

IFCT0118 BIG DATA PARA INGENIERÍAS

Metodología

Teleformación

Duración

100 horas

Objetivos Didácticos

Construir y desarrollar bases de datos orientadas a la toma de decisiones y a la extracción de conocimiento.

Temario del Curso

1. Fundamentos y preparación de datos

- Comprensión de la Minería de datos
 - Conceptos básicos
 - Definición y objetivos del Data Mining
 - Diferencia entre minería de datos, análisis de datos y machine learning.
 - Ciclo de la vida del proceso KDD (Descubrimiento de Conocimiento en Bases de datos).
 - Fundamentos teóricos
 - Introducción a técnicas de clasificación, clustering y asociación.
 - Fundamentos matemáticos: probabilidad, álgebra lineal y estadísticas básicas.
- Manejo y manipulación de Datos
 - Bases de datos
 - Fundamentos de bases de datos relacionales.
 - Introducción a SQL para extracción y manipulación de datos.
 - Conexión con bases de datos no relacionales (NoSQL)
 - Preprocesamiento y limpieza de datos
 - Identificación y tratamiento de valores faltantes y atípicos.
 - Normalización y escalado de datos.
 - Reducción de dimensionalidad (PCA y selección de características)
- Uso de herramientas de análisis de datos
 - Introducción a las herramientas
 - Python y R: Comparación y uso para Data Mining
 - Hadoop y Spark: procesamiento en grandes volúmenes de datos.
 - Implementación práctica
 - Configuración de entornos y uso de bibliotecas clave (pandas, scikit-learn, tidyverse, etc.)
 - Lectura, manipulación y visualización básica de datos.
- Búsqueda de la eficiencia energética y desarrollo sostenible:
 - En el desarrollo de software y diseño
 - En la selección de proveedores y servidores

2. Modelado, Evaluación y Aplicaciones Avanzadas

- Desarrollo y Evaluación de Modelos
 - Modelos predictivos
 - Construcción de modelos de regresión y clasificación.
 - Evaluación y optimización de modelos: métricas (precisión, recall, AUC, etc.)
 - Validación cruzada y ajuste de hiperparámetros.
 - Técnicas avanzadas
 - Algoritmos de clustering: k-means, DBSCAN, jerárquicos
 - Análisis de reglas de asociación: A priori y FP-Growth
 - Evaluación e interpretación
 - Interpretación de resultados de modelos



- Métodos para mejorar la precisión y robustez de los modelos
- Identificación de las aplicaciones prácticas
 - Minería de datos en sectores clave
 - Aplicaciones en marketing, salud, finanzas, e-commerce, etc.
 - Casos reales de uso y análisis de impacto.
 - Proyecto práctico
 - Implementación de un pipeline completo de minería de datos desde la preparación hasta el modelo final.
 - Documentación y presentación de resultados