

Comp. 2000  
196

UNIVERSIDAD DE BUENOS AIRES  
FACULTAD DE CIENCIAS EXACTAS Y NATURALES

1. DEPARTAMENTO: Computación
  2. CUATRIMESTRE: 1ero. del 2000
  3. ASIGNATURA: Data Mining
  4. CARRERA: Licenciatura en Ciencias de la Computación
  5. CARÁCTER DE LA MATERIA: optativa
  6. NUMERO DE CODIGO DE CARRERA: 18
  7. NUMERO DE CODIGO DE MATERIA: C039
  8. PUNTAJE: 2
  9. PLAN DE ESTUDIOS AÑO: 1993 y 1987
  10. DURACION DE LA MATERIA: Cuatrimestral
  11. HORAS DE CLASE SEMANAL:
    - a) TEORICAS/PRACTICA: 6hs
    - b) LABORATORIO:
    - c) PROBLEMAS:
    - d) SEMINARIOS: -----
  12. CARGA HORARIA TOTAL: 6hs
  13. ASIGNATURAS CORRELATIVAS: Base de Datos e Inteligencia Artificial (Plan 87 y 93)
  14. FORMA DE EVALUACION: trabajos prácticos y final
  15. PROGRAMA Y BIBLIOGRAFIA: Adjuntas a esta hoja
- FECHA: mayo del 2000

Lic. Osvaldo GONZALEZ  
Firma y Aclaración  
del Profesor

Firma del Director  
y Sello Aclaratorio

D.ª MARTA MEJAIL  
ESTERA ADJUNTA  
DEPT. DE COMPUTACION  
F. C. E. y N. UBA

Materia del Depto. de Computacion  
materia para el Segundo Cuatrimestre de 1998.-

# Data Mining

DEPARTAMENTO: Computación .....

ASIGNATURA: Data Mining .....

CARRERA/S: Licenciatura en Ciencias de la Computacion  
(PLAN 87 PLAN 93) .....

CARACTER: OPTATIVA .....

PUNTAJE: . puntos .....

DURACION DE LA MATERIA: cuatrimestral .....

HORAS DE CLASE: a) TEORICAS.....3 Hrs. b) PROBLEMAS .....  
c) LABORATORIO .... 3 Hrs. d) SEMINARIOS .....  
e) TOTALES ..... 6 Hrs

ASIGNATURAS CORRELATIVAS: Bases de Datos .....

FORMA DE EVALUACION: Estudio de Trabajos Teóricos y entrega de T.P.....

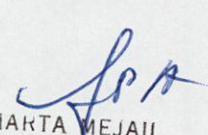
PROGRAMA:

## 1. Conceptos fundacionales de la lógica Inductiva

- Definición de Teoria
- Definición de correctitud
- Definición de "Excesivamente general" y "Excesivamente específico"
- Definición formal de ILP
- Algoritmo para generar teoría correcta
- Top-down y Bottom-up
- Aprendizaje de predicados simples y múltiples
- Métodos de Inferencia para la Lógica Inductiva
- Resolución Inversa
- Unfolding

## 2. El proceso de mining: KDD

- Introducción
- Una Definición aproximada: KDD en bases de datos
- Definición de data Mining

  
Dra. MARTA MEJAIL  
DIRECTORA ADJUNTA  
DEPTO. DE COMPUTACION  
F. C. E. y N. UBA

Síntesis de los métodos de Data Mining  
Métodos de Data Mining Populares  
Componentes de los Algoritmos de Data Mining  
Antecedentes: Algoritmo ID3  
El proceso de descubrir Conocimiento

### **3. Data Mining y Data Warehouse.**

Que es un Data Warehouse y porque es necesario  
Diseño de Decision Support  
Integración con data Mining

### **4. El proceso de descubrir Conocimiento. KDD Process**

Data Selection  
Cleaning  
Enrichment  
Coding  
OLAP tools  
Data Mining  
Ejemplos

### **4. Modelos de Data Mining**

Decision Trees  
Association rules  
Neural Network  
Sequential Patterns  
Time Series  
Clustering  
Algoritmos Genéticos

### **5. Componentes de los algoritmos de Data Mining**

Model Representation  
Modelo Evaluation  
Search Method

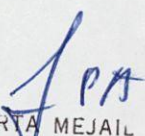
### **6. Algunos aspectos formales de los algoritmos de aprendizaje**

Aprendizaje como compresion de conjunto de datos  
Ruido y Redundancia  
Fuzzy Databases  
Denormalización  
Primitivas data mining

### **7. Temas Asociados**

Spatial Data Mining  
Sampling de Bases de Datos  
Active Data Mining  
Data Mining & WEB

## **Parte Practica**

  
Dra. MARTA MEJAIL  
D. RECTORA ADJUNTA  
DEPTO. DE COMPUTACION  
F. C. E. y N. UBA

La materia consta además de una entrega de Trabajos tanto teórico como práctico donde cada alumno debe desarrollar un modelo de Data Mining para un problema específico. Además se entregaran papers para que sean debatidos en clase.

Lic. Osvaldo Omar González  
Profesor Adjunto

  
Dra. MARTA MEJAIL  
DIRECTORA ADJUNTA  
DEPTO. DE COMPUTACION  
F. C. E. y N. UBA

## Bibliografía Asociada a la materia de data mining:

**Adriaans P & Zantinge D.** "Data Mining"  
Addison-Wesley [1996]

**Adriaans P & Knobbe A.** "Discovering Foreign key relations in relational databases"  
Workshop on Statics, Machine Learning and Knowledge Discovery in Databases [1995]

**Agrawal et Al** "In advances in Knowledge Discovery and Data Mining"  
AAAI Press [1995]

**Agrawal & Srikant** "Mining Sequential Pattern"  
IBM Research [1995]

**Agrawal & Sarawagi & Gupta** "On computing the data Cube"  
IBM Research [1996]

**Ali K. & Manganaris S.** "Partial Classification using Association rules"  
IBM Research [1996]

**Apt K & Bol R.** "Logic programming and Negation"  
A Survey Journal of Logic Programming 19 [1994]

**Apt K & Doets R.** "A new definition of SLDNF-Resolution"  
Journal of Logic Programming 18 [1994]

**Apte Ch. & June Hong S.** "Predicting Equity Returns from Securities Data with Minimal Rule Generation"  
MIT Press [1996]

**Brachman R.** "The Process of Knowledge Discovery in Databases"  
MIT Press [1996]

**Berry Michael & Linoff Gordon** "Data Mining techniques"  
Wiley Computer publishing [1997]

**Bostrom & Idestam-Almquist** "Specialization of logic programs by pruning SLD trees"  
Ellis Horwood Ney York [1994]

**Cabena & Hadjinian et al** "Discovering data mining: From the concept to implementation"  
International Technical Support Organization [1997]

**Ceri, Gottlob, Tanca** "Logic Programming and databases"  
Springer-Verlag [1990]

**Clark P. & Niblett T** "The CN2 Induction Algorithm"  
Machine Learning 3, 261-283 [1996]

**Date C.J.** "An Introduction to database Systems"  
Addison-Wesley Publishikng Company [1995]

**Date** "A guide to the SQL standard"  
Addison-Wesley [1994]

**Dzeroski S.** "Inductive Logic Programming and Knowledge Discovery in DataBases"  
MIT Press [1996]

**Fayyad et al** "Advances in Knowledge Discovery and data mining"  
MIT Press [1996]

**Fayyad U. & Haussler & Stolorz P** "Mining scientific Data"  
COMMUNICATION OF ACM 51-57 [1996]

  
**Dra. MARTA MEJAIL**  
DIRECTORA ADJUNTA  
DEPTO. DE COMPUTACION  
F. C. E. y N. UBA

**Fayyad U. & Shapiro G & Smyth P.** "From Data Mining to Knowledge Discovery in databases"  
AI Magazine 17 , 37-54 [1996]

**Fayad U. & Uthurusamy R.** "Data Mining and Knowledge Discovery in Databases"  
COMMUNICATION OF ACM - November 1996/Vol 39. N 11

**Fayyad U. & Shapiro G & Smyth P.** "The KDD Process for extracting Useful Knowledge from Volumes of Data"  
COMMUNICATION OF ACM - November 1996/Vol 39. N 11

**Fayyad U & Piatetsky-Shapiro G** "From Data Mining to Knowledge Discovery"  
en "In advances in Knowledge Discovery and Data Mining"  
AAAI Press [1995]

**Heinonen O. & Mannila H.** "Attribute-Oriented Induction and Conceptual Clustering"  
Dep. Computer Science University of Helsinki (FINLANDIA)

**González Osvaldo** "Sistemas Basados en Conocimiento: Razonamiento Nomonótono"  
Informe III UBACyT – Proyecto UBACyT [1993]

**González Osvaldo** "Bases de Datos Inteligentes"  
Departamento de computación UBA [1996]

**González Osvaldo** "Knowledge based Data Mining"  
Data Mining & Knowledge Discovery [1998]

**González Osvaldo** "Curso de Data Mining"  
Departamento de computación UBA [1997]

**Groth Robert** "Data Mining: A hands-on approach for business professionals"  
Prentice Hall [1997]

**Helman Paul** "The Science of database Management"  
IRWIN [1994]

**Hughes John** "Object-Oriented databases"  
Prentice-Hall International editions [1991]

**Khoshafian Setrag** "Object- Oriented Databases"  
Willey Professional Computing [1993]

**Komorowski Jan, Zytkow Jan** "Principles of Data Mining and Knowledge Discovery"  
Springer – Lecture Notes in Artificial Intelligence (1263) [1967]

**Komrowski H** "Partial evaluation as a means for inferencing data estructures in an applicative language"  
Proceedings of 9<sup>th</sup> ACM sumposium on Principles of Programming Languages ACM[1982].

**Kowalski Robert** "Predicate Logic as a Programming Language"  
Information processing 74 [1974]

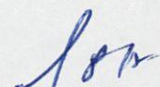
**Kowalski Robert** "Logic for Problem Solving"  
North-Holland, New York [1979]

**Kumar Das Subrata** "Deductive Databases and Logic Programming"  
Addison-Wesley Publishikng Company [1992]

**Levi Isaac** "For the sake of the argument"  
Cambridge University [1997]

**Loomis Mary** "Object Databases: The essentials"  
McGraw-Hill [1995]

**McGabe Francis** "Logic and Objects"  
Prentice Hall [1994]

  
Dra. MARTA MEJAIL  
DIRECTORA ADJUNTA  
DEPTO. DE COMPUTACION  
F. C. E. Y N. UBA