

### Actividad 3.5 – Clasificación de vinos

El objeto de esta actividad es poner en práctica los conocimientos adquiridos hasta el momento para ellos vamos a utilizar el siguiente dataset que contiene una serie de características físico-químicas que determina la calidad del vino en una escala de valores del 1 al 10.

El enlace donde se encuentran los dataset es el siguiente:

<https://archive.ics.uci.edu/ml/datasets/Wine+Quality>

Como proyecto de partida se puede utilizar el ejemplo:

**Título:** Ejemplo\_3\_3\_Clasificación\_con\_Naive\_Bayes\_(Heart\_Diseases)

**Url:** [https://colab.research.google.com/drive/1hwri6X-N\\_cHmpZs31-zyK2XwRyfA4EGN?usp=sharing](https://colab.research.google.com/drive/1hwri6X-N_cHmpZs31-zyK2XwRyfA4EGN?usp=sharing)

En el cuaderno se ha de incluir los siguientes pasos:

1. **(5%)** Importación de los datasets (**utilizar el dataset winequality-red.csv**)
2. **(5%)** Mostrar la matriz de correlación de variables.
3. **(5%)** Aplicar cualquier otra técnica de selección de características que consideres adecuados y justificar tu propuesta.
4. **(10%)** Realizar una comparativa de la precisión en el entrenamiento de los diferentes modelos de NaivaBayes y KNN. Aplicando Cross Validation.
5. Una vez decides el modelo que consideras mejor, entonces realizar las siguientes tareas:
  - a. **(10%)** Entrenarlo y obtener la matriz de confusión.
  - b. **(10%)** Exportar a un fichero los parámetros del modelo entrenado.
  - c. **(10%)** Importar los parámetros del modelo.
  - d. **(10%)** Aplicar el modelo (predict) a todos los datos del dataset y obtener la matriz de confusión.
6. **(10%)** Comparar el resultado obtenido con el valor de calidad indicado en el dataset por medio de una matriz de confusión
  - a. Obtener la precisión del resultado obtenido, para determinar si coincide con la precisión que se calculó durante el entrenamiento
7. **(10%)** Probar a utilizar el cuaderno con el dataset de los vinos blancos y realizar captura de los resultados obtenidos. (**utilizar el dataset winequality-white.csv**)

#### **(5%) Formato:**

- El pdf final tiene una portada.
- Se utiliza un índice, apartado de conclusiones y referencias (web).
- Se hace mención a referencias externas, no recogidas en el material suministrado.

**(10%) Aportaciones personales:**

- Se enriquece la actividad con aportaciones personales distintas a las solicitadas en la propia actividad: de opinión, estrategia, herramientas utilizadas en la resolución de la actividad o elementos gráficos (dibujos, fórmulas, etc) en el propio cuaderno de Google Colab

Cada pregunta se evaluará entre 0 y 10 atendiendo a los siguientes criterios:

Puntos	Clasificador
De 0 a 1	Nada adecuado
De 2 a 3	Mínimamente adecuado
De 4 a 5	Algo adecuado
De 6 a 7	Moderadamente adecuado
De 8 a 9	Muy adecuado
10	Excelente

## Formato de entrega

- Nombrar el archivo siguiendo el siguiente patrón:

SNS\_ACT3\_5\_NombreApellidos.pdf

- Entregar un fichero en formato pdf con el siguiente contenido:
  - Incluir en el pie de cada página el nombre y apellidos del autor/a, así como el número de página y el total de páginas que contiene el documento.
  - Las imágenes capturadas han de tener la resolución necesaria para una buena visualización.
  - Indicar la url del GitHub donde se encuentran el/los diferentes cuadernos que has utilizado, con el objeto de consultarlos para descargarlo y verificar su funcionamiento. **En el caso de no indicar la url de Github se restará un punto en la nota final de esta actividad.**
  - **Importante:** Respetar la estructura de las preguntas, con el objeto de permitir una corrección homogénea para todas las actividades. De lo contrario se valorará con zeros puntos.
  - Al comienzo del notebook poner el nombre y apellidos del autor/a.