

DATOS DE IDENTIFICACIÓN			
Nombre de la asignatura		Ingeniería de Características	
Unidad Regional		Centro	
División		Ciencias Exactas y Naturales	
Departamento		Matemáticas	
Programa		Maestría en Ciencia de Datos	
Carácter		Obligatorio (X) Optativo ()	
Horas teoría	3	Horas práctica	4
Valor en créditos		10	
OBJETIVO GENERAL			
<p>Conocer y aplicar las herramientas necesarias para mejorar la confianza de las organizaciones en sus datos garantizando la calidad (exactitud, integridad, consistencia, uniformidad, densidad y unicidad) de los mismos. Para esto, el estudiante debe contar con la capacidad y entrenamiento necesario para descubrir, corregir o eliminar registros de datos erróneos de una o varias fuentes de información.</p>			
OBJETIVOS ESPECÍFICOS			
<ol style="list-style-type: none"> 1. Conocer y aplicar los métodos de detección de datos faltantes 2. Conocer y aplicar la detección de outlier e inliers, así como su tratamiento 3. Aprender las técnicas y herramientas software para la integración de los datos 4. Conocer y aplicar los métodos de transformación y reducción de datos 			
CONTENIDO SINTÉTICO			
Orden	Tema		
1	Técnicas y herramientas para limpieza de datos: inconsistencias, datos faltantes, outliers e inliers.		
2	Técnicas y herramientas para integración de datos: fuentes de datos, redundancia, heterogeneidad semántica, herramientas integración.		
3	Técnicas y herramientas para transformación de datos: generalización, suavizado, agregación, construcción de atributos, normalización Z-core.		
4	Técnicas y herramientas para reducción de datos: agregado a cubos, reducción de atributos, compresión de datos, discretización y jerarquización.		
MODALIDADES O FORMAS DE CONDUCCIÓN DE LOS PROCESOS DE ENSEÑANZA Y APRENDIZAJE			
<ol style="list-style-type: none"> 1. Lectura previa de materiales y referencias de apoyo. 2. Análisis presencial de casos de estudio. 3. Discusión presencial de la aplicación de las técnicas vistas en clase a fuentes de datos diversas. 4. Ejercicios de implementación de las herramientas vistas en clase. 5. Solución en forma individual de un problema de adquisición, normalización y limpieza de datos a partir de fuentes de datos reales. 			
MODALIDADES DE EVALUACIÓN Y ACREDITACIÓN			
Aspecto	Ponderación		
Participación de la discusión presencial	10%		
Ejercicios	30%		
Código de la solución del problema asignado	40%		
Presentación oral de la solución al problema asignado	20%		

BIBLIOGRAFÍA, DOCUMENTACIÓN Y MATERIALES DE APOYO

Autor	Título	Editorial	Año
Foreman, John	Data smart: Using Data Science to Transform Information into Insight	Wiley	2013
Kazil, Jacqueline y Jarmul, Katharine	Data Wrangling with Python: Tips and tools to make your life easier	O'Reilly	2016
Kimball, Ralph y Caserta, Joe	The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data	Wiley	2004
Field, Cady	The Data Science Handbook	Wiley	2017

PERFIL ACADÉMICO DESEABLE DEL RESPONSABLE DE LA ASIGNATURA

Profesional con maestría o doctorado en ciencias de la computación o áreas afines con experiencia en el desarrollo de sistemas basados en grandes volúmenes de datos. Se utilizará como criterio de selección la experiencia en el manejo de bases de datos relacionales y no relacionales.

NOMBRE Y FIRMA DE QUIEN DISEÑÓ LA CARTA DESCRIPTIVA

Juan Pablo Soto Barrera