

## Aprendizaje de Maquina No Supervisado

# KNN & Market Basket

## KNN

Dada la base de datos 'wine-clustering' y una **muestra de un vino externo**, que se muestra en las imágenes. Se aplico un algoritmo **KNN** con un numero de vecinos ajustado a **5**, y enfocado al nivel de Alcohol de las observaciones.

```
wines_data = pd.read_csv('wine-clustering.csv')
wines_data
```

	Alcohol	Malic_Acid	Ash	Ash_Alcanity	Magnesium	Total_Phenols	Flavanoids	Nonflavanoid_Phenols	Proanthocyanins	Color_Intensity	Hue	OD280
0	14.23	1.71	2.43	15.6	127	2.80	3.06	0.28	2.29	5.64	1.04	3.92
1	13.20	1.78	2.14	11.2	100	2.65	2.76	0.26	1.28	4.38	1.05	3.40
2	13.16	2.36	2.67	18.6	101	2.80	3.24	0.30	2.81	5.68	1.03	3.17
3	14.37	1.95	2.50	16.8	113	3.85	3.49	0.24	2.18	7.80	0.86	3.45
4	13.24	2.59	2.87	21.0	118	2.80	2.69	0.39	1.82	4.32	1.04	2.93
...	...	...	...	...	...	...	...	...	...	...	...	...
173	13.71	5.65	2.45	20.5	95	1.68	0.61	0.52	1.06	7.70	0.64	1.74
174	13.40	3.91	2.48	23.0	102	1.80	0.75	0.43	1.41	7.30	0.70	1.56
175	13.27	4.28	2.26	20.0	120	1.59	0.69	0.43	1.35	10.20	0.59	1.56
176	13.17	2.59	2.37	20.0	120	1.65	0.68	0.53	1.46	9.30	0.60	1.62
177	14.13	4.10	2.74	24.5	96	2.05	0.76	0.56	1.35	9.20	0.61	1.60

Wine Clustering

```
test = pd.DataFrame({
    'Alcohol': [14], 'Malic_Acid': [2], 'Ash': [2.5], 'Ash_Alcanity': [16], 'Magnesium': [115],
    'Total_Phenols': [3], 'Flavanoids': [2.5], 'Nonflavanoid_Phenols': [0.4],
    'Proanthocyanins': [2], 'Color_Intensity': [9], 'Hue': [1], 'OD280': [3.5], 'Proline': [800]
})
```

Vino de prueba

Los datos arrojados por el algoritmo nos muestran que los vecinos mas cercanos a esta observación tienen un promedio de **13.39** de nivel de Alcohol.

```
alcohol_mean = neighbors['Alcohol'].mean()
print('Concentracion del Alcohol de los 5 vecinos mas cercanos: ')
for i, value in enumerate(neighbors['Alcohol'], start=1):
    print(f'{i}. {value}')

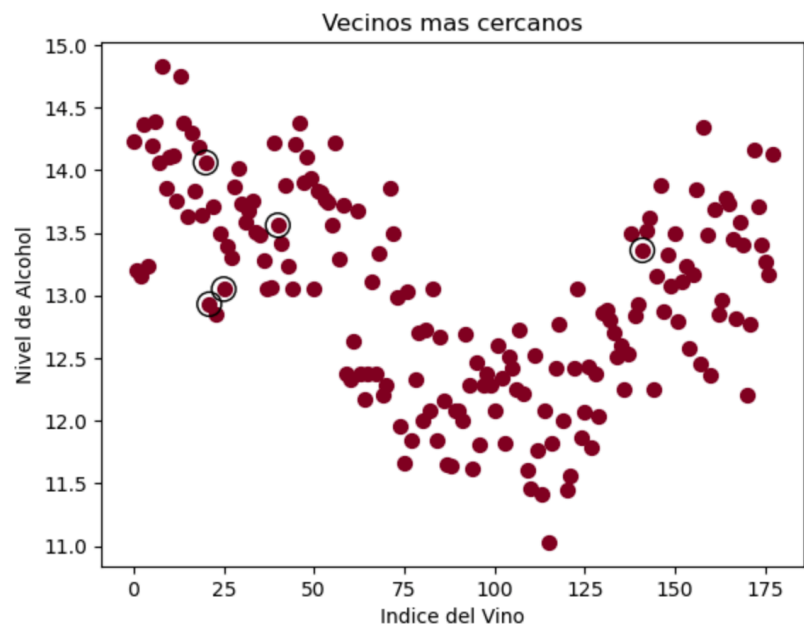
print(f'Promedio de Alcohol de los {k} vecinos mas cercanos: {alcohol_mean:.2f}')
Concentracion del Alcohol de los 5 vecinos mas cercanos:
1. 13.56
2. 14.06
3. 13.05
4. 12.93
5. 13.36
Promedio de Alcohol de los 5 vecinos mas cercanos: 13.39
```

Vecinos mas cercanos

Ubicando en un scatter plot, que grafica el nivel de alcohol y el indice del vino de acuerdo al dataset, obtenemos la siguiente observación.

```
neighbor_indices = indices[0]
neighbors = wines_data.iloc[neighbor_indices]

# Visualizacion de los 5 vecinos mas cercanos
plt.figure()
plt.title('Vecinos mas cercanos')
plt.scatter(wines_data.index, wines_data['Alcohol'], marker = 'o', s = 50, color = '#800020')
plt.scatter(neighbors.index, neighbors['Alcohol'], marker = 'o', s = 150, color = 'black', facecolors = 'none')
plt.xlabel('Indice del Vino')
plt.ylabel('Nivel de Alcohol')
plt.show()
```



Grafica de los vecinos mas cercanos encerrados en negro

Si bien la grafica puede ser un tanto confusa, pues no muestra los grupos separados correctamente, ya que se esta graficando de acuerdo al indice, debido a que la base de datos es de multiple dimension, y habría que reducirla mediante *PCA*, al menos nos da una idea de donde se ubican los vecinos, y donde podría estar ubicada nuestra muestra externa, por encima del **13** y debajo del **14**.

## Market Basket

En base a un dataset pequeño que contiene las compras de 11 clientes, al **binarizar** los datos, y aplicar un análisis **A priori** y de **reglas de asociación**, notamos como la compra de cerveza (**beer**) lleva a la compra de papas (**chips**),

siendo la **combinación líder**, y justo después tenemos pan (**bread**) con mantequilla (**butter**) y cebollas (**onions**) con tomates (**tomatoes**).

```
association_rules(articulos_frecuentes, metric = 'lift',
                  min_threshold = 0.5).sort_values('lift', ascending = False).reset_index(drop = True)
```

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	representativity	leverage	conviction	zhangs_metric	jaccard	certa
0	(beer)	(chips, milk)	0.363636	0.272727	0.272727	0.750000	2.750000	1.0	0.173554	2.909091	1.000000	0.750000	0.656
1	(chips, milk)	(beer)	0.272727	0.363636	0.272727	1.000000	2.750000	1.0	0.173554	inf	0.875000	0.750000	1.000
2	(chips)	(beer)	0.454545	0.363636	0.363636	0.800000	2.200000	1.0	0.198347	3.181818	1.000000	0.800000	0.685
3	(beer)	(chips)	0.363636	0.454545	0.363636	1.000000	2.200000	1.0	0.198347	inf	0.857143	0.800000	1.000
4	(chips)	(milk, beer)	0.454545	0.272727	0.272727	0.600000	2.200000	1.0	0.148760	1.818182	1.000000	0.600000	0.450
5	(milk, beer)	(chips)	0.272727	0.454545	0.272727	1.000000	2.200000	1.0	0.148760	inf	0.750000	0.600000	1.000
6	(butter)	(bread)	0.545455	0.363636	0.363636	0.666667	1.833333	1.0	0.165289	1.909091	1.000000	0.666667	0.476
7	(onions)	(tomatoes)	0.272727	0.545455	0.272727	1.000000	1.833333	1.0	0.123967	inf	0.625000	0.500000	1.000
8	(tomatoes)	(onions)	0.545455	0.272727	0.272727	0.500000	1.833333	1.0	0.123967	1.454545	1.000000	0.500000	0.312
9	(bread)	(butter)	0.363636	0.545455	0.363636	1.000000	1.833333	1.0	0.165289	inf	0.714286	0.666667	1.000
10	(butter)	(carrots)	0.545455	0.272727	0.272727	0.500000	1.833333	1.0	0.123967	1.454545	1.000000	0.500000	0.312
11	(carrots)	(butter)	0.272727	0.545455	0.272727	1.000000	1.833333	1.0	0.123967	inf	0.625000	0.500000	1.000
12	(chips, beer)	(milk)	0.363636	0.454545	0.272727	0.750000	1.650000	1.0	0.107438	2.181818	0.619048	0.500000	0.541
13	(beer)	(milk)	0.363636	0.454545	0.272727	0.750000	1.650000	1.0	0.107438	2.181818	0.619048	0.500000	0.541
14	(milk)	(beer)	0.454545	0.363636	0.272727	0.600000	1.650000	1.0	0.107438	1.590909	0.722222	0.500000	0.371
15	(milk)	(chips, beer)	0.454545	0.363636	0.272727	0.600000	1.650000	1.0	0.107438	1.590909	0.722222	0.500000	0.371
16	(butter)	(coffee)	0.545455	0.363636	0.272727	0.500000	1.375000	1.0	0.074380	1.272727	0.600000	0.428571	0.214
17	(coffee)	(butter)	0.363636	0.545455	0.272727	0.750000	1.375000	1.0	0.074380	1.818182	0.428571	0.428571	0.450
18	(chips)	(milk)	0.454545	0.454545	0.272727	0.600000	1.320000	1.0	0.066116	1.363636	0.444444	0.428571	0.266
19	(milk)	(chips)	0.454545	0.454545	0.272727	0.600000	1.320000	1.0	0.066116	1.363636	0.444444	0.428571	0.266
20	(tomatoes)	(milk)	0.545455	0.454545	0.272727	0.500000	1.100000	1.0	0.024793	1.090909	0.200000	0.375000	0.083
21	(milk)	(tomatoes)	0.454545	0.545455	0.272727	0.600000	1.100000	1.0	0.024793	1.136364	0.166667	0.375000	0.120

Reglas de asociación

Esto nos da claridad de como en un mercado se pueden acomodar los anaqueles para **propiciar la compra de estos productos**, o bien, dar promociones. En el caso de la cerveza, por ejemplo, en la compra de 3 productos se regala 1 paquete de papas, o similar, beneficiando las necesidades de los consumidores, y por supuesto del negocio.