

IFCT166PO. ALGORITMOS DE BIGDATA PARA INGENIERÍAS

Código: IFCT166PO

Horas: 80

Descripción del curso:

Acerca de este curso

Cod: IFCT166PO

Unidad 1. MINERÍA DE DATOS O DATA MINING Y EL APRENDIZAJE AUTOMÁTICO.

1.1. Introducción a la minería de datos y el aprendizaje automático

1.1.1. Proceso KDD

1.1.2. Modelos y Técnicas de Data Mining

1.1.3. Áreas de aplicación

1.1.4. Minería de textos y Web Mining

1.1.5. Data mining y marketing

Unidad 2. R COMO HERRAMIENTA PARA BIG DATA

2.1. Instalación de R y RStudio

2.1.1. Introducción al lenguaje

2.1.2. Historia e Introducción a R

2.1.3. Operaciones Básicas y Números

2.1.4. Atributos, Entrada y Coerción

2.1.5. Matrices

- 2.1.6. Precedencia Operaciones Vectoriales
- 2.1.7. Manejo de fechas y tiempo
- 2.1.8. Listas, Factores, Valores Faltantes y Dataframes
- 2.1.9. Subconjuntos de Datos
- 2.1.10. Leer y Escribir Datos
- 2.2. Uso del lenguaje
 - 2.2.1. Estructuras de Control
 - 2.2.2. Funciones
 - 2.2.3. Reglas de Alcance
- 2.3. Sistema de gráficos
 - 2.3.1. Funciones *apply: apply
 - 2.3.2. Funciones *apply: lapply / sapply
 - 2.3.3. Funciones *apply: mapply / rep
 - 2.3.4. Graficación con el Sistema de Base de Gráficos
 - 2.3.5. Algunas Funciones Gráficas de Alto Nivel
 - 2.3.6. Parámetros en el Sistema de Gráficos
 - 2.3.7. Colores en el Sistema de Gráficos
 - 2.3.8. Graficación con Notación Matemática
 - 2.3.9. Graficación con texto y notación matemática
 - 2.3.10. Creación de Gráficas en 3D
- 2.4. Expresiones regulares. Graficas con ggplot2 y Simulación
 - 2.4.1. Expresiones Regulares
 - 2.4.2. Paquete de gráficos ggplot2
 - 2.4.3. Simulación
- 2.5. R en el mundo real
 - 2.5.1. Estadística Descriptiva y Predictiva con R
 - 2.5.2. Integración de R en Hadoop

Unidad 3. PRE-PROCESAMIENTO &, PROCESAMIENTO DE DATOS

- 3.1. Obtención y limpieza de los datos (ETL)
- 3.2. Inferencia estadística
- 3.3. Pruebas de hipótesis
- 3.4. Modelos de regresión
- 3.5. Árboles de Decisión
- 3.6. Algoritmos de Clasificación / Regresión (J48/C5.0, M5P)
- 3.7. Normalización, Tipos de distancia, Correlación
- 3.8. Machine Learning
- 3.9. Comparar Artículos (k-NN) .
- 3.10. Modelo predictivo de profit (k-NN, M5P...) .
- 3.11. Modelo predictivo de clasificación (J48, k-NN).

Unidad 4. ANÁLISIS DE LOS DATOS

- 4.1. Inteligencia Analítica de negocios
- 4.2. La teoría de grafos y el análisis de redes sociales
- 4.3. Presentación de resultados

Unidad 5. PROCESAMIENTO DISTRIBUIDO DE DATOS CON HADOOP

- 5.1. ¿Qué es Hadoop?
- 5.2. El sistema de archivos HDFS
- 5.3. Algunos comandos de referencia
- 5.4. Procesamiento MapReduce con Hadoop
- 5.5. El concepto de los clusters en Hadoop

Unidad 6. WEKA Y DATA MINING

6.1. ¿Qué es Weka?

6.2. Técnicas de Data Mining en Weka

6.3. Interfaces de Weka

6.4. Selección de atributos

¿Qué aprenderás?

Aplicar los algoritmos utilizados en Big Data, aprender el lenguaje R para su uso en Big Data, realizar un desarrollo con R y Hadoop, así como realizar Data Mining con Weka.