

Actividad 4: *Machine learning* o cómo aprenden las máquinas

PROPÓSITO

La primera actividad introduce a los estudiantes en la idea de cómo los computadores pueden manejar los datos de acuerdo con determinadas características, para buscar una relación o si hay patrones entre ellos. Posteriormente, se los invita a trabajar un número mayor de datos con un algoritmo de *machine learning* (ML) llamado K-Means. Se trata de un proceso guiado que emplea pocos datos en Excel; ello permite bajar la complejidad para que puedan implementar un algoritmo de ML que se aplica en procesos reales y muy difundidos.

Objetivos de Aprendizaje

OA 3 Desarrollar y programar algoritmos para ejecutar procedimientos matemáticos, realizar cálculos y obtener términos definidos por una regla o patrón.

OA i Buscar, seleccionar, manejar y producir información matemática/cuantitativa confiable a través de la web.

Actitudes

- Aprovechar las herramientas disponibles para aprender y resolver problemas.

Duración: 18 horas pedagógicas

DESARROLLO

LA TENDENCIA EN POCOS DATOS

Los sistemas de Amazon no saben qué es leer un libro ni qué se siente al leer uno en particular; sin embargo, pueden hacer recomendaciones: “Las personas que han comprado este libro también han comprado estos otros”. Por su parte, Netflix da acceso a una gran cantidad de películas, series, documentales, entre otros; también hace recomendaciones. Ambos sistemas son capaces de hacerlo, porque reconocen patrones que se dan en los datos y de algoritmos que usan esa información. Decimos que los sistemas “aprenden”.

Conexión interdisciplinaria:
Educación Ciudadana.
OA b, 3° y 4° medio

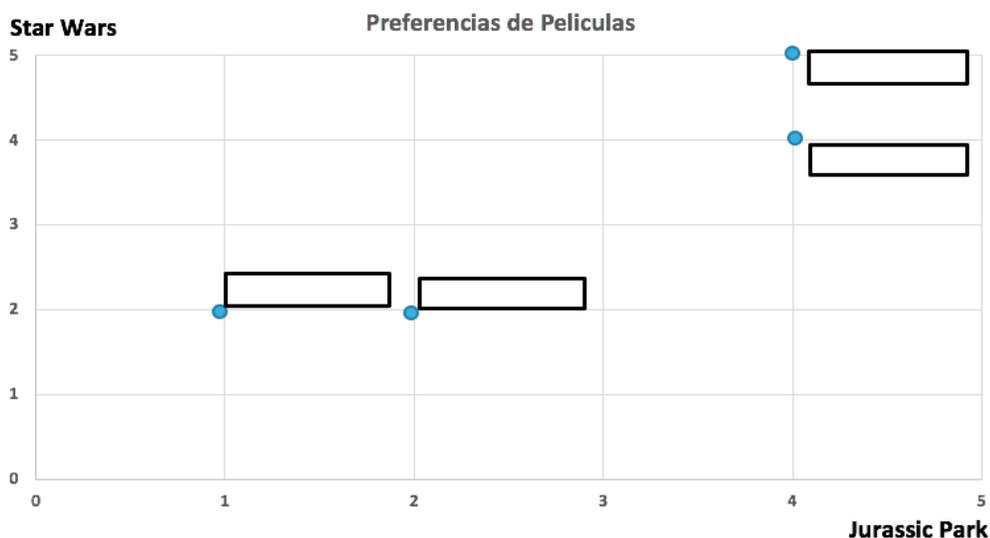
1. ¿Cómo pueden los sistemas o algoritmos, recomendar libros o películas?

2. La siguiente tabla muestra las preferencias de distintas personas sobre algunas películas.

Personas \ Películas	Star Wars	Jurassic Park	Toy Story	Avenger
Sebastián	****	****	*	**
Paola	*****	****	**	*
Ignacio	**	**	****	***
Javiera	**	*	***	****
Sandra	*****	?	?	**

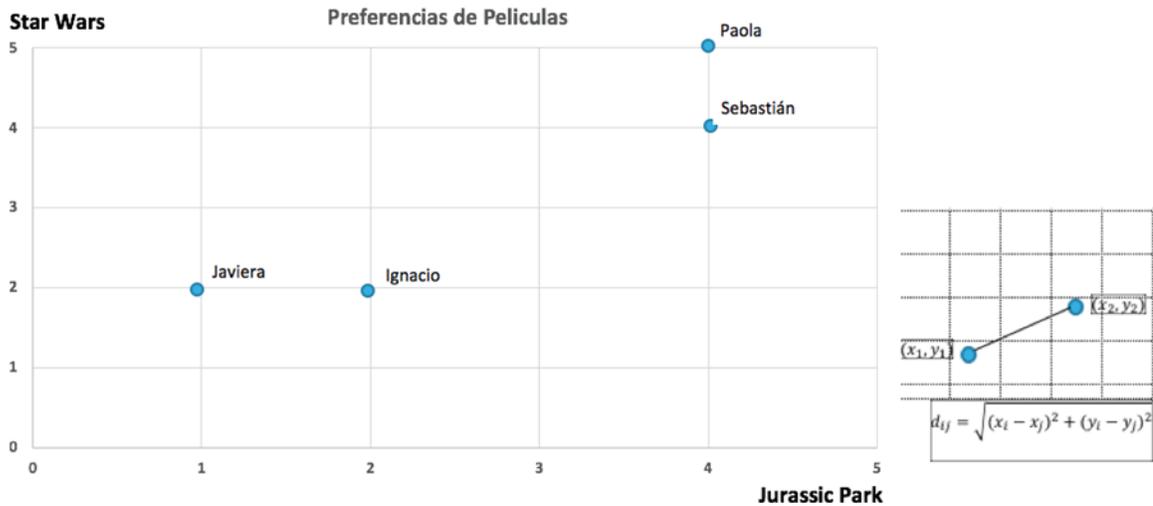
¿Cómo puede un algoritmo utilizar estos datos para recomendar películas? ¿Cómo lo harías tú?

3. Los puntos en las siguientes gráficas representan las preferencias de Sebastián, Paola, Ignacio y Javiera para las películas Star Wars y Jurassic Park.
- a. Completa los rectángulos con los nombres de estas personas, según los datos de la tabla anterior.



- b. ¿Cómo están ubicadas las personas a quienes les gustó Star Wars respecto de las que prefirieron Jurassic Park?

4. Observa la siguiente gráfica:



- a. Según esta nueva forma de representar los datos, ¿cómo están ubicadas las personas a quienes les gustó Star Wars respecto de las que les prefirieron Jurassic Park?

Sugerencia: Para calcular la distancia euclidiana entre los puntos similares, considera la distancia entre dos puntos (x_i, y_i) , (x_j, y_j) :

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}.$$

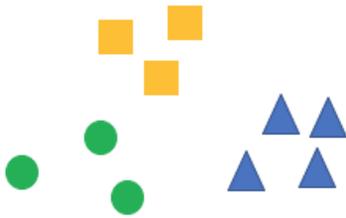
5. Sandra no ha visto Jurassic Park, pero dio 5 puntos a Star Wars. ¿Qué película le puedes recomendar: Jurassic Park o Toy Story? ¿Por qué?

K-MEANS Y EL MACHINE LEARNING

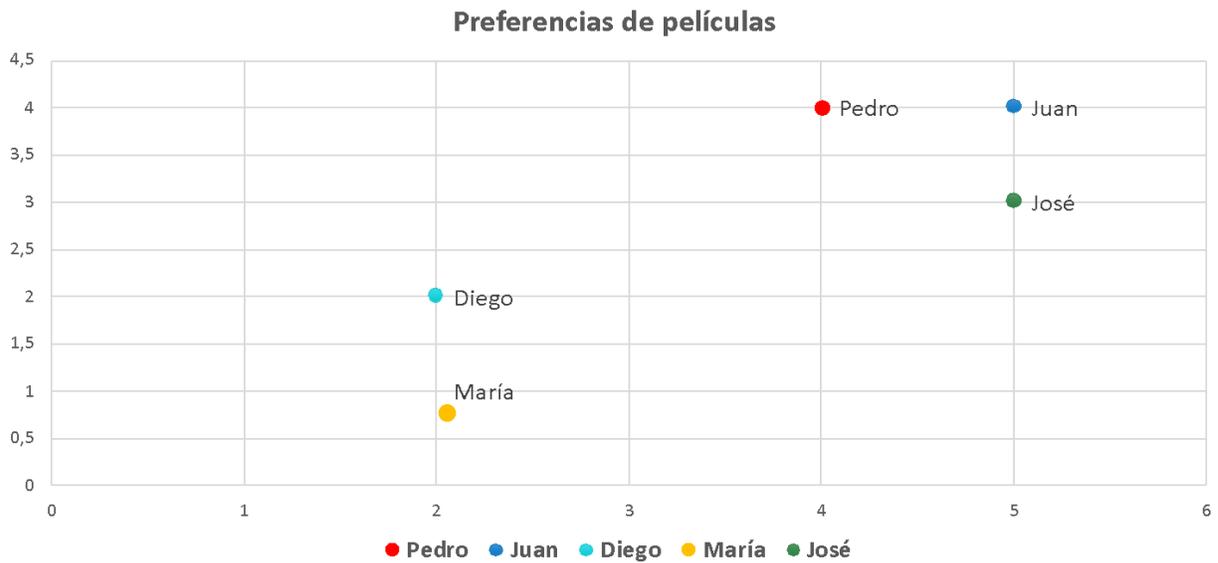
1. Anteriormente trabajaron con pocos datos agrupados. ¿Qué pasa si hay muchos más datos (pueden ser miles o millones)? Para saberlo, se puede usar aprendizaje automático o *machine learning* (ML).
 - a. ¿Saben qué es ML?
 - b. ¿Para qué se usa?
 - c. ¿Saben si alguna aplicación o servicio que usan implementa ML? ¿Cuáles?
 - d. Discutan sobre lo que cada uno sabe o imagina.
 - e. Observen el video <https://youtu.be/KytW151dpqU> y complementen sus ideas y descubrimientos previos.

2. ¿Qué significa agrupamiento?

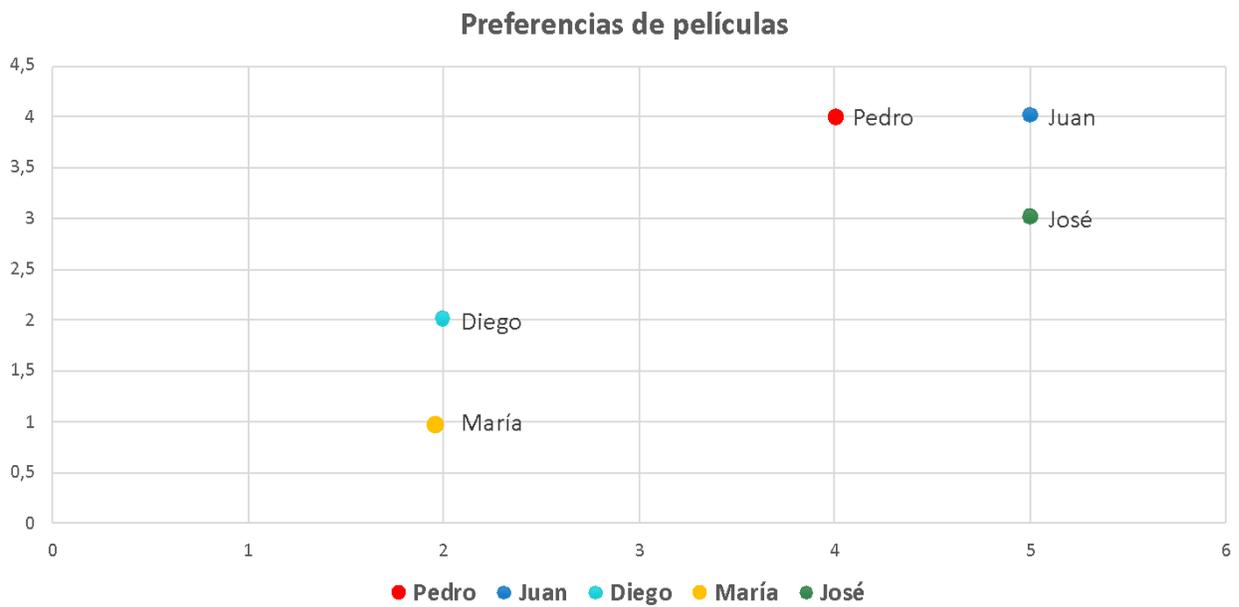
a. Consideren las siguientes figuras, ¿cómo las agruparían?



b. Organicen los puntos de la siguiente gráfica en tres grupos.



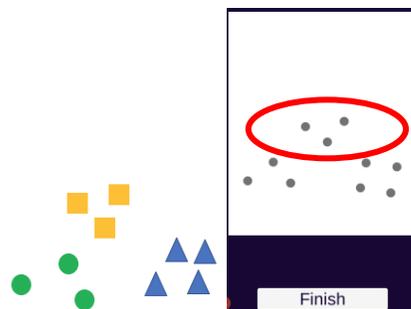
c. Para la misma gráfica, organicen los puntos en dos grupos.



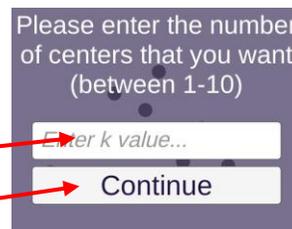
- d. ¿Qué número de agrupación representa mejor las preferencias de las personas respecto de las películas graficadas: la de tres grupos o la de dos? ¿Por qué?
3. En uno de los celulares del grupo, descarguen la app Kmedia . Está en inglés, pero los orientaremos para que puedan ejecutarla y hacer la actividad. De la ventana principal, seleccionen la opción "START".



4. Presionen dentro del cuadrado blanco del celular para colocar las figuras de círculos, cuadrados y triángulos, en posiciones similares a las de la actividad inicial. Luego presionen "Finish".



Ingresen el número de centros correspondiente al número de grupos que hicieron; es decir, si hicieron tres grupos, ingresen ahora el 3 como número de centros.



Luego presionen "Continue".

Presionen dentro del cuadrado blanco que deseen para colocar el número de centros que señalaron.

- ¿Cuántos grupos quedaron?
- ¿Cómo quedaron representados?
- ¿Tuvieron que hacer iteraciones después de ingresar los puntos de centros?
- ¿Cuántas hicieron hasta que terminó el proceso?
- ¿Cómo se agruparon finalmente los puntos respecto de los centros?

5. Al terminar, presionen “Restart All” e ingresen la siguiente organización de puntos.
Luego, ingresen el número 2 como el número de centros y posteriormente ubíquenlos donde deseen.

Antes de presionar “Start”, observen y anoten cómo se organizan los puntos en torno a cada centro.

En las siguientes iteraciones, observen y anoten:

- Cómo se mueven los centros respecto de los grupos ya armados
- Si hay puntos que cambien de color, ¿por qué lo hacen?
- ¿Pueden observar algún comportamiento regular, algún patrón?
- ¿Cuándo piensan que termina de hacerse las iteraciones?



6. Luego de terminarlas, seleccionen la opción “Redo Centers” e ingresen dos centros en posiciones diferentes de como las pusieron en el punto anterior. ¿Cambian las organizaciones de los grupos? ¿Y el número de iteraciones? ¿Y la disposición final? ¿Qué pueden decir respecto de la selección de la posición inicial de estos centros?
Prueben haciendo varios ingresos de nuevos centros; también cambien los grupos de puntos y el número de centros, experimenten y pongan a prueba sus hipótesis e ideas.
7. Existen muchos algoritmos de *machine learning*; en esta actividad, tuvieron una aproximación al K-means. Investiguen algo más al respecto en <http://www.aprendemachinelearning.com/K-means-en-python-paso-a-paso/> leyendo el algoritmo K-means y el ejemplo presentado en la página.
- Describan la secuencia de pasos en el algoritmo K-means, con sus propias palabras o haciendo un diagrama.
 - ¿Cómo cambiaron sus respuestas con el conocimiento nuevo?
 - ¿Para qué se usa la distancia euclidiana en el algoritmo de K-means?
 - ¿Por qué caracterizar más y tener más datos de los problemas nos permite tener mejores resultados con los algoritmos utilizados?
 - ¿Qué pasa si se tiene datos “pobres” o malos, o si se tiene datos con sesgo, por ejemplo, de raza, religión, políticos, de sexo, entre otros?
8. Aunque solo vieron un algoritmo trabajado en temas de inteligencia artificial, como *machine learning*, ¿piensan que la matemática que se utiliza en inteligencia artificial es muy complicada y está fuera de sus posibilidades?

ORIENTACIONES PARA EL DOCENTE

1. Este tema que puede interesante y altamente motivador para los jóvenes. Está en muchas aplicaciones que manejan día a día, por lo que se puede aprovechar el espacio para discutir desde áreas de desarrollo profesional hasta temas éticos sobre cómo se usa la información de las personas y si ellas saben qué información se está usando.
2. Hay mucha información de estos temas en internet, y parte importante está en inglés. Que esto no sea un problema: trate de que comiencen a familiarizarse con estos espacios, que sean persistentes, que sepan realizar búsquedas y seguir “hilos” de información.
3. Se requiere que, al menos, un estudiante por grupo tenga un celular con acceso a internet y pueda descargar una aplicación en él.
4. Se sugiere los siguientes indicadores para evaluar formativamente los aprendizajes:
 - Ejecutan programas que incluyen diferentes tipos de datos.
 - Programan algoritmos que permiten obtener un término definido al ejecutar procedimientos matemáticos.

RECURSOS Y SITIOS WEB

Sitios web sugeridos para estudiantes y profesores

- *How K-means clustering groups data: A Simple Example*
<https://www.youtube.com/watch?v=wE8H-MEHSks>
- *Aplicaciones machine learning*
<http://www.aprendemachinelearning.com/aplicaciones-del-machine-learning/>
- *7 pasos del machine learning para construir tu máquina*
<http://www.aprendemachinelearning.com/7-pasos-machine-learning-construir-maquina/>