Diplomatura en Data Analytics con R y Python

Ciclo introductorio

Definición de conceptos

- Introducción a la Ciencia de Datos
- Niveles a los que opera la ciencia de datos
- Introducción a Data Warehouse
- Introducción a Data Mining
- Introducción a Knowledge Discovery
- Introducción a Herramientas OLAP y Tableros de comando
- Repaso de herramientas disponibles
- Taxonomía de las competencias de un científico de datos
- Primeros pasos en R
- Instalando y cargando paquetes en R

Introducción a R

- Variables, vectores y matrices
- Dataframes
- Paquetes y librerías
- Manejo de archivos csv
- Conversiones de tipos
- Conexión a bases de datos
- Ejecución condicional
- Bucles
- Funciones en R
- Medidas estadísticas
- Correlaciones
- Funciones estadísticas

Introducción a Python

- Descarga e instalación
- Principales librerías
- Variables y tipos de datos
- Listas, tuplas y diccionarios
- Ejecución condicional

- Ciclos definidos e indefinidos
- Manejo de Excepciones
- Funciones y Generadores
- Clases y objetos
- Manejo de archivos y directorios
- Conexión a archivos planos
- Conexión a Excel
- Conexión a Bases de Datos

Repaso de probabilidad y estadística

- Repaso de conceptos de probabilidad y estadística
- Test de Hipótesis
- Correlaciones
- AB Test

Ciclo Regular:

Arboles de decisión

- Algoritmo básico en Excel
- División en entrenamiento y prueba
- Uso de Rpart y cp
- Predicción y valoración de la solución
- Uso de Party
- Aplicación al problema del call center
- Instalación y uso del paquete en Python
- Ejemplo en Python
- Principales parámetros de ajuste y control

Clusters

- Algoritmo básico en Excel
- Uso de kmeans en R
- Ejemplo de aplicación real en R
- Otros algoritmos de agrupamiento en R
- Ejercicio de aplicación de agrupamiento de mascotas
- Uso de Scikit-Learn en Python



- Análisis de componentes principales en Python
- K-Medias en Python
- Clustering jerárquico en Python

Reglas de Asociación

- Algoritmo básico en Access
- División en entrenamiento y prueba en R
- Uso de arules en R
- Ejemplo de una aplicación real votación de reglas en R
- Ejemplo de una aplicación real a datos de ventas en supermercados en R
- Instalación y uso del paquete en Python
- Ejemplo en Python
- Principales parámetros de ajuste y control
- Esquema de votación de reglas en Python

Redes Neuronales

- Algoritmo básico en Excel
- División en entrenamiento y prueba en R
- Uso de neuralnet
- Predicción y valoración de la solución en R
- Discusión del problema de legibilidad de los resultados desde el negocio
- Competencia entre árboles y redes en un caso concreto en R
- Instalación y uso del paquete en Python
- Ejemplo en Python
- Principales parámetros de ajuste y control en Python
- Problema concreto en Python

Series temporales

- Taxonomía
- Separación de componentes



- Predicciones
- ARIMA implementado en R
- Predicción en series con un único período
- Predicción en series con múltiples períodos
- Predicción en series con períodos variables: renormalización
- Implementación en Python

Minería de textos

- Clasificación supervisada de piezas de texto
- Construcción de una red semántica
- Discusión de las posibilidades de reconocimiento de voz
- Ejemplos de aplicación real del algoritmo de clasificación
- Aplicación a la detección de sentimientos
- Ejemplo de implementación en Python

Vecinos Cercanos (Knn)

- Instalación y uso del paquete en R
- Ejemplo conceptual en Excel
- Ejemplo en R
- Instalación y uso del paquete en Python
- Ejemplo en Python
- Principales parámetros de ajuste y control

Bayes Ingenuo

- Implementación en Excel
- Ejemplo de aplicación en R
- Comparación de 4 algoritmos predictivos (Bayes ingenuo, árboles, redes neuronales y regression logística)
- Ejemplo de aplicación en Python



Random Forest

- Bootstrap y bagging
- Descripción conceptual del método random forest
- Paquete randomForest para R
- Ejemplo de aplicación
- Comparación con otras técnicas
- Implementación en Python

Métodos bayesianos avanzados

- Descripción conceptual del método:
- Regresión lineal bayesiana
- Regresión logística bayesiana
- Inferencia bayesiana
- Red bayesiana
- Ejemplos de aplicación en R
    - Paquete BAS
    - Paquete brms
    - Paquete arm
    - Paquete bnlearnd
- Comparación con otras técnicas
- Ejemplos de aplicación en Python

Máquina de soporte vectorial

- Descripción conceptual del método
- Paquete e1071 para R
- Ejemplo de aplicación en R
- Ejemplo de aplicación en Python
- Comparación con otras técnicas

Discriminante lineal y cuadrático

- Instalación y uso del paquete



- Ejemplo conceptual en Excel
- Ejemplo en R
- Ejemplo en Python
- Principales parámetros de ajuste y control en R
- Principales parámetros de ajuste y control en Python
- Problema concreto

Diseño de Datawarehouses

- Diferencias entre los DW y los OLTP
- Tipos de datos y soportes
- Dimensiones y jerarquías
- Estimación de recursos y tiempos según tipos de datos
- Ejemplos de staging
- Interacción entre equipos (DB-ETL-Soporte-FrontEnd-Usuarios)
- Ejercicios de diseño de DW

Diseño y construcción de ETL

- Procesos ETL: generalidades, buenas prácticas, esquemas de ETL
- Detalles de cada etapa: extracción, limpieza, normalización, transformación, carga
- Estrategias de update
- Ejercicios de diseño de ETL para la carga inicial
- Ejercicios de diseño de ETL para la actualización
- Licencia, descarga e instalación de Open Refine
- Uso general como herramienta de limpieza de datos
- Pre-procesado de los datos
- Casos de interés y ejemplos
- Licencias, descarga e instalación de Pentaho-Kettle
- Pentaho Kettle: funciones, características, utilización
- Auditoría y documentación
- Ejemplos de uso