



UNIVERSIDAD DE BOGOTÁ
JORGE TADEO LOZANO

FACULTAD DE CIENCIAS NATURALES E INGENIERÍA
ESPECIALIZACIÓN EN DESARROLLO BASES DE DATOS

Nombre de la Asignatura	MINERÍA DE DATOS
-------------------------	------------------

Código	751105	Prerrequisitos				
Fundamentación	Específica		Actividad académica		Teórica	
No. de Créditos	2	IH (presencial)	32 horas	IH (no presencial)	64 Horas	
Fecha de actualización	Febrero 2018					

Programas que requieren el servicio	PROGRAMA
	Especialización en Desarrollo Bases de Datos

Justificación

La creciente adopción de tecnología para la automatización y control de procesos en prácticamente todas las áreas existentes, ha ocasionado una generación importante de datos heterogéneos en origen, estructura, volumen y contenido. Hoy las organizaciones son conscientes de la importancia de un manejo correcto de la información para anticiparse a la competencia y poder llegar de una manera mas atractiva y sin perturbación al cliente, mejorando el servicio, producto, o lo que ofrezca al usuario. De ahí que sea importante acudir a métodos y técnicas para encontrar información en estos datos que no es evidente pero si de gran valor. La ciencia de datos (minería de datos) provee un camino para este fin.

Objetivo general:

Al finalizar esta asignatura el alumno conocerá y aprenderá algunas de las técnicas para analizar datos provenientes de bodegas y bases de datos, utilizando para ello modelos de estadística con el fin desarrollar aplicaciones que apoyen la toma de decisiones.

Objetivos específicos

Conocer y aprender a utilizar las principales técnicas de análisis de información sobre bases de datos y en especial de bodegas de datos, tanto del punto de vista estadístico como del de la computación.

Descripción de los contenidos

- Introducción a ciencia de datos
 - Conceptos
 - Aplicaciones
 - Repaso de Álgebra Lineal y conceptos que se necesitan
- Metodología en ciencia de datos
- Análisis exploratorio
- Preparación de datos
 - Tratamiento de Valores faltantes



UNIVERSIDAD DE BOGOTÁ
JORGE TADEO LOZANO.

FACULTAD DE CIENCIAS NATURALES E INGENIERÍA
ESPECIALIZACIÓN EN DESARROLLO BASES DE DATOS

<ul style="list-style-type: none">○ Ruido○ Transformación de datos <ul style="list-style-type: none">• Analisis de Canasta - ecommerce data• Analisis de churn via Aprendizaje Supervisado Técnicas de aprendizaje supervisado• Customer Segmentation via clustering<ul style="list-style-type: none">○ Conceptos.○ Métricas de similitud, k-means, k-medoids.○ Clustering jerárquico○ Clustering incremental
Metodología
Clases magistrales Exposiciones a cargo de los estudiantes, de acuerdo a temáticas establecidas previamente. Laboratorios Los participantes desarrollan un proyecto aplicando los diferentes tópicos abordados en clase. Evaluación final
Criterios de evaluación:
<ul style="list-style-type: none">• 50 % Trabajo del alumno• 20 % Participación en clase y presentaciones• 30 % Examen o proyecto
Bibliografía básica para los estudiantes
<ol style="list-style-type: none">1. Layton Robert (2015) Learning Data Mining with Python. PACKT Publishing.2. Mc Kinney Wes (2013) Python for data analysis. Orreily3. Downey Allen B (2014). Think stats. Green Tea Press.4. Kantardzic, Mehmed. (2003). <i>Data Mining concepts, models, methods and techniques</i>. Wiley-Interscience.5. Michael J.A. Berry, Gordon S. Linoff. (2004). <i>Data Mining Techniques For Marketing, Sales, and Customer Relationship Management</i>. Second Edition. Online.



UNIVERSIDAD DE BOGOTÁ
JORGE TADEO LOZANO.

FACULTAD DE CIENCIAS NATURALES E INGENIERÍA
ESPECIALIZACIÓN EN DESARROLLO BASES DE DATOS

6. Witten, Ian H. (2011). *Data mining : practical machine learning tools and techniques* Morgan Kaufmann,

Bibliografía complementaria y digital

SITIOS WEB

- www.kdnuggets.com
- www.kdd.org
- Datacamp.com
- Citeseer.pst.edu
- **Otros**
- Knime.org