

FACULTAD DE INGENIERIA
DEPARTAMENTO DE INGENIERIA DE SISTEMAS

Nombre de la asignatura	GESTIÓN DE DATOS	
Código de la asignatura		
Fecha de Actualización		
Intensidad horaria semanal	Horas Contacto: 3	Horas Trabajo Independiente: 6
Créditos Académicos	2	
Pre-requisitos	NA	
Período Académico de Vigencia	2022-I	
Horario y Lugar	Viernes 6:00PM – 9:00PM ó Sala Virtual	

Objetivos
<p>Este curso se dirige al aprendizaje relacionado con la gestión de datos que faciliten posteriormente su análisis en entornos organizacionales. El objetivo del curso es presentar de forma integrada los diferentes conceptos, métodos y herramientas para la extracción y preparación de datos. Como objetivos específicos se encuentran los siguientes:</p> <ol style="list-style-type: none"> 1. Aprender los mecanismos de recolección y consolidación de información de distintas fuentes (bases de datos empresariales, fuentes abiertas, fuentes web)

<p>2. Realizar preparación y análisis básico de datos, incluyendo transformación y visualización, y opciones de manejo de datos perdidos y atípicos</p> <p>3. Comprender las implicaciones del manejo de grandes volúmenes de datos y cómo las tecnologías de Big data contribuyen a su manejo apropiado</p>
<p>Habilidades que contribuye a desarrollar</p>
<p>Temáticas:</p> <ul style="list-style-type: none"> • Consolidación de información en datos estructurados • Consolidación de información en datos no estructurados • Análisis básico de datos • Preparación de datos <p>Formación Integral:</p> <ul style="list-style-type: none"> • Trabajo en grupo • Investigación guiada • Expresión escrita • Trabajo autónomo
<p>Estrategias Pedagógicas</p>
<p>Aprendizaje directivo: el profesor realizará exposición de los temas para explicitar, complementar y clarificar los temas de la asignatura. Cuando el tema lo permita se harán ejemplos prácticos sobre cada tema.</p> <p>Auto-aprendizaje: Los estudiantes deberán preparar previamente la clase mediante la lectura crítica y activa de los temas del día. Así mismo, desarrollarán lecturas posteriores de profundización.</p> <p>Componente práctico: Los contenidos del curso se irán trabajando al mismo tiempo con talleres y ejemplos prácticos; adicionalmente, se presentan ejemplos empleando software especializado. Grupalmente, fuera del salón, los estudiantes realizarán proyectos o talleres cuyo enunciado es dado por el profesor.</p>

Evaluación	
<p>La asignatura se evaluará de la siguiente manera:</p> <ul style="list-style-type: none"> • 2 evaluaciones parciales de 20% cada una • proyectos-talleres con valor de 60% 	
BIBLIOGRAFIA	
REFERENCIAS BIBLIOGRAFICAS	<ol style="list-style-type: none"> 1. Feature engineering for Machine Learning. Zheng Alice, Casari Amanda. OReilly 2018 2. Data Mining and Predictive Analytics. Larose Daniel, Larose Chantal. Wiley, 2015. 3. Data Preparation for Data Mining, Dorian Pyle, Morgan Kaufmann Publishers, 1999 4. Practical applications of data mining Suh, Sang C. Jones & Bartlett Learning, 2012. 5. Introducción a la minería de datos, José Hernández Orallo, Prentice Hall, 2005 6. Data integration blueprint and modeling techniques for a scalable and sustainable architecture, Anthony Giordano, IBM Press Pearson, 2011 7. Introduction to Information Retrieval, Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Cambridge University Press. 2009.

Contenidos Temáticos			
TEMA 1: Análisis básico de datos			
TEMA 2: Preparación de datos			
TEMA 3: Consolidación de información en datos estructurados			
TEMA 4: Consolidación de información en datos no estructurados			
TEMA 5: Manejo de grandes volúmenes de datos			
Sem	Temas	Actividades en Clase	Actividades de Trabajo Independiente para la Siguiente Clase
1	Proceso de extracción. Tipos de fuentes de información. Extracción de DBMS y CSV	Ejercicios de Extracción	Taller ETL Básico
2	Extracción de JSON, XML Transformaciones básicas	Ejercicios de Extracción	Taller ETL Intermedio Preparación Parcial 1
3	REST APIs	Ejercicios de Extracción Parcial 1	
4	Introducción a la preparación de datos Exploración y selección de datos	Ejercicios exploración de datos	Lectura detección de datos inconsistentes Taller parte 1 selección de datos y exploración inicial
5	Limpieza de datos	Ejercicios limpieza de datos	Lectura Vista Minable. Taller parte 2 exploración con conclusiones y limpieza de datos
6	Preparación vista minable Conclusiones vista minable	Ejercicios vista minable	Preparación parcial Taller parte 3 Generación de Vista minable

7	Introducción al análisis de datos no estructurados	Parcial 2	Entrega de Taller (un día antes por la plataforma) y presentación durante la clase.
8	Introducción a Big Data.		Lectura Data Lakes Taller ETL avanzado
9	Mongo DB / Elasticsearch Streaming	Taller en clase	Entrega Taller