


Undergraduate Council  
 New Course  Course Change *2005*  
 Core Category: *P2016* Effective Fall *2008*

or

Graduate/Professional Studies Council  
 New Course  Course Change  
 Effective Fall \_\_\_

RECEIVED MAR 06 2008

APPROVED OCT 22 2008

- Department: CHE ENG College: ENGR
- Person Submitting Form: Dr. Michael P. Harold Telephone: 34307
- Course Information on New/Revised course:
  - Instructional Area / Course Number / Long Course Title:  
PETR / 5300 / Data Mining and Database Management
  - Instructional Area / Course Number / Short Course Title (30 characters max.)  
PETR / 5300 / DATA MINING & DATABASE MGMT
  - SCH: 3.00 Level: SR CIP Code: 14.2501.00 Lect Hrs: 3 Lab Hrs: 0
- Justification for adding/changing course: To provide for new discipline areas
- Was the proposed/revised course previously offered as a special topics course?  Yes  No  
 If Yes, please complete:
  - Instructional Area / Course Number / Long Course Title:  
\_\_\_ / \_\_\_ / \_\_\_
  - Content ID: \_\_\_ Start Date (yyyy3): \_\_\_
- Authorized Degree Program(s): BS PETR ENGR
  - Does this course affect major/minor requirements in the College/Department?  Yes  No
  - Does this course affect major/minor requirements in other Colleges/Departments?  Yes  No
  - Are special fees attached to this course?  Yes  No
  - Can the course be repeated for credit?  Yes  No
- Grade Option: Letter (A, B, C ...) Instruction Type: lecture ONLY (Note: Lect/Lab info. must match item 3, above.)
- If this form involves a change to an existing course, please obtain the following information from the course inventory: Instructional Area / Course Number / Long Course Title  
 \_\_\_ / \_\_\_ / \_\_\_
  - Start Date (yyyy3): \_\_\_ Content I.D.: \_\_\_
- Proposed Catalog Description: (If there are no prerequisites, type in "none".)  
Cr: 3. (3-0). Prerequisites: MATH 3321 and MECE 2334. Description (30 words max.): Data mining goals, data quality, data preprocessing, OLAP, exploratory data analysis, classification, regression, clustering, dimensionality reduction, association rules, post processing, data mining case studies. *Petroleum data.*
- Dean's Signature:  Date: 3/6/08

Print/Type Name: Joseph Tedesco, Dean

**UC 9786 08F**  
**Page 2 of 7**

**Cullen College of Engineering**  
**CBM003 Supplement - B Fe**  
**(New Course)**

*Must be attached to CBM003 form*

Course: <u>PETR</u>	<u>5300</u>
<i>Subject Prefix</i>	<i>Course Number</i>

1. **Course Title:** Data Mining and Database Management  
*Print course inventory screen using RARCAS/CATM and attach.*
2. **Pre-requisite/Co-requisite:** Math 3321 and MECE 2334.
3. **Rational for Course Format:** Standard university course format
4. **Rational for Course Content:** Ability to handle large volume of data
5. **ABET Constituents consulted:** Petroleum Engineering Advisory Board, several industry focus groups
6. **State Course Outcomes:** Students learn how to find, integrate and analyze large volumes of data from multiple sources
7. **Course Performance after implementing format and content changes:** 1
8. **Is course required?**                      X Yes                       No
9. **Required course outline attached?**   X Yes                       No
10. **Estimated student demand**   50   per semester
11. **Similar courses in other departments:**  Yes                      X No  
     a. *If yes, list course(s)* \_\_\_\_\_
12. **Is course part of a sequence?**                      Yes                      XNo  
     a. *If Yes, identify the sequence and comment on the relation to prior and subsequent courses:*  
         **Textbook(s) and other required materials:** Jiawei Han and Micheline Kamber, *Data Mining: Concepts and Techniques*  
         Morgan Kaufman Publishers, second edition, 2006

**Note:** Special Fees: If special fees requested, **Course Related Fee Request Form** will be required.

<sup>1</sup> Department reports will be requested about the effects of your new course on your curriculum both 12 and 24 months after the effective date for this new course.

Cullen College of Engineering **UC 9786 08F**  
CBM003 Supplement - *B F* Page 3 of 7  
(New Course)

	<u>2/21/08</u>	<input checked="" type="checkbox"/> Approved
Chair of Initiating Dept. Signature	Date	

COLLEGE OF NATURAL SCIENCES & MATHEMATICS

HTTP://NSM.UH.EDU

COURSE TITLE/SECTION: Data Mining for Engineering (new course)

TIME: TBA

FACULTY: Carlos Ordonez

OFFICE HOURS:

E-mail: cordonez@uh.edu

Phone: 31013

FAX: 33335

---

**I. Course: PETR 5300 *Data Mining and Database Management***

**Prerequisites:** MATH 3321 and MECE 2334.

**A. Catalog Description ()**

Data mining goals, data quality, data preprocessing, OLAP, exploratory data analysis, classification, regression, clustering, dimensionality reduction, association rules, post processing and data mining case studies.

**B. Purpose**

Data mining is about building models and finding patterns on large amounts of data. Data mining is an important research field combining statistical, machine learning and database techniques. The course covers most of the important data mining techniques and provides background knowledge on how to conduct a data mining project. An introduction to data mining will be given then the course will present statistical analysis on one variable. Data mining techniques such as classification, regression, clustering and association rules will be discussed in detail. Techniques to preprocess a data set for a data mining task will be discussed. Also basic visualization techniques and statistical methods will be introduced. More advanced topics including stream mining, class decomposition through clustering and classification techniques will also be introduced.

**II. Course Objectives**

Upon completion of this course, students;

1. will understand data mining
2. will have experience on how to conduct a data mining project
3. will understand elementary statistical analysis like scatter plots, histograms and the normal distribution
4. will know popular classification techniques, such as decision trees, support vector machines and nearest-neighbor approaches.
5. will be able to apply dimensionality reduction techniques such as PCA
6. will know the most important association rule techniques
7. will have exposure to research topics

**III. Course Content**

1. Introduction to Data Mining
2. Statistical analysis on one variable
3. Data quality
4. Data preprocessing
5. OLAP and cubes
6. Clustering
7. Dimensionality reduction
8. Classification
9. Regression
10. Scoring
11. Association rules
12. Advanced topics

**IV. Course Structure**

24 lectures

2 exams

8 homeworks

## V. Textbooks

### Required Text:

Jiawei Han and Micheline Kamber, *Data Mining: Concepts and Techniques*.  
Morgan Kaufman Publishers, second edition, 2006.

### Auxiliary Text:

T. Mitchell, *Machine Learning*, MIT Press, 1997.

## VI Course Requirements

There will be two exams: one midterm and one final. There will be eight homeworks. Students will need to learn a data mining tool working on a relational database system.

## VII. Evaluation and Grading

Homeworks: 45%

Exams: 50%

Class Participation: 5%

Each student has to have a **weighted average of 74.0 or higher in the exams of the course** in order to receive a grade of "B-" or better for the course. Students will be responsible for material covered in the lectures and assigned in the readings. Exams will be closed book. Written assignments are individual. Home works using data mining software will be done by teams of two students.

We encourage students to help each other understand course material to clarify the meaning of homework problems or to discuss problem-solving strategies, but it is **not** permissible for one student to help or be helped by another student in working through homework problems and in the course project. If, in discussing course materials and problems, students believe that their like-mindedness from such discussions could be construed as collaboration on their assignments, students must cite each other, briefly explaining the extent of their collaboration. Any assistance that is not given proper citation may be considered a violation of the Honor Code, and might result in obtaining a grade of F in the course, and in further prosecution.

**Policy on grades of I (Incomplete):** A grade of 'I' will only be given in extreme emergency situations.

## VIII. Consultation

Instructor: Dr. Carlos Ordonez  
office hours (571 PGH)  
e-mail: cordonez@uh.edu

**IX. Bibliography**

The following conferences center on data mining and database systems:

1. Data mining:
  - Conference proceedings: KDD, ICDM, DMKD, PKDD.
  - Journal: Data Mining and Knowledge Discovery
2. Database Systems:
  - Conference proceedings: SIGMOD, CIKM, ICDE
  - Journals: ACM-TODS, IEEE-TKDE
3. Machine Learning:
  - Conference proceedings: ICML, ECML, AAI, IJCAI, etc.
  - Journals: Machine Learning, KAIS etc.

**Addendum:** Whenever possible, and in accordance with 504/ADA guidelines, the University of Houston will attempt to provide reasonable academic accommodations to students who request and require them. Please call 713-743-5400 for more assistance.