**PREWORK**

**SESIÓN 03**

**Objetivo:**

En esta sesión veremos cómo funcionan los algoritmos no supervisados y cómo construir un sistema de recomendaciones con base en los algoritmos no supervisados.

**Pre-requisitos técnicos:**

Antes de comenzar, es importante que tengas instalados los siguientes paquetes para continuar:

* Pandas: Es una herramienta que te permite manipular datos de múltiples fuentes de forma sencilla, no importando si vienen de un archivo CSV, de una base de datos, o de un archivo de Excel. Lo utilizaremos para cargar datos rápidamente sin preocuparnos demasiado de su origen. Se instala con:
  + pip3 install pandas
* NumPy: Es una herramienta que te permite manejar álgebra lineal de forma acelerada. Los algoritmos de NumPy están diseñados para utilizar la mínima capacidad de cómputo posible y ser lo más rápidos y ágiles posibles. Se instala con:
  + pip3 install numpy
* SciPy: Es una herramienta que te permite manejar paquetes de operaciones científicas. Lo utilizaremos para calcular la moda de un conjunto. Se instala con:
  + pip3 install scipy
* MatPlotLib: es una herramienta de graficación muy utilizada por data scientists. Lo podemos utilizar para mostrar datos en forma de gráficas de barras, poligonales, etc. Se instala con:
  + pip3 install matplotlib
* Jupyter Notebook: Es nuestra herramienta base para construir algoritmos rápidamente con python. Te permite colocar código en python y puedes correr las celdas de forma escalonada, de tal manera que si algo falla, solo tienes que corregir la celda que falla, y no todo el código. Se instala con:
  + pip3 install jupyter

**Pre-requisito 1: Dataframes**

En esta sesión esperamos que tengas los conocimientos de manejo de datasets fresco, por lo que sería bueno que demos un repaso rápido de cómo manejar dataframes de pandas:

1. Un Dataframe es una tabla de dos dimensiones que tienen ejes etiquetados. Es una estructura de datos que te permite manipular los datos que necesites o filtrar los datos conforme lo requieras. Los Dataframes pueden venir de muchas fuentes, como son archivos CSV, archivos de Excel, bases de datos SQL, entre muchos otros.
2. Una de las funciones que vas a requerir en esta sesión será el leer un archivo CSV. Un archivo CSV significa “*comma separated values*” y es un archivo de texto que puedes generar con Excel, que te permite guardar tablas en un archivo de texto simple.
3. Puedes leer un archivo CSV con la línea de Python 3:

*variabledf = pandas.read\_csv(“archivo.csv”)*

1. Y puedes transformar una serie o dataframe de pandas a un arreglo de NumPy con la función:

*variabledf.to\_numpy()*

**Pre-requisito 2: Módulos de python.**

En programación, un módulo o librería es un fragmento de código que se utiliza muy a menudo, pero no vale la pena escribirla una y otra vez. Un buen programador no solamente escribe programas, sino que los escribe de tal manera que puedan servir en un futuro también, sin necesidad de estar constantemente reescribiendolos.

Nosotros crearemos en este caso módulos de python, que son programas y funciones dentro de un archivo “.py” el cual podemos copiar y pegar en cada uno de nuestros proyectos. Es importante saber que los módulos de Python tienen que estar en la misma carpeta que el proyecto principal, puesto que el intérprete de python los buscará en la misma carpeta a todos los proyectos.

Para generar tu módulo de Python, realizas lo siguiente:

* Creas un archivo con extensión .py (como por ejemplo MLUtilities.py)
* Dentro del archivo, pones todas las funciones que te interesan.
* Para mandar llamar al módulo, solo tienes que escribir el módulo con un import:

*import MLUtilities as utils*

**Introducción a los conceptos de la sesión:**

**Entrenamiento no supervisado:**

El entrenamiento no supervisado es una técnica de Machine Learning en la cual los usuarios no necesitan supervisar el modelo, en cambio, dejan que el modelo trabaje por su propia cuenta para descubrir patrones e información que pasó desapercibida en un principio.

Si bien es cierto que permite generar procesos complejos, el algoritmo no supervisado puede ser impredecible, comparado con otros métodos de aprendizaje natural. Es importante recordar esto cuando manejamos los algoritmos no supervisados.

**Sistema de recomendación:**

Un sistema de recomendaciones es un sistema que intenta predecir la preferencia que un usuario le dará a un cierto producto o servicio. Esto es utilizado en aplicaciones comerciales, y en esta sesión construiremos uno basado en algoritmos no supervisados.

Sin embargo, aun cuando los algoritmos de Machine Learning se utilizan como mecanismos para construirlos, hay muchas maneras diferentes de generar un sistema de recomendaciones. Una vez que veamos las bases del sistema de recomendaciones, el resto se desarrollará de manera natural, así que no tengas miedo en aplicar lo aprendido y crear tu propio sistema de recomendaciones si estás envuelto en un proyecto -digamos- de atención al cliente.

**Quiz:**

**¿Por qué los algoritmos no supervisados reciben ese nombre?**

1. No requieren de control humano para funcionar
2. No requieren ser programados por un humano
3. **No requieren supervisión para dar resultados**
4. No requieren supervisión para programarse

**¿Qué es un Dataframe?**

1. Un archivo de excel con datos
2. Un archivo de texto simple con datos
3. Una estructura de datos en forma abstracta
4. **Una estructura de datos en forma de tabla**

**¿Qué significa CSV?**

1. Correspondent Searching Values
2. **Comma Separated Values**
3. Correlated Separation Validator
4. Comma Searching Validator

**¿Para qué sirve un sistema de recomendación?**

1. Para recomendarte un producto
2. Para recomendarte un servicio
3. Para recomendarte una solución
4. **Todas las anteriores**