



Universidad Tecnológica de la Mixteca

Clave DGP: 200089

Maestría en Ciencia de Datos

PROGRAMA DE ESTUDIOS

NOMBRE DE LA ASIGNATURA

Limpieza y Manipulación de Datos

SEMESTRE	CLAVE DE LA ASIGNATURA	TOTAL DE HORAS
Primer semestre	371014	35 Mediación docente 65 Estudio independiente

OBJETIVO(S) GENERAL(ES) DE LA ASIGNATURA

Que los alumnos obtengan conocimiento, habilidades y las herramientas necesarias para limpiar, manipular y transformar un conjunto de datos de manera efectiva para el uso de análisis y ciencia de datos.

TEMAS Y SUBTEMAS

- 1. Introducción**
 - 1.1. Calidad de los datos
 - 1.2. Fuentes de errores en conjuntos de datos
 - 1.3. Herramientas y técnicas para la limpieza de datos
- 2. Manipulación y transformación**
 - 2.1. Ordenación, filtrado, agregación, selección
 - 2.2. Escalado
 - 2.3. Codificación
 - 2.4. Creación de nuevas variables
- 3. Integración de datos**
 - 3.1. Fusión, vinculación, reconciliación de datos
 - 3.2. Técnicas de integración de datos
 - 3.3. Herramientas para la integración de datos
- 4. Evaluación de la calidad de los datos**
 - 4.1. Métricas: precisión, completitud, consistencia, relevancia
 - 4.2. Herramienta para la evaluación de la calidad de los datos
 - 4.3. Buenas prácticas para la calidad de los datos: limpieza y manipulación
- 5. Casos de estudio y aplicaciones**
 - 5.1. Aplicación de limpieza de datos
 - 5.2. Análisis de casos de éxito
 - 5.3. Proyectos de diferentes dominios

ACTIVIDADES DE APRENDIZAJE

Sesiones dirigidas por el profesor. Las sesiones se desarrollarán utilizando medios de apoyo didáctico como son la computadora, pizarrón y proyector. Asimismo, se desarrollarán prácticas sobre los temas del curso.

CRITERIOS Y PROCEDIMIENTOS DE EVALUACIÓN Y ACREDITACIÓN

Exámenes parciales y final. Tareas Simulaciones en computadora. Proyectos. Esto tendrá una equivalencia del 100% en la calificación final del semestre

BIBLIOGRAFÍA (TIPO, TÍTULO, AUTOR, EDITORIAL Y AÑO)

Básica:

- Ilyas, I. F., & Chu, X. (2019). Data cleaning. Morgan & Claypool.
- Osborne, J. W. (2012). Best practices in data cleaning: A complete guide to everything you need to do before and after collecting your data. Sage publications.

Consulta:

- C. C. Aggarwal, Data mining: The textbook, Springer, 2015.
- S. S. Skiena, The Data Science Design Manual, Springer, 2017

PERFIL PROFESIONAL DEL DOCENTE

Doctorado en Ciencias de la Computación, o áreas afines, con especialidad en Inteligencia artificial y/o Ciencia de datos.

Vo.Bo

M.T.C.A. ERIK GERMÁN RAMOS PÉREZ
COORDINADOR DE LA UNIVERSIDAD VIRTUAL

AUTORIZÓ

L.I. MARIO ALBERTO MORENO ROCHA
VICE-RECTOR ACADÉMICO