# Guion

## Práctica 3 – Naive Bayes

### (FRAN) Descripción del dataset

Nº de casos utilizados (filas), atributos (columnas), tipos de datos de cada columna, valores posibles.

Mostrar imágenes o gráficas con el % de valores de cada columna y su frecuencia

#### Set 1

#### Set 2

### (FRAN) Exploración y preprocesamiento de datos

Definir atributos objetivo (target) y características (features), limpieza que hemos hecho, atributos que hemos quitado porque restaban importancia al modelo, modificación de valores, normalización.

### (LUIS) Procedimiento de entrenamiento del modelo

Hablar de vectorización de los atributos objetivo para facilitar la tarea al modelo y del encoding

Hablar de overfitting y underfitting

Hablar de tasa del error y precisión del modelo

Hablar de hiperparámetros

### (LUIS) Procedimiento de predicción del modelo

### (LUIS) Estadísticas generales y conclusiones extraídas

Mostrar gráficas finales

Mostrar distintas precisiones en función de la variación en el cjto de datos añadidos a cada dataset (modificar el dataset 2 de 1000 a 5000 y así)

Decir del dataset 2 que se han juntado dos target (Piso y Edificio) en un solo atributo para marcar el atributo objetivo e individualizarlo

## Práctica 4 – Arboles de decisión

### (FRAN) Descripción del dataset (Set 1)

Nº de casos utilizados (filas), atributos (columnas), tipos de datos de cada columna, valores posibles.

Mostrar imágenes o gráficas con el % de valores de cada columna y su frecuencia

Poner fuente de origen del dataset (Kaggle) y echarle flores a Kaggle

### (FRAN) Exploración y preprocesamiento de datos

Definir atributos objetivo (target) y características (features), limpieza que hemos hecho, atributos que hemos quitado porque restaban importancia al modelo, modificación de valores, normalización.

### (LUIS) Procedimiento de entrenamiento del modelo y uso de scikit-learn y funcionamiento del algoritmo RandomForest

Hablar de vectorización de los atributos objetivo para facilitar la tarea al modelo y del encoding

Hablar del gridsearch

Hablar de overfitting y underfitting

Hablar de tasa del error y precisión del modelo

Hablar de hiperparámetros

Hablar de scikit-learn y echarle flores y a su facilidad para implementar algoritmos y crear modelos de aprendizaje supervisado desde 0 (para aquellos que no son data scientists)

### (LUIS) Procedimiento de predicción del modelo y árbol generado

#### Profundidad 2

#### Profundidad 3

#### Profundidad 4

#### Profundidad n

### (LUIS) Estadísticas generales y conclusiones extraídas

Mostrar gráficas finales

Mostrar distintas precisiones en función de la variación en el cjto de datos añadidos a cada dataset