those with physical, learning and psychological disabilities. University policy states that it is the responsibility of students with disabilities to contact the Office of Disabled Student Services (DSS) in Prichard Hall Room 117, phone 304 696-2271 to provide documentation of their disability. Following this, the DSS Coordinator will send a letter to each of the student's instructors outlining the academic accommodation they will need to ensure equality in classroom experiences, outside assignment, testing and grading. The instructor and student will meet to discuss how the accommodation(s) requested will be provided. For more information, please visit http://www.marshall.edu/disabled or contact Disabled Student Services Office at Prichard Hall Room 117, phone 304-696-2271.

University Computing Services Acceptable Use Policy:

Students are expected to read, understand, and follow the Acceptable Use Policy which can be found at http://www.marshall.edu/ucs/CS/accptuse.asp.

Additional References

1) Data Mining (i.e., Knowledge Discovery in Data [KDD]) Conferences
   a) ACM SIGKDD Int. Conf. on Knowledge Discovery in Databases and Data Mining (KDD)
   b) SIAM Data Mining Conf. (SDM)
   c) (IEEE) Int. Conf. on Data Mining (ICDM)
   d) European Conf. on Machine Learning and Principles and practices of Knowledge Discovery and Data Mining (ECML-PKDD)
   e) Pacific-Asia Conf. on Knowledge Discovery and Data Mining (PAKDD)
2) Other related conferences
   a) ACM SIGMOD, VLDB, ICDE, EDBT, ICDT
   b) Web and IR conferences: WWW, SIGIR
   c) ML conferences: ICML
3) Journals
   a) Data Mining and Knowledge Discovery (DAMI or DMKD)
   b) IEEE Trans. On Knowledge and Data Eng. (TKDE)
   c) KDD Explorations (http://www.kdd.org/explorations)
   d) ACM Trans. on KDD
4) Tools
   a) WEKA: http://www.cs.waikato.ac.nz/ml/weka/
   b) R Project: http://www.r-project.org/
5) T. Dasu and T. Johnson. Exploratory Data Mining and Data Cleaning. John Wiley & Sons, 2003
6) U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy. Advances in Knowledge Discovery and Data Mining. AAAI/MIT Press, 1996
7) U. Fayyad, G. Grinstein, and A. Wierse, Information Visualization in Data Mining and Knowledge Discovery, Morgan Kaufmann, 2001
8) D. J. Hand, H. Mannila, and P. Smyth, Principles of Data Mining, MIT Press, 2001
9) T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed., Springer-Verlag, 2009
10) B. Liu, Web Data Mining, 2nd ed., Springer 2011.
11) T. M. Mitchell, Machine Learning, McGraw Hill, 1997
12) G. Piatetsky-Shapiro and W. J. Frawley. Knowledge Discovery in Databases. AAAI/MIT Press, 1991
13) P.-N. Tan, M. Steinbach and V. Kumar, Introduction to Data Mining, Wiley, 2005
14) I. H. Witten and E. Frank, Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann, 3rd ed. 2011
15) T. Dasu and T. Johnson. Exploratory Data Mining and Data Cleaning. John Wiley, 2003
16) U. Fayyad, G. Grinstein, and A. Wierse. Information Visualization in Data Mining and Knowledge Discovery, Morgan Kaufmann, 2001
17) L. Kaufman and P. J. Rousseeuw. Finding Groups in Data: an Introduction to Cluster Analysis. John Wiley & Sons, 1990.
18) H. V. Jagadish, et al., Special Issue on Data Reduction Techniques. Bulletin of the Tech. Committee on Data Eng., 20(4), Dec. 1997
19) D. A. Keim. Information visualization and visual data mining, IEEE trans. on Visualization and Computer Graphics, 8(1), 2002
20) D. Pyle. Data Preparation for Data Mining. Morgan Kaufmann, 1999
21) S. Santini and R. Jain," Similarity measures", IEEE Trans. on Pattern Analysis and Machine Intelligence, 21(9), 1999
22) D. P. Ballou and G. K. Tayi. Enhancing data quality in data warehouse environments. Comm. of ACM, 42:73-78, 1999

23) A. Bruce, D. Donoho, and H.-Y. Gao. Wavelet analysis. *IEEE Spectrum*, Oct 1996
24) M. Hua and J. Pei. Cleaning disguised missing data: A heuristic approach. *KDD'07*
25) H. V. Jagadish, et al., Special Issue on Data Reduction Techniques. Bulletin of the Technical Committee on Data Engineering, 20(4), Dec. 1997
26) H. Liu and H. Motoda (eds.). *Feature Extraction, Construction, and Selection: A Data Mining Perspective*. Kluwer Academic, 1998
27) J. E. Olson. *Data Quality: The Accuracy Dimension*. Morgan Kaufmann, 2003
28) D. Pyle. Data Preparation for Data Mining. Morgan Kaufmann, 1999
29) R. Wang, V. Storey, and C. Firth. A framework for analysis of data quality research. IEEE Trans. Knowledge and Data Engineering, 7:623-640, 1995
30) R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases. SIGMOD'93
31) R. Agrawal and R. Srikant. Mining sequential patterns. ICDE'95
32) R. Agrawal and R. Srikant. Fast algorithms for mining association rules. VLDB'94
33) H. Mannila, H. Toivonen, and A. I. Verkamo. Efficient algorithms for discovering association rules. KDD'94
34) A. Savasere, E. Omiecinski, and S. Navathe. An efficient algorithm for mining association rules in large databases. VLDB'95
35) J. S. Park, M. S. Chen, and P. S. Yu. An effective hash-based algorithm for mining association rules. SIGMOD'95
36) H. Toivonen. Sampling large databases for association rules. VLDB'96
37) S. Brin, R. Motwani, J. D. Ullman, and S. Tsur. Dynamic itemset counting and implication rules for market basket analysis. SIGMOD'97
38) S. Sarawagi, S. Thomas, and R. Agrawal. Integrating association rule mining with relational database systems: Alternatives and implications. SIGMOD'98
39) S. Brin, R. Motwani, and C. Silverstein. Beyond market basket: Generalizing association rules to correlations. SIGMOD'97.
40) M. Klemettinen, H. Mannila, P. Ronkainen, H. Toivonen, and A. I. Verkamo. Finding interesting rules from large sets of discovered association rules. CIKM'94.
41) R. J. Hilderman and H. J. Hamilton. *Knowledge Discovery and Measures of Interest*. Kluwer Academic, 2001.
42) E. Omiecinski. Alternative Interest Measures for Mining Associations. TKDE'03.
43) T. Wu, Y. Chen, and J. Han, "Re-Examination of Interestingness Measures in Pattern Mining: A Unified Framework", Data Mining and Knowledge Discovery, 21(3):371-397, 2010
44) Manish Mehta, Rakesh Agrawal, Jorma Rissanen: SLIQ: A Fast Scalable Classifier for Data Mining. EDBT 1996: 18-32
45) J. Shafer et al., SPRINT: A Scalable Parallel Classifier for Data Mining. VLDB'96
46) Gehrke, Ramakrishnan & Ganti, RainForest - A Framework for Fast Decision Tree Construction of Large Datasets. VLDB'98
47) Jain AK, Dubes RC, Algorithms for clustering data, Prentice-Hall, 1988.