

## Práctica en Clase 2 - PC2

# Digit Recognizer

### 1. URL

<https://www.kaggle.com/competitions/digit-recognizer/>

### 2. Descripción del trabajo

Desde el entorno de kaggle crear un notebook completando las siguientes tareas:

2.1. Descripción del problema. Acceder a la URL, leer la información y describir brevemente en que consiste el problema. Comente también sobre los valores del atributo etiqueta.

2.2. *Selection*. Completar tareas de la etapa selección.

2.3. *Preprocessing*. Completar tareas de la etapa preprocesamiento.

2.4. *Transformation*. Completar tareas de la etapa transformación.

2.5. *Clustering*. Con fines ilustrativos, agrupar las características ( $X$ ) del subconjunto de entrenamiento con el número real de clases, creando de esta forma nuevos atributos de agrupamiento para cada instancia. Utilizar los algoritmos *K-Means* y *Spectral Clustering*.

2.6. *Dimesionality reduction*. Reducir el subconjunto de datos de entrenamiento para 2 dimensiones ( $X^m \Rightarrow X^p, p = 2$ ). Utilizar las siguientes técnicas: PCA, T-SNE ó LSP.

2.7. *Visualization*. Utilizando un *scatterplot*, visualizar los resultados de agrupamiento con el subconjunto reducido en 2 dimensiones. Discuta brevemente cuál técnica de agrupamiento se aproxima a los grupos verdaderos asignados según el atributo etiqueta ( $y$ ).

2.8. Proponer una solución para entrenar un modelo ML y hacer la *submission* de la solución en el portal de kaggle.