



UNIVERSIDAD SANTO TOMÁS
PRIMER CLAUSTRO UNIVERSITARIO DE COLOMBIA

SECCIONAL TUNJA

VIGILADA MINEDUCACIÓN - SNIES 1732

APRENDIZAJE SUPERVISADO



APRENDIZAJE AUTOMÁTICO

Machine Learning es una rama de la Informática, más específicamente de la Ciencia de Datos, que aplica técnicas matemáticas y estadísticas para implementar sistemas que aprenden basados en el procesamiento de un conjunto de datos.

Dentro de las principales actividades se encuentran la detección y la extracción.

¡Siempre
hacia lo alto!



APRENDIZAJE AUTOMÁTICO

Un conjunto de datos está conformado por varias **instancias** (filas) y cada una de ellas contiene un grupo de **atributos** (columnas).

Los atributos pueden ser de tipo **categoricos**. Estos pueden almacenar un conjunto prefijado de valores. Ejemplo:

- season : season (1:springer, 2:summer, 3:fall, 4:winter)
- yr : year (0: 2011, 1:2012)
- mnth : month (1 to 12)
- hr : hour (0 to 23)
- holiday : weather day is holiday or not (extracted from <http://dchr.dc.gov/page/holiday-schedule>)
- weekday : day of the week
- workingday : if day is neither weekend nor holiday is 1, otherwise is 0.
- + weathersit :
 - 1: Clear, Few clouds, Partly cloudy, Partly cloudy
 - 2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist
 - 3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds
 - 4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog

```
cols =['season', 'month', 'holiday','weekday','workingday','weather']
for col in cols:
    datos_bici[col]=datos_bici[col].astype('category')
datos_bici.info()
```

iSiempre
hacia lo alto!



APRENDIZAJE AUTOMÁTICO

Atributos categóricos:

- Nominales: no hay orden entre los valores. Ejemplo: comida favorita de una persona (italiana, comida rápida, japonesa, colombiana, mexicana, peruana).
- Ordinales: los valores si tienen un orden y noción de lejanía. Ejemplo: clase de pasaje en un crucero (1^{era}, 2^{da} o 3^{ra})



APRENDIZAJE AUTOMÁTICO

Los atributos también pueden ser de tipo **continuos**. Estos pueden almacenar un conjunto de valores numéricos arbitrarios en un rango dado. Ejemplo: la edad, número de bicicletas rentadas, velocidad.



¡Siempre
hacia lo alto!

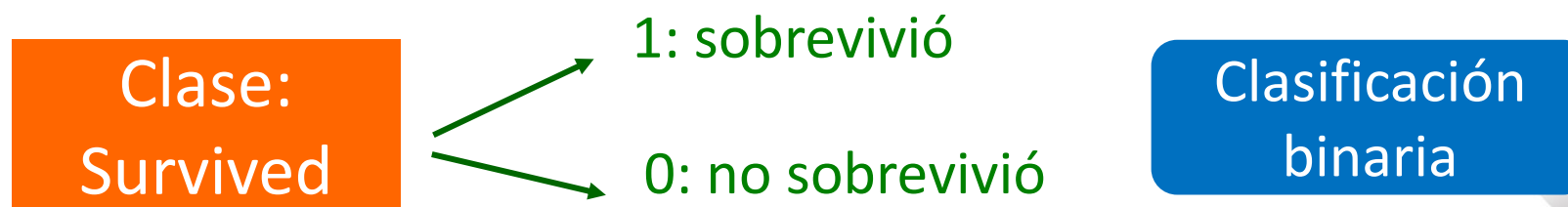


APRENDIZAJE AUTOMÁTICO

Aprendizaje Supervisado:

Los conjuntos de datos tienen un atributo clase. Es decir, se busca predecir una clase a partir de los valores del resto de los atributos.

- Clasificación: la clase es un atributo categórico. Ejemplo: predecir si un pasajero del Titanic sobrevive o no según el valor de los otros atributos.



- Regresión: la clase es un atributo continuo.

¡Siempre
hacia lo alto!



APRENDIZAJE AUTOMÁTICO

Pasos a realizar:

- Importar librerías.
- Importar datos.
- Entender los datos.
- Preprocesamiento de los datos.
- Aplicar algoritmos de Machine Learning.
- Predicción a partir de los modelos usados.

¡Siempre
hacia lo alto!



APRENDIZAJE AUTOMÁTICO

Taller en clase: Titanic

Importar el conjunto de datos de entrenamiento (891) y hacer el análisis que corresponda. Separe las actividades según se muestra y proporcione una conclusión del análisis realizado.

- Importar librerías.
- Importar datos.
- Entender los datos.
- Preprocesamiento de los datos.
- Aplicar algoritmos de Machine Learning.
- Predicción a partir de los modelos usados.

```
print(TrainData.shape)
```

```
(891, 12)
```

¡Siempre
hacia lo alto!



APRENDIZAJE AUTOMÁTICO

Taller en clase: Titanic

1. Importar librerías.
2. Importar datos.
3. Entender los datos.
 - a) Cantidad de instancias
 - b) Significado de atributos
 - c) Mostrar estadísticas
 - d) Mostrar nulos
 - e) Mostrar tipos de datos



APRENDIZAJE AUTOMÁTICO

Taller en clase: Titanic

4. Preprocesamiento de datos.

¡Siempre
hacia lo alto!



APRENDIZAJE AUTOMÁTICO

Taller en clase: Titanic

4. Preprocesamiento de datos.
 - a) Cambio de sexo a número
 - b) Cambio de embarque a número
 - c) Tratamiento a edad faltante (eliminar nulos)
 - d) Eliminar cabina
 - e) Eliminar atributos innecesarios
 - f) Eliminar filas sin datos



APRENDIZAJE AUTOMÁTICO

Taller en clase: Titanic

5. Aplicar algoritmos de ML.
 - a) Importar las librerías propias de los métodos a usar (regresión logística)
 - b) Dividir etiquetas de datos (del conjunto de entrenamiento tomamos el 80% para entrenar y el otro 20% para pruebas)
 - c) Aplicar el algoritmo de entrenamiento de regresión logística
 - d) Comprobar métricas y graficar la matriz de confusión



APRENDIZAJE AUTOMÁTICO

Taller en clase: Titanic

5. Aplicar algoritmos de ML y hacer predicciones.
 - e) Aplicar los modelos de árboles de decisión, Soporte de Vectores, Vecinos más cercanos.
 - f) Hacer una tabla comparativa de la precisión de cada método y de los resultados de la matriz de confusión.



APRENDIZAJE AUTOMÁTICO

Taller por parejas: Potabilidad del Agua

Realizar el procedimiento y análisis, guiados por el ejercicio anterior

**¡Siempre
hacia lo alto!**



REFERENCIAS

- Caballero, R., Martín, E. y Riesco, A. (2019). Big Data con Python. Recolección, almacenamiento y proceso.