

Nombre de la asignatura: **MINERÍA DE DATOS**

Línea de Trabajo: **Inteligencia Artificial**

Tiempo de dedicación del estudiante a las actividades de:

DOC	TIS	TPS	Horas Totales	Créditos
48	20	100	168	6

**DOC:** Docencia; **TIS:** Trabajo independiente significativo; **TPS:** Trabajo profesional supervisado

### 1. Historial de la asignatura

Fecha revisión/actualización	Participantes	Observaciones, cambios o justificación
15/Diciembre/2023	Dr. Juan Antonio Vargas Enríquez Dr. Jesus Carlos Carmona Frausto	Versión inicial, actualización incluyendo el uso de IA.

### 2. Pre-requisitos y Correquisitos

Ninguno

### 3. Objetivo de la asignatura

Permitir comprender de una manera más contextualizada y precisa los comportamientos de los grandes volúmenes de información que son generados y conocer algunas técnicas para obtener valor agregado de esta información.

### 4. Aportación al perfil del graduado

El alumno será capaz de:

- Conocer los conceptos básicos de los algoritmos representativos de Minería de Datos.
- Conocer las técnicas de evaluación de algoritmos representativos de la Minería de Datos.
- Aplicación de algoritmos representativos de Minería de Datos a problemas reales.
- Comprender los fundamentos del clasificador de árbol de decisión.
- Establecer los conjuntos de entrenamiento, la complejidad del modelo, la selección del modelo y los errores comunes en el modelo y en su evaluación.
- Aplicar los fundamentos de redes bayesianas, de vector de soporte, de máquinas y de redes neuronales artificiales en la solución de problemas reales de clasificación.

### 5. Contenido temático

Unidad	Temas	Subtemas
1	Introducción.	1. Introducción 2. ¿Qué es la minería de datos? 3. Retos Motivadores 4. Los orígenes de la minería de datos 5. Tareas de minería de datos
2	Datos	1. Tipos de datos 2. Calidad de los datos 3. Procesamiento de los datos 4. Medidas de similitud y disimilitud
3	Análisis exploratorio	1. Introducción al análisis exploratorio 2. Ejercicio de análisis exploratorio

4	Clasificación: conceptos básicos y técnicas	<ol style="list-style-type: none"> <li>1. Conceptos básicos</li> <li>2. Marco general de clasificación</li> <li>3. Clasificador de árbol de decisión</li> <li>4. Sobreajuste del modelo</li> <li>5. Selección de modelo</li> <li>6. Evaluación del modelo</li> <li>7. Presencia de hiperparámetros</li> <li>8. Errores de la selección y evaluación de modelos</li> <li>9. Comparación de modelos</li> </ol>
5	Técnicas avanzadas de clasificación	<ol style="list-style-type: none"> <li>1. Tipos de clasificadores</li> <li>2. Clasificador basado en reglas</li> <li>3. Clasificadores de vecinos más cercanos</li> <li>4. Clasificador ingenuo de Bayes</li> <li>5. Redes bayesianas</li> <li>6. Regresión logística</li> <li>7. Red neuronal artificial (RNA)</li> <li>8. Aprendizaje profundo</li> <li>9. Máquina de vectores de soporte (SVM)</li> <li>10. Métodos de conjunto</li> <li>11. Problema de desequilibrio de clases</li> <li>12. Problema multiclase</li> </ol>
6	Análisis de asociación: conceptos básicos y algoritmos	<ol style="list-style-type: none"> <li>1. Preliminares</li> <li>2. Generación frecuente de conjuntos de elementos</li> <li>3. Generación de reglas</li> <li>4. Representación compacta de conjuntos de elementos frecuentes</li> <li>5. Métodos alternativos para generar conjuntos de elementos frecuentes*</li> <li>6. Algoritmo de crecimiento FP*</li> <li>7. Evaluación de patrones de asociación</li> <li>8. Efecto de la distribución sesgada del apoyo</li> </ol>
7	Análisis de asociación: conceptos avanzados	<ol style="list-style-type: none"> <li>1. Manejo de atributos categóricos</li> <li>2. Manejo de atributos continuos</li> <li>3. Manejo de una jerarquía de conceptos</li> <li>4. Patrones secuenciales</li> <li>5. Patrones de subgrafos</li> <li>6. Patrones poco frecuentes</li> </ol>
8	Análisis de grupos: conceptos y algoritmos	<ol style="list-style-type: none"> <li>1. Descripción general</li> <li>2. K-means</li> <li>3. Agrupación jerárquica aglomerativa</li> <li>4. DBSCAN</li> <li>5. Evaluación de conglomerados</li> </ol>
9	Análisis de grupos: conceptos avanzados	<ol style="list-style-type: none"> <li>1. Características de datos, conglomerados y algoritmos de agrupamiento</li> <li>2. Agrupación basada en prototipos</li> <li>3. Agrupación basada en densidad</li> <li>4. Agrupación basada en gráficos</li> <li>5. Algoritmos de agrupación escalables</li> <li>6. ¿Qué algoritmo de agrupación?</li> </ol>
10	Detección de anomalías	<ol style="list-style-type: none"> <li>1. Características de los problemas de detección de anomalías</li> <li>2. Características de los métodos de detección de anomalías</li> <li>3. Enfoques estadísticos</li> <li>4. Enfoques basados en la proximidad</li> <li>5. Enfoques basados en agrupaciones</li> <li>6. Enfoques basados en la reconstrucción</li> </ol>

		7. Clasificación de una clase 8. Enfoques teóricos de la información 9. Evaluación de la detección de anomalías
--	--	---

**6. Metodología de desarrollo del curso**

- Clases impartidas por el profesor.
- Examen de medio término
- Examen final
- Realización de prácticas para dar solución a problemas reales que involucren:
  - ✓ Diseño de algoritmos de Minería de Datos.
  - ✓ Establecimiento de conjuntos de entrenamiento, complejidad, selección, errores y evaluación del modelo.
  - ✓ Uso de redes bayesianas, de vector de soporte, de máquinas y de redes neuronales artificiales.

**7. Sugerencias de evaluación**

Se recomienda la siguiente ponderación:

Examen de medio término	20%
Examen Final	20%
Ejercicios prácticos	60%

**8. Bibliografía y software de Apoyo**

- "Introduction to data mining", 2<sup>nd</sup> edition, Pang-Ning, T., Michael, S., Vipin, K., & Anuj, K., 2016, Pearson Education India, 1459 pages.

**9. Actividades propuestas**

Unidad	Actividad
1,2	Investigación y exposición sobre los conceptos básicos de Minería de Datos y tipos de datos
3,4,5,6,7,8,9,10	Resolver ejercicios prácticos de diseño de algoritmos de Minería de Datos, conjuntos de entrenamiento, complejidad, selección, errores y evaluación de modelos, redes bayesianas, vector de soporte, máquinas y de redes neuronales artificiales. Conocer las capacidades y limitaciones de agentes de inteligencia artificial generadores de código para generar cada uno de estos modelos.

**10. Catedrático (s) responsable (s)**



Dr. Juan Antonio Vargas Enríquez



Dr. Jesús Carlos Carmona Frausto