



Análisis de datos con scikitlearn y pandas

Isabel Cenamor

isabelrosario.cenamor@campusviu.es



- Repositorios de datos online
- Introducción a pandas
- Scikit-learn
 - Clasificación
 - Regresión

> Clasificación

- Importar algoritmos de clasificación: `DecisionTreeClassifier`
 - Definir atributos de entrada / salida
- Visualización del árbol
- Validación de los resultados
 - Split vs validación cruzada
- Balanceo de datos

> Regresión

- Generación de datos con **numpy**
- Algoritmos de regresión
- Impresión de datos estadísticos
- Representación de atributos de entrada
- Comparativa entre atributos
- Problemas con los tipos de datos
- Predicción y visualización

Ficheros

- **Clasificación**
 - **iris.data**
 - **badges.csv**
- **Regresión**
 - **adult-data.csv**
 - **adult-test.csv**

Colaboratory

- Clasificación

<https://colab.research.google.com/drive/1K-dcGVsGB6zZagoPu1eeABT0zZND0vTs>

- Regresión

https://colab.research.google.com/drive/1mA3D_pTuv2DCYkCTbeSEtyifGo-SGrD

Gracias