

# Clasificación y Regresión

# El plan

1. Lectura de datos
2. Limpieza y preprocesamiento
3. Entrenamiento de los modelos
4. Evaluación del modelo resultante.

# Dataset de Iris

1. Cargue los datos de Iris usando sklearn
2. Entrene un predictor de variedad de especie con Random Forest y SVM. Utilice KFold (con  $k=5$ ) para el testing y validación.  
Determine un rango de parámetros adecuado y seleccione los mejores parámetros de cada método.
3. Obtenga las matrices de confusión para el mejor modelo de cada método.

# Dataset de pasajeros del Titanic

1. Cargue los datos de pasajeros del Titanic usando pandas.
2. Seleccione y procese adecuadamente los datos para la predicción de la supervivencia de cada pasajero.
3. Entrene un predictor de supervivencia con Random Forest y SVM. Utilice KFold (con  $k=5$ ) para el testing y validación. Determine un rango de parámetros adecuado y seleccione los mejores parámetros de cada método.
4. Obtenga las matrices de confusión para el mejor modelo de cada método.

# Dataset de calidades de vinos tintos

1. Cargue los datos de calidad de vinos usando pandas, separe la columna de "quality" como la clase.
2. Normalice las columnas.
3. Entrene un predictor de calidad de vinos con Random Forest y SVM. Utilice KFold (con  $k=5$ ) para el testing y validación. Determine un rango de parámetros adecuado y seleccione los mejores parámetros de cada método.
4. Obtenga las matrices de confusión para el mejor modelo de cada método.
5. Repita los ejercicios anteriores pero sin normalizar las columnas. ¿Se obtienen resultados similares?

# Dataset de cancelación de clientes

1. Cargue los datos de abandono de clientes usando pandas: elimine las columnas 'State','Area Code','Phone', separe la columna de "Churn?" cómo la clase.
2. Procese y normalice las columnas.
3. Entrene un predictor que indique si el cliente dejará de serlo con Random Forest y SVM. Utilice KFold (con k=5) para el testing y validación. Determine un rango de parámetros adecuado y seleccione los mejores parámetros de cada método.
4. Obtenga las matrices de confusión para el mejor modelo de cada método.
5. Repita los ejercicios anteriores pero sin normalizar las columnas. ¿Se obtienen resultados similares?

# Dataset de ventas de Rossmann

1. Cargue los datos de Rossmann usando panda.
2. Estime la cantidad de ventas en la tienda 150 con un predictor lineal, utilizando como entrenamiento todas las demás. Prepare adecuadamente los datos usando `get_dummies`. No utilice los campos: 'Sales', 'Store', 'Date' y 'Customers'.
3. Grafique las ventas de la tienda 150 y su estimación . ¿Qué puede observar de los datos estimados? ¿Qué características de los datos de ventas se estimaron bien y qué otras mal?