**Data set extraído de: kaggle competitions**

**Título del Proyecto**: “Modelo de predicción para determinar los pasajeros que sobrevivieron al naufragio del Titanic”

**I – Definición del problema**

El 15 de abril de 1912, durante su viaje inaugural, el RMS Titanic, ampliamente considerado como "insumergible", se hundió después de chocar con un iceberg. Desafortunadamente, no había suficientes botes salvavidas para todos a bordo, lo que resultó en la muerte de 1502 de los 2224 pasajeros y tripulantes.

Teniendo en cuenta la sucesión de los hechos, todo indicaría que algunos grupos de personas en el barco tenían mayores probabilidades de sobrevivir que otros.

**II - Objetivo**

Construir un modelo predictivo que responda a la pregunta: "¿Qué tipo de personas tenían más probabilidades de sobrevivir?" utilizando datos de pasajeros.

**III - Metodología**

**Características del dataset:**

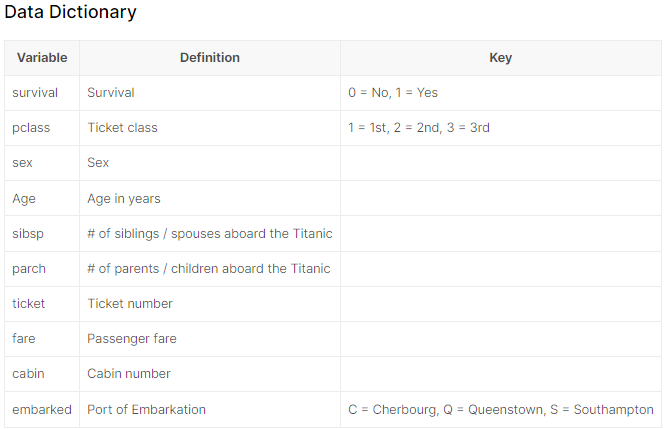
El dataset comprende dos conjuntos de datos (train.csv y test.csv) los cuales incluyen la

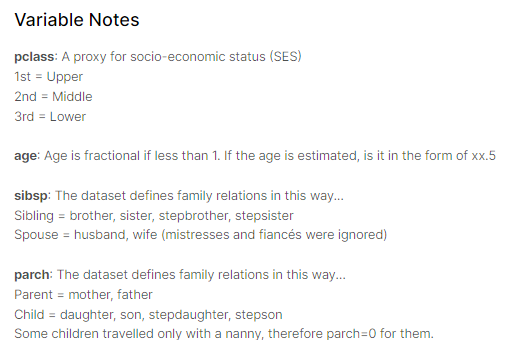
información de los pasajeros como: nombre, edad, género, clase socioeconómica, etc.

*Train-data*: características de 891 pasajeros.

*Test-data*: contiene información similar, de 418 pasajeros, pero no revela la verdad sobre terreno para cada pasajero.

Predicción (producto): Utilizando los datos train.csv, predecir si los otros 418 pasajeros a bordo sobrevivieron.





**Fases metodológicas:**

**Generalidades:**

Se utilizará el data-train para construir los modelos de aprendizaje automático, con su verdad de terreno para cada pasajero. Por otro lado, el data-test se utilizará para verificar la efectividad del modelo en datos no vistos.

Procedimiento

1. Definir el problema:

-Qué se pretende predecir?

- De qué datos se dispone?

- Qué datos es necesario conseguir?

1. Importar las librerías necesarias
2. Importar los datos

*url\_train/ url\_test*

1. Exploración de los datos:

* Verificar la cantidad y el tipo de datos en ambos dataset
* Verificar los datos faltantes y sus estadísticas

1. Preprocesamiento de los datos
2. Predicciones

##### *Etapas de un problema de machine learning[¶](http://localhost:8890/notebooks/Desktop/APRENDIZAJE_AUTOMATICO_Gulich/Coloquio/Machine-learning%20-con-Python-y-Scikit-learn.ipynb" \l "Etapas-de-un-problema-de-machine-learning)*

* **Definir el problema:**
  + ¿Qué se pretende predecir?
  + ¿De qué datos se dispone?
  + ¿Qué datos es necesario conseguir?
* **Explorar y entender los datos** que se van a emplear para crear el modelo.
* **Métrica de éxito:** definir una forma apropiada de cuantificar cómo de buenos son los resultados obtenidos.
* **Preparar la estrategia para evaluar el modelo:** separar las observaciones en un conjunto de **entrenamiento**, un conjunto de **validación** (o validación cruzada) y un conjunto de **test**.
  + Es muy importante asegurar que ninguna información del conjunto de test participa en el proceso de entrenamiento del modelo.
* **Preprocesar los datos:** aplicar las **transformaciones** necesarias para que los datos puedan ser interpretados por el algoritmo de machine learning seleccionado.
* **Ajustar un primer modelo capaz de superar unos resultados mínimos**. Por ejemplo,
  + en problemas de **clasificación**, el mínimo a superar es el porcentaje de la clase mayoritaria - **la moda**-.
  + en un modelo de **regresión**, la **media** de la variable respuesta.
* **Gradualmente, mejorar el modelo** incorporando
  + creando nuevas variables u
  + optimizando los hiperparámetros.
* **Evaluar la capacidad del modelo final** con el conjunto de test para tener una estimación de la capacidad que tiene el modelo cuando predice nuevas observaciones.
* **Entrenar el modelo final con todos los datos disponibles**.

**IV - Resultados esperados del Proyecto**. Son los resultados factibles alcanzables en el tiempo fijado y contando con los recursos mencionados en la Metodología. Especificarlos en función del tiempo cronológico para cada etapa, y con indicadores de logro para cada aspecto.

IV - Resultados esperados

Resultados claramente asociados a la resolución del problema propuesto, establecer conexiones con Proyectos en que se está contribuyendo de Curos previos, posteriores y con actividades simultáneas de la formación, lograr ampliamente las competencias planteadas por la Curso.

<https://www.youtube.com/watch?v=g0ZV-neSM7E>