

ROUTE SHEET
PERMANENT COURSE CHANGE/APPROVAL
 (Attach course change request form)

Prefix & Number CS 553 Course Title Data Mining and Data Warehousing

Abbreviation for Schedule (20 characters): Data Min/Data Warehs

Nature of course request (Mark all that apply)

- | | | |
|--|--|---|
| <input checked="" type="checkbox"/> Add a course | <input type="checkbox"/> Prerequisite change | <input type="checkbox"/> LACC course |
| <input type="checkbox"/> Delete a course | <input type="checkbox"/> Number/Prefix change | <input type="checkbox"/> Undergraduate course |
| <input type="checkbox"/> Title change | <input type="checkbox"/> Description change | <input checked="" type="checkbox"/> Graduate course |
| <input type="checkbox"/> Writing Intensive (WI) | <input type="checkbox"/> Multicultural Diversity (D) | <input type="checkbox"/> 400/500 course |
| <input type="checkbox"/> Quantitative Literacy (Q) | <input type="checkbox"/> Honors course (H) | <input type="checkbox"/> Other: _____ |

- 1) Faculty Sponsor Signature *[Signature]* Date 3.7.14
- 2) Dept./Program Coordinator *[Signature]* Date 3/7/14
- 3) Division Chair *[Signature]* Date 3/7/14
- Curriculum Chair *[Signature]* Date 3/7/14

- 4) Faculty Senate Committees: The Curriculum Committee reviews all course proposals except for honors and graduate courses, which are reviewed instead by the Honors Committee or Graduate Committee. All 400/500 "split" courses must be approved by both the Curriculum and Graduate Committees. All curriculum committee decisions are forwarded to the Senate Executive Committee.

a) Curriculum Committee Chair _____ Date _____

___ N/A ___ Approved ___ NOT Approved

b) Graduate Committee Chair _____ Date _____

___ N/A ___ Approved ___ NOT Approved

c) Honors Committee Chair _____ Date _____

___ N/A ___ Approved ___ NOT Approved

5) Faculty Senate President _____ Date _____

___ Approved by the Senate Executive Committee

___ Approved by the Senate ___ NOT Approved (Return to sponsor)

6) Appropriate Dean _____ Date _____

___ Approved ___ NOT Approved (Return to Faculty Senate President)

7) Provost/VPAA _____ Date _____

___ Approved ___ NOT Approved (Return to Faculty Senate President)

REQUEST FORM PERMANENT COURSE APPROVAL

Initiated by (print): Jie Liu Date: 2.25.2014

ADDING A COURSE

Prefix/Number	Descriptive Title	Cr. Hours
CS 553	Data Mining and Data Warehousing	4

Catalog Description:

This course discusses the techniques most commonly used in the analysis of large volumes of data, often referred to as Big Data issue or BI, in the extraction of knowledge from this data, and in making decisions based on the knowledge acquired. The course is a hands-on one because we will learn how to build a data warehouse and how to conduct data mining.

Course Goals and Objectives:

Upon completion of this course, students will be able to:

- Contrast the differences between a data mining project with a data ware house project.
- Classify data mining algorithms and rate the appropriate algorithms to software a problem on hand.
- Apply the tools and algorithms to solve real world problems by identifying the correct data source, formulate the data source, and interpret the mining results, and assess the effectiveness and correctness of the mining results.
- Plan a data mining project.
- Students will be exposed with the current trends in BI and related field.

Justification for adding the course (e.g. alignment with other institutions, program revision, etc.):

Big data is one of the cornerstones of the technology advancement shaping our world. We do not have a class covering the topics.

Briefly describe other WOU faculty/programs consulted (attach additional sheet(s) if necessary).

Faculty in the Computer Science Division has been consulted. No outside faculty or programs are affected by the addition of this course.

Faculty and Facilities Needed:

1 Smart Classroom, 1 Instructor

Attach brief course outline

1. Introduction to data mining and data warehousing
 - a. Why data mining and data warehousing
 - b. The main differences between data mining and data warehousing
 - c. The history of data mining and data warehousing
 - d. What is BI
 - e. Major steps of a BI project
2. On data warehousing (OLAP)
 - a. Architecture for supporting a data warehousing
 - b. Data preparation
 - c. Star schema
 - d. Measures and dimension
 - e. Using SQL Server 2012 and its data tools to create a cube
 - i. Define data source
 - ii. Define data source views
 - iii. Define dimensions and measures
3. Data mining techniques
 - a. Architecture for supporting a data warehousing
 - b. Decision trees
 - c. Naïve Bayes
 - d. Association rules
 - e. Clustering
 - f. Model Evaluation

CS553 Data Mining and Data Warehouse

Course Syllabus, Winter 2014

Time/Place : TR 9:300 – 10:50 AM /ITC301
Instructor/Office/Phone : Jie Liu / ITC308B / 8-8989
Office Hours : See www.wou.edu/~liuj
Email Address : liuj@wou.edu

Required Text:

Delivering Business Intelligence with Microsoft SQL Server 2012 by Brian Larson, 3/E, ISBN-13: 978-0071759380

Course Goal:

Big data is one of the cornerstones of the technology advancement for the next few years. In the never-ending quest for a competitive advantage, organizations are turning to large repositories of corporate and external data to uncover trends, statistics and other actionable information to help them determine their next move. CS553 introduces data mining and data warehouse using SQL Server 2012 Analysis Services. The courses covers creation of data warehouse (Cube), star schema, GUI of Microsoft Data Tool and SQL Server 2012 Analysis Services and various data mining algorithms including the following algorithm types:

- Classification predicts one or more discrete variables, based on the other attributes in the dataset.
- Regression predicts one or more continuous variables, such as profit or loss, based on other attributes in the dataset.
- Segmentation divides data into groups, or clusters, of items that have similar properties.
- Association finds correlations between different attributes in a dataset. The most common application of this kind of algorithm is for creating association rules, which can be used in a market basket analysis.
- Sequence analysis summarizes frequent sequences or episodes in data, such as a Web path flow.

Students have several research topics to manage as well. The first is to research and install open source tools especially the most popular tool Hadoop. Students will also need to master the mathematic foundations behind each of the algorithms, be able to evaluate the fitting algorithms for problems on hand, and be able to analyze the final reports from running different algorithms to reach accurate conclusions.

Course Outcomes:

Students will be able to contrast the differences between a data mining project with a data ware house project. Students will be able to classify data mining algorithms and rate the appropriate algorithms to software a problem on hand.

Students will be able to apply the tools and algorithms to solve real world problems by identifying the correct data source, formulate the data source, and interpret the mining results, and assess the effectiveness and correctness of the mining results.

Students will be able to plan a data mining project.

Students will be exposed with the current trends in BI and related field.

Grading:

Labs and Exercises	25%	
Quizzes		15%
Midterm		20%
Project	10%	
Final	30%	

The cut off for letter grades are: A 90%, B 80%, C 70%, D 60%

A passing grade (60%) on the final is required to pass the course. Students who score 80% of the possible on the final and have the highest score will receive an A. Students who pass MS certification Exam 70-463 or Exam 70-468 during finals week receive an A automatically.

Labs and Exercises:

Labs and exercises have to be done individually although discussions are strong encouraged. We may have additional reading and research assignments. However, we will do most of the hands on exercises in the book.

Quizzes and Exam:

There will be four quizzes. The lowest quiz score will be discarded. As a result, students do not need to make up a missed quiz. Midterm will be given the Thursday of week 5. Final is comprehensive. You may bring one sheet of notes to the midterm and two sheets of notes to the final. Otherwise, the exams are close book and notes.

Project:

More to come!

Class Philosophy:

I would like very much for students to contribute to the overall learning process. If a student has a question, an idea, an answer to a question, a good article to share, or a suggestion, please let us all hear it. Hopefully, others may learn something from it.

Academic Dishonesty:

Academic dishonesty refers to cheating: a serious ethical issue. You are encouraged to learn from each other and to help each other on concepts discussed in class; however, not directly on questions that are part of labs and exercises. Copying from others' labs and exercises and cheating on quizzes and exams are grounds for a zero on the exercise/lab/exam/quiz for both parties involved and possibly a failure on the course. Written work that appears to be copies of each other will not be given credit. If I suspect an Academic Dishonesty issue, I will call you into my office and discuss the options you have. "Helping" or "being helped by" another student or the appearance of doing so during a quiz or an exam will be considered academic dishonesty. This will be grounds for a zero on the quiz or exam for all parties involved, no questions asked. (This section is modified based on Dr. Broeg's similar statements)

Tentative Schedule (as of 4/3/2013):

Date	Chapter	Misc.
Week 1	1	
Week 2	2, 3	Q1
Week 3	3, 4, 5	
Week 4	6, 9	Q2
Week 5	Midterm	Midterm
Week 6	15,	
Week 7	15,	Q3
Week 8	16	No class on 5/21
Week 9	17	Q4

Week 10	Project Presentations	
Week 11	Final	

Note:

1. All the exercises and labs are due at the beginning of the class on the due day. All exercises and labs must be handed in on time ready or not. Exercises or labs that are ONE week late will not be accepted for credit. Exercises or labs turned in late will receive a 20% reduction in credit for each working day. Exceptions to this rule are rare and they must be cleared with the professor first.
2. Actual class schedule may vary considerably from the estimated -- pay attention.
3. Most of the labs and exercises will take many hours to complete. Start early on each lab and allow about twice as much time as you think you will need.
4. The expected attendance rate is 99%. I give pop quizzes to award students attending lectures. I do not repeat my lectures during my office hours
5. The detailed schedule is given so you can preview the materials.