



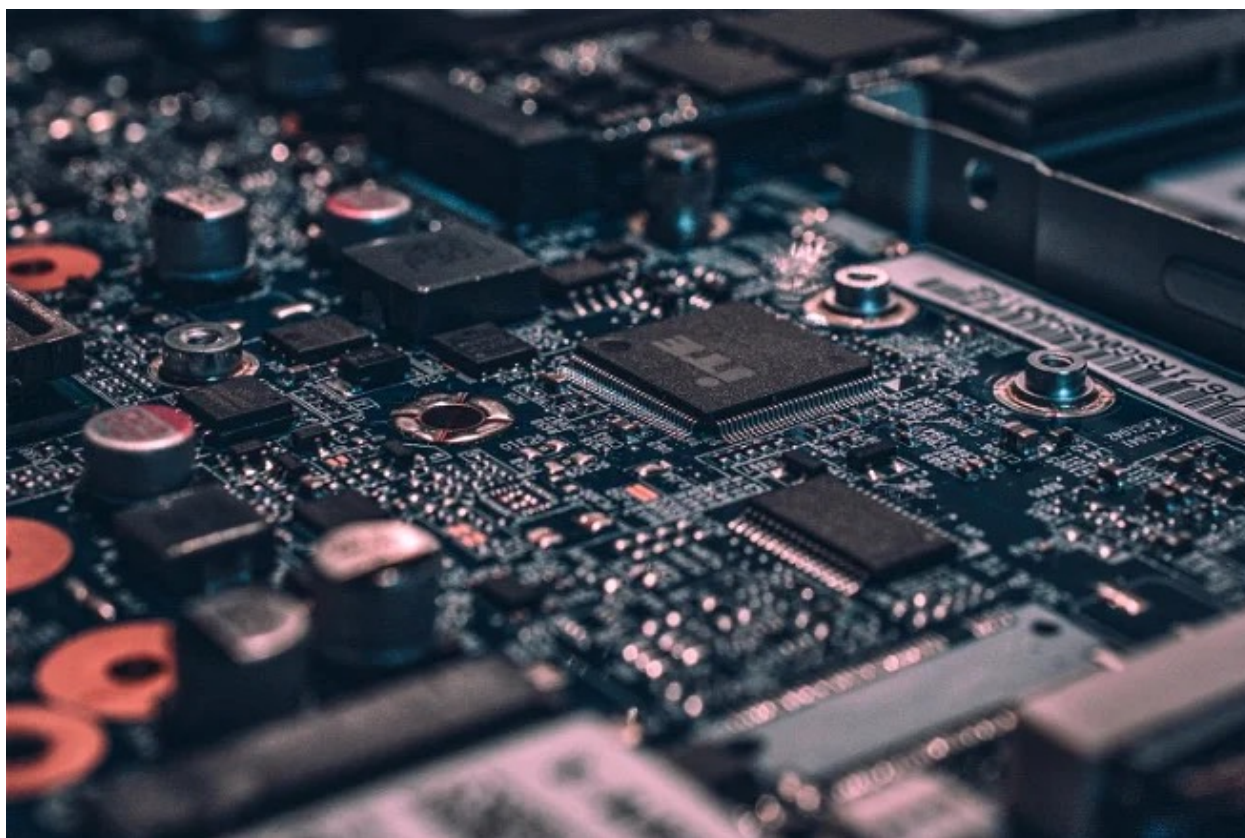
UNIVERSIDAD NACIONAL
AUTÓNOMA DE MÉXICO

FACULTAD DE INGENIERÍA



Inteligencia Artificial **Práctica 1**

Apriori R Asociación



ALUMNOS:

ALCÁNTARA GUERRERO ALFREDO GUADALUPE

RECINOS HERNÁNDEZ LUIS MARIO

Índice

Introducción.....	2
Desarrollo.....	2
Reflexión Final.....	4
Bibliografía.....	5

Introducción

La inteligencia artificial es una disciplina que asimila conocimientos, directrices y técnicas de ingeniería junto con métodos científicos para poder generarnos procedimientos automatizados con un cierto grado de conciencia y resiliencia a su entorno con el objetivo de resolver conflictos. Cuando se comienza a hablar de procesos inteligentes, conscientes, debemos de considerar que para que un algoritmo pueda comprender cada vez más de su entorno, será menester la integración de un proceso de aprendizaje. Conforme más avances se han hecho en el campo de la inteligencia artificial, la importancia de la preparación de mayores volúmenes de información para su aprendizaje ha tomado más importancia. En la inteligencia artificial moderna, se tiene un eje en la información que se le puede suministrar a un algoritmo para que, por medio del reconocimiento de patrones y reforzamiento de conductas favorables nos quedemos con procesos automatizados más capaces, que poco a poco puedan pasar de resolver problemas específicos, a convertirse en solucionados a escenarios más generales.

Como primer práctica para el laboratorio de la materia, daremos los primeros pasos en esta relación de desarrollador o desarrolladora con una maquina que procesa información en aras de la formación de alguna inteligencia. Nos comenzaremos a familiarizar con plataformas de desarrollo y herramientas capaces de facilitar este proceso para podernos fundar las bases para el proceso y análisis de información en materia de inteligencia artificial.

De igual forma, dentro de esta práctica se comenzará a analizar la información como tuplas, o datos transaccionales por medio de tres expresiones estadísticas:

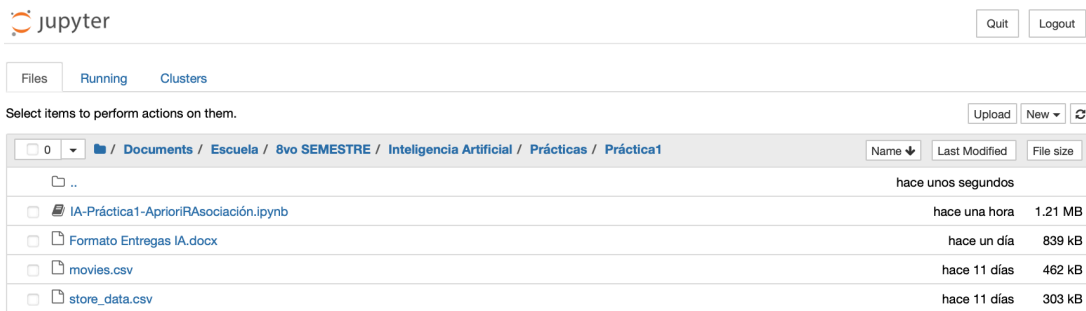
$$\text{Soporte} = \frac{\text{Frecuencia}(X, Y)}{\text{NumeroCasos}N}$$

$$\text{Confianza} = \frac{\text{Frecuencia}(X, Y)}{\text{Frecuencia}(X)}$$

$$\text{Elevacion} = \frac{\text{Soporte}}{\text{Soporte}(X) \times \text{Soporte}(Y)}$$

Desarrollo

El desarrollo de la práctica consistió en la carga de datos de una base a un editor de textos potenciado por python para poder realizar búsquedas, consultas y cálculos con los datos que el profesor nos dio al inicio de la práctica. El software que será empleado para esta y futuras prácticas de laboratorio fue Anaconda: Jupyter.



La carga de la información al software y hoja de texto preparada por el profesor para la primera práctica empleó la terminal de nuestras computadoras para encontrar y lanzar los archivos encontrados en la máquina.

```
Mare — jupyter_mac.command — python - jupyter_mac.command — 80x24
[I 2023-03-27 22:46:14.432 LabApp] JupyterLab extension loaded from /Users/Mare/
opt/anaconda3/lib/python3.9/site-packages/jupyterlab
[I 2023-03-27 22:46:14.432 LabApp] JupyterLab application directory is /Users/Ma
re/opt/anaconda3/share/jupyter/lab
[I 22:46:14.444 NotebookApp] The port 8888 is already in use, trying another por
t.
[I 22:46:14.445 NotebookApp] Serving notebooks from local directory: /Users/Mare
[I 22:46:14.445 NotebookApp] Jupyter Notebook 6.4.12 is running at:
[I 22:46:14.446 NotebookApp] http://localhost:8889/?token=f1405ee0b05328ff3e973a
128f1cacb16768aab08b2d630a
[I 22:46:14.446 NotebookApp] or http://127.0.0.1:8889/?token=f1405ee0b05328ff3e
973a128f1cacb16768aab08b2d630a
[I 22:46:14.446 NotebookApp] Use Control-C to stop this server and shut down all
kernels (twice to skip confirmation).
[C 22:46:14.485 NotebookApp]

To access the notebook, open this file in a browser:
file:///Users/Mare/Library/Jupyter/runtime/nbsrvr-4919-open.html
Or copy and paste one of these URLs:
http://localhost:8889/?token=f1405ee0b05328ff3e973a128f1cacb16768aab08b2
d630a
or http://127.0.0.1:8889/?token=f1405ee0b05328ff3e973a128f1cacb16768aab08b2
d630a

```

```
Mare — jupyter_mac.command — python - jupyter_mac.command — 80x24
Inteligencia Artificial/Prácticas/Práctica1/IA-Práctica1-AprioriRAsociación.ipynb
[I 21:19:24.388 NotebookApp] Saving file at /Documents/Escuela/8vo SEMESTRE/Inte
ligencia Artificial/Prácticas/Práctica1/IA-Práctica1-AprioriRAsociación.ipynb
[I 21:22:24.384 NotebookApp] Saving file at /Documents/Escuela/8vo SEMESTRE/Inte
ligencia Artificial/Prácticas/Práctica1/IA-Práctica1-AprioriRAsociación.ipynb
[I 21:23:24.383 NotebookApp] Saving file at /Documents/Escuela/8vo SEMESTRE/Inte
ligencia Artificial/Prácticas/Práctica1/IA-Práctica1-AprioriRAsociación.ipynb
[I 21:25:24.380 NotebookApp] Saving file at /Documents/Escuela/8vo SEMESTRE/Inte
ligencia Artificial/Prácticas/Práctica1/IA-Práctica1-AprioriRAsociación.ipynb
[I 21:27:24.382 NotebookApp] Saving file at /Documents/Escuela/8vo SEMESTRE/Inte
ligencia Artificial/Prácticas/Práctica1/IA-Práctica1-AprioriRAsociación.ipynb
[I 21:29:24.379 NotebookApp] Saving file at /Documents/Escuela/8vo SEMESTRE/Inte
ligencia Artificial/Prácticas/Práctica1/IA-Práctica1-AprioriRAsociación.ipynb
[I 21:36:05.723 NotebookApp] Saving file at /Documents/Escuela/8vo SEMESTRE/Inte
ligencia Artificial/Prácticas/Práctica1/IA-Práctica1-AprioriRAsociación.ipynb
[I 21:38:05.714 NotebookApp] Saving file at /Documents/Escuela/8vo SEMESTRE/Inte
ligencia Artificial/Prácticas/Práctica1/IA-Práctica1-AprioriRAsociación.ipynb
[I 21:42:05.735 NotebookApp] Saving file at /Documents/Escuela/8vo SEMESTRE/Inte
ligencia Artificial/Prácticas/Práctica1/IA-Práctica1-AprioriRAsociación.ipynb
[I 21:44:05.709 NotebookApp] Saving file at /Documents/Escuela/8vo SEMESTRE/Inte
ligencia Artificial/Prácticas/Práctica1/IA-Práctica1-AprioriRAsociación.ipynb
[I 21:46:05.710 NotebookApp] Saving file at /Documents/Escuela/8vo SEMESTRE/Inte
ligencia Artificial/Prácticas/Práctica1/IA-Práctica1-AprioriRAsociación.ipynb

```

A partir del momento en que el archivo se abre dentro del navegador, el proceso de guardado y control de versiones podrá ser monitoreado en la misma terminal abierta.

El análisis de datos dentro de anaconda requerirá de algunos preparativos empleando ciertos paquetes de python como los que se muestran a continuación:

1) Importar las bibliotecas necesarias

```
In [1]: !pip install apyori # pip es un administrador de paquetes de Python. Se instala el paquete Apyori
Requirement already satisfied: apyori in /Users/Mare/opt/anaconda3/lib/python3.9/site-packages (1.1.2)
```

```
In [3]: import pandas as pd # Para la manipulación y análisis de los datos
import numpy as np # Para crear vectores y matrices n dimensionales
import matplotlib.pyplot as plt # Para la generación de gráficas a partir de los datos
from apyori import apriori
```

2) Importar los datos

Fuente de datos: movies.csv

```
In [4]: DatosMovies = pd.read_csv('movies.csv')
DatosMovies
```

Out[4]:

	The Revenant	13 Hours	Allied	Zootopia	Jigsaw	Achorman	Grinch	Fast and Furious	Ghostbusters	Wolverine	Mad Max	John Wick	La La Land	The Good Dinosaur	Ninja Turtles	The Good Dinosaur Bad Moms
0	Beirut	Martian	Get Out	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	Deadpool	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	X-Men	Allied	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	Ninja Turtles	Moana	Ghost in the Shell	Ralph Breaks the Internet	John Wick	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	Mad Max	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...
7454	Big Sick	Looper	Hulk	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

En este caso, el archivo con los datos listos para usar se llamaba «movies.csv» a partir de este, se podrán ocupar una serie de comandos para consultar la información de distintas maneras, a continuación se muestra un fragmento de la libreta de Jupyter donde se realizaron las configuraciones 3 y 4:

Configuración 3

Obtener reglas para aquellas películas que se hayan rentado al menos 77 veces en un día (539 veces en una semana):

i) El soporte mínimo se calcula de $539/7460 = 0.072252$ (7.2%).

ii) La confianza mínima para las reglas de 20%

iii) Considerando una elevación de 1.5

Observación: Estos valores se eligen arbitrariamente, por lo que, se recomienda probar valores y analizar la diferencia en las reglas.

```
In [18]: ReglasC3 = apriori(MoviesLista,
min_support=0.072,
min_confidence=0.2,
min_lift = 1.5)
```

```
In [19]: ResultadosC3 = list(ReglasC3)
print(len(ResultadosC3))
```

0

Observación de ejercicio

Al colocar una serie de parámetros para la tercera configuración de nuestro algoritmo, podemos observar que los datos relacionales encontrados que satisficieran los criterios numéricos de relación establecidos en la entrada [18] son nulos, por lo que será necesario probar con otros parámetros (unos menos exigentes, más probables) para podernos hacer con datos para leer.

Configuración 4

Obtener reglas para aquellas películas que se hayan rentado al menos 40 veces en un día (280 veces en una semana):

- i) El soporte mínimo se calcula de $280/7460 = 0.0375335$ (3.7%).□
- ii) La confianza mínima para las reglas de 20%
- iii) Considerando una elevación de 1.5□

Observación: Estos valores se eligen arbitrariamente, por lo que, se recomienda probar valores y analizar la diferencia en las reglas.

```
In [22]: ReglasC4 = apriori(MoviesLista,
                        min_support=0.037,
                        min_confidence=0.2,
                        min_lift = 1.5)
```

```
In [23]: ResultadosC4 = list(ReglasC4)
print(len(ResultadosC4))

3
```

Ahora que podemos ver que sí hay datos transaccionales que cumplan con los criterios establecidos, podemos continuar procesando la información para quedarnos con datos fáciles de leer.

```
In [26]: ResultadosC4
```

```
Out[26]: [RelationRecord(items=frozenset({'Jumanji', 'Ninja Turtles'}), support=0.04115281501340483, ordered_statistic
s=[OrderedStatistic(items_base=frozenset({'Jumanji'}), items_add=frozenset({'Ninja Turtles'}), confidence=0.4
171195652173913, lift=1.742279930863236)]),
RelationRecord(items=frozenset({'Jumanji', 'Tomb Rider'}), support=0.03941018766756032, ordered_statistics=[
OrderedStatistic(items_base=frozenset({'Jumanji'}), items_add=frozenset({'Tomb Rider'}), confidence=0.3994565
217391304, lift=2.283483258370814), OrderedStatistic(items_base=frozenset({'Tomb Rider'}), items_add=frozenset
({'Jumanji'}), confidence=0.22528735632183905, lift=2.283483258370814)]),
RelationRecord(items=frozenset({'Moana', 'Ninja Turtles'}), support=0.04825737265415549, ordered_statistics=[
OrderedStatistic(items_base=frozenset({'Moana'}), items_add=frozenset({'Ninja Turtles'}), confidence=0.37075
18022657054, lift=1.5486049523528347), OrderedStatistic(items_base=frozenset({'Ninja Turtles'}), items_add=fro
zenset({'Moana'}), confidence=0.20156774916013437, lift=1.5486049523528347)])]
```

Visualización de datos de la información encontrada

Se pueden alcanzar a observar los datos de las tres coincidencias encontradas en el sistema de información, cada una empieza con la leyenda: "RelationRecord", seguida de las películas que guardan la relación y los datos estadísticos de cada uno que hicieron posible el match con los criterios de asociación establecidos

```
In [27]: pd.DataFrame(ResultadosC4)
```

```
Out[27]:
```

	items	support	ordered_statistics
0	(Jumanji, Ninja Turtles)	0.041153	[(('Jumanji'), (Ninja Turtles)), 0.41711956521739...
1	(Jumanji, Tomb Rider)	0.039410	[(('Jumanji'), (Tomb Rider)), 0.3994565217391304,...
2	(Moana, Ninja Turtles)	0.048257	[(('Moana'), (Ninja Turtles)), 0.3707518022657054...

Presentación de Información

Por más que seamos capaces de interpretar compactos cúmulos de información con los de la salida [26], una buena forma de trabajar con estas grandes bases de datos, será el conocer y emplear herramientas para el ordenamiento y selección de datos relevantes, de esta forma, podemos hacer de la información obtenida no solo más fácil de leer para nosotros mismos, sino que también accesible para personas que no se especialicen en el análisis de información, en este perfil podremos encontrar desde a nuestros usuarios y usuarias, hasta otros integrantes de nuestro equipo multidisciplinario.

En la tabla mostrada en la slaida número [27], hicimos uso del comando "DataFrame" con la librería "pd" para presetnar de forma simple y legible, las coincidencias correspondientes a la entrada [26].

```
In [28]: for item in ResultadosC4:
        #El primer índice de la lista
        Emparejar = item[0]
        items = [x for x in Emparejar]
        print("Regla: " + str(item[0]))

        #El segundo índice de la lista
        print("Soporte: " + str(item[1]))

        #El tercer índice de la lista
        print("Confianza: " + str(item[2][0][2]))
        print("Elevación: " + str(item[2][0][3]))
        print("=====")
```



```

Regla: frozenset({'Jumanji', 'Ninja Turtles'})
Soporte: 0.04115281501340483
Confianza: 0.4171195652173913
Elevación: 1.742279930863236
=====

```

```

Regla: frozenset({'Jumanji', 'Tomb Rider'})
Soporte: 0.03941018766756032
Confianza: 0.3994565217391304
Elevación: 2.283483258370814
=====

```

```

Regla: frozenset({'Moana', 'Ninja Turtles'})
Soporte: 0.04825737265415549
Confianza: 0.3707518022657054
Elevación: 1.5486049523528347
=====

```

Análisis de la información

Ahora que sabemos que la información relacionada de nuestro interés puede ser ordenada y colocada de distintas maneras para su cómoda lectura, comencemos a analizar los datos que la configuración 4 nos dejó para el caso de relación entre las películas "Jumanji" y "Ninja Turtles":

In [36]: `print(ResultadosC4[0])`

```

RelationRecord(items=frozenset({'Jumanji', 'Ninja Turtles'}), support=0.04115281501340483, ordered_statistics=[OrderedStatistic(items_base=frozenset({'Jumanji'}), items_add=frozenset({'Ninja Turtles'}), confidence=0.4171195652173913, lift=1.742279930863236)])

```

Con esta tupla de datos, podemos hacernos con una conclusión similar a la vista en la configuración 1; dado que ambas películas están centradas en un mismo público familiar, poseen cierta relación que se puede observar en sus valores calculados de soporte, confianza y elevación:

```

soporte: 0.04115 (4.1%)
confianza: 0.417119 (41.71%)
elevación: 1.74

```

run cell, select below

Estos datos nos dan la información necesaria para afirmar que existen 2 veces más probabilidad de que una persona que rente una de las películas de la tupla, haga lo mismo con la otra.

Reflexión Final

• Recinos Hernández Luis Mario

Con el desarrollo de esta práctica pude dar los primeros pasos para familiarizarme con el entorno y herramientas de trabajo que se utilizarán a lo largo de las sesiones de laboratorio para procesar y analizar distintos tipos de datos para establecer una relación entre desarrollador(a) y máquina en aras de fomentar un aprendizaje regularizado de nuestros procesos mediante la implementación de algoritmos y métodos de análisis de información. En esta primera práctica tuve la oportunidad de obtener información desde un archivo .csv con información normalizada y colocarla apropiadamente en el software de anaconda para su aprovechamiento afectivo. De igual forma, pude empedar dentro de este mismo software diferentes aplicaciones del algoritmo a priori y de este, obtener conclusiones sobre las relaciones que existían entre la información de renta de películas dentro del archivo dado por el profesor. Por todos los conocimientos y técnicas aplicadas para la resolución de esta práctica, puedo afirmar que los objetivos de esta práctica fueron satisfactoriamente cumplidos.

- **Alcántara Guerrero Alfredo Guadalupe**

El objetivo de la práctica se cumplió debido a que por medio de los comandos de Python entendí como leer datos a partir de un dataset y guardarlos en una variable para manipularlos, luego a partir de que generamos una gráfica para saber la distribución de los datos, sabiendo la distribución establecimos los parámetros de soporte, confianza y elevación para podar las reglas de asociación dejando únicamente las reglas más significativas, luego de esas reglas escogimos las primeras 3 para darles una interpretación.

Podemos hacer uso de la información obtenida por las reglas de asociación para reorganizar las películas, ya sea para dejar cerca las películas que suelen rentarse juntas o dejarlas separadas para obligar a los clientes a recorrer la tienda.

Bibliografía

- Pacini, H. (2011). Issues in the Path to the Singularity. Journal of Consciousness Exploration & Research. Volumen (2), pp. 691-705
- Braga, A. Y Logan, R. (2019). AI and the Singularity: A Fallacy or a Great Opportunity? Departamento de Comunicación Social de la Universidad Católica de Rio de Janeiro.