

Plan de curso Minería de datos

PLANIFICACIÓN DE CURSO

Primer Semestre académico 2025 - Docencia Presencial

I. ACTIVIDAD CURRICULAR Y CARGA HORARIA

Asignatura:	Minería de datos	Código:	COM4001-1
Semestre de la Carrera:	Otoño		
Carrera:	Ingeniería Civil en Computación		
Escuela:	Instituto de Ciencias de la Ingeniería		
Docente(s):	Luis Cossio		
Ayudante(s):	Gabriel Díaz V.		
Horario:	Catedra Miércoles 8:30-10:00 y Jueves 12:00-13:30. Ayudantías 12:00-13:30 Viernes		

Créditos SCT:	6
Carga horaria semestral ¹ :	150 horas
Carga horaria semanal:	10 horas

Tiempo de trabajo sincrónico semanal:	70 horas
Tiempo de trabajo asincrónico semanal:	110 horas

II. RESULTADOS U OBJETIVOS DE APRENDIZAJE ESPERADOS ESTE SEMESTRE

1)	Construye y valida hipótesis de generación de información, a través de la aplicación de un conjunto de técnicas básicas de minería de datos.
2)	Interpreta los resultados del proceso de minería de datos en relación a las hipótesis planteadas, distinguiendo patrones novedosos y significativos de información.
3)	Manejo de herramientas computacionales para el análisis de datos

¹ Considere que 1 crédito SCT equivale a 30 horas de trabajo total (presencial/sincrónico y autónomo/asincrónico) en el semestre.

4)

III. UNIDADES, CONTENIDOS Y ACTIVIDADES

UNIDAD 1: <i>Proceso y Análisis de Bases de Datos</i>				
Semana	Contenidos	Actividades de enseñanza y aprendizaje		Actividades de evaluación diagnóstica, formativa y/o sumativa
		Tiempo sincrónico	Tiempo asincrónico (trabajo autónomo del o la estudiante)	
1	<i>Introducción a la Minería de Datos</i>	4,5	5,5	1. Introducción a conceptos de datos y terminologías 2. Ayudantía: Repaso de python como herramienta para el semestre
2	Estadística, varianza y correlación	4,5	5,5	1. Estadística y distribución de datos. 2. Introducción a la correlación de variables 3. Ayudantía: Exploración y visualización de bases de datos

UNIDAD 2: <i>Análisis Supervisado</i>				
Semana	Contenidos	Actividades de enseñanza y aprendizaje		Actividades de evaluación diagnóstica, formativa y/o sumativa
		Tiempo sincrónico	Tiempo asincrónico (trabajo autónomo del o la estudiante)	
3	<i>Regresión: Predictores lineales, polinomiales, discretos</i>	4,5	5,5	1. Regresión lineal, no lineal y métricas de error 2. Ayudantía: Estadística y correlación 3. Tarea 1: Exploración y Correlación de datos.
4	Clasificadores (árboles de decisión, KNN), preprocesamiento métricas de evaluación	4,5	5,5	1. Clasificadores árboles de Decisión 2. Clasificadores KNN 3. Ayudantía: Regresión

5	Metodologías clásicas de clasificación (naive bayes, decision tree, SVM)	4,5	5,5	<ol style="list-style-type: none"> 1. Clasificador Naive Bayes 2. Clasificador SVM 3. Tarea 2: Regresión y clasificación
5,5	Evaluación del desempeño de clasificadores (bootstrap, cross validation) y métodos ensemble	4,5	5,5	<ol style="list-style-type: none"> 1. Métricas de clasificación (conjuntos no balanceados) 2. Técnicas de entrenamiento y evaluación (Bootstrap y cross-validation) 3. Ayudantía: Clasificadores Naive Bayes y SVM.
7	Redes neuronales para clasificación (fully connected)	4,5	5,5	<ol style="list-style-type: none"> 1. Redes neuronales: Concepto y operaciones básicas en redes Fully Connected. 2. Ayudantía: Evaluación de desempeño (Bootstrap, splits de conjuntos y buenas prácticas)
8	Redes convolucionales	4,5	5,5	<ol style="list-style-type: none"> 1. Introducción al procesamiento de imágenes 2. Procesamiento de imágenes con redes convolucionales. 3. Tarea 3: Redes neuronales fully connected

UNIDAD 3: <i>Análisis no-supervisado</i>				
Semana	Contenidos	Actividades de enseñanza y aprendizaje		Actividades de evaluación diagnóstica, formativa y/o sumativa
		Tiempo sincrónico	Tiempo asincrónico (trabajo autónomo del o la estudiante)	
9	<i>Encoding y Reducción de dimensionalidad (PCA y espacio latente)</i>	4,5	5,5	<ol style="list-style-type: none"> 1. Reducción de dimensionalidad 2. Ayudantía: Redes convolucionales

10	Aprendizaje no supervisado (KNN)	4,5	5,5	<ol style="list-style-type: none"> 1. Aprendizaje No supervisado Clusters (KNN's) 2. Ayudantía: PCA
11	Aprendizaje no supervisado (DBSCAN, hierarchical clustering)	4,5	5,5	<ol style="list-style-type: none"> 1. Aprendizaje No supervisado Clusters (KNN's y Hierarchical clustering. 2. Métricas en algoritmos no supervisados. 3. Tarea 4: Control 2: Aprendizaje No supervisado, Redes neuronales
12	Más en redes neuronales	4,5	5,5	<ol style="list-style-type: none"> 1. Redes neuronales recursivas Ayudantía 2. Ayudantía: Comparación entre algoritmos no supervisados.
13	Repaso	4,5	5,5	

IV. CONDICIONES Y POLÍTICAS DE EVALUACIÓN

- Se realizarán 4 tareas
 - Tarea 1: 07/04/25
 - Tarea 2: 05/04/25
 - Tarea 3: 02/06/25
 - Tarea 4: 23/06/25
- Se realizaran 2 controles
 - Control 1: 21/04/25
 - Control 2: 16/06/25
- Se realizaran un examen
 - Examen: 14/07/25

Condiciones de evaluaciones:

- Los controles de cátedra serán evaluados de manera presencial, al igual que el examen.
- Controles y tareas deben ser aprobadas por separado (Nota mayor o igual a 4.0).
- La Nota de Controles (NC) está definida por los dos Controles de Cátedra (CC) y el Examen (E). El promedio simple de los dos Controles de Cátedra define la Nota de Presentación (NP), que permite eximirse del examen en caso de ser mayor o igual a 5.5. La Nota de Controles está definida según la fórmula:
 $NC=50\%NP+50\%E.$
 - El examen no reemplaza el peor Control de Cátedra.
- En caso de aprobarse el curso (es decir, si tanto la Nota de Controles como la Nota de tareas son ambas mayores o iguales a 4.0), la Nota Final (NF) se calcula siguiendo la fórmula:
 $NF=60\%NC+40\%AC.$

V. BIBLIOGRAFÍA Y RECURSOS OBLIGATORIOS

P. Tan, M. Steinbach, V. Kumar. Introduction to Data Mining. Addison-Wesley. 2006

Ian H. Witten, Eibe Frank, Mark A. Hall. Data Mining: Practical Machine Learning Tools and Techniques, Third Edition (Morgan Kaufmann Series in Data Management Systems). 2011

An Introduction to Statistical Learning with applications in python. <https://www.statlearning.com/> . 2021

VI. BIBLIOGRAFÍA Y RECURSOS COMPLEMENTARIOS

Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition.
O'Reilly Media, Inc. 2019