

# CIENCIA DE DATOS

Área personal / Mis cursos / [CIENCIA DE DATOS](#) / [UNIDAD 2: ANALÍTICA DE DATOS](#)  
/ [FORO 2-2: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA](#)



Buscar en los foros



## FORO 2-2: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA

Vencimiento: Tuesday, 6 de June de 2023, 23:50

✓ Hecho

⚙️ Configuraciones ▾

Mostrar respuestas anidadas

Se ha alcanzado la fecha límite para publicar en este foro, por lo que ya no puede publicar en él.

### FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA

Monday, 5 de June de 2023, 19:06

[Enlace permanente](#)



**Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA**

de [ANGELO DAVID PAZMIÑO QUIMI](#) - Monday, 5 de June de 2023, 19:45

A la hora de comparar y evaluar la similitud o diferencia entre distintos conjuntos de datos, los algoritmos de similitud y distancia son herramientas esenciales en la ciencia de datos. Estos métodos se emplean con frecuencia en tareas como la clasificación, la agrupación, la búsqueda y las recomendaciones de artículos relacionados. La asociación entre dos conjuntos de datos se mide mediante técnicas de similitud, y la diferencia entre dos conjuntos de datos se mide mediante técnicas de distancia.

[Enlace permanente](#) [Mostrar mensaje anterior](#)



**Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA**

de [FERNANDO ANDRE SANCHEZ MOREIRA](#) - Monday, 5 de June de 2023, 19:48

Las [técnicas de similitud y distancia](#) son fundamentales en la analítica de datos, ya que permiten comparar y medir la similitud entre diferentes elementos. Estas técnicas ayudan a identificar patrones, agrupar datos similares y encontrar relaciones ocultas en conjuntos de datos. Las medidas de similitud y distancia, como la distancia euclídea, la similitud del coseno o la distancia de Hamming, proporcionan métricas cuantitativas que permiten cuantificar y comparar la similitud o la diferencia entre elementos individuales o conjuntos de datos completos. Estas técnicas son ampliamente utilizadas en áreas como la clasificación, la segmentación de clientes, la recomendación de productos y el descubrimiento de patrones en grandes volúmenes de datos. Al aplicar [técnicas de similitud y distancia](#) en la analítica de datos, los profesionales pueden obtener una comprensión más profunda de los datos y generar insights valiosos para la toma de decisiones informadas.

[Enlace permanente](#) [Mostrar mensaje anterior](#)



**Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA**

de [VILMA PATRICIA RAMIREZ AGILA](#) - Monday, 5 de June de 2023, 19:50

Las [técnicas de similitud y distancia](#) son métodos utilizados para medir la relación o proximidad entre objetos o conjuntos de datos. Estas técnicas son ampliamente utilizadas en campos como la minería de datos, el aprendizaje automático, la recuperación de información y la inteligencia artificial. Algunas de las técnicas más comunes incluyen

- Métrica de similitud
- Distancia euclídea
- Distancia de Hamming
- Similitud del coseno, entre otras.

Estas técnicas y métricas son solo una muestra de las muchas herramientas disponibles para medir la similitud y distancia entre

objetos. La elección de la técnica adecuada depende del dominio de aplicación y los requisitos específicos del problema.

[Enlace permanente](#) [Mostrar mensaje anterior](#)



#### Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA

de [ARTURO ENRIQUE ROSERO ESPINOZA](#) - Monday, 5 de June de 2023, 19:51

Las [técnicas de similitud y distancia](#) son métodos utilizados en diversos campos, como la minería de datos, la recuperación de información y el aprendizaje automático, para medir la relación entre dos objetos o conjuntos de datos.

Métrica de similitud

Distancia euclíadiana

Distancia de Hamming

Similitud del coseno

Distancia de Jaccard

Indexación basada en similitud

Clustering basado en similitud

Análisis de similitud de texto

[Enlace permanente](#) [Mostrar mensaje anterior](#)



#### Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA

de [DAHIANA LISSI TIERRA QUINTO](#) - Monday, 5 de June de 2023, 19:53

Las [técnicas de similitud y distancia](#) son importantes dentro de la ciencia de datos porque a partir de aquellas técnicas los científicos de datos pueden medir la similitud o la diferencia entre objetos o conjuntos de datos. Estas técnicas son esenciales en una amplia gama de aplicaciones, incluyendo la clasificación de datos, la agrupación (clustering), la búsqueda de información y la recomendación de elementos similares.

Entre las técnicas de similitud que más se utilizan se encuentran: coeficiente de correlación, similitud coseno, similitud Jaccard y distancia de Levenshtein. Por otra parte, las técnicas de distancia más empleadas son: euclíadiana, Manhattan, Mahalanobis y Hamming. Todas estas técnicas ayudan a la minería de datos, identificar patrones, recomendar productos y más. Sin embargo, tiene sus desafíos como poder identificar qué método emplear, así como saber elegir las variables adecuadas.

Fuera de los desafíos, se puede demostrar como las [técnicas de similitud y distancia](#) desempeñan un papel fundamental en la ciencia de datos, puesto que proporcionan medidas objetivas para comparar y analizar datos, lo que es crucial en diversas aplicaciones, desde la clasificación y agrupación de datos hasta la búsqueda de información y la recomendación de elementos similares.

[Enlace permanente](#) [Mostrar mensaje anterior](#)



#### Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA

de [OLIVER MICHAEL TUBAY ZAMBRANO](#) - Monday, 5 de June de 2023, 19:53

Las [técnicas de similitud y distancia](#) se utilizan para medir qué tan similares o diferentes son dos puntos de datos. Estas técnicas se utilizan en una variedad de tareas de ciencia de datos, como sistemas de recomendación, agrupación y detección de anomalías. Hay muchas técnicas diferentes de similitud y distancia, cada una con sus propias fortalezas y debilidades. Algunas de las técnicas más populares incluyen:

- **Distancia euclíadiana:** Esta es la métrica de distancia más común. Mide la distancia entre dos puntos en un espacio euclidiano.
- **Distancia de Manhattan:** Esta métrica de distancia es menos sensible a los valores atípicos que la distancia euclíadiana. Mide la distancia entre dos puntos en un espacio de Manhattan.
- **Similitud de Jaccard:** Esta métrica de similitud mide la similitud entre dos conjuntos calculando el tamaño de su intersección dividido por el tamaño de su unión.

La elección de la técnica de similitud o distancia depende de la tarea específica de ciencia de datos y de las características de los datos.

[Enlace permanente](#) [Mostrar mensaje anterior](#)



#### Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA

de [KARELYS ALEXANDRA BOHORQUEZ DUARTE](#) - Monday, 5 de June de 2023, 19:53

Cuando hemos creado una matriz de similitud o distancia describiendo todas las parejas de actores, podemos estudiar la similitud de las diferencias entre las relaciones de "casos" de la misma manera que estudiaríamos las similitudes entre atributos. Existen varias técnicas para medir la similitud y la distancia entre diferentes elementos, como documentos, palabras o vectores.

[Enlace permanente](#) [Mostrar mensaje anterior](#)



**Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA**  
de [OSCAR JOEL MORAN CEDEÑO](#) - Monday, 5 de June de 2023, 19:55

Las [técnicas de similitud y distancia](#) en la ciencia de datos son fundamentales para medir la similitud y diferencia entre objetos. La similitud del coseno evalúa la similitud direccional entre vectores, siendo útil para procesamiento de texto y recomendaciones. Por otro lado, la distancia euclíadiana calcula la distancia directa entre puntos en un espacio euclidiano, siendo empleada en análisis numérico y agrupamiento. Estas técnicas son utilizadas en tareas como clasificación, agrupamiento y detección de anomalías, facilitando la toma de decisiones y la generación de predicciones basadas en datos. La elección de la técnica adecuada depende de la naturaleza de los datos y la tarea específica que se esté abordando.

[Enlace permanente](#) [Mostrar mensaje anterior](#)



**Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA**  
de [PAUL ALEXANDER GUARANDA MERO](#) - Monday, 5 de June de 2023, 19:56

En ciencia de datos, se emplean [técnicas de similitud y distancia](#) para evaluar la similitud o distancia entre dos objetos o conjuntos de datos. Estas técnicas desempeñan un papel fundamental en diversas aplicaciones, como la recuperación de información, la agrupación de datos, la clasificación y la detección de anomalías, como ejemplos tenemos distancia euclíadiana, la distancia de manhattan, la distancia de hamming, coeficiente de similitud de jaccard, distancia de levenshtein y similitud de coseno. La similitud del coseno es una medida comúnmente empleada en la recuperación de información y la agrupación de documentos. Se calcula como el coseno del ángulo entre dos vectores, donde los vectores representan los documentos y las características son ponderaciones de términos.

[Enlace permanente](#) [Mostrar mensaje anterior](#)



**Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA**  
de [PEDRO ANDRES DELGADO FLORES](#) - Monday, 5 de June de 2023, 19:56

Las [técnicas de similitud y distancia](#) permiten el análisis exploratorio en la medición de grupos de datos para identificar sus proximidades y diferencias dentro del área de la ciencia de datos. En similitud cuenta con diferentes técnicas que sirven para reconocer las relaciones y comparaciones que mantiene dos conjuntos de variables como son la Coeficiente de correlación, Similitud Jaccard, Distancia de Levenshtein. En distancia utiliza varias técnicas que sirven para identificar el trayecto o longitud que mantiene dos puntos de datos como son la Distancia euclíadiana, Distancia de Manhattan, Distancia de Mahalanobis, entre otros. Se los pueden aplicar para la identificación de patrones en imágenes, recomendación de productos, minería de datos, análisis de redes sociales. A pesar de que hay muchas técnicas se debe de elegir la técnica adecuada dependiendo de la situación e intención del análisis y de igual forma saber escoger las variables prioritarias para que aporten un conocimiento óptimo a los resultados obtenidos.

[Enlace permanente](#) [Mostrar mensaje anterior](#)



**Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA**  
de [JAIME ANDRES VELEZ VERA](#) - Monday, 5 de June de 2023, 19:58

Las [técnicas de similitud y distancia](#) son métodos utilizados en diversos campos, como la minería de datos, la recuperación de información, el aprendizaje automático y la estadística, para medir la similitud o la distancia entre objetos o conjuntos de datos. Algunas de las técnicas más comunes incluyen:

- \*Métrica de similitud
- \*Distancia euclíadiana
- \*Distancia de Hamming
- \*Similitud del coseno, entre otras.

[Enlace permanente](#) [Mostrar mensaje anterior](#)



**Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA**  
de [ANDRICK STEVEN VIZUETA LOPEZ](#) - Monday, 5 de June de 2023, 20:03

La ciencia de datos utiliza [técnicas de similitud y distancia](#) para analizar grandes cantidades de datos y extraer información significativa. Sin embargo, estas técnicas presentan desafíos, como la elección del método adecuado para cada situación, la selección de variables relevantes y la combinación con otras herramientas y enfoques analíticos. La ciencia de datos es un enfoque multidisciplinario que combina principios y prácticas del campo de las matemáticas, la estadística, la inteligencia artificial y la ingeniería de computación para analizar grandes cantidades de datos. La información que genera la ciencia de datos ayuda a las organizaciones a aumentar la eficiencia operativa, identificar nuevas oportunidades comerciales y mejorar los programas de marketing y ventas, entre otros beneficios. La ciencia de datos incorpora varias disciplinas, como la ingeniería de datos, la preparación de datos, la minería de datos, el análisis predictivo, el aprendizaje automático y la visualización de datos, así como estadísticas, matemáticas y programación de software.

**Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA**de [HENRY ANTONIO JARAMILLO OLIVARES](#) - Monday, 5 de June de 2023, 20:03

Las técnicas de similitud y distancia se utilizan para medir la similitud o la distancia entre objetos o conjuntos de datos. Estas técnicas son fundamentales en muchas aplicaciones, como la recuperación de información, la agrupación de datos y la clasificación.

Entre las más comunes se tiene:

**Distancia euclíadiana:** Es una medida comúnmente utilizada para calcular la distancia entre dos puntos en un espacio euclidiano. Para dos vectores n-dimensionales, se calcula como la raíz cuadrada de la suma de las diferencias al cuadrado de cada dimensión.

**Distancia Manhattan:** También conocida como distancia de la ciudad o distancia L1, mide la distancia entre dos puntos sumando las diferencias absolutas entre las coordenadas de los puntos a lo largo de cada dimensión.

**Coeficiente de similitud de Jaccard:** Es una medida de similitud utilizada para conjuntos finitos. Se calcula dividiendo el tamaño de la intersección de los conjuntos entre el tamaño de la unión de los conjuntos.

**Distancia de Hamming:** Se utiliza para medir la diferencia entre dos cadenas de igual longitud. Mide el número de posiciones en las que los símbolos difieren entre las dos cadenas.

**Coeficiente de correlación de Pearson:** Mide la similitud entre dos variables continuas. Calcula la covarianza entre las variables y las divide por el producto de sus desviaciones estándar.

**Distancia coseno:** Mide el ángulo entre dos vectores y se utiliza comúnmente para comparar la similitud entre documentos de texto. Se calcula como el producto escalar de los vectores dividido por el producto de sus magnitudes.

**Distancia de Levenshtein:** También conocida como distancia de edición, se utiliza para medir la diferencia entre dos cadenas. Mide el número mínimo de operaciones (inserciones, eliminaciones o sustituciones) requeridas para transformar una cadena en otra.

**Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA**de [KAYSY MARCELA BARRERA PEREZ](#) - Monday, 5 de June de 2023, 20:04

Las [técnicas de similitud y distancia](#) son herramientas utilizadas en el análisis de datos y la inteligencia artificial para medir la similitud o la diferencia entre dos objetos o conjuntos de datos. La similitud se refiere a la medida en que dos objetos son parecidos o tienen características en común, mientras que la distancia se refiere a la medida en que dos objetos son diferentes o tienen características diferentes.

Algunas de las [técnicas de similitud y distancia](#) más utilizadas en ciencia de datos son:

- Distancia euclíadiana: mide la distancia entre dos puntos en un espacio n-dimensional.
- Distancia de Manhattan: mide la distancia entre dos puntos sumando las diferencias absolutas de sus coordenadas.
- Distancia de Hamming: mide la distancia entre dos cadenas de bits contando el número de bits diferentes.
- Similitud del coseno: mide la similitud entre dos vectores de características calculando el coseno del ángulo entre ellos.

Estas técnicas se utilizan en una amplia variedad de aplicaciones de ciencia de datos, como la clasificación de imágenes, la agrupación de datos, la recomendación de productos, la detección de fraudes, entre otras.

**Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA**de [CARLOS DAVID GARCIA CEDEÑO](#) - Monday, 5 de June de 2023, 20:05

Las [técnicas de similitud y distancia](#) son utilizadas en diversos campos, como la minería de datos, la recuperación de información, la inteligencia artificial y el procesamiento de lenguaje natural, entre otros. Estas técnicas permiten medir la similitud o la diferencia entre objetos o conjuntos de datos, lo que resulta útil para realizar tareas como la clasificación, agrupación o búsqueda de información relevante. Entre las técnicas de similitud que más se utilizan se encuentran: coeficiente de correlación, similitud coseno, similitud Jaccard y distancia de Levenshtein. Por otra parte, las técnicas de distancia más empleadas son: euclíadiana, Manhattan, Mahalanobis y Hamming. La elección de la técnica adecuada depende del contexto y de los datos con los que se esté trabajando.

**Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA**de [JOSE CARLOS PAYE ARROYO](#) - Monday, 5 de June de 2023, 20:05

Las [técnicas de similitud y distancia](#) son herramientas importantes en ciencia de datos para analizar grupos de datos y descubrir cómo se relacionan y difieren entre sí. En términos de similitud, hay diversas técnicas que permiten reconocer las relaciones y comparaciones entre conjuntos de variables, como el coeficiente de correlación, la similitud Jaccard y la distancia de Levenshtein. Por otro lado, en cuanto a la distancia, existen varias técnicas que ayudan a medir la distancia o la separación entre dos puntos de datos, como la distancia euclíadiana, la distancia de Manhattan y la distancia de Mahalanobis, entre otras. Estas técnicas pueden ser aplicadas en diversas áreas, como la identificación de patrones en imágenes, la recomendación de productos, la minería de datos y el análisis de redes sociales. Sin embargo, es importante elegir la técnica adecuada según la situación y el objetivo del análisis, así como seleccionar las variables más relevantes para obtener resultados óptimos y un conocimiento enriquecedor.

[Enlace permanente](#)[Mostrar mensaje anterior](#)**Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA**de [NAYELI SOLIS CHERE](#) - Monday, 5 de June de 2023, 20:07

Una de las herramientas importantes para la ciencia de datos son las [técnicas de similitud y distancia](#) que permiten cuantificar la relación entre los diferentes datos. Las técnicas de similitud más comunes son:

- **Coeficiente de correlación:** se emplea para determinar si existe una relación entre dos variables y qué tan fuerte es esa relación.
- **Similitud coseno:** determina la similitud o la distancia entre dos vectores numéricos. Cuanto menor es el ángulo, mayor es la similitud entre los vectores y cuanto mayor es el ángulo, menor es la similitud.
- **Similitud Jaccard:** puede oscilar entre 0 y 1, donde un valor de 1 indica que los dos conjuntos son idénticos y un valor de 0 indica que no tienen elementos en común.
- **Distancia de Levenshtein:** es una medida de similitud entre dos cadenas de texto que se utiliza en procesamiento de lenguaje natural y en problemas de búsqueda de texto.

Por otra parte, las técnicas de distancia más comunes son:

- **Distancia euclíadiana:** se utiliza comúnmente en el análisis de datos y en el aprendizaje automático como una medida de similitud entre dos instancias.
- **Distancia de Manhattan:** se utiliza a menudo en la planificación de rutas y en la optimización de la logística.
- **Distancia de Mahalanobis:** se usa también para detectar casos atípicos y para identificar patrones en datos multivariados.
- **Distancia de Hamming:** mide la cantidad de elementos que deben cambiarse en una de las cadenas para que se igualen.

La [técnicas de similitud y distancia](#) son aplicables para distintos procesos como la recomendación de productos, identificación de patrones en imágenes con el fin de detectar enfermedades, en la detección de plagio, minería de datos, y en los motores de búsqueda para proporcionar la información que mayormente se relacione con la consulta especificada.

Hay que tener en cuenta que estas técnicas son empleadas según el conjunto de datos y el problema a resolver. En ocasiones es necesario complementarlas con otras herramientas para mayor efectividad.

[Enlace permanente](#)[Mostrar mensaje anterior](#)**Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA**de [LEONARDO ADRIAN OLVERA ORELLANA](#) - Monday, 5 de June de 2023, 20:16

Las [técnicas de similitud y distancia](#) son ampliamente utilizadas en diversas áreas de la ciencia de datos, como clasificación, agrupación, recomendación y búsqueda de información. Algunas de las técnicas más comunes incluyen:

Distancia Euclidiana: Se emplea para medir la distancia entre datos numéricos. Se calcula tomando la raíz cuadrada de la suma de los cuadrados de las diferencias entre las coordenadas de dos puntos en un espacio euclidiano.

Distancia Manhattan: También conocida como distancia de la ciudad, se utiliza para medir la distancia entre dos puntos en un espacio bidimensional. Consiste en sumar las diferencias absolutas de las coordenadas de los puntos. Es útil cuando se trabaja con datos que no se ajustan a una estructura euclidiana.

Distancia de Jaccard: Esta técnica mide la similitud entre conjuntos. Se calcula dividiendo el tamaño de la intersección de dos conjuntos por el tamaño de su unión. Es especialmente útil para comparar conjuntos de elementos no numéricos, como palabras o categorías.

Distancia del coseno: Se utiliza para comparar vectores en un espacio multidimensional. Mide el ángulo entre dos vectores y es comúnmente empleado en la comparación de documentos de texto o perfiles de usuarios en sistemas de recomendación. Un valor cercano a cero indica una mayor similitud entre los vectores.

Coeficiente de correlación de Pearson: Es una medida de la correlación lineal entre dos conjuntos de datos numéricos. Proporciona información sobre la relación lineal entre los datos y varía entre -1 y 1. Un valor cercano a 1 indica una fuerte correlación positiva, mientras que un valor cercano a -1 indica una fuerte correlación negativa.

Índice de Jaccard: Similar a la distancia de Jaccard, el índice de Jaccard se utiliza para medir la similitud entre conjuntos. Se calcula

dividiendo el tamaño de la intersección de dos conjuntos por el tamaño de su unión. A diferencia de la distancia de Jaccard, este índice proporciona una medida de similitud en lugar de distancia.

Estas técnicas son esenciales para comprender y analizar la similitud y la diferencia entre conjuntos de datos, lo que permite tomar decisiones informadas en diversos contextos de la ciencia de datos.

[Enlace permanente](#) [Mostrar mensaje anterior](#)



#### Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA

de [GABRIELA SARAY QUIMIS ESPINOZA](#) - Monday, 5 de June de 2023, 20:17

- Para medir la similitud de objetos o instancias en la ciencia de datos, se deben usar técnicas de similitud y de distancia.
- Tareas como la clasificación, el agrupamiento, la recomendación y la búsqueda de similitudes son posibles gracias a estas técnicas.
- **Las técnicas de similitud** incluyen el coeficiente de correlación, la similitud del coseno, la similitud de Jaccard y la distancia de Levenshtein.
- **Las técnicas de distancia** incluyen la distancia euclídea, de Manhattan, de Mahalanobis y de Hamming.
- Las aplicaciones de estas técnicas incluyen la identificación de patrones en imágenes médicas, la recomendación de productos, la agrupación de datos, la recuperación de información, el procesamiento del lenguaje natural y el análisis de redes sociales.

[Enlace permanente](#) [Mostrar mensaje anterior](#)



#### Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA

de [NICOLE ANGELA HOLGUIN SANCAN](#) - Monday, 5 de June de 2023, 20:24

Las [técnicas de similitud y distancia](#) son herramientas utilizadas en diversas disciplinas para medir la semejanza o diferencia entre objetos o conjuntos de datos. A continuación, se describen como se basan en algoritmos que comparan características o atributos de los elementos a evaluar. La distancia euclídea es una medida común que calcula la distancia entre dos puntos en un espacio vectorial. Por otra parte, los enfoques incluyen el coeficiente de similitud de Jaccard para conjuntos, la correlación de Pearson para variables numéricas y el coeficiente de Levenshtein para cadenas de texto. Estas técnicas tienen aplicaciones en campos como la clasificación de documentos, la recomendación de productos y el análisis de datos.

[Enlace permanente](#) [Mostrar mensaje anterior](#)



#### Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA

de [XAVIER ROBERTO CRUZ LADINES](#) - Monday, 5 de June de 2023, 20:25

Las [técnicas de similitud y distancia](#) son fundamentales en la ciencia de datos para comparar y medir la similitud entre diferentes objetos o instancias. Estas técnicas se utilizan en diversas tareas de minería de datos, como clasificación, agrupamiento, recomendación y búsqueda de similitud.

Algunas técnicas comunes de similitud incluyen el coeficiente de correlación, similitud coseno, similitud Jaccard y distancia de Levenshtein. Por otro lado, las técnicas de distancia, como la euclídea, Manhattan, Mahalanobis y Hamming, miden la diferencia entre conjuntos de datos.

Estas técnicas tienen diversas aplicaciones en ciencia de datos, como la detección de enfermedades en imágenes médicas, recomendación de productos, agrupación de datos, recuperación de información, procesamiento del lenguaje natural, minería de datos, reconocimiento de patrones y visión por computadora, y análisis de redes sociales.

Sin embargo, utilizar estas técnicas presenta desafíos, como la elección adecuada del método según el tipo de datos y problema, la selección de variables relevantes, y la necesidad de combinarlas con otras herramientas y enfoques analíticos para capturar la complejidad de los conjuntos de datos.

[Enlace permanente](#) [Mostrar mensaje anterior](#)



#### Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA

de [AARON JOEL ACOSTA MONTIEL](#) - Monday, 5 de June de 2023, 20:26

Las [técnicas de similitud y distancia](#) son métodos utilizados para medir la relación o proximidad entre objetos o conjuntos de datos. Estos métodos se emplean con frecuencia en tareas como la clasificación, la agrupación, la búsqueda y las recomendaciones de artículos relacionados, he aquí algunas de las técnicas más comunes:

- Métrica de similitud
- Distancia euclídea
- Distancia de Hamming

- Similitud del coseno

[Enlace permanente](#)[Mostrar mensaje anterior](#)**Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA**de [FREDDY GREGORY URETA VARGAS](#) - Monday, 5 de June de 2023, 20:27

Las técnicas de distancia y similitud son herramientas fundamentales en diversas aplicaciones de ciencia de datos y permiten comparar conjuntos de datos para determinar su diferencia o similitud en función de las necesidades específicas. Algunas de las técnicas más comunes incluyen:

- **Distancia euclíadiana:** Utilizada para calcular la distancia entre dos puntos en un espacio euclíadiano y cuantificar la proximidad o lejanía entre ellos.
- **Distancia de Manhattan:** También conocida como distancia de ciudad, que calcula la distancia entre dos puntos sumando las diferencias absolutas entre las coordenadas de los puntos en un espacio bidimensional o tridimensional.
- **Similitud del coseno:** Una técnica utilizada para medir la orientación y dirección relativa entre dos vectores de características.
- **Distancia de Mahalanobis:** Tiene en cuenta la relación entre las variables al calcular la diferencia entre dos puntos, siendo especialmente útil cuando los datos presentan diferentes unidades o escalas.

[Enlace permanente](#)[Mostrar mensaje anterior](#)**Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA**de [ANTHONY ELIAN MONCAYO FAJARDO](#) - Monday, 5 de June de 2023, 20:35

Las [técnicas de similitud y distancia](#) son fundamentales para el análisis y la interpretación de datos, por lo que permiten la cuantificación de la similitud o diferencia entre objetos o conjunto de objetos y esto a su vez proporcionan información valiosa para diversos procesos, como es la agrupación, clasificación y recuperación de información.

Técnicas y métricas para calcular la similitud y distancia.

- Distancia euclíadiana: se mide la distancia entre dos puntos en un espacio de n dimensión, en el que se usa el teorema de Pitágoras.
- Distancia Hamming: utilizada para la medición de las diferencias entre dos secuencias de igual longitud en especial en los tipos de datos binarios o categóricos.
- Distancia de Levenshtein: se calcula la distancia entre dos cadenas de texto midiendo el número mínimo de operaciones que se requieren para poder transformar una cadena en otra.

[Enlace permanente](#)[Mostrar mensaje anterior](#)**Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA**de [ADRIANA GEOVANINA VELASTEGUI SANDOVAL](#) - Monday, 5 de June de 2023, 20:37

Las [técnicas de similitud y distancia](#) son métodos utilizados en diversos campos, incluyendo la ciencia de datos y el análisis de datos. Estas técnicas permiten cuantificar la relación o la diferencia entre dos puntos de datos. Es importante seleccionar la técnica adecuada para el problema en cuestión, teniendo en cuenta las características de los datos y los objetivos del análisis, estos son métodos utilizados en diversos campos, incluyendo la ciencia de datos y el análisis de datos. Estas técnicas permiten cuantificar la relación o la diferencia entre dos puntos de datos. Es importante seleccionar la técnica adecuada para el problema en cuestión, teniendo en cuenta las características de los datos y los objetivos del análisis.

[Enlace permanente](#)[Mostrar mensaje anterior](#)**Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA**de [DARLY YENEDY MORAN ESTUPIÑÁN](#) - Monday, 5 de June de 2023, 20:47

Las [técnicas de similitud y distancia](#) son métodos utilizados en el campo de la inteligencia artificial y la minería de datos para medir la similitud o la distancia entre dos objetos o conjuntos de datos. Estas técnicas son esenciales en una amplia gama de aplicaciones, como la recuperación de información, la clasificación de documentos, el filtrado colaborativo, la agrupación de datos y más. La similitud y la distancia se definen en función de las características y propiedades de los objetos o datos que se están comparando. En general, la similitud se refiere a cuánto similares o cercanos están dos objetos entre sí, mientras que la distancia mide cuánto se separan o diferencian.

La similitud del coseno es muy utilizada para medir la similitud entre vectores en espacios multidimensionales. La similitud de

Jaccard se utiliza para medir la similitud entre conjuntos. La distancia euclíadiana se utiliza para medir la distancia entre puntos en un espacio euclíadiano. La distancia de Hamming se utiliza para medir la distancia entre cadenas de bits de igual longitud. La distancia de Levenshtein se utiliza para medir la distancia entre cadenas de caracteres y se basa en la cantidad mínima de operaciones (inserciones, eliminaciones y sustituciones) requeridas para transformar una cadena en otra.

[Enlace permanente](#) [Mostrar mensaje anterior](#)

**Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA**

de [CZISKA WALESKA MORAN ARMIJOS](#) - Monday, 5 de June de 2023, 20:51

Las [técnicas de similitud y distancia](#) son métodos utilizados en diversos campos, como la estadística, la minería de datos y la inteligencia artificial, para comparar la similitud o la distancia entre elementos.

Estas técnicas se basan en medidas numéricas que cuantifican la diferencia o la similitud entre dos objetos, como la distancia euclíadiana, la distancia de Hamming o el coeficiente de similitud de Jaccard.

Se aplican en tareas como la clasificación de datos, la búsqueda de información relevante y la agrupación de elementos similares. Estas técnicas son fundamentales para el análisis y la interpretación de grandes conjuntos de datos en diversas disciplinas.

[Enlace permanente](#) [Mostrar mensaje anterior](#)

**Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA**

de [ANTHONY DOUGLAS FLORES TOMALA](#) - Monday, 5 de June de 2023, 20:51

Las [técnicas de similitud y distancia](#) son importantes en la ciencia de datos y aprendizaje automático, son utilizadas para la recuperación de información, la agrupación de datos, la clasificación y recomendaciones.

Estos mismos tienen características o atributos que describen los objetos o los conjuntos de datos esto quiere decir mientras más similares sean dos objetos, más cerca estarán en términos de similitud y más pequeña será la distancia entre ellos.

Las [técnicas de similitud y distancia](#) son herramientas fundamentales en ciencia de datos y aprendizaje automático. Proporcionan medidas objetivas, flexibles y fácilmente interpretables de la similitud y diferencia entre objetos o conjuntos de datos, lo que las hace valiosas en diversas aplicaciones y escenarios.

[Enlace permanente](#) [Mostrar mensaje anterior](#)

**Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA**

de [ALEXANDER ISRAEL POVEDA GARCES](#) - Monday, 5 de June de 2023, 21:12

Dentro de la ciencia de datos, las técnicas son un recurso vital para el desarrollo de sus diferentes actividades. Teniendo esto en cuenta, las [técnicas de similitud y distancia](#) permiten a los científicos de datos medir de forma concreta la igualdad (similitud) o diferencia (distancia) entre dos objetos. Entre las técnicas de similitud encontramos coeficiente de correlación, similitud de Jaccard, similitud de coseno y distancia de Levenshtein. Por otro lado, entre las técnicas de distancia tenemos la euclíadiana, Manhattan, Mahalanobis y Hamming.

La aplicación de estas técnicas nos permite en primer lugar agrupar y analizar conjuntos de datos de forma ordenada, identificando dentro de este las anomalías o patrones que pueden ser empleados con la finalidad de mejorar algún proceso o servicio. De igual manera, ayuda a procesar y recuperar información, otorgándonos la capacidad de analizar fuentes de datos como las redes sociales para encontrar indicadores que resulten importantes en la toma de decisiones. Estas técnicas pueden ser aplicadas en áreas como la medicina, marketing, informática, matemáticas, entre otras.

Finalmente, si bien la similitud y distancia son de gran utilidad dentro de la ciencia de datos, el principal desafío al emplearlas es la elección de un método adecuado, ya que estos variarán según los datos y la problemática. De igual forma, la selección de variables puede afectar en gran medida nuestros procesos, debido a esto es imprescindible llevar a cabo un estudio para determinar todos sus factores y así conseguir resultados más favorables.

[Enlace permanente](#) [Mostrar mensaje anterior](#)

**Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA**

de [JOHNNY JOSE CRUZ BAQUERIZO](#) - Monday, 5 de June de 2023, 21:29

Las [técnicas de similitud y distancia](#) son herramientas utilizadas en diversos campos para medir y comparar la similitud o diferencia entre objetos, datos o características.

Son herramientas que ayudan a cuantificar y comparar la similitud o diferencia entre objetos o características. Son fundamentales en diversas disciplinas para el análisis de datos y la toma de decisiones informadas.

- Distancia euclíadiana: Calcula la distancia directa entre dos puntos en un espacio euclíadiano.
- Distancia de Manhattan: Calcula la distancia sumando las diferencias absolutas de las coordenadas de dos puntos en un espacio

bidimensional o tridimensional.

- Similitud del coseno: Mide la similitud entre dos vectores de características basándose en el ángulo entre ellos.
- Distancia de Mahalanobis: Calcula la distancia entre dos puntos teniendo en cuenta la relación entre las variables y considerando diferentes unidades o escalas.

[Enlace permanente](#)[Mostrar mensaje anterior](#)
**Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA**

de [JOSELYN TATIANA RIVERA BAQUE](#) - Monday, 5 de June de 2023, 21:46

## FORO 2-2: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA

Técnicas de similitud: Son métodos utilizados para medir la similitud o proximidad entre objetos o datos. Estas técnicas se basan en comparar las características o propiedades de los objetos y evaluar qué tan similares son. Se pueden utilizar para clasificar, agrupar o recomendar elementos similares en diversas aplicaciones, como recomendaciones de productos o análisis de texto.

Técnicas de distancia: Son métodos que calculan la distancia o diferencia entre dos objetos o puntos en un espacio. Estas técnicas ayudan a cuantificar la separación o lejanía entre los objetos en función de sus características. Algunas técnicas comunes son:

- Distancia euclídea: Calcula la distancia directa entre dos puntos en un espacio.
- Distancia de Manhattan: Calcula la distancia sumando las diferencias absolutas de las coordenadas de los puntos.
- Distancia de Mahalanobis: Considera la relación entre las variables al calcular la diferencia entre dos puntos.

[Enlace permanente](#)[Mostrar mensaje anterior](#)
**Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA**

de [GENESIS LISSET SANCHEZ MORAN](#) - Monday, 5 de June de 2023, 21:51

Los [técnicas de similitud y distancia](#) son la base de la ciencia de datos Comparar y medir la similitud de diferentes objetos o instancias. Estas tecnologías son para varias tareas de minería de datos como la clasificación, Clustering, sugerencia y búsqueda de similitud.

Algunas de las técnicas más comunes incluyen

- Métrica de similitud
- Distancia euclídea
- Distancia de Hamming
- Similitud del coseno, etc.

[Enlace permanente](#)[Mostrar mensaje anterior](#)
**Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA**

de [BETZABETH MADELINE MUÑOZ VILLEGRAS](#) - Monday, 5 de June de 2023, 22:35

En el campo de la ciencia de datos, las [técnicas de similitud y distancia](#) son herramientas importantes para llevar a cabo diversas tareas. Estas técnicas permiten a los científicos de datos cuantificar de manera precisa la igualdad (similitud) o diferencia (distancia) entre dos objetos. Algunas técnicas comunes de similitud incluyen el coeficiente de correlación, la similitud de Jaccard, la similitud de coseno y la *distancia de Levenshtein*. Por otro lado, las técnicas de distancia populares son la *euclídea*, *Manhattan*, *Mahalanobis* y *Hamming* y aplicar estas técnicas tiene varios beneficios.

1. En primer lugar, nos ayuda a organizar y analizar conjuntos de datos de manera ordenada, lo que nos permite identificar patrones o anomalías que pueden mejorar procesos o servicios.
2. Además, nos permite procesar y recuperar información, lo que nos brinda la capacidad de analizar fuentes de datos como las redes sociales para encontrar indicadores relevantes para la toma de decisiones.
3. Estas técnicas son aplicables en diversos campos, como medicina, marketing, informática, matemáticas, entre otros.

Sin embargo, al utilizar estas técnicas, surge el desafío de elegir el método adecuado, ya que esto puede variar según los datos y la problemática específica. También es importante tener en cuenta la selección de variables, ya que puede afectar significativamente nuestros procesos. Por lo tanto, es fundamental llevar a cabo un estudio exhaustivo para determinar todos los factores relevantes y obtener resultados más favorables.

Calificación máxima:10 (1)

[Enlace permanente](#)[Mostrar mensaje anterior](#)
**Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA**

de [FRANCISCO JAVIER VALVERDE CHAGUAY](#) - Monday, 5 de June de 2023, 22:37

En el ámbito de la ciencia de datos, resulta esencial emplear [técnicas de similitud y distancia](#) con el fin de evaluar y cuantificar la semejanza o la disparidad entre objetos o conjuntos de datos. En ciencia de datos, existen varias técnicas de distancia, algunas de las más comunes son las siguientes:

- Distancia Euclidiana: Es una medida de distancia que se calcula utilizando la fórmula de la distancia entre dos puntos en un espacio euclíadiano.
- Distancia de Manhattan: También conocida como distancia de ciudad o distancia L1, se calcula sumando las diferencias absolutas entre las coordenadas de dos puntos.
- Distancia de Minkowski: Es una generalización de las distancias euclidiana y de Manhattan.
- Distancia de Hamming: Se utiliza para medir la diferencia entre dos cadenas de igual longitud.
- Distancia de Jaccard: Se utiliza para medir la similitud entre dos conjuntos. Se calcula dividiendo el tamaño de la intersección de los conjuntos por el tamaño de su unión.

Así mismo existen diversas técnicas utilizadas para medir la similitud, algunas de las técnicas más comunes incluyen:

- Similitud de coseno: Esta técnica se utiliza para calcular la similitud entre dos vectores en un espacio vectorial.
- Coeficiente de correlación de Pearson: Se utiliza para medir la correlación lineal entre dos variables.
- Índice de similitud de Jaccard: Similar a la distancia de Jaccard, mide la similitud entre conjuntos pero se calcula como el cociente entre la cardinalidad de la intersección.

[Enlace permanente](#) [Mostrar mensaje anterior](#)



#### Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA

de [ISAAC ROBERTO GOMEZ JUNCO](#) - Monday, 5 de June de 2023, 23:09

Las [técnicas de similitud y distancia](#) son esenciales en ciencia de datos para medir la proximidad o diferencia entre objetos. Estas medidas desempeñan un papel crucial en diversas tareas, como clustering, clasificación y recomendación. La similitud coseno, la distancia euclidiana, la distancia de Levenshtein, la distancia de Hamming, el índice de Jaccard y la distancia de Mahalanobis son algunas de las técnicas ampliamente utilizadas. Estas medidas permiten cuantificar la similitud o la distancia entre datos, lo que facilita la comparación y agrupación de objetos en función de sus características y atributos. Al seleccionar la medida adecuada, se puede lograr un análisis de datos más efectivo y obtener resultados significativos en diferentes aplicaciones de ciencia de datos.

[Enlace permanente](#) [Mostrar mensaje anterior](#)



#### Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA

de [ELVIN DANIEL IBARRA RONQUILLO](#) - Monday, 5 de June de 2023, 23:23

Las [técnicas de similitud y distancia](#) son métodos matemáticos que permiten medir el grado de semejanza o diferencia entre dos objetos, como especies, rasgos funcionales o patrones de vínculos. Estas técnicas se pueden aplicar a diferentes tipos de datos, como binarios, cualitativos o cuantitativos, y se pueden expresar en términos de similitud o distancia. La similitud es un valor que indica cuánto se parecen dos objetos, mientras que la distancia es un valor que indica cuánto se alejan.

[Enlace permanente](#) [Mostrar mensaje anterior](#)



#### Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA

de [ALBA DENIS RUIZ GUAMAN](#) - Tuesday, 6 de June de 2023, 08:16

Similitud y distancia son dos herramientas comúnmente utilizadas para visualizar patrones de relaciones entre variables también son muy útiles para estudiar datos de redes sociales. Una vez que hemos construido una matriz de similitud o distancia que describe todos los pares de participantes, podemos estudiar la similitud de las diferencias entre las relaciones "casuales", del mismo modo que estudiamos la similitud de los atributos. En las próximas dos secciones, mostraremos un ejemplo muy breve de cómo usar la escala multidimensional y el análisis de conglomerados jerárquicos para identificar patrones en la matriz de similitud/distancia de cada participante. Ambas herramientas son ampliamente utilizadas para el análisis sin red; hay una literatura grande y excelente sobre muchas complejidades importantes del uso de estas técnicas. Nuestro objetivo aquí es simplemente proporcionar una introducción muy simple.

[Enlace permanente](#) [Mostrar mensaje anterior](#)



#### Re: FORO 2-1: RESUMEN SOBRE TECNICAS DE SIMILITUD Y DISTANCIA

de [JULIAN STEVEN LAVAYEN GOMEZ](#) - Tuesday, 6 de June de 2023, 20:31

Las técnicas se utilizan en la clasificación, agrupación, búsqueda de información y recomendación de elementos similares. Las técnicas más comunes incluyen el coeficiente de correlación, similitud coseno, similitud Jaccard, distancia euclidiana, distancia Manhattan, Mahalanobis y Hamming. Elegir el método adecuado y seleccionar las variables correctas son desafíos importantes. Estas técnicas son cruciales en diversas aplicaciones y proporcionan medidas objetivas para comparar y analizar datos.

[Enlace permanente](#) [Mostrar mensaje anterior](#)

[◀ FORO 2-1: RESUMEN SOBRE CONCEPTOS DE DISTANCIA Y SIMILITUD](#)[Ir a...](#)[FORO 2-3: RESUMEN SOBRE MÉTODOS DE CLUSTER ►](#)

Usted se ha identificado como BETZABETH MADELINE MUÑOZ VILLEGRAS (Cerrar sesión)

Reiniciar tour para usuario en esta página

15150059\_15150091\_1515811

Español - Internacional (es)

English (en)

Español - Internacional (es)

Resumen de retención de datos

Descargar la app para dispositivos móviles