



**Escuela Superior
de Ingeniería y Tecnología**
Universidad de La Laguna

Gestión del Conocimiento en las Organizaciones

Sistemas de recomendación. Métodos de
filtrado colaborativo

Alberto Oria Martín

(alu0101415598@ull.edu.es)

Izan Pérez González

(alu0101543345@ull.edu.es)

Airam Prieto González

(alu0101546377@ull.edu.es)



Matrices empleadas.....	3
Comportamiento general.....	3
Matrices pequeñas (5×10 y 10×25).....	3
Matrices medianas y grandes (25×100, 50×250, 100×1000).....	3
Ejemplos de ejecución.....	4
Matriz 5x10.....	5
Pearson.....	5
Distancia coseno.....	6
Distancia euclídea.....	7
Matriz 10x25.....	8
Pearson.....	8
Distancia coseno.....	9
Distancia euclídea.....	10
Comparativa entre métodos de predicción.....	11
Resultados y tendencias generales.....	12
Tabla descriptiva.....	12
Conclusión final.....	13



Matrices empleadas

- [utility-matrix-5-10-1.txt](#)
- [utility-matrix-10-25-1.txt](#)
- [utility-matrix-25-100-1.txt](#)
- [utility-matrix-50-250-1.txt](#)
- [utility-matrix-100-1000-1.txt](#)

Comportamiento general

Matrices pequeñas (5×10 y 10×25)

En estas matrices con pocos ítems por usuario, el **coeficiente de correlación de Pearson** mostró resultados inestables: similitudes con signo negativo y predicciones fuera de rango si no se filtran las correlaciones negativas. Esto ocurre porque **la correlación requiere un número suficiente de ítems comunes**; con pocos valores compartidos, la varianza es poco representativa y el resultado se vuelve ruidoso.

La **distancia coseno** produjo similitudes positivas y más coherentes con los valores originales. Al basarse en la orientación de los vectores, ignora las diferencias de escala entre usuarios, lo que genera predicciones más estables ante la falta de datos o valores atípicos.

La **distancia euclídea**, al transformarse en similitud inversa, arrojó similitudes bajas y homogéneas. Sus predicciones fueron conservadoras y amortiguadas, evitando extremos y reduciendo el efecto de ratings aislados, aunque con menor capacidad de discriminación.

Matrices medianas y grandes (25×100, 50×250, 100×1000)

A medida que la matriz crece, la **distancia coseno** se mantiene estable y ofrece similitudes coherentes entre usuarios con comportamientos parecidos.

El **coeficiente de Pearson** mejora su fiabilidad cuando hay densidad moderada (por ejemplo, en 25×100), pero se degrada con el aumento de sparsity en 50×250 y 100×1000, mostrando correlaciones cercanas a cero debido a la falta de coincidencias suficientes.

La **distancia euclídea**, aunque estable, sufre el fenómeno de *concentración de*



distancias en espacios de alta dimensión, lo que reduce su capacidad de distinguir entre usuarios realmente distintos.

Nota: Las matrices 25×100, 50×250 y 100×1000 no se incluyen visualmente por limitaciones de espacio, pero sus resultados siguieron las mismas tendencias observadas: mejora progresiva de Pearson con densidad moderada y estabilidad de Coseno y Euclídea en entornos dispersos.

Ejemplos de ejecución

Las capturas mostradas a continuación muestran los resultados de cada métrica, la primera con la predicción simple y la segunda usando mean difference.

Matriz 5x10

Pearson

Matriz de utilidad completada:										
	Ítem 0	Ítem 1	Ítem 2	Ítem 3	Ítem 4	Ítem 5	Ítem 6	Ítem 7	Ítem 8	Ítem 9
Usuario 0	3.14	2.65	1.65	-1.51	1.12	0.88	0.42	3.98	1.8	3.14
Usuario 1	3.41	0.31	3.8	4.23	2.16	4.51	2.39	0.87	2.47	-0.44
Usuario 2	4.41	4.5	2.05	-1.45	0.05	0.42	3.35	3.74	4.08	2.86
Usuario 3	1.73	1.18	-0.19	1.51	4.87	2.22	3	2.9	2.11	-1.32
Usuario 4	0.56	4.89	1.22	0.8	3.8	4.88	2.83	0.99	4.49	0.44

Tabla de similitudes entre usuarios según la métrica pearson:					
	Usuario 0	Usuario 1	Usuario 2	Usuario 3	Usuario 4
Usuario 0	0	-0.48	0.56	-0.33	-0.58
Usuario 1	-0.48	0	-0.26	-0.55	-0.24
Usuario 2	0.56	-0.26	0	-0.97	-0.05
Usuario 3	-0.33	-0.55	-0.97	0	0.34
Usuario 4	-0.58	-0.24	-0.05	0.34	0



Matriz de utilidad completada:

	Ítem 0	Ítem 1	Ítem 2	Ítem 3	Ítem 4	Ítem 5	Ítem 6	Ítem 7	Ítem 8	Ítem 9
Usuario 0	3.14	2.65	1.65	3.23	1.12	0.88	0.42	3.98	3.01	3.14
Usuario 1	3.41	0.31	3.8	4.23	2.16	4.51	2.39	0.87	2.47	4.74
Usuario 2	4.41	4.5	2.05	4.14	0.05	1.78	3.35	3.74	4.08	4.04
Usuario 3	1.73	3.68	2.44	1.51	4.87	2.22	3	2.9	2.11	1.27
Usuario 4	0.56	4.89	1.22	0.8	3.8	4.88	2.83	0.99	4.49	0.44

Tabla de similitudes entre usuarios según la métrica pearson:

	Usuario 0	Usuario 1	Usuario 2	Usuario 3	Usuario 4
Usuario 0	0	-0.48	0.56	-0.33	-0.58
Usuario 1	-0.48	0	-0.26	-0.55	-0.24
Usuario 2	0.56	-0.26	0	-0.97	-0.05
Usuario 3	-0.33	-0.55	-0.97	0	0.34
Usuario 4	-0.58	-0.24	-0.05	0.34	0

- Predicciones a veces negativas.
- Alta dispersión entre usuarios.
- Inestable por escasez de ítems en común.

Distancia coseno

Matriz de utilidad completada:

	Ítem 0	Ítem 1	Ítem 2	Ítem 3	Ítem 4	Ítem 5	Ítem 6	Ítem 7	Ítem 8	Ítem 9
Usuario 0	3.14	2.65	1.65	1.51	1.12	0.88	0.42	3.98	3.23	3.14
Usuario 1	3.41	0.31	3.8	4.23	2.16	4.51	2.39	0.87	2.47	-
Usuario 2	4.41	4.5	2.05	1.19	0.05	2.71	3.35	3.74	4.08	1.9
Usuario 3	1.73	2.7	2.45	1.51	4.87	2.22	3	2.9	2.11	0.44
Usuario 4	0.56	4.89	1.22	0.8	3.8	4.88	2.83	0.99	4.49	0.44

Tabla de similitudes entre usuarios según la métrica coseno:

	Usuario 0	Usuario 1	Usuario 2	Usuario 3	Usuario 4
Usuario 0	0	0.62	0.9	0.69	0.54
Usuario 1	0.62	0	0.75	0.78	0.69
Usuario 2	0.9	0.75	0	0.69	0.76
Usuario 3	0.69	0.78	0.69	0	0.85
Usuario 4	0.54	0.69	0.76	0.85	0



Matriz de utilidad completada:

	Ítem 0	Ítem 1	Ítem 2	Ítem 3	Ítem 4	Ítem 5	Ítem 6	Ítem 7	Ítem 8	Ítem 9
Usuario 0	3.14	2.65	1.65	1.01	1.12	0.88	0.42	3.98	2.3	3.14
Usuario 1	3.41	0.31	3.8	4.23	2.16	4.51	2.39	0.87	2.47	2.68
Usuario 2	4.41	4.5	2.05	1.85	0.05	3.48	3.35	3.74	4.08	2.72
Usuario 3	1.73	2.73	2.5	1.51	4.87	2.22	3	2.9	2.11	1.55
Usuario 4	0.56	4.89	1.22	0.8	3.8	4.88	2.83	0.99	4.49	0.44

Tabla de similitudes entre usuarios según la métrica coseno:

	Usuario 0	Usuario 1	Usuario 2	Usuario 3	Usuario 4
Usuario 0	0	0.62	0.9	0.69	0.54
Usuario 1	0.62	0	0.75	0.78	0.69
Usuario 2	0.9	0.75	0	0.69	0.76
Usuario 3	0.69	0.78	0.69	0	0.85
Usuario 4	0.54	0.69	0.76	0.85	0

- Predicciones positivas y consistentes con el rango original.
- Suaviza el efecto de valores extremos.
- Mide la orientación de los ratings, no su magnitud.



Distancia euclídea

Matriz de utilidad completada:										
	Ítem 0	Ítem 1	Ítem 2	Ítem 3	Ítem 4	Ítem 5	Ítem 6	Ítem 7	Ítem 8	Ítem 9
Usuario 0	3.14	2.65	1.65	1.51	1.12	0.88	0.42	3.98	3.23	3.14
Usuario 1	3.41	0.31	3.8	4.23	2.16	4.51	2.39	0.87	2.47	-
Usuario 2	4.41	4.5	2.05	1.51	0.05	1.41	3.35	3.74	4.08	3.14
Usuario 3	1.73	3.84	1.42	1.51	4.87	2.22	3	2.9	2.11	1.7
Usuario 4	0.56	4.89	1.22	0.8	3.8	4.88	2.83	0.99	4.49	0.44

Tabla de similitudes entre usuarios según la métrica euclidean:					
	Usuario 0	Usuario 1	Usuario 2	Usuario 3	Usuario 4
Usuario 0	0	0.16	0.26	0.2	0.13
Usuario 1	0.16	0	0.16	0.19	0.14
Usuario 2	0.26	0.16	0	0.17	0.16
Usuario 3	0.2	0.19	0.17	0	0.23
Usuario 4	0.13	0.14	0.16	0.23	0

Matriz de utilidad completada:										
	Ítem 0	Ítem 1	Ítem 2	Ítem 3	Ítem 4	Ítem 5	Ítem 6	Ítem 7	Ítem 8	Ítem 9
Usuario 0	3.14	2.65	1.65	1.01	1.12	0.88	0.42	3.98	2.3	3.14
Usuario 1	3.41	0.31	3.8	4.23	2.16	4.51	2.39	0.87	2.47	2.68
Usuario 2	4.41	4.5	2.05	2.12	0.05	2.19	3.35	3.74	4.08	4.06
Usuario 3	1.73	4.19	1.96	1.51	4.87	2.22	3	2.9	2.11	2.14
Usuario 4	0.56	4.89	1.22	0.8	3.8	4.88	2.83	0.99	4.49	0.44

Tabla de similitudes entre usuarios según la métrica euclidean:					
	Usuario 0	Usuario 1	Usuario 2	Usuario 3	Usuario 4
Usuario 0	0	0.16	0.26	0.2	0.13
Usuario 1	0.16	0	0.16	0.19	0.14
Usuario 2	0.26	0.16	0	0.17	0.16
Usuario 3	0.2	0.19	0.17	0	0.23
Usuario 4	0.13	0.14	0.16	0.23	0

- Predicciones más conservadoras.
- Similitudes bajas pero consistentes.
- Reduce el impacto de outliers individuales.

Análisis:

Pearson muestra su limitación en conjuntos pequeños; Coseno ofrece



predicciones razonables y Euclídea mantiene estabilidad pero poca sensibilidad entre usuarios.

Matriz 10x25

Pearson

Matriz de utilidad completada:																									
	Item 0	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14	Item 15	Item 16	Item 17	Item 18	Item 19	Item 20	Item 21	Item 22	Item 23	Item 24
Usuario 0	0.82	0.41	0.49	1.31	2.90	0.48	0.03	0.37	0.48	1.3	1.08	0.34	0.34	1.4	0.25	2.72	1.28	0.33	0.48	0.39	1.99	0.28	1.37	0.33	1.32
Usuario 1	1.11	0.72	1.19	1.02	0.95	0.4	1.5	3.11	0.99	2.05	3.08	0.75	2.8	0.85	2.71	2.28	0.43	0.28	0.42	1.03	0.38	0.76	3.97	1.26	2.84
Usuario 2	1.02	1.27	1.6	3.12	0.25	0.29	0.4	0.88	0.58	0.91	2.62	3.21	3.21	2.19	2.27	2.44	0.43	0.48	0.89	2.07	3.22	0.57	0.51	0.48	1.22
Usuario 3	0.91	0.44	0.79	3.18	0.23	1.99	0.01	0.79	0.8	0.82	0.28	0.01	1.18	2.08	0.2	1.4	0.76	0.42	2.02	1.93	3.08	1.5	1.34	0.37	
Usuario 4	0.74	3.17	2.22	2.22	0.89	2.08	1.59	2.09	0.11	0.8	0.08	1.07	2.72	2.72	1.91	3.28	0.48	3.09	0.48	0.92	1.1	3.11	2.15	2.48	2.12
Usuario 5	1.01	3.98	3.38	3.11	3.91	0.86	0.11	0.44	1.98	1.47	3.21	3.34	1.58	0.82	0.44	2.11	2.82	0.99	0.05	1.99	0.26	0.7	0.62	0.99	
Usuario 6	3.18	0.71	3.08	2.84	2.84	1.88	0.97	2.83	0.52	0.25	0.34	0.41	2.95	2.92	0.99	2.84	1.93	2.7	0.98	1.35	1.99	1.54	2.73	0.43	3.08
Usuario 7	0.97	0.14	3.17	0.48	3.09	1.82	2.21	1.24	1.95	0.77	1.25	0.39	0.32	0.99	0.72	0.4	2.25	1.38	0.85	0.42	0.36	0.88	2.1	1.12	0.44
Usuario 8	1.44	0.88	1.38	2.51	0.14	1.6	2.35	3.41	1.73	0.74	3.38	0.5	0.5	1.84	2.0	1.89	0.22	1.54	0.9	1.31	0.21	0.33	0.88	2.89	1.98
Usuario 9	3.08	3.18	2.99	2.59	0.83	2.12	0.18	0.5	0.84	3.82	0.28	0.45	3.01	3.93	2.43	1.48	1.88	1.18	0.68	2.05	0.51	2.53	0	1.78	0.33
Tabla de similitudes entre usuarios según la métrica pearson:																									
	Usuario 0	Usuario 1	Usuario 2	Usuario 3	Usuario 4	Usuario 5	Usuario 6	Usuario 7	Usuario 8	Usuario 9															
Usuario 0	0	0.21	0.01	0.22	0.22	0.17	0.01	0.01	-0.12	-0.23															
Usuario 1	0.21	0	-0.84	0.05	-0.88	0.05	0.13	0.09	-0.02	-0.14															
Usuario 2	0.01	-0.84	0	-0.05	-0.17	-0.14	-0.14	0.05	0.13	0.14															
Usuario 3	0.22	0.05	-0.05	0	-0.2	-0.01	-0.2	0.29	-0.11	0.13															
Usuario 4	0.22	-0.88	-0.17	-0.2	0	0.13	0.22	-0.09	-0.14	-0.17															
Usuario 5	0.17	0.05	-0.14	-0.01	0.13	0	0.14	-0.01	-0.15	0.25															
Usuario 6	0.01	0.13	-0.14	-0.2	0.22	0.14	0	-0.14	-0.05	-0.12															
Usuario 7	0.01	0.09	0.05	0.29	-0.09	-0.01	-0.14	0	-0.09	-0.45															
Usuario 8	-0.12	-0.02	0.13	-0.11	-0.24	-0.15	-0.05	-0.09	0	0.12															
Usuario 9	-0.23	-0.14	0.14	0.13	-0.17	0.25	-0.12	-0.45	0.12	0															

- Muy sensible a valores atípicos o vecinos con datos faltantes.
- Puede dar valores indefinidos o erráticos (altos o negativos).
- Útil sólo cuando existe suficiente solapamiento entre usuarios.
- Funciona mejor con la fórmula de predicción basada en desviaciones respecto a la media.



Distancia coseno

Matriz de utilidos completada:																									
	Item 0	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14	Item 15	Item 16	Item 17	Item 18	Item 19	Item 20	Item 21	Item 22	Item 23	Item 24
Usuario 0	0	0.81	0.72	0.77	0.79	0.8	0.89	0.73	0.68	0.68															
Usuario 1	0.81	0	0.8	0.76	0.8	0.77	0.87	0.79	0.79	0.74															
Usuario 2	0.72	0.8	0	0.75	0.71	0.69	0.77	0.77	0.84	0.82															
Usuario 3	0.77	0.76	0.75	0	0.71	0.7	0.71	0.81	0.65	0.78															
Usuario 4	0.79	0.8	0.71	0.71	0	0.76	0.86	0.7	0.71	0.75															
Usuario 5	0.8	0.77	0.69	0.7	0.76	0	0.83	0.69	0.68	0.8															
Usuario 6	0.89	0.87	0.77	0.71	0.88	0.83	0	0.74	0.8	0.8															
Usuario 7	0.73	0.79	0.77	0.81	0.7	0.69	0.74	0	0.7	0.63															
Usuario 8	0.68	0.79	0.84	0.65	0.71	0.68	0.8	0.7	0	0.8															
Usuario 9	0.68	0.74	0.82	0.78	0.75	0.8	0.8	0.63	0.8	0															

Matriz de utilidos completada:																									
	Item 0	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14	Item 15	Item 16	Item 17	Item 18	Item 19	Item 20	Item 21	Item 22	Item 23	Item 24
Usuario 0	0.82	4.41	4.49	3.33	2.98	4.48	0.83	4.37	3.48	3.3	3.68	3.95	0.34	2.4	4.25	2.75	2.28	0.33	2.25	4.39	2.88	0.38	3.37	0.33	3.32
Usuario 1	3.31	4.72	3.39	3.02	2.88	4.4	3.5	3.33	3.89	2.05	3.48	4.75	2.8	0.35	2.75	2.38	2.43	3.28	0.42	3.33	3.38	0.93	3.97	3.38	2.84
Usuario 2	3.82	3.27	3.6	3.35	3.32	4.25	4.39	4.4	0.58	0.58	4.91	2.43	3.21	2.35	2.37	2.44	4.83	0.89	2.07	3.35	4.57	0.01	4.48	3.22	4.78
Usuario 3	2.4	4.44	3.79	3.18	3.23	3.99	4.02	4.79	2.68	4.85	4.36	0.29	0.03	3.38	2.48	4.5	3.4	0.76	0.42	2.02	2.93	3.86	3.5	3.34	0.47
Usuario 4	3.03	3.98	3.38	3.11	3.88	0.88	0.33	0.44	3.98	3.47	3.23	3.58	3.27	4.04	2.33	2.18	2.29	2.82	0.99	4.05	3.99	0.26	4.7	0.62	0.09
Usuario 5	3.38	4.73	3.98	3.43	2.84	3.88	0.07	3.83	3.28	0.55	4.34	4.43	2.95	3.92	4.99	2.84	3.93	2.7	4.98	3.55	3.18	3.54	2.73	0.78	3.08
Usuario 6	0.97	4.34	2.95	4.45	3.09	3.82	2.21	3.24	3.95	4.77	3.25	4.39	0.52	0.72	4.72	4.4	2.23	3.99	0.05	0.42	0.38	4.08	2.72	3.3	4.04
Usuario 7	3.44	0.08	3.38	2.51	4.24	3.8	2.35	3.43	3.73	4.74	3.38	4.5	4.0	3.84	2.0	3.09	4.22	3.54	4.9	3.33	0.23	0.33	4.08	2.89	3.98
Usuario 8	3.68	3.18	2.99	2.59	4.03	2.32	4.38	0.5	4.04	3.82	4.28	0.45	3.03	3.93	2.43	3.48	3.08	3.38	4.68	2.05	4.51	2.53	4	3.78	0.33
Tabla de similitudes entre usuarios según la métrica coseno:																									
	Usuario 0	Usuario 1	Usuario 2	Usuario 3	Usuario 4	Usuario 5	Usuario 6	Usuario 7	Usuario 8	Usuario 9															
Usuario 0	0	0.81	0.72	0.77	0.79	0.8	0.89	0.73	0.68	0.68															
Usuario 1	0.81	0	0.8	0.76	0.8	0.77	0.87	0.79	0.79	0.74															
Usuario 2	0.72	0.8	0	0.75	0.71	0.69	0.77	0.77	0.84	0.82															
Usuario 3	0.77	0.76	0.75	0	0.71	0.7	0.71	0.81	0.65	0.78															
Usuario 4	0.79	0.8	0.71	0.71	0	0.76	0.86	0.7	0.71	0.75															
Usuario 5	0.8	0.77	0.69	0.7	0.76	0	0.83	0.69	0.68	0.8															
Usuario 6	0.89	0.87	0.77	0.71	0.88	0.83	0	0.74	0.8	0.8															
Usuario 7	0.73	0.79	0.77	0.81	0.7	0.69	0.74	0	0.7	0.63															
Usuario 8	0.68	0.79	0.84	0.65	0.71	0.68	0.8	0.7	0	0.8															
Usuario 9	0.68	0.74	0.82	0.78	0.75	0.8	0.8	0.63	0.8	0															

- Ignora la magnitud absoluta, centrando la comparación en la dirección.
- Genera resultados más estables ante sparsity.
- No distingue patrones inversos (usuarios con preferencias opuestas).



Distancia euclídea

Matriz de utilidad completada:																									
	Item 0	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14	Item 15	Item 16	Item 17	Item 18	Item 19	Item 20	Item 21	Item 22	Item 23	Item 24
Usuario 0	0	0.12	0.1	0.11	0.12	0.12	0.16	0.11	0.09	0.09															
Usuario 1	0.12	0	0.11	0.11	0.12	0.11	0.13	0.12	0.11	0.1															
Usuario 2	0.1	0.11	0	0.1	0.09	0.09	0.1	0.1	0.12	0.11															
Usuario 3	0.11	0.11	0.1	0	0.1	0.1	0.09	0.12	0.08	0.11															
Usuario 4	0.12	0.12	0.09	0.1	0	0.12	0.13	0.1	0.1	0.1															
Usuario 5	0.12	0.11	0.09	0.1	0.12	0	0.12	0.11	0.09	0.11															
Usuario 6	0.16	0.13	0.1	0.09	0.11	0.12	0	0.1	0.1	0.11															
Usuario 7	0.11	0.12	0.1	0.12	0.1	0.11	0.1	0	0	0.09	0.08														
Usuario 8	0.09	0.11	0.12	0.08	0.1	0.09	0.1	0.09	0	0.1															
Usuario 9	0.09	0.1	0.11	0.11	0.1	0.11	0.11	0.08	0.1	0															

Matriz de utilidad completada:																									
	Item 0	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14	Item 15	Item 16	Item 17	Item 18	Item 19	Item 20	Item 21	Item 22	Item 23	Item 24
Usuario 0	0.82	4.01	4.49	3.31	2.96	4.48	0.01	4.37	4.48	3.3	3.08	**2.76**	0.16	3.4	4.25	2.75	3.28	0.13	**2.82**	4.19	**2.18**	0.16	3.57	0.13	3.12
Usuario 1	3.31	4.72	3.19	3.02	**2.03**	4.4	3.5	3.31	3.89	2.05	3.48	4.75	2.8	0.25	2.71	2.26	2.43	3.26	0.42	3.33	3.18	**0.07**	3.07	1.26	2.84
Usuario 2	4.02	3.27	3.6	3.15	4.25	4.39	4.4	0.56	0.58	4.91	2.43	3.31	3.21	2.35	2.37	2.44	4.83	0.89	2.07	3.15	4.57	0.01	**4.46**	1.22	4.78
Usuario 3	**0.97**	4.44	3.79	3.18	4.23	3.99	4.02	4.79	**0.49**	4.05	4.26	0.29	0.03	3.18	2.46	4.5	3.4	0.76	0.42	2.02	2.93	3.06	3.5	1.34	0.47
Usuario 4	0.74	3.37	3.26	2.23	0.09	2.08	3.56	2.09	**0.12**	3.4	0.08	1.67	2.72	3.75	3.91	3.18	0.48	3.09	4.48	4.91	3.1	3.11	**2.6**	2.49	3.43
Usuario 5	1.03	3.06	3.38	3.11	3.91	0.86	0.33	0.44	3.96	3.47	3.23	**3.28**	4.04	2.11	**2.29**	2.79	2.82	0.99	4.08	1.99	0.26	4.7	0.62	0.09	
Usuario 6	3.38	4.71	3.98	**3.03**	2.04	3.88	0.07	3.03	**0.02**	0.55	4.34	4.41	2.95	3.02	4.09	2.84	3.95	2.7	4.96	3.55	**0.09**	3.54	2.73	**3.44**	3.06
Usuario 7	0.97	4.34	**2.53**	4.46	3.09	3.82	2.21	3.24	3.95	4.77	3.25	4.39	**0.52**	**0.76**	4.72	4.4	2.23	**1.05**	0.05	0.42	0.36	4.08	**3.03**	**3.19**	4.04
Usuario 8	3.44	0.06	3.38	2.51	4.34	3.6	2.35	3.43	3.73	4.74	3.28	4.5	4.6	3.84	2.6	3.09	4.22	3.54	4.9	3.31	0.23	0.13	4.06	2.89	3.96
Usuario 9	3.08	3.18	2.99	2.59	4.03	2.12	4.18	0.5	4.04	3.82	4.28	0.45	3.01	3.93	2.43	3.46	3.00	3.16	4.08	2.05	4.51	2.53	4	3.76	0.33
Tabla de similitudes entre usuarios según la métrica euclídea:																									
	Usuario 0	Usuario 1	Usuario 2	Usuario 3	Usuario 4	Usuario 5	Usuario 6	Usuario 7	Usuario 8	Usuario 9															
Usuario 0	0	0.12	0.1	0.11	0.12	0.12	0.16	0.11	0.09	0.09															
Usuario 1	0.12	0	0.11	0.11	0.12	0.11	0.13	0.12	0.11	0.1															
Usuario 2	0.1	0.11	0	0.1	0.09	0.09	0.1	0.1	0.12	0.11															
Usuario 3	0.11	0.11	0.1	0	0.1	0.1	0.09	0.12	0.08	0.11															
Usuario 4	0.12	0.12	0.09	0.1	0	0.12	0.13	0.1	0.1	0.1															
Usuario 5	0.12	0.11	0.09	0.1	0.12	0	0.12	0.11	0.09	0.11															
Usuario 6	0.16	0.13	0.1	0.09	0.11	0.12	0	0.1	0.1	0.11															
Usuario 7	0.11	0.12	0.1	0.12	0.1	0.11	0.1	0	0.09	0.08															
Usuario 8	0.09	0.11	0.12	0.08	0.1	0.09	0.1	0.09	0	0.1															
Usuario 9	0.09	0.1	0.11	0.11	0.1	0.11	0.11	0.08	0.1	0															

- Basada en magnitud, sensible a la escala de ratings.
- Estable incluso con pocos vecinos.
- Suaviza predicciones, pero pierde capacidad de discriminación en altas dimensiones.

Análisis:

Pearson es útil con datos densos y correlaciones claras, pero falla con datos dispersos o vecinos aislados. Distancia coseno y distancia euclídea proporcionan predicciones más estables y confiables en escenarios con sparsity (datos faltantes) o ratings extremos.

Comparativa entre métodos de predicción

En los experimentos se aplicaron dos métodos de predicción sobre las distintas métricas de similitud:



1. Predicción simple:

$$\hat{r}(u, i) = \frac{\sum_{v \in N_u^k} \text{sim}(u, v) \cdot r(v, i)}{\sum_{v \in N_u^k} |\text{sim}(u, v)|}$$

Se basa directamente en las puntuaciones de los vecinos. Como se ha podido observar, tiende a favorecer métricas como **distancia coseno** y **euclídea**, que producen valores siempre positivos y homogéneos. Con matrices pequeñas o dispersas, este método ofrece predicciones más estables y dentro del rango original.

2. Predicción con diferencia respecto a la media:

$$\hat{r}(u, i) = \bar{r}(u) + \frac{\sum_{v \in N_u^k} \text{sim}(u, v) \cdot (r(v, i) - \bar{r}(v))}{\sum_{v \in N_u^k} |\text{sim}(u, v)|}$$

Corrige sesgos individuales (usuarios que puntúan sistemáticamente alto o bajo).

Este enfoque se ajusta mejor al **coeficiente de Pearson**, ya que ambas fórmulas se basan en desviaciones respecto a la media. En matrices densas, produce estimaciones más precisas y con menor error promedio.



Resultados y tendencias generales

- En **matrices pequeñas (5×10, 10×25)**, la predicción simple funciona mejor con **coseno** y **euclídea**, porque Pearson generaba correlaciones poco fiables y, al aplicar la fórmula con media, amplificaba el ruido.
- En **matrices medianas (25×100)**, la diferencia por media mejoró la coherencia de Pearson, reduciendo el sesgo y produciendo valores más realistas.
- En **matrices grandes (50×250, 100×1000)**, la predicción con media estabiliza Pearson, pero su rendimiento se ve limitado por la falta de coincidencias entre usuarios. En cambio, **coseno** se mantiene robusto con el método simple, y **euclídea** conserva su consistencia, aunque con menor discriminación.
- En general, las métricas angulares (como el coseno) responden mejor a entornos dispersos, mientras que las centradas en desviaciones (como Pearson) se benefician de mayor densidad y del uso de la media.

Tabla descriptiva

Métrica	Predicción simple	Predicción con media
Pearson	Inestable en datos escasos; puede dar valores fuera de rango.	Más coherente; corrige sesgos individuales. Ideal en matrices densas.
Coseno	Predicciones estables y consistentes. Ideal en matrices pequeñas o dispersas.	Mejora poco con la media, ya que ignora magnitud.
Euclídea	Predicciones suaves y conservadoras.	Resultados similares; poco sensible a sesgos.



Conclusión final

- **Pearson** es el más adecuado en matrices densas o moderadamente completas, especialmente junto a la predicción con diferencia respecto a la media.
- **Distancia coseno** ofrece resultados más consistentes en entornos dispersos o con pocos ítems comunes, siendo ideal con la predicción simple.
- **Distancia euclídea** es la más estable, pero su capacidad de discriminación disminuye al aumentar la dimensionalidad.
- En conjunto, la elección óptima depende del nivel de sparsity:
 - Para datos **densos** → **Pearson + media**.
 - Para datos **dispersos** → **Coseno + predicción simple**.
 - Para entornos **mixtos o de gran tamaño** → **Euclídea + simple**.