

Data science para geociencias

Enero, 2021

Calendario y objetivos de aprendizaje semanales

El calendario es tentativo y sujeto a cambios. Los objetivos de aprendizaje mencionados deben tomarse como los conceptos clave y ejercicios que deberán quedar claros al final de cada semana. **Tomaremos un periodo de pausa para ponernos al corriente con los proyectos, del 09/03/2021 al 11/04/2021.**

▪ Semana 1

- **11/01/2021** Tema 1: Introducción a la ciencia de datos.
 - Comprender los conceptos fundamentales
- **15/01/2021** Tema 2: Preparación de los datos.
 - Comprender el flujo de la limpieza, transformación y visualización de los datos.
 - Ejercicio de limpieza de datos de calidad de aire.
 - Ejercicio de visualización de datos de intensidad de sismos históricos.

▪ Semana 2

- **18/01/2021** Tema 3: Técnicas de validación.
 - Comprender la diferencia entre las técnicas de validación para modelos de inferencia.
 - Ejercicio de Hold Out validation de incendios forestales.
 - Ejercicio de K-fold cross validation de incendios forestales.
- **22/01/2021** Tema 4: Selección de características I.
 - Comprender la importancia de la selección de características.
 - Ejercicio de transformación de imágenes satelitales a formato estructurado
 - Ejercicio de análisis de correlación.

■ Semana 3

- **25/01/2021** Tema 4: Selección de características II.
 - Ejercicio de reducción de dimensionalidad por PCA.
 - Ejercicio de selección de características.
- **29/01/2021** Tema 5: Rendimiento de modelos
 - Aprender a evaluar un modelo de inferencia.
 - Ejercicio de métricas de evaluación para el dataset de IRIS.

■ Semana 4

- **01/02/2021** Tema 6 - 6.1: Modelos de inferencia y Regresión lineal I
 - Comprender conceptos fundamentales del shallow learning
 - Comprender como funciona la regresión lineal
 - Ejercicio de regresión lineal univariada
- **05/02/2021** Tema 6.1: Regresión lineal II
 - Ejercicio de regresión lineal multivariada: Cálculo de esfuerzos en concreto.
 - Ejercicio de regresión lineal multivariada para series de tiempo: Cálculo de temperatura a partir de datos de calidad de aire.

■ Semana 5

- **08/02/2021** Tema 6.2: Regresión Logística
 - Comprender el funcionamiento de la regresión logística
 - Ejercicio de predicción de lluvia al siguiente día.
- **12/02/2021** Tema 6.3: Árbol de decisión
 - Comprender el funcionamiento de los árboles de decisión
 - Ejercicio de clasificación de suelos con imágenes satelitales con un árbol de decisión.

■ Semana 6

- **15/02/2021** Tema 6.4: Naive Bayes
 - Comprender el funcionamiento de Naive Bayes
 - Ejercicio de clasificación de suelo urbanos con Naive Bayes.
- **19/02/2021** Tema 6.5: Clustering
 - Comprender el funcionamiento de Clustering
 - Ejercicio de clusterización de sismos.

■ **Semana 7**

- **22/02/2021** Tema 6.6: K-vecinos cercanos
 - Comprender el funcionamiento de KNN
 - Ejercicio de detección de sargazo con KNN
- **26/02/2021** **Descanso**

■ **Semana 8**

- **01/03/2021** Tema 6.7: Máquinas de soporte vectorial
 - Comprender el funcionamiento de las SVM
 - Ejercicio de clasificación de detección de asteroides de la NASA
- **05/03/2021** Tema 7 - 7.1: Modelos de Bagging
 - Comprender las ventajas de los modelos de ensamble y entender el esquema de Bagging.
 - Ejercicio de clasificación de cultivos con Bagging

■ **Semana 9**

- **08/03/2021** Tema 7.2: Modelos de Ensamble y Boosting
 - Comprender el esquema de Boosting.
 - Ejercicio de clasificación de imágenes de fracturas con AdaBoost