

PROGRAMA DE CURSO

Código	Nombre			
IN6531	INTRODUCCIÓN A LA MINERÍA DE DATOS			
Nombre en Inglés				
Introduction to Data Mining				
Créditos	Unidades Docentes	Horas de Cátedra	Horas Docencia Auxiliar	Horas de Trabajo Personal
6	10	3	1,5	5,5
Requisitos			Carácter del Curso	
MA3403 Probabilidades y Estadísticas			Electivo para la Carrera de Ingeniería Civil Industrial	
Resultados de Aprendizaje				
Al finalizar el curso, los estudiantes:				
<ol style="list-style-type: none"> 1. Conocen modelos de Bases de Datos, Bodegas de Datos, OLAP, entendiendo sus usos en aplicaciones en las organizaciones. 2. Entienden los principales objetivos y métodos de Data Mining. 3. Conocen aplicaciones típicas de Data Mining y tiene nociones sobre el uso de herramientas computacionales para la solución y aplicación de distintos métodos de Minería de Datos. 4. Tienen el conocimiento y la capacidad para elegir el método adecuado en una aplicación real y presentar una solución a dicha problemática. 				

Metodología Docente	Evaluación General
<p>Los estudiantes, en forma adicional a las clases de cátedra, a lo largo del semestre desarrollarán seis laboratorios, un proyecto de minería de datos, CTP's y un control.</p> <p>Todos los laboratorios se desarrollarán con la herramienta Python.</p> <p>Laboratorio 1: Python El objetivo de este laboratorio es introducir a los estudiantes en la herramienta Python que utilizarán a lo largo del semestre.</p> <p>Laboratorio 2: Preprocesamiento de datos El objetivo de esta laboratorio es entender la necesidad del preprocesamiento y la transformación de los datos y aplicar algunas técnicas básicas para éste.</p> <p>Laboratorio 3: Clustering El objetivo de este laboratorio es adquirir el conocimiento para aplicar técnicas de</p>	<p>Los laboratorios y el proyecto semestral deberán ser realizadas en grupos de tres personas.</p> <p>La nota proyecto semestral se calcula de la siguiente manera: Nota proyecto semestral = 0.6*Nota Informe Final + 0.1*Entrega I + 0.1*Entrega II + 0.2*Video</p> <p>En conjunto con el proyecto se realizarán CTPs.</p> <p>Al final del semestre todos los estudiantes rendirán un control.</p> <p>Para aprobar el curso se necesita lo siguiente: Nota control ≥ 4.0 Nota promedio laboratorios ≥ 4.0 Nota proyecto semestral ≥ 4.0 Nota CTP ≥ 5.0</p>

segmentación. Haciendo un análisis crítico de los resultados de los modelos con el fin de hallar el número óptimo de segmentos.

Laboratorio 4: Clasificación

El objetivo de este laboratorio es adquirir el conocimiento necesario para la aplicación de técnicas de clasificación tales como Regresión Logística, Árboles de Decisión y Support Vector Machine. Haciendo evaluación de los resultados de los modelos a través de métricas de desempeño tales como Recall, Precision, Accuracy y AUC, con el fin de obtener una mejor solución.

Laboratorio 5: Regresión

El objetivo de este laboratorio es conocer y aplicar distintas técnicas de regresión y evaluar los resultados de los modelos a través de métricas de desempeño tales como MAE, RMSE Y R-cuadrado, con el fin de obtener una mejor solución.

Laboratorio 6: Text Mining

El objetivo de este laboratorio es que los estudiantes conozcan algunas técnicas para llevar el texto de un documento a un vector para así poder aplicar técnicas de Minería de Datos. Además, adquirir el conocimiento necesario para la aplicación de técnicas de modelado de tópicos (ej: Latent Dirichlet Allocation).

Proyecto Semestral

El objetivo es que los estudiantes resuelvan un problema real desde la fase de la definición del problema hasta la presentación final de la solución. La idea de este proyecto es la simulación de un proyecto real, teniendo la capacidad de elegir el mejor método para la resolución del problema.

La nota final del curso se determina en la siguiente manera:

Nota control	20%
Nota promedio laboratorios	40%
Nota proyecto semestral	20%
Nota CTP	20%

UNIDADES TEMÁTICAS

Número	Nombre de la Unidad	Duración en Semanas
1	INTRODUCCIÓN Y MOTIVACIÓN	1.0
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1. Introducción al curso. 2. Presentación externa: un profesional del área presenta su experiencia trabajando con los métodos y conceptos del curso.	Los estudiantes: Reconocen el impacto que el contenido de este curso puede tener en empresas privadas y organizaciones públicas.	

Número	Nombre de la Unidad	Duración en Semanas
2	CONCEPTOS PARA GUARDAR Y ACCEDER A LOS DATOS	1.0
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1. Revisión del modelo de Bases de Datos, y la evolución de las tecnologías asociadas a Bases de Datos, dando nociones básicas sobre Data Warehouse, Data Marts y OLAP.	Los estudiantes: Son capaces de construir una base de datos simple y manejan las técnicas básicas para la manipulación de datos.	

Número	Nombre de la Unidad	Duración en Semanas
3	PROCESO KDD	1.0
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1. Introducción al proceso KDD (Knowledge Discovery Databases) y su relación con la Minería de Datos. Diferentes formas de aprendizaje automático.	Los estudiantes: Adquieren una vista general del proceso para analizar datos y encontrar información.	Shmueli et al. (2010)

Número	Nombre de la Unidad	Duración en Semanas
4	CALIDAD DE DATOS	1.0
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1. Los diferentes conceptos de errores en los datos disponibles.	Los estudiantes: Son capaces de reconocer errores en los datos y manejar las herramientas correspondientes.	

Número	Nombre de la Unidad	Duración en Semanas
5	SELECCIÓN DE DATOS	1.0
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1. Métodos para selección de atributos	Los estudiantes: Reconocen la importancia de la selección de atributos. Son capaces de identificar los métodos más indicados para la selección de atributos y maneja las herramientas correspondientes.	

Número	Nombre de la Unidad	Duración en Semanas
6	TRANSFORMACIÓN DE DATOS	1.0
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1. Conceptos para la transformación de los datos	Los estudiantes: Adquieren experiencia en la transformación de datos y reconoce su impacto en los resultados del posterior análisis.	Shmueli et al. (2010)

Número	Nombre de la Unidad	Duración en Semanas
7	MÉTODOS DE LA MINERÍA DE DATOS	3.0
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1. Regresión logística, árboles de decisión, métodos de clustering, redes neuronales, Support Vector Machines.	Los estudiantes: Son capaces de identificar el método más adecuado en una tarea dada y tienen experiencia con el manejo de herramientas computacionales.	Shmueli et al. (2010)

Número	Nombre de la Unidad	Duración en Semanas
8	APLICACIONES	3.0
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1. Casos de segmentación de clientes, detección de fraude, análisis de riesgo, entre otros.	Los estudiantes: Adquieren experiencia con la aplicación de la minería de datos a través del análisis de diversas aplicaciones.	Shmueli et al. (2010)

Número	Nombre de la Unidad	Duración en Semanas
9	DIRECCIONES DE INVESTIGACIÓN RECIENTES	1.0
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1. Minería de datos y teoría de juegos, selección de atributos, entre otros.	Los estudiantes: Reconocen el estado del arte actual y tiene una noción de la investigación actual del área.	

Bibliografía General	
<u>Bibliografía obligatoria:</u>	
Será publicada en UCursos.	
<u>Bibliografía complementaria:</u>	
Shmueli, G., Patel, N. R., Bruce, P. C. (2010): Data Mining for Business Intelligence. 2 nd ed., John Wiley and Sons, Hoboken, New Jersey	

Vigencia desde:	Otoño 2020
Elaborado por:	Richard Weber
Aprobado por:	Comisión de Docencia DII