



**Universidad Autónoma del Estado de México**

**Material didáctico multimedia**

**Sólo visión**

**La minería de datos en el proceso de KDD ( Knowledge  
Discovery and Data Mining)**

**Unidad de Aprendizaje Minería de Datos  
Licenciatura de Ingeniería en Computación  
Facultad de Ingeniería**

**Elaborado por M en I Sara Vera Noguez  
Durante el período intensivo verano 2016**

## ***Justificación académica***

### **Introducción**

El curso de Minería de Datos corresponde a una Unidad de Aprendizaje (L41093) de la línea de acentuación de desarrollo de software, de la licenciatura de Ingeniería en Computación, que se imparte en la Facultad de Ingeniería, así como en los Centros Universitarios: Atlacomulco, Ecatepec, Texcoco, Valle de Chalco, Valle de México, Valle de Teotihuacán y Zumpango.

Se trata de un curso teórico - práctico de 6 créditos, con una carga de 2 horas teóricas y horas prácticas a la semana.

### **Competencias**

De acuerdo a lo plasmado en el programa de estudios vigente y avalado por los HH Consejos, las competencias a desarrollar en el curso son las siguientes:

El alumno:

- Conocerá las bases metodológicas para la extracción de conocimiento a partir de grandes repositorios de información.
- Conocerá las bases teóricas de las técnicas de minería más usuales que permiten la obtención de modelos y patrones a partir de la vista minable.
- Conocerá el software disponible para realizar minería de datos.
- Podrá conjuntar los conocimientos adquiridos para realizar un proyecto de minería de datos en un contexto real.
- Podrá interpretar y evaluar los resultados obtenidos para trasladarlos en acciones recomendadas.
- Conocerá la metodología CRISP-DM y la minería de datos complejos.

## Estructura del curso

De acuerdo al programa actual oficial, la unidad de aprendizaje se encuentra estructurada por las siguientes cinco unidades de aprendizaje:

1. Introducción a la minería de datos
2. La minería de datos en el proceso de KDD (Knowledge Discovery in Databases)
3. Tareas y métodos en minería de datos
4. Herramientas de software disponibles para minería de datos
5. Metodología CRISP-DM y minería de datos complejos

## Importancia del material en el contexto del curso

El curso tiene un contenido muy amplio, particularmente en la unidad 3, pero se debe hacer énfasis con información básica suficiente en la unidad 2.

Conocer y entender el proceso KDD (Knowledge Discovery in Databases) es de gran importancia para comprender la finalidad y procesos a seguir en un proyecto de Minería de Datos.

El tener claridad en el proceso propicia que se encuentre utilidad y aplicación al gran cúmulo de tareas y métodos estudiados en la unidad 3.

Con la intención de lograr que los estudiantes conozcan el proceso KDD es conveniente contar un material que ayude a conjuntar información relativa a dicho proceso

También es útil en este sentido relacionar el proceso que rige a la Minería de Datos con el proceso de desarrollo de software, con el que los estudiantes de esta línea de acentuación, de forma general, están bastante familiarizados.

De acuerdo a diversos autores como Biggs, Marroquín, Demarchi, entre otros; es aconsejable que la información sea presentada a los estudiantes mediante diversos canales perceptivos, y es con esta intención que se ha desarrollado el material; para ser usado durante las sesiones presenciales con la intención de que los estudiantes puedan escuchar la explicación del profesor a la vez que se refuerza la información con el material proyectable.

En la elaboración del material se aborda el proceso KDD de lo general a lo particular, partiendo de un diagrama en el que se ve el proceso completo, en el que identifican los distintos momentos que lo conforman y que para cada momento y cada fin se cuenta con conjunto de herramientas y técnicas específicas, de esto se hace una analogía con los cubierto y vajilla de una mesa, en la que para cada fin y momento en particular se cuenta con un instrumento adecuado.

Si bien el material está pensado para la adquisición de información por parte del estudiante y en este sentido dentro del material en mayor medida se presenta información, se han incluido algunas imágenes que sirva de ejemplo o analogía.

## Recomendaciones de uso

Se aconseja proyectar el material a la vez que el profesor explica el contenido de la unidad 2. Si la clase se realiza en una sala de cómputo se puede publicar en una plataforma como moodle, para que los estudiantes puedan seguir el material desde el equipo en el que estén trabajando, así como intercalar el contenido del material, que se encuentra dividido en tres archivos, con algunas actividades que eviten la monotoneidad de la clase y que permitan a los estudiantes procesar la información de forma gradual.

En la imagen de la figura 1, se muestra una captura de pantalla de moodle, de la forma en que durante el curso se utilizó el material que presento (en este caso un archivo se dividió en dos por ser muy pesado para la plataforma).

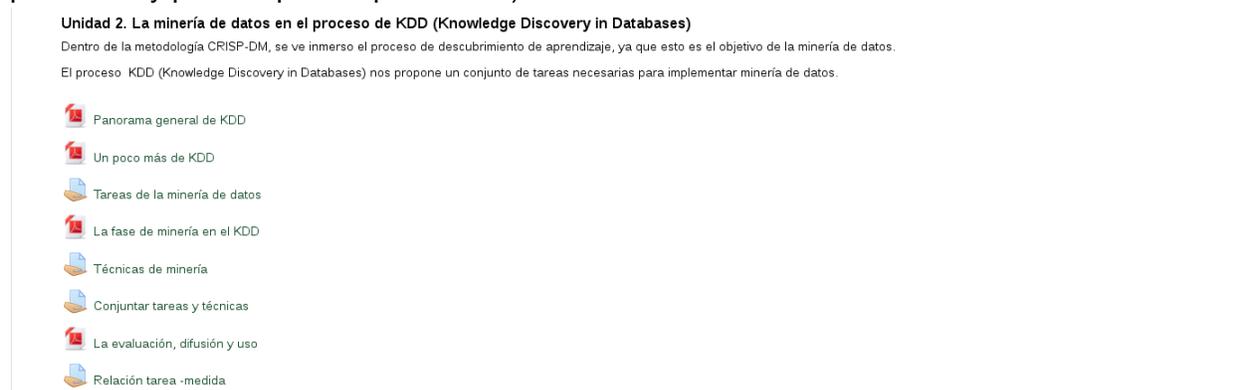


Figura 1. Uso del material didáctico de proyección en moodle con algunas actividades propuestas.

En el primer paquete de diapositivas, incluido en el archivo 2.PanoramaGeneralDelProcesoKDD.pdf se busca dar una primera aproximación general y somera del proceso KDD, así como algunas definiciones básicas, este se puede dejar que los estudiantes lo revisen antes de comenzar el tema.

En el segundo paquete, que se encuentra en el archivo 2.2DetallesDelProcesoKDD.pdf se incluye información relativa a las fases del proceso, tareas de la minería de datos y se

concluye con la propuesta de algunas actividades en equipo para procesar la información incluida, así como la recomendación de lectura uno de los libros utilizados.

En el último paquete, incluido en el archivo 2.2.3DetallesDelProcesoKDDFasesFinales.pdf, se mencionan algunos detalles de las últimas fases de procesos.

Ya que el material que se presenta se ha utilizado en curso en el periodo intensivo de verano, no se hacen recomendaciones en cuanto al número de sesiones para su uso, ya que la distribución de clases en los períodos regulares es muy diferente a la de los intensivos.

Por otro lado se tiene la certeza de que los profesores de la unidad de aprendizaje podrán enriquecer la propuesta presentada y adecuarla al estilo de aprendizaje de sus estudiantes, sí como a la conformación del grupo, y escenario.

## Bibliografía

Biggs, J., & Biggs, J. B. (2005). "Calidad del aprendizaje universitario" (Vol. 7). España: Narcea ediciones.

De la Rosa Reyes, María de los Ángeles (2004), El Desarrollo de Competencias Comunicativas: uno de los Principales Retos en la Educación Superior". México, UNAM.

Demarchi, Claudia. (2009). "Estrategias de comunicación en entornos virtuales". Argentina: Escuela de Ciencias de la Información

Han, D. J. (2007). Principles of Data Mining, EEUU: MIT Press.

Hernández Orallo, J., M. J. Ramírez Quintana, et al. (2004). Introducción a la Minería de Datos. España, Pearson Educación SA.

Introducción a la Minería de Datos. España, Pearson Educación SA.

Maimon, O. Z. and L. Rokach (2005). Data mining and knowledge discovery handbook. USA, Springer.

Marroquín Pérez, Manue. (1995). "Importancia de la Comunicación Interpersonal en *La Comunicación Interpersonal*". España: Ediciones Mensajero

Pérez López, C. and D. Santín Gonzalez (2006). Data Mining- Soluciones Con Enterprise Miner. México, Alfaomega, Ra-Ma.

Sumathi, S. and S. N. Sivanandam (2006). Introduction to data mining and its applications. Berlin, Germany, Springer-Verlag New York Inc.

Tan, P. N., M. Steinbach, et al. (2005). Introduction to data mining, Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA. Maimon, O. Z. and L. Rokach (2005). Data mining and knowledge discovery handbook. USA, Springer.