



**UNIVERSIDAD NACIONAL DE INGENIERÍA
FACULTAD DE CIENCIAS
ESCUELA PROFESIONAL DE CIENCIA DE LA COMPUTACIÓN**

SÍLABO

INFORMACIÓN GENERAL

ASIGNATURA	:	MINERÍA DE DATOS
CÓDIGO	:	CC442
CRÉDITOS	:	04 (CUATRO)
PRE-REQUISITO	:	CC721 - INTELIGENCIA ARTIFICIAL
CONDICIÓN	:	OBLIGATORIO
HORAS POR SEMANA	:	06 (TEORÍA: 02, LABORATORIO: 04)
SISTEMA DE EVALUACIÓN	:	G

SUMILLA

La asignatura se ubica en la intensificación de Computación, en el ámbito de las asignaturas de la Inteligencia Artificial. La Minería de datos, y el aprendizaje automático en general, están ligados al campo de la estadística y de la algorítmica, y abordan las técnicas para la extracción de conocimiento implícito en conjuntos de datos.

La minería de datos es una herramienta poderosa para el proceso de descubrir nuevas correlaciones, patrones y tendencias mediante el análisis de grandes cantidades de datos almacenados en repositorios. En el curso, los estudiantes aprenden a aplicar los principios de data mining para manejar y analizar conjuntos grandes y complejos de datos, incluyendo aquellos que se encuentran en la web.

COMPETENCIAS

- Detección, interpretación y predicción de patrones cuantitativos y cualitativos en los datos.
- Proceso de extraer información o patrones interesantes (no triviales, implícitos, previamente desconocidos y potencialmente útiles) desde grandes repositorios.
- Comprender los conocimientos base de datos y los principios de la ciencia de la computación.
- Decidir ante un problema práctico concreto qué tarea de minería de datos conviene utilizar, qué modelo se quiere obtener, qué técnica resultaría más adecuada de utilizar y cómo evaluar los resultados obtenidos.

PROGRAMA ANALÍTICO

1. **CAPÍTULO 1:** Introducción a la minería de datos, Objetivos de minería de datos, Etapas del proceso de minería de datos.
2. **CAPÍTULO 2:** Reducción de Dimensión: Análisis de Componentes Principales (PCA), Análisis de Componentes Principales *Regularizados* (SparsePCA)-- Regresión Lineal Múltiple**
3. **CAPÍTULO 4:** Relaciones entre registros: Association Rules, *Collaborative Filtering, Clustering (k-means, k-medians)*.
4. **CAPÍTULO 3:** Métodos de clasificación/Predicción: *k-nearest neighbor, Naive Bayes, CART/Random Forest, Regresión Logística, Análisis Discriminante*
5. **CAPÍTULO 5:** *Forecasting Time Series: Moving Average. Simple Exponential Smoothing*

BIBLIOGRAFÍA

1. Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. *Introduction to Data Mining*. Addison- Wesley. Longman Publishing Co. 0321321367. 2005
2. Witten, Frank & Hall. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan & Kauffmann. 978-0-12-374856-0. 2011
3. Xindong Wu, Vipin Kumar *The Top Ten Algorithms in Data Mining* Chapman and Hall/CRC. 9781420089646. 2009.
4. Shmueli G., Bruce P., Gedeck P., Patel N. *Data Mining for Business Analytics* concepts, techniques and applications in Python. John Wiley & Son, 2020
5. D. Hand, P. Smyth, H. Mannila. *Principles of Data Mining*. MIT press. 2001
6. D. Pyle. *Data preparation for data mining*. Morgan Kauffmann. 1999
7. Ian H. Witten y Eibe Frank *Data Mining: Practical machine tools and techniques*. Elsevier. 2005.
8. Jiawey Han y Micheline Kamber *Data Mining: Concepts and techniques*. Morgan Kauffmann. 2006..

SEMANA	UNIDAD	CALENDARIO DE TEMAS/EVALUACIONES
1	Introducción a la minería de datos, Objetivos de minería de datos, Etapas del proceso de minería de datos	Clases (Evaluación/ Prueba Entrada)
2	Técnicas de Reducción de Dimensión. Análisis de Componentes Principales (PCA)	Clases
3	Algoritmos de Predicción. Regresión Lineal Múltiple: Análisis de Componentes Principales Sparse (SparsePCA)	Clases +/1er Laboratorio Dirigido
4	Análisis de Conglomerados. K-means, k-medians	Clases + /1er Laboratorio Calificado
5	Algoritmos de Asociación. Reglas de Asociación	Clases + /2do laboratorio Dirigido
6	Algoritmos de Asociación. Collaborative Filtering	Clases + /2do Laboratorio Dirigido
7	Algoritmos de clasificación/Predicción. Métodos nearest neighbor	Clases + /2do Laboratorio Calificado
8		Examen Parcial
9	Algoritmos de clasificación/Predicción. Clasificador Naive Bayes	Clases + /3er Laboratorio Dirigido
10	Algoritmos de clasificación/Predicción. Análisis Discriminante	Clase +/ 3er Laboratorio Calificado
11	Algoritmos de clasificación/Predicción. árboles de Regresión/Random Forest	Clases +/ 4to Laboratorio Dirigido
12	Algoritmos de clasificación/Predicción Regresión Logística	Clases +/ 4to Laboratorio Calificado
13	Forecasting Time Series.	Clases +/ 5to Laboratorio Dirigido
14	Forecasting Time Series. Métodos de Suavización: <i>Simple Exponential Smoothing</i>	Clases +/ 5to Laboratorio Dirigido
15		Clases +/ 5to Laboratorio Calificado
16		Examen Final
17		
18		Examen Sustitutorio