**UNIVERSIDAD TÉCNICA NACIONAL**

**CARRERA: INGENIERIA DEL SOFTWARE**

**CURSO: MINERÍA DE DATOS**

**CÓDIGO: ISW-911**

**NATURALEZA DEL CURSO: TEÓRICO-PRÁCTICO**

**NIVEL: IX**

**CRÉDITOS: 3**

**HORAS PRESENCIALES / SEMANA: 5 (1 HORAS TEORÍA, 4 HORAS PRÁCTICA)**

**TIEMPO DE ESTUDIO INDEPENDIENTE POR SEMANA: 04 horas, 39 minutos**

**MODALIDAD: CUATRIMESTRAL**

**REQUISITOS: ISW-811**

**CO-REQUISITOS: NINGUNO**

**I. DESCRIPCIÓN DEL CURSO**

Este curso está orientado para estudiantes del IX nivel de Carrera. En él, se presentan las técnicas de Almacenes de Datos y OLAP para facilitar el procesamiento analítico de apoyo a la toma de decisiones estratégicas.

En segundo lugar, y con más extensión, se presenta la necesidad del análisis inteligente y automático de la información para el descubrimiento del conocimiento útil.

Para abordarlo, se presentarán las técnicas de aprendizaje automático más habituales en minería de datos. Se estudiará la idoneidad de cada una de ellas para diferentes problemas.

La metodología que se sigue para el desarrollo de los contenidos temáticos del curso es variada, utilizando diferentes técnicas didácticas que faciliten el proceso enseñanza – aprendizaje; con ello se busca una participación activa por parte del estudiante, en constante interacción con su medio y los recursos disponibles en él. En lo que respecta a la evaluación, se hará de ésta, un proceso de aprendizaje, significativo, donde el estudiante logre encontrarle funcionalidad con su diario quehacer.

**II. OBJETIVO GENERAL**

* Aplicar las técnicas de minería de datos apropiadas para problemas concretos de extracción de conocimiento útil para el análisis o la toma de decisiones.

**III. OBJETIVOS ESPECÍFICOS**

1. Reconocer el potencial y la problemática del análisis de sistemas de información para la toma de decisiones.
2. Conocer las ventajas y casos donde es aconsejable recopilar información interna y externa en un Almacén de Datos.
3. Conocer el modelo multidimensional de datos de los almacenes de datos y los operadores de refinamiento asociados: drill, roll, slice & dice, pivot.
4. Conocer la arquitectura y diferentes implementaciones (ROLAP, MOLAP) de Almacenes de Datos.
5. Reconocer pautas para el diseño y mantenimiento de ADs.
6. Reconocer los beneficios de su uso sistemático de técnicas de extracción de conocimiento para la obtención de modelos y patrones predictivos o descriptivos.
7. Conocer las fases del Descubrimiento de Conocimiento de Bases de Datos y la importancia de las mismas en el éxito del proceso (en especial las de limpieza y selección de datos).

**IV. CONTENIDOS**

**UNIDAD I. Introducción**

* Finalidades y Evolución de los Sistemas de Información.
* Herramientas para la Toma de Decisiones: diferencias e interrelación.
* Almacenes de Datos, OLAP y Minería de Datos: definición e interrelación.

**UNIDAD II. Almacenes de Datos**

* Introducción a los almacenes de datos: motivación definición y características.
* Arquitectura de un sistema de almacén de datos.
* Explotación de un almacén de datos: herramientas OLAP.
* Sistemas ROLAP y MOLAP.
* Carga y Mantenimiento de un Almacén de Datos.
* Diseño de un almacén de datos.
* Líneas de investigación abiertas.

**UNIDAD III. Minería de Datos**

* Introducción a la Minería de Datos (DM)
  + Motivación
  + Problemas tipo y aplicaciones
* Relación de DM con otras disciplinas
  + El proceso de KDD
  + Las Fases del KDD
* Tipología de Técnicas de Minería de Datos
  + Sistemas Comerciales
  + Visualización
  + Técnicas de Minería de Datos
  + El Problema de la Extracción Automática de Conocimiento.
  + Evaluación de Hipótesis
  + Técnicas no supervisadas y descriptivas.
  + Técnicas supervisadas y predictivas.
* Web Mining
  + Los Problemas de la Información No Estructurada.
  + Extracción de Conocimiento a partir de Documentos HTML y texto.
  + Extracción de Información semi-estructurada (XML).
  + Líneas de Investigación Abiertas

**V. METODOLOGÍA**

Clases participativas donde el estudiante debe crear, investigar y evacuar dudas durante el curso, así mismo comprender aspectos no relacionados directamente con el área de estudio pero que le serán útiles en su trabajo diario conocerlos.

**VI. EVALUACIÓN**

Este curso se evalúa así:

|  |  |
| --- | --- |
| Proyecto | 40% |
| Examen 1 | 15% |
| Examen 2 | 20% |
| Exposición | 15% |
| Tareas | 5% |
| Quices | 5% |
| **Total** | **100%** |

**VII. BIBLIOGRAFÍA**

*Inmon, W.H. "Building the Data Warehouse", John Wiley, 1992*

*Inmon, W.H. et al. "Managing the Data Warehouse", John Wiley, 1997*

*Inmon, W.H. et al. "Data Warehouse Performance", John Wiley, 1999*

*Kimball, R. "The Data Warehouse Toolkit", John Wiley, 1996*

*Kimball, R et al. "The Data Warehouse Lifecycle Toolkit", John Wiley, 1998*

*Giovinazzo, W. "Object-Oriented Data Warehouse Design", Prentice-Hall, 2000.*

*Jarke, M. et al. "Fundamentals of Data Warehouses", Springer, 2000.*

**VIII. CRONOGRAMA**

|  |  |  |  |
| --- | --- | --- | --- |
| **Clase** | **Fecha** | **Unidad** | **Actividades** |
| 1 | 10/09/18 | **Introducción** | 1. Bienvenida al curso. 2. Presentación de los alumnos. 3. Lectura de Programa, Evaluación y Cronograma. 4. Presentación finalidades y Evolución de los Sistemas de Información. 5. Presentación herramientas para la Toma de Decisiones: diferencias e interrelación. 6. Presentación almacenes de Datos, OLAP y Minería de Datos: definición e interrelación. |
| 2 | 17/10/18 | **Almacenes de Datos** | 1. Introducción a los almacenes de datos: motivación definición y características. 2. Arquitectura de un sistema de almacén de datos. 3. Explotación de un almacén de datos: herramientas OLAP. 4. Sistemas ROLAP y MOLAP. |
| 3 | 24/09/18 | **Almacenes de Datos** | 1. Carga y Mantenimiento de un Almacén de Datos. 2. Diseño de un almacén de datos. 3. Líneas de investigación abiertas. |
| 4 | 01/10/18 | **Almacenes de Datos** | 1. Practica |
| 5 | 08/10/18 | **Minería de Datos** | 1. Introducción a la Minería de Datos (DM)    * Motivación    * Problemas tipo y aplicaciones |
| 6 | 15/10/18 | **Minería de Datos** | 1. Relación de DM con otras disciplinas    * El proceso de KDD    * Las Fases del KDD |
| 7 | 22/10/18 | **Examen 1** |  |
| 8 | 29/10/18 | **Minería de Datos** | 1. Tipología de Técnicas de Minería de Datos    * Sistemas Comerciales    * Visualización    * Técnicas de Minería de Datos    * El Problema de la Extracción Automática de Conocimiento. |
| 9 | 05/11/18 | **Minería de Datos** | * + Evaluación de Hipótesis   + Técnicas no supervisadas y descriptivas.   + Técnicas supervisadas y predictivas. |
| 10 | 12/11/18 | **Minería de Datos** | 1. Web Mining    * Los Problemas de la Información No Estructurada. |
| 11 | 19/11/18 | **Minería de Datos** | * + Extracción de Conocimiento a partir de Documentos HTML y texto. |
| 12 | 26/11/18 | **Minería de Datos** | * + Extracción de Información semi-estructurada (XML).   + Líneas de Investigación Abiertas |
| 13 | 3/12/18 | **Examen 2** |  |
| 14 | 10/12/18 | **Defensa Proyecto** |  |