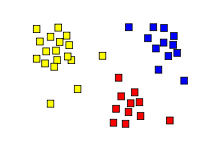
Análisis de conglomerados

De Wikipedia, la enciclopedia libre

*Para el enfoque de*[*aprendizaje supervisado*](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Statistical_classification&usg=ALkJrhihMHjJB29QEOc8lvZR8PilJ9HHOQ)*, véase*[*Clasificación estadística*](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Statistical_classification&usg=ALkJrhihMHjJB29QEOc8lvZR8PilJ9HHOQ)*.*

|  |
| --- |
| [**Aprendizaje mecánico**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Machine_learning&usg=ALkJrhiqyl1gcPCrbxDJThh3e7E51ejcTg)**y** [**Minería de datos**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Data_mining&usg=ALkJrhgvrkYUd-C50pyD-jwLcxzfWmQRfQ) |
| [kernel Machine.svg](https://en.wikipedia.org/wiki/File:Kernel_Machine.svg) |
| **Problemas**  [Clasificación](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Statistical_classification&usg=ALkJrhihMHjJB29QEOc8lvZR8PilJ9HHOQ)  **Clustering**  [Regresión](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Regression_analysis&usg=ALkJrhjqeCnhhLpfE7CNtzHVizv0tK30Yg)  [Detección de anomalías](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Anomaly_detection&usg=ALkJrhhQJ0-GweI0Xv6IEQwqj4P8y7Ar4g)  [reglas de asociación](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Association_rule_learning&usg=ALkJrhjgRIEhJNVTynV4A8KP_eVRgUCfYQ)  [Aprendizaje reforzado](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Reinforcement_learning&usg=ALkJrhjvcYHNtPIKaYNS9Tgqn5t02Z98RA)  [Predicción estructurada](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Structured_prediction&usg=ALkJrhg38ixFY8woLPqZCEqXgaFJW91VFA)  [Ingeniería de funciones](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Feature_engineering&usg=ALkJrhibpnU-b3mXSgxdwsz4F_FFvYhlbg)  [Aprendizaje de funciones](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Feature_learning&usg=ALkJrhhIJlUz_9n8G8IpE3sgDk5cVmavuQ)  [Aprender en línea](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Online_machine_learning&usg=ALkJrhi6BvRqbrH1QXJ52ZcrEVLRsxjACg)  [Aprendizaje semi-supervisado](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Semi-supervised_learning&usg=ALkJrhg6PeVW_B3KALw0ORLAb2pvoGgVXg)  [Aprendizaje sin supervisión](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Unsupervised_learning&usg=ALkJrhigynAdzL5e_YZb3Kaug5ykPhOaog)  [Aprender a clasificar](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Learning_to_rank&usg=ALkJrhgJSVrGxxY7Bgyry6WkimE6T5SeEg)  [Inducción gramatical](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Grammar_induction&usg=ALkJrhgfrMU_5sP99mi9kWYt-o3MHY7E4w) |
| [**Aprendizaje supervisado**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Supervised_learning&usg=ALkJrhjQnvhmMZPbQB24Mtj2eCD38C8oIA)( [**Clasificación**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Statistical_classification&usg=ALkJrhihMHjJB29QEOc8lvZR8PilJ9HHOQ) • [**regresión**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Regression_analysis&usg=ALkJrhjqeCnhhLpfE7CNtzHVizv0tK30Yg) )  [Árboles de decisión](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Decision_tree_learning&usg=ALkJrhilkQT3BrCYCNdDhu83SoH7V5d_BA)  [Conjuntos](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Ensemble_learning&usg=ALkJrhiOnX39WBsOYQ7lLNP5nRQpjl3o6w) ( [embolsado, refuerzo](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Ensemble_learning&usg=ALkJrhiOnX39WBsOYQ7lLNP5nRQpjl3o6w) , [bosque al azar](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Random_forest&usg=ALkJrhi_vGnvzb-enA0mMErSaqcr5RlQSg) )  [KNN](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm&usg=ALkJrhigSXQl4M7Um6e8AALiaOCwRYqfjA)  [Regresión lineal](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Linear_regression&usg=ALkJrhjMiEclG7Ket46n2Q5QoVsHCjj07Q)  [Naive Bayes](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Naive_Bayes_classifier&usg=ALkJrhi0sdmDr9yRG1q3y9Upwqm0T1kCxw)  [Redes neuronales](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Artificial_neural_network&usg=ALkJrhjrru950dvqm72cdwXCx2H2IS_vLg)  [Regresión logística](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Logistic_regression&usg=ALkJrhizhBKgUpfJNWIDA1Ks90wuUJJUDQ)  [Perceptron](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Perceptron&usg=ALkJrhi5tW3SvX8ApM3Tx9xes3l4TI7jDQ)  [Relevancia vector máquina (RVM)](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Relevance_vector_machine&usg=ALkJrhgQXh_PWcXGXBY8ciEAqb9Uatj_Aw)  [Máquina de vector de apoyo (SVM)](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Support_vector_machine&usg=ALkJrhg3meQuNz0LM7Y1yv8kqr0mF3wVXw) |
| **Clustering**  [ABEDUL](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/BIRCH&usg=ALkJrhidLJ_AuyBYEATW6fSV1G6p53Jw1w)  [Jerárquico](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Hierarchical_clustering&usg=ALkJrhhP79NNtd9SRZS9LyKVX0KcKW9ukA)  [*K-* medios](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/K-means_clustering&usg=ALkJrhiyUH_pwirgtWP1N9GqWWA1gduRng)  [Expectativa-maximización (EM)](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Expectation-maximization_algorithm&usg=ALkJrhhbhcqCPd7wWL9JgGZY5xRPuu_pdw)  [DBSCAN](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/DBSCAN&usg=ALkJrhjRaTKu8cpbIDTI13z_zVmUycbpiQ)  [ÓPTICA](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/OPTICS_algorithm&usg=ALkJrhib6DU3e0hcx6PelkSkhqKE1JRK9w)  [Cambio medio](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Mean-shift&usg=ALkJrhi6aad6Qc8wtfqMqHdKoJj-XLkSvw) |
| [**Reducción de dimensionalidad**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Dimensionality_reduction&usg=ALkJrhgxEP7uSix2J3lH2ztZYEp-m9MZ3g)  [Análisis factorial](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Factor_analysis&usg=ALkJrhiu4J-58jGemSi7Hs3oqAr9gNGqig)  [CCA](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Canonical_correlation_analysis&usg=ALkJrhgO1ZS6fzOU4j2URSVIQ9O2MXMSAw)  [ICA](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Independent_component_analysis&usg=ALkJrhjOki4gxbQtUor3P28ehPhopUkjPg)  [LDA](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Linear_discriminant_analysis&usg=ALkJrhhEx_Lep3PRY96qhl3XKV4Oo76wVg)  [NMF](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Non-negative_matrix_factorization&usg=ALkJrhg7biNv73WgRMRoK1vWQOtLRizc3Q)  [PCA](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Principal_component_analysis&usg=ALkJrhiHNcosX0QB8uyyLew1EyMoaP77DQ)  [Tnee](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/T-distributed_stochastic_neighbor_embedding&usg=ALkJrhg3tbDqfyAGwA_ZTF7AuyNTVZ_kCQ) |
| [**Predicción estructurada**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Structured_prediction&usg=ALkJrhg38ixFY8woLPqZCEqXgaFJW91VFA)  [Modelos gráficos](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Graphical_model&usg=ALkJrhjehFGpNZikm8w38DNSv9_eX9lnVA) ( [Bayes net](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Bayesian_network&usg=ALkJrhi1O_hd72LrxqUoIjzONGC54aEEvQ) , [CRF](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Conditional_random_field&usg=ALkJrhjWAYPM_AHlb6Iqi2N_5dxMp1Pg0g) , [HMM](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Hidden_Markov_model&usg=ALkJrhi5bnDkFqKplp1ugOgzIlNlw4qVUQ) ) |
| [**Detección de anomalías**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Anomaly_detection&usg=ALkJrhhQJ0-GweI0Xv6IEQwqj4P8y7Ar4g)  [KNN](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/K-nearest_neighbors_classification&usg=ALkJrhj0qxIDQmX46daqOs7XhRxIGTDkNA)  [Factor atípico local](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Local_outlier_factor&usg=ALkJrhhvgbzj2Sj5mRAzkmAIiKml2WQA1w) |
| [**Redes neurales**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Artificial_neural_network&usg=ALkJrhjrru950dvqm72cdwXCx2H2IS_vLg)  [Autoencoder](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Autoencoder&usg=ALkJrhhmB2GK178DsO0Y6p6971Ru-oC6FQ)  [Aprendizaje profundo](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Deep_learning&usg=ALkJrhhbpYyNaruDl5WYx-gQd7ateHKnqw)  [Perceptron multicapa](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Multilayer_perceptron&usg=ALkJrhiIe7lp6H56Q7lLAJsn5o42bIAu4Q)  [RNN](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Recurrent_neural_network&usg=ALkJrhgfIL2kr22K_8qQPEcddn02Wjvd2Q)  [Máquina Boltzmann restringida](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Restricted_Boltzmann_machine&usg=ALkJrhjkgj-TWyMJGeOM9YUaC00NYEwosA)  [SOM](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Self-organizing_map&usg=ALkJrhhKnSi1I8Husm90X2GvYkNzuifw-A)  [Red neuronal convolucional](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Convolutional_neural_network&usg=ALkJrhhL5zywO_KCoWmPUb1bO6RWqnXWJg) |
| [**Aprendizaje reforzado**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Reinforcement_Learning&usg=ALkJrhhDRjLkLOjqo8aUDz4uUeObs5Xg6Q)  [Q-Aprendizaje](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Q-Learning&usg=ALkJrhiOTsWevFXzt4EPSoav6CdzwFS-Bg)  [SARSA](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/State-Action-Reward-State-Action&usg=ALkJrhgGPTvtgtrmQcRAIyuUuqhpB7vbwA)  [Diferencia Temporal (TD)](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Temporal_Difference_Learning&usg=ALkJrhjonV3IVmYjVY2zSdJM5iFoZBswHg) |
| **Teoría**  [Dilema de sesgo-varianza](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Bias-variance_dilemma&usg=ALkJrhgOtp4MJpQd5xTdXPQBYo3Y-S3Miw)  [Teoría del aprendizaje computacional](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Computational_learning_theory&usg=ALkJrhhgFDK5uERfQhm7fELe8FO7uuQ3nw)  [Minimización empírica del riesgo](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Empirical_risk_minimization&usg=ALkJrhgqFNE7_xAXU3lqZea1Zm0q6cdA8Q)  [Aprendizaje Occam](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Occam_learning&usg=ALkJrhhC9oxzAH5DO_gkJxWeoOTK_zjgww)  [Aprendizaje PAC](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Probably_approximately_correct_learning&usg=ALkJrhhKclalpOFxgDjuly8ZJ7N9awQkeQ)  [Aprendizaje estadístico](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Statistical_learning_theory&usg=ALkJrhjDhVdvuLWT1n3S379s91QZnZrTaQ)  [Teoría de la CV](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Vapnik%25E2%2580%2593Chervonenkis_theory&usg=ALkJrhiaawsvNPsPGnMFV6wuHpu4jpafig) |
| **Lugares de aprendizaje de máquinas**  [NIPS](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Conference_on_Neural_Information_Processing_Systems&usg=ALkJrhif3GRsVYpys2awh1uBUPVX8b9w4A)  [ICML](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/International_Conference_on_Machine_Learning&usg=ALkJrhhUA48hc00RAQsuwahj6U3aQMzLZQ)  [IJMLC](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/w/index.php%3Ftitle%3DInternational_Journal_of_Machine_Learning_and_Cybernetics%26action%3Dedit%26redlink%3D1&usg=ALkJrhiTvOHhLR7Dcf6ipEu3P4c_6q0bVQ)  [ML](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Machine_Learning_(journal)&usg=ALkJrhgswB8fi2RxMAVq6zo58vzOcFhcIg)  [JMLR](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Journal_of_Machine_Learning_Research&usg=ALkJrhis1ykmfe2RibamJ9ln-DFFw3z8tg)  [ArXiv: cs.LG](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=http://arxiv.org/list/cs.LG/recent&usg=ALkJrhjkHyuKWs80sqX1_93vMsekFbYTpA) |
| **[Portal-puzzle.svg](https://en.wikipedia.org/wiki/File:Portal-puzzle.svg)**[**Portal de aprendizaje automático**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Portal:Machine_learning&usg=ALkJrhgLVFplRPZzUoKDzorHdVL0aURMyw) |
| [V](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Template:Machine_learning_bar&usg=ALkJrhj-HNydbgrcXJI791H4kPbnNUjE1Q)  [T](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Template_talk:Machine_learning_bar&usg=ALkJrhgk9_dLfWfSnEu8MpDzbP2Z7A-sXQ)  [mi](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/w/index.php%3Ftitle%3DTemplate:Machine_learning_bar%26action%3Dedit&usg=ALkJrhgJhn1M5Z48qJqoTp4VTgKhWvn8xQ) |

[](https://en.wikipedia.org/wiki/File:Cluster-2.svg)

El resultado de un análisis de agrupamiento se muestra como la coloración de los cuadrados en tres grupos.

**El análisis de clúster** o **agrupación** es la tarea de agrupar un conjunto de objetos de tal manera que los objetos en el mismo grupo (llamados un **grupo** ) son más similares (en algún sentido u otro) a los demás que a los de otros grupos (clusters) . Es una tarea principal de exploración exploratoria de [datos](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Data_mining&usg=ALkJrhgvrkYUd-C50pyD-jwLcxzfWmQRfQ) y una técnica común para [el análisis de datos](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Data_analysis&usg=ALkJrhixRDf_Oltsq_wo0Bkty2658kHTZw) [estadísticos](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Statistics&usg=ALkJrhguZx67dNUFb80jqo4lYoQKDNlUHw) , utilizada en muchos campos, incluyendo el aprendizaje automático, el reconocimiento [de](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Machine_learning&usg=ALkJrhiqyl1gcPCrbxDJThh3e7E51ejcTg) [patrones](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Pattern_recognition&usg=ALkJrhgDyhK7mXHT_bTTudPbSo1brtNmnA) , [el análisis de imágenes](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Image_analysis&usg=ALkJrhjVJCD1ogezpi4HS0mGNkNyhSS-RA) , la [recuperación de información](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Information_retrieval&usg=ALkJrhia3GakkxmmGFV5TjYALbi-kZoJEA) , la [bioinformática](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Bioinformatics&usg=ALkJrhjsGDYOyj92RtpwhLgggBiR7PVEPg) , [la compresión de datos](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Data_compression&usg=ALkJrhjNWWHBxl81rj5fjdaEYu4DDobOHg) y [la computación](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Data_compression&usg=ALkJrhjNWWHBxl81rj5fjdaEYu4DDobOHg) [gráfica](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Computer_graphics&usg=ALkJrhjG4KFZTAxzvfXsKVflKRyH9OGKWQ) .

El análisis de [clústeres en](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Algorithm&usg=ALkJrhgJWxARrTPHxFmXM_nAgNDJLzlhYw) sí no es un [algoritmo específico](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Algorithm&usg=ALkJrhgJWxARrTPHxFmXM_nAgNDJLzlhYw) , sino la tarea general a resolver. Puede lograrse mediante diversos algoritmos que difieren significativamente en su noción de lo que constituye un clúster y cómo encontrarlos eficientemente. Las nociones [populares de agrupaciones](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Distance_function&usg=ALkJrhhU0kLNSnfGoxGbq3psHQR_4AnF9Q) incluyen grupos con [distancias](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Distance_function&usg=ALkJrhhU0kLNSnfGoxGbq3psHQR_4AnF9Q) pequeñas entre los miembros del grupo, áreas densas del espacio de datos, intervalos o distribuciones [estadísticas particulares](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Statistical_distribution&usg=ALkJrhjPqrKCljJiCLgiDWxADPHM8V_l4A) . Por lo tanto, la agrupación puede ser formulada como un problema [de optimización multiobjetivo](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Multi-objective_optimization&usg=ALkJrhh_w498F5VVgNJsQ3glk7u5Z1PiJg) . El algoritmo de agrupación y la configuración de parámetros apropiados (incluidos los valores como la [función](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Metric_(mathematics)&usg=ALkJrhhSlJFN2M5ddcH8AzGbLyD2ke4MSA) de distancia a utilizar, el umbral de densidad o el número de clústeres esperados) dependen del conjunto de datos individuales y del uso previsto de los resultados. El análisis de clústeres como tal no es una tarea automática, sino un proceso [iterativo](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Knowledge_discovery&usg=ALkJrhiYu3_h84rANz1wEJ_y7JV53FG3dA) de [descubrimiento de conocimiento](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Knowledge_discovery&usg=ALkJrhiYu3_h84rANz1wEJ_y7JV53FG3dA) o una optimización multiobjetiva interactiva que implica juicio y fracaso. A menudo es necesario modificar el preprocesamiento de datos y los parámetros del modelo hasta que el resultado obtenga las propiedades deseadas.

Además del término *agrupación* , hay una serie de términos con significados *similares* , incluyendo *clasificación*[*automática*](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Statistical_classification&usg=ALkJrhihMHjJB29QEOc8lvZR8PilJ9HHOQ) , [*taxonomía numérica*](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Numerical_taxonomy&usg=ALkJrhhKXyczBrG2UFlvQ-Ie5_NMZEu-KQ) , *botryology* (del griego βότρυς "uva") y *análisis tipológico* . Las sutiles diferencias son a menudo en el uso de los resultados: mientras que en la minería de datos, los grupos resultantes son la cuestión de interés, en la clasificación automática de la potencia discriminativa resultante es de interés.

El análisis del racimo fue originado en la antropología por el conductor y Kroeber en 1932 y introducido a la psicología por Zubin en 1938 y [Roberto Tryon](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Robert_Tryon&usg=ALkJrhjeO2mb1oLThcPacs61FmYI_iHP0A) en 1939 y famoso utilizado por Cattell a partir de 1943 para la clasificación de la teoría del rasgo en psicología de la personalidad.

**Contenido**

[1Definición](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#Definition)

[2Algoritmos](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#Algorithms)

[2.1Clústeres basados ​​en conectividad (agrupación jerárquica)](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#Connectivity-based_clustering_.28hierarchical_clustering.29)

[2.2Centralización basada en Centroid](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#Centroid-based_clustering)

[2.3Distribución basada en la agrupación](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#Distribution-based_clustering)

[2.4Agrupación basada en densidad](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#Density-based_clustering)

[2.5Evolución reciente](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#Recent_developments)

[2.6Otros métodos](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#Other_methods)

[3Evaluación y evaluación](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#Evaluation_and_assessment)

[3.1Evaluación interna](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#Internal_evaluation)

[3.2Evaluación externa](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#External_evaluation)

[4Aplicaciones](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#Applications)

[5Ver también](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#See_also)

[5.1Tipos especializados de análisis de agrupaciones](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#Specialized_types_of_cluster_analysis)

[5.2Técnicas utilizadas en el análisis de conglomerados](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#Techniques_used_in_cluster_analysis)

[5.3Proyección de datos y preprocesamiento](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#Data_projection_and_preprocessing)

[5.4Otros](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#Other)

[6Referencias](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#References)

[7Enlaces externos](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#External_links)

Definición [ [editar](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/w/index.php%3Ftitle%3DCluster_analysis%26action%3Dedit%26section%3D1&usg=ALkJrhjgd6H2L7QRCvGFYPQeckqI0W7zGw" \o "Editar sección: Definición) ]

La noción de un "cluster" no se puede definir con precisión, que es una de las razones por las que hay tantos algoritmos de agrupación. [[4]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-estivill-4)Existe un denominador común: un grupo de objetos de datos. Sin embargo, diferentes investigadores utilizan diferentes modelos de cluster, y para cada uno de estos modelos de clúster de nuevo se pueden dar diferentes algoritmos. La noción de un conglomerado, tal como se encuentra por diferentes algoritmos, varía significativamente en sus propiedades. La comprensión de estos "modelos de clúster" es clave para entender las diferencias entre los diversos algoritmos. Los modelos de clúster típicos incluyen:

Modelos de conectividad: por ejemplo, la [agrupación jerárquica construye](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Hierarchical_clustering&usg=ALkJrhhP79NNtd9SRZS9LyKVX0KcKW9ukA) modelos basados ​​en la conectividad a distancia.

Modelos Centroid: por ejemplo, el [algoritmo k-means](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/K-means_algorithm&usg=ALkJrhgOsB5DwkyOmb9hxcOYPfHEEpBA6g) representa cada clúster por un solo vector medio.

Modelos de distribución: los clústeres se modelan utilizando distribuciones estadísticas, tales como [distribuciones normales multivariadas utilizadas](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Multivariate_normal_distribution&usg=ALkJrhigmmlgkvReAHHtFK0Fv8ZRLtYJiQ) por el [algoritmo Expectativa-maximización](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Expectation-maximization_algorithm&usg=ALkJrhhbhcqCPd7wWL9JgGZY5xRPuu_pdw) .

Modelos de densidad: por ejemplo, [DBSCAN](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/DBSCAN&usg=ALkJrhjRaTKu8cpbIDTI13z_zVmUycbpiQ) y [OPTICS](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/OPTICS&usg=ALkJrhjLkot0aTxqf-A7r-P0BGSFU1M66A) definen clusters como regiones densas conectadas en el espacio de datos.

Modelos de [subspace](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Biclustering&usg=ALkJrhie_M_qGePNvOW21ajIe07rbw53xQ) : en [Biclustering](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Biclustering&usg=ALkJrhie_M_qGePNvOW21ajIe07rbw53xQ) (también conocido como Co-clustering o clustering de dos modos), los clústeres se modelan con los miembros del clúster y los atributos relevantes.

Modelos de grupo: algunos algoritmos no proporcionan un modelo refinado para sus resultados y sólo proporcionan la información de agrupación.

Modelos basados ​​en gráficos: una [agrupación](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Clique_(graph_theory)&usg=ALkJrhhCD0UUe4laCHYxgKi-dNn-epz_aQ) , es decir, un subconjunto de nodos en una [gráfica](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Graph_(discrete_mathematics)&usg=ALkJrhhcw0QI0lZOYSH9_GDXhEIvBGvxOg) tal que cada dos nodos en el subconjunto están conectados por un borde puede considerarse como una forma prototípica de agrupación. Las relajaciones del requisito de conectividad completa (una fracción de los bordes pueden faltar) se conocen como cuasi-cliques, como en el [algoritmo de agrupación HCS](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/HCS_clustering_algorithm&usg=ALkJrhhKGthz-OmNB-bfDdPzqs31IlBEFw) .

Un "agrupamiento" es esencialmente un conjunto de tales agrupaciones, que por lo general contienen todos los objetos del conjunto de datos. Además, puede especificar la relación de los clústeres entre sí, por ejemplo, una jerarquía de clústeres incrustados entre sí. Los agrupamientos se pueden distinguir como:

Clustering duro: cada objeto pertenece a un clúster o no

Clustering suave (también: [clustering fuzzy](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Fuzzy_clustering&usg=ALkJrhh2hFJGK-_rILTO3O8lR-TX5OTG9w) ): cada objeto [pertenece](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Fuzzy_clustering&usg=ALkJrhh2hFJGK-_rILTO3O8lR-TX5OTG9w) a cada grupo a un cierto grado (por ejemplo, una probabilidad de pertenecer al cluster)

También hay distinciones más finas posibles, por ejemplo:

Clustering de particionamiento estricto: aquí cada objeto pertenece exactamente a un cluster

Clustering de particionamiento estricto con outliers: los objetos también pueden pertenecer a ningún cluster y se consideran [valores atípicos](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Anomaly_detection&usg=ALkJrhhQJ0-GweI0Xv6IEQwqj4P8y7Ar4g) .

Superposición de clústeres (también: clustering alternativo, multi-view clustering): mientras que normalmente un clustering duro, los objetos pueden pertenecer a más de un cluster.

Clúster jerárquico: los objetos que pertenecen a un clúster secundario también pertenecen al clúster principal

[Agrupación de subespacios](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Subspace_clustering&usg=ALkJrhjDRR1FDZbRUUWPs32wiNW0nEJ-aw) : mientras que una agrupación superpuesta, dentro de un subespacio definido de forma única, no se espera que los conglomerados se superpongan.

Algoritmos [ [editar](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/w/index.php%3Ftitle%3DCluster_analysis%26action%3Dedit%26section%3D2&usg=ALkJrhhW25wrz4D0aaKfmXkHZyYWXMF-YQ" \o "Editar sección: Algoritmos) ]

[*Categoría principal*](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Category:Data_clustering_algorithms&usg=ALkJrhgDp5mcCsDDx-eAkmvxvGSNe2Yg9A)*:*[*Algoritmos de agrupación de datos*](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Category:Data_clustering_algorithms&usg=ALkJrhgDp5mcCsDDx-eAkmvxvGSNe2Yg9A)

Los algoritmos de agrupación pueden categorizarse basándose en su modelo de clúster, como se menciona anteriormente. La siguiente descripción sólo enumerará los ejemplos más destacados de algoritmos de agrupación, ya que posiblemente hay más de 100 algoritmos de agrupación publicados. No todos ofrecen modelos para sus clusters y, por lo tanto, no se pueden categorizar fácilmente. Un resumen de los algoritmos explicados en Wikipedia se puede encontrar en la [lista de algoritmos de estadística](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/List_of_algorithms&usg=ALkJrhgurszSqC1kr-fWnK4qAYA2tWsP2A#Statistics) .

No existe un algoritmo de agrupación objetivamente "correcto", pero como se observó, "la agrupación está en el ojo del observador". [[4]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-estivill-4) El algoritmo de agrupamiento más apropiado para un problema particular a menudo necesita ser elegido experimentalmente, a menos que exista una razón matemática para preferir un modelo de agrupamiento sobre otro. Cabe señalar que un algoritmo que está diseñado para un tipo de modelo no tiene ninguna posibilidad en un conjunto de datos que contiene un tipo radicalmente diferente de modelo. [[4]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-estivill-4) Por ejemplo, k-means no puede encontrar clusters no convexos. [[4]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-estivill-4)

**Clustering basado en conectividad (clustering jerárquico)**[ [editar](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/w/index.php%3Ftitle%3DCluster_analysis%26action%3Dedit%26section%3D3&usg=ALkJrhjI9gQ8pte0dfI-RdJhOns0sH_GaA" \o "Sección de edición: agrupación basada en conectividad (agrupación jerárquica)) ]

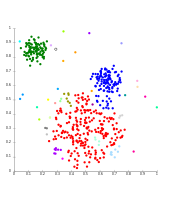
*Artículo*[*principal*](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Hierarchical_clustering&usg=ALkJrhhP79NNtd9SRZS9LyKVX0KcKW9ukA)*:*[*agrupación jerárquica*](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Hierarchical_clustering&usg=ALkJrhhP79NNtd9SRZS9LyKVX0KcKW9ukA)

El clustering basado en conectividad, también conocido como [*clustering jerárquico*](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Hierarchical_clustering&usg=ALkJrhhP79NNtd9SRZS9LyKVX0KcKW9ukA) , se basa en la idea central de que los objetos están más [*relacionados*](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Hierarchical_clustering&usg=ALkJrhhP79NNtd9SRZS9LyKVX0KcKW9ukA) con objetos cercanos que con objetos más alejados. Estos algoritmos conectan "objetos" para formar "clusters" basados ​​en su distancia. Un grupo se puede describir en gran medida por la distancia máxima necesaria para conectar las partes del grupo. A distancias diferentes, se formarán diferentes clusters, que se pueden representar usando un [dendrograma](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Dendrogram&usg=ALkJrhgVMxBDtmgqHJpQxwKwuWIpmC6FYw) , que explica donde se [origina](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Dendrogram&usg=ALkJrhgVMxBDtmgqHJpQxwKwuWIpmC6FYw) el nombre común "clustering jerárquico": estos algoritmos no proporcionan un solo particionamiento del conjunto de datos, sino que proporcionan una extensa jerarquía de Racimos que se combinan entre sí a ciertas distancias. En un dendrograma, el eje y marca la distancia a la que se agrupan los cúmulos, mientras que los objetos se colocan a lo largo del eje x de tal manera que los cúmulos no se mezclan.

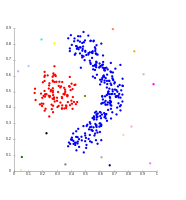
El clustering basado en conectividad es una familia entera de métodos que difieren por el modo en que se calculan las distancias. Aparte de la elección habitual de [funciones de distancia](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Distance_function&usg=ALkJrhhU0kLNSnfGoxGbq3psHQR_4AnF9Q) , el usuario también tiene que decidir sobre el criterio de vinculación (ya que un grupo consta de varios objetos, hay varios candidatos para calcular la distancia a) para usar. Las [elecciones populares](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Single-linkage_clustering&usg=ALkJrhjnAna-fukWJld4XteePxVPEee6hQ) se conocen como [clustering de](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Single-linkage_clustering&usg=ALkJrhjnAna-fukWJld4XteePxVPEee6hQ) un [solo enlace](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Single-linkage_clustering&usg=ALkJrhjnAna-fukWJld4XteePxVPEee6hQ) (el mínimo de distancias de objetos), [clustering de enlace completo](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Complete_linkage_clustering&usg=ALkJrhh3HnLsTwTpfPx7bf-8h-6jBbzVbg) (el máximo de distancias de objetos) o [UPGMA](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/UPGMA&usg=ALkJrhh9QH2RaITIOiTwImP4LQ0GhX599A) ("Método de grupo de pares no ponderado con media aritmética"). Además, el agrupamiento jerárquico puede ser aglomerado (comenzando con elementos únicos y agregándolos en grupos) o divisivo (comenzando con el conjunto de datos completo y dividiéndolo en particiones).

Estos métodos no producirán una partición única del conjunto de datos, sino una jerarquía a partir de la cual el usuario todavía debe elegir clústeres apropiados. No son muy robustos con los valores atípicos, que se presentan como grupos adicionales o incluso hacen que otros clústeres se fusionen (conocido como "fenómeno de encadenamiento", en particular con el agrupamiento [de enlace único](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Single-linkage_clustering&usg=ALkJrhjnAna-fukWJld4XteePxVPEee6hQ) ). En el caso general, la complejidad  {\ Displaystyle {\ mathcal {O}} (n ^ {3})} Para aglomeración aglomerada y  {\ Displaystyle {\ mathcal {O}} (2 ^ {n-1})} Para [clustering divisivo](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Divisive_clustering&usg=ALkJrhju1T-6lAxOXssOQRlQ7yhhs8KxXQ) , [[5]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-5) lo que los hace [demasiado](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Divisive_clustering&usg=ALkJrhju1T-6lAxOXssOQRlQ7yhhs8KxXQ) lento para grandes [conjuntos de](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Divisive_clustering&usg=ALkJrhju1T-6lAxOXssOQRlQ7yhhs8KxXQ) datos. Para algunos casos especiales, los métodos óptimos eficientes (de complejidad  {\ Displaystyle {\ mathcal {O}} (n ^ {2})} ) Son conocidos: SLINK [[6]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-6) para single-linkage y CLINK [[7]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-7) para clustering de enlace completo. En la comunidad de [minería de datos,](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Data_mining&usg=ALkJrhgvrkYUd-C50pyD-jwLcxzfWmQRfQ) estos métodos se reconocen como una base teórica del análisis de conglomerados, pero a menudo se consideran obsoletos. Sin embargo, proporcionaron inspiración para muchos métodos posteriores, como la agrupación basada en densidad.

**Ejemplos de agrupamiento de enlaces**

[](https://en.wikipedia.org/wiki/File:SLINK-Gaussian-data.svg)

Single-linkage en datos gaussianos. En 35 clusters, el clúster más grande comienza a fragmentarse en partes más pequeñas, mientras que antes de que todavía estaba conectado a la segunda mayor debido al efecto de enlace único.

[](https://en.wikipedia.org/wiki/File:SLINK-density-data.svg)

Single-linkage en clusters basados ​​en densidad. 20 clusters extraídos, la mayoría de los cuales contienen elementos únicos, ya que el agrupamiento de enlaces no tiene una noción de "ruido".

**Centroid basado en la agrupación**[ [editar](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/w/index.php%3Ftitle%3DCluster_analysis%26action%3Dedit%26section%3D4&usg=ALkJrhjW5aFLSc-VRD_CmjlUY_EgIIkmcA" \o "Sección de edición: agrupación basada en Centroid) ]

*Artículo*[*principal*](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/K-means_clustering&usg=ALkJrhiyUH_pwirgtWP1N9GqWWA1gduRng)*:*[*k-means clustering*](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/K-means_clustering&usg=ALkJrhiyUH_pwirgtWP1N9GqWWA1gduRng)

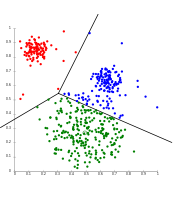
En agrupaciones basadas en centroides, los clusters están representados por un vector central, que no necesariamente puede ser un miembro del conjunto de datos. Cuando el número de [clústeres](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/K-means_clustering&usg=ALkJrhiyUH_pwirgtWP1N9GqWWA1gduRng) se fija a k, el [agrupamiento *k* -means](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/K-means_clustering&usg=ALkJrhiyUH_pwirgtWP1N9GqWWA1gduRng) da una definición formal como un problema de optimización:  {\ Displaystyle k} Agrupar centros y asignar los objetos al centro de clúster más cercano, de tal manera que las distancias al cuadrado del clúster se minimicen.

El problema de optimización en sí se conoce como [NP-duro](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/NP-hard&usg=ALkJrhhtHoM8OKkMq_Jh8XPGBhB4pTt-cQ) , y por lo tanto el enfoque común es buscar sólo soluciones aproximadas. Un método aproximativo particularmente bien conocido es [el algoritmo de Lloyd](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Lloyd%27s_algorithm&usg=ALkJrhjL1R1KdVwXsJjIkBkxBJQcgKxubQ) , [[8]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-lloyd-8) a menudo conocido como " *algoritmo k-means* ". Sin embargo, sólo encuentra un [óptimo local](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Local_optimum&usg=ALkJrhj-5kXytWscneu6pY3ryhU5LPhlew) , y se ejecuta comúnmente varias veces con diferentes inicializaciones aleatorias. Las variaciones de k-medios a menudo incluyen optimizaciones tales como elegir el mejor de múltiples ejecuciones, pero también restringir los centroides a los miembros del conjunto de datos ( [k-medoides](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/K-medoids&usg=ALkJrhiP1NUCSk-sF1PIWFJTKlOSersTNw) ), la elección de [medianas](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Median&usg=ALkJrhhAZT62ZGMTBkUx2tKYZF27NP58BQ) ( [k-medianas agrupación](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/K-medians_clustering&usg=ALkJrhh4197n9TOQjHwkORAmQW0NX6e1UA) ), la elección de los centros [iniciales](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/K-medians_clustering&usg=ALkJrhh4197n9TOQjHwkORAmQW0NX6e1UA) menos al azar [K-means ++](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/K-means%252B%252B&usg=ALkJrhgcHf5vJE8DD5JTkDsdgMBsq6ovIQ) ) o permitir una asignación de clúster difusa ( [Fuzzy c-means](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Fuzzy_clustering&usg=ALkJrhh2hFJGK-_rILTO3O8lR-TX5OTG9w) ).

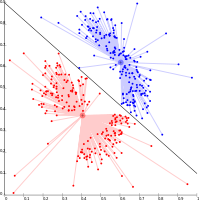
La mayoría de los algoritmos k-means-type requieren el [número de clusters](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Determining_the_number_of_clusters_in_a_data_set&usg=ALkJrhgQSWstVWTu5CHZIojQ_CWwM7ELbA) -  {\ Displaystyle k} - que se especificará de antemano, que se considera uno de los mayores inconvenientes de estos algoritmos. Además, los algoritmos prefieren clústeres de tamaño aproximadamente similar, ya que siempre asignarán un objeto al centroide más cercano. Esto a menudo conduce a cortar incorrectamente bordes entre clústeres (lo cual no es sorprendente, ya que el algoritmo optimizó los centros de agrupación, no los bordes de los clústeres).

K-means tiene una serie de propiedades teóricas interesantes. En primer lugar, divide el espacio de datos en una estructura conocida como [diagrama de Voronoi](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Voronoi_diagram&usg=ALkJrhjEGkYaMgqfWGLX-CXcR_K4ZIoyTQ) . En segundo lugar, está conceptualmente cerca de la clasificación del vecino más cercano, y como tal es popular en el aprendizaje automático. En tercer lugar, puede verse como una variación de la clasificación basada en modelos, y el algoritmo de Lloyd como una variación del [algoritmo de maximización de expectativas](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Expectation-maximization_algorithm&usg=ALkJrhhbhcqCPd7wWL9JgGZY5xRPuu_pdw) para este modelo que se analiza a continuación.

**Ejemplos de clustering k-Means**

[](https://en.wikipedia.org/wiki/File:KMeans-Gaussian-data.svg)

K-means separa los datos en células de Voronoi, lo que supone clusters de igual tamaño (no adecuados aquí)

[](https://en.wikipedia.org/wiki/File:KMeans-density-data.svg)

K-means no puede representar clusters basados ​​en la densidad

**Distribución basada en la agrupación**[ [editar](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/w/index.php%3Ftitle%3DCluster_analysis%26action%3Dedit%26section%3D5&usg=ALkJrhj8gotaFrgm12tUaq83bJi3uGnYVQ" \o "Sección de edición: agrupación basada en distribución) ]

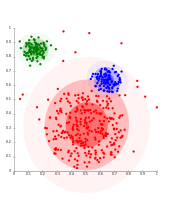
El modelo de agrupamiento más estrechamente relacionado con las estadísticas se basa en [modelos de distribución](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Probability_distribution&usg=ALkJrhjKvNPGo2Eytbap4gF8x0U6P0KI9A) . Los clústeres pueden definirse fácilmente como objetos pertenecientes a la misma distribución. Una propiedad conveniente de este enfoque es que se parece mucho a la manera en que se generan conjuntos de datos artificiales: mediante el muestreo de objetos aleatorios de una distribución.

Si bien el fundamento teórico de estos métodos es excelente, sufren de un problema clave conocido como [superacondicionamiento](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Overfitting&usg=ALkJrhibd9y-tlJuhYGZblokS11eeM5e-A) , a menos que se pongan restricciones a la complejidad del modelo. Un modelo más complejo normalmente será capaz de explicar mejor los datos, lo que hace que la elección de la complejidad del modelo apropiado sea intrínsecamente difícil.

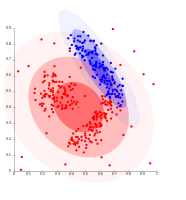
Un método prominente es conocido como modelos de mezcla [gaussiana (utilizando](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Expectation-maximization_algorithm&usg=ALkJrhhbhcqCPd7wWL9JgGZY5xRPuu_pdw) el [algoritmo de expectativa-maximización](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Expectation-maximization_algorithm&usg=ALkJrhhbhcqCPd7wWL9JgGZY5xRPuu_pdw) ). En este caso, el conjunto de datos suele ser modelado con un número fijo (para evitar sobreesfuerzos) de [distribuciones gaussianas](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Gaussian_distribution&usg=ALkJrhhQw6BF-4eGNra_dig1tx67GBE7kw) que se inicializan al azar y cuyos parámetros se optimizan iterativamente para ajustarse mejor al conjunto de datos. Esto convergerá a un [óptimo local](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Local_optimum&usg=ALkJrhj-5kXytWscneu6pY3ryhU5LPhlew) , por lo que múltiples ejecuciones pueden producir resultados diferentes. Con el fin de obtener un clustering duro, los objetos a menudo se asignan a la distribución gaussiana a la que probablemente pertenecen; Para las agrupaciones suaves, esto no es necesario.

El clustering basado en distribución produce modelos complejos para clusters que pueden capturar [correlación y dependencia](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Correlation_and_dependence&usg=ALkJrhh_cxora0g02AHMUP0CllRndMns3Q) entre atributos. Sin embargo, estos algoritmos suponen una carga adicional para el usuario: para muchos conjuntos de datos reales, no puede haber un modelo matemático definido concisamente (por ejemplo, suponiendo que las distribuciones gaussianas son una suposición bastante fuerte en los datos).

**Expectación-Maximización (EM) ejemplos de agrupación**

[](https://en.wikipedia.org/wiki/File:EM-Gaussian-data.svg)

En los datos Gaussian-distribuidos, EM trabaja bien, puesto que utiliza Gaussians para los racimos de modelado

[](https://en.wikipedia.org/wiki/File:EM-density-data.svg)

Los clusters basados ​​en densidad no pueden ser modelados usando distribuciones gaussianas

**Densidad basada en la agrupación**[ [editar](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/w/index.php%3Ftitle%3DCluster_analysis%26action%3Dedit%26section%3D6&usg=ALkJrhh2jO1IxXQZjyhxC8iCjYJxeTcjPg" \o "Sección de edición: agrupación basada en densidad) ]

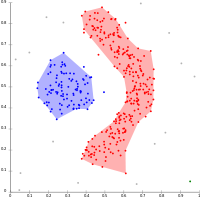
En la agrupación basada en la densidad, [[9] los](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-9) grupos se definen como áreas de mayor densidad que el resto del conjunto de datos. Los objetos en estas áreas dispersas -que son requeridos para separar clústeres- suelen considerarse ruidos y puntos fronterizos.

El método de agrupación basado en densidad más popular [[10]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-10) es [DBSCAN](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/DBSCAN&usg=ALkJrhjRaTKu8cpbIDTI13z_zVmUycbpiQ) . [[11]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-11) En contraste con muchos métodos más nuevos, ofrece un modelo bien definido del racimo llamado "densidad-accesibilidad". Similar al agrupamiento basado en enlaces, se basa en puntos de conexión dentro de ciertos umbrales de distancia. Sin embargo, sólo conecta puntos que satisfagan un criterio de densidad, en la variante original definida como un número mínimo de otros objetos dentro de este radio. Un clúster consiste en todos los objetos conectados a la densidad (que pueden formar un grupo de una forma arbitraria, en contraste con muchos otros métodos) más todos los objetos que están dentro del alcance de estos objetos. Otra propiedad interesante de DBSCAN es que su complejidad es bastante baja -se requiere un número lineal de consultas de rango en la base de datos- y que descubrirá esencialmente los mismos resultados (es [determinista](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Deterministic_algorithm&usg=ALkJrhjefHsf-lB-jK3l8akNzytskICXLQ) para núcleos y puntos de ruido, pero no para puntos [fronterizos](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Deterministic_algorithm&usg=ALkJrhjefHsf-lB-jK3l8akNzytskICXLQ) ) En cada ejecución, por lo tanto no hay necesidad de ejecutarlo varias veces. [OPTICS](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/OPTICS_algorithm&usg=ALkJrhib6DU3e0hcx6PelkSkhqKE1JRK9w)[[12]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-12) es una generalización de DBSCAN que elimina la necesidad de elegir un valor apropiado para el parámetro de rango  {\ Displaystyle \ varepsilon} , Y produce un resultado jerárquico relacionado con el [agrupamiento](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Hierarchical_clustering&usg=ALkJrhhP79NNtd9SRZS9LyKVX0KcKW9ukA) de [enlaces](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Hierarchical_clustering&usg=ALkJrhhP79NNtd9SRZS9LyKVX0KcKW9ukA) . DeLi-Clu, [[13]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-13) Density-Link-Clustering combina ideas de clustering de [un solo enlace](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Single-linkage_clustering&usg=ALkJrhjnAna-fukWJld4XteePxVPEee6hQ) y OPTICS, eliminando la  {\ Displaystyle \ varepsilon} Parámetro completamente y ofreciendo mejoras de rendimiento sobre OPTICS mediante el uso de un índice de [árbol R.](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/R-tree&usg=ALkJrhjFzfd3-p711fiNrBibDW6Om8Ps9Q)

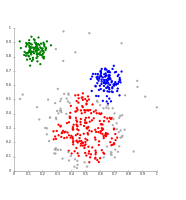
El [inconveniente](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/DBSCAN&usg=ALkJrhjRaTKu8cpbIDTI13z_zVmUycbpiQ) clave de [DBSCAN](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/DBSCAN&usg=ALkJrhjRaTKu8cpbIDTI13z_zVmUycbpiQ) y [OPTICS](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/OPTICS&usg=ALkJrhjLkot0aTxqf-A7r-P0BGSFU1M66A) es que esperan algún tipo de caída de densidad para detectar los bordes de los [clústeres](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/DBSCAN&usg=ALkJrhjRaTKu8cpbIDTI13z_zVmUycbpiQ) . Además, no pueden detectar estructuras de clúster intrínsecas que son prevalentes en la mayoría de los datos de la vida real. Una variación de DBSCAN, [EnDBSCAN](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/w/index.php%3Ftitle%3DEnDBSCAN_algorithm%26action%3Dedit%26redlink%3D1&usg=ALkJrhjukifTQDIL1z4KKHXkd0AZFUlExg) , [[14]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-14) detecta eficientemente este tipo de estructuras. En conjuntos de datos con, por ejemplo, distribuciones Gaussianas solapadas - un caso de uso común en datos artificiales - los bordes de clúster producidos por estos algoritmos a menudo parecerán arbitrarios, debido a que la densidad de clúster disminuye continuamente. En un conjunto de datos que consiste en mezclas de gaussianos, estos algoritmos son casi siempre superados por métodos como [EM clustering](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Expectation%25E2%2580%2593maximization_algorithm&usg=ALkJrhjpPeh2ys5liYOQtGXtGJOfnx1JUw) que son capaces de modelar con precisión este tipo de datos.

[El cambio medio](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Mean-shift&usg=ALkJrhi6aad6Qc8wtfqMqHdKoJj-XLkSvw) es un enfoque de agrupamiento en el que cada objeto se mueve al área más densa en su vecindad, basado en la [estimación de la densidad del núcleo](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Kernel_density_estimation&usg=ALkJrhiPJLMYnOcc7-qXqF46zouX9vBS_g) . Eventualmente, los objetos convergen a los máximos locales de densidad. Similar a la agrupación de k-means, estos "atractores de densidad" pueden servir como representantes para el conjunto de datos, pero el cambio de media puede detectar clústeres de forma arbitraria similares a DBSCAN. Debido al costoso procedimiento iterativo ya la estimación de la densidad, el cambio de media suele ser más lento que DBSCAN o k-Means.

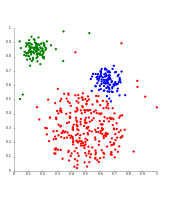
**Ejemplos de clustering basado en densidad**

[](https://en.wikipedia.org/wiki/File:DBSCAN-density-data.svg)

Densidad basada en la agrupación con [DBSCAN](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/DBSCAN&usg=ALkJrhjRaTKu8cpbIDTI13z_zVmUycbpiQ) .

[](https://en.wikipedia.org/wiki/File:DBSCAN-Gaussian-data.svg)

[DBSCAN](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/DBSCAN&usg=ALkJrhjRaTKu8cpbIDTI13z_zVmUycbpiQ) supone [agrupaciones](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/DBSCAN&usg=ALkJrhjRaTKu8cpbIDTI13z_zVmUycbpiQ) de densidad similar, y puede tener problemas para separar los clusters cercanos

[](https://en.wikipedia.org/wiki/File:OPTICS-Gaussian-data.svg)

[OPTICS](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/OPTICS_algorithm&usg=ALkJrhib6DU3e0hcx6PelkSkhqKE1JRK9w) es una variante DBSCAN que maneja diferentes densidades mucho mejor

**Desarrollos recientes**

En los últimos años se han realizado esfuerzos considerables para mejorar el rendimiento de los algoritmos existentes. [[15]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-15) [[16]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-16) Entre ellos están [*CLARANS*](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/w/index.php%3Ftitle%3DCLARANS%26action%3Dedit%26redlink%3D1&usg=ALkJrhixJMM9xIEodw3UiFls0pwktxNpJg) (Ng y Han, 1994), [[17]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-17) y [*BIRCH*](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Birch_(data_clustering)&usg=ALkJrhikx6PFTrwSyIBfknAUinE9cF0OwA) (Zhang et al., 1996). [[18]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-18) Con la reciente necesidad de procesar conjuntos de datos cada vez mayores (también conocidos como [grandes datos](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Big_data&usg=ALkJrhhbdKt-ouD_qXRX_1RwjUQQAobjNA) ), la voluntad de intercambiar el significado semántico de los clusters generados por el desempeño ha ido en aumento. Esto llevó al desarrollo de métodos de pre-clustering, tales como [clustering de copa](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Canopy_clustering_algorithm&usg=ALkJrhi1mwek-rTfUU4B-SOXdgspt5rGkw) , que puede procesar enormes conjuntos de datos de manera eficiente, pero los "clusters" resultantes son simplemente una pre-partición aproximada del conjunto de datos para luego analizar las particiones con los métodos más lentos existentes Como [k-significa la agrupación](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/K-means_clustering&usg=ALkJrhiyUH_pwirgtWP1N9GqWWA1gduRng) . Varios otros enfoques a la agrupación se han intentado, como la semilla basada en la agrupación. [[19]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-19)

Para [los datos de alta dimensión](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/High-dimensional_space&usg=ALkJrhhZhAV__kVJbWdvjXdpAleQM7EyCw) , muchos de los métodos [existentes](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Curse_of_dimensionality&usg=ALkJrhgzYWiTmfQahnDaXbcXIPYGYi5sAg) fracasan debido a la [maldición de la dimensionalidad](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Curse_of_dimensionality&usg=ALkJrhgzYWiTmfQahnDaXbcXIPYGYi5sAg) , lo que hace que las funciones particulares de distancia sean problemáticas en los espacios de alta dimensión. Esto condujo a nuevos [algoritmos de agrupamiento para datos de alta dimensión](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Clustering_high-dimensional_data&usg=ALkJrhi1KfryEhiBeqKxVOANr41oKW7Mqg) que se enfocan en el [agrupamiento subespacial](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Subspace_clustering&usg=ALkJrhjDRR1FDZbRUUWPs32wiNW0nEJ-aw) (donde sólo se usan algunos atributos, y los modelos de clúster incluyen los atributos relevantes para el clúster) y [clustering de correlación](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Correlation_clustering&usg=ALkJrhjFehrQc5CPPDXlELzCyQIx_lwwnw) que también busca subespacio girado arbitrario (correlacionado) Clusters que pueden ser [modelados](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Correlation&usg=ALkJrhgZ6_-x9C8NhFVu5vn3BfYW_q_KFQ) dando una [correlación](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Correlation&usg=ALkJrhgZ6_-x9C8NhFVu5vn3BfYW_q_KFQ) de sus atributos. [[20]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-20) Los ejemplos de estos algoritmos de agrupamiento son CLIQUE [[21]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-21) y [SUBCLU](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/SUBCLU&usg=ALkJrhizdWiRTSXBfyE7Kg46EtkdN7c6Xw) . [[22]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-22)

Se han adoptado ideas a partir de métodos de agrupación basados ​​en densidad (en particular la familia de algoritmos [DBSCAN](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/DBSCAN&usg=ALkJrhjRaTKu8cpbIDTI13z_zVmUycbpiQ) / [OPTICS](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/OPTICS&usg=ALkJrhjLkot0aTxqf-A7r-P0BGSFU1M66A) ) para clasificar en subespacio (HiSC, [[23]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-23)clustering subespacial jerárquico y DiSH [[24]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-24) ) y correlación jerárquica Clustering, 4C [[26]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-26) utilizando "correlación de conectividad" y ERiC [[27]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-27) exploración jerárquica basada en la densidad de clusters de correlación).

Se han propuesto varios sistemas de agrupamiento basados ​​en [información mutua](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Mutual_information&usg=ALkJrhjD98sS21F2_tLjeOW-v5vmmK5dog) . Uno es la métrica de [*información de*](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Variation_of_information&usg=ALkJrhibiCTVsYnZ77Lewpu_KV2nir45pg) Marina Meilă; [[28]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-28) otro proporciona clustering jerárquico. [[29]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-29)Utilizando algoritmos genéticos, se puede optimizar una amplia gama de diferentes funciones de ajuste, incluida la información mutua. [[30]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-30) También [los](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Message_passing&usg=ALkJrhh3PYUeInvxvsoPS30UBIyaodcrpg) algoritmos de [paso de](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Message_passing&usg=ALkJrhh3PYUeInvxvsoPS30UBIyaodcrpg) mensajes, un desarrollo reciente en [Ciencias de la](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Computer_Science&usg=ALkJrhh5jijrzGJDAqTQ4kcpOO2yUKCFEQ) [Computación](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Statistical_Physics&usg=ALkJrhi5ZfEXOqAPWnKVU0DvVoO7WeCEYg) y [Física Estadística](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Statistical_Physics&usg=ALkJrhi5ZfEXOqAPWnKVU0DvVoO7WeCEYg) , ha llevado a la creación de nuevos tipos de algoritmos de agrupamiento. [[31]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-31)

**Otros métodos**[ [editar](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/w/index.php%3Ftitle%3DCluster_analysis%26action%3Dedit%26section%3D8&usg=ALkJrhgRC9cOn5SjT8Y3taPCowX7GPr5mA" \o "Editar sección: Otros métodos) ]

[Esquema algorítmico secuencial básico](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Basic_sequential_algorithmic_scheme&usg=ALkJrhjGZsRDEaYlok2qMQK2BxpV10xXRw) (BSAS)

Evaluación y evaluación [ [editar](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/w/index.php%3Ftitle%3DCluster_analysis%26action%3Dedit%26section%3D9&usg=ALkJrhghpN03GX7EK6Ah_ZKHoylIJ8WoRA" \o "Editar sección: Evaluación y evaluación) ]

A veces, la evaluación de los resultados de agrupamiento se denomina validación de clúster.

Ha habido varias sugerencias para una medida de similitud entre dos agrupaciones. Una medida de este tipo puede utilizarse para comparar la eficacia de los diferentes algoritmos de agrupación de datos en un conjunto de datos. Estas medidas suelen estar ligadas al tipo de criterio que se considera en la evaluación de la calidad de un método de agrupamiento.

**Evaluación interna**[ [editar](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/w/index.php%3Ftitle%3DCluster_analysis%26action%3Dedit%26section%3D10&usg=ALkJrhgFIPXSu8jUFPu1Pv_1yNNQFzfo5w" \o "Sección de edición: Evaluación interna) ]

*Véase también:*[*Determinación del número de clústeres en un conjunto de datos*](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Determining_the_number_of_clusters_in_a_data_set&usg=ALkJrhgQSWstVWTu5CHZIojQ_CWwM7ELbA)

Cuando se evalúa un resultado de agrupación en función de los datos que se agruparon a sí mismo, se denomina evaluación interna. Estos métodos usualmente asignan la mejor puntuación al algoritmo que produce clusters con alta similitud dentro de un cluster y baja similitud entre clusters. Un inconveniente de la utilización de criterios internos en la evaluación de grupos es que las altas puntuaciones en una medida interna no necesariamente resultan en aplicaciones de recuperación de información eficaces. [[32]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-Christopher_D._Manning.2C_Prabhakar_Raghavan_.26_Hinrich_Schutze-32) Además, esta evaluación está sesgada hacia los algoritmos que utilizan el mismo modelo de cluster. Por ejemplo, el agrupamiento de k-Means optimiza naturalmente las distancias de objeto, y un criterio interno basado en la distancia probablemente sobrevalorará el agrupamiento resultante.

Por lo tanto, las medidas de evaluación interna son las más adecuadas para obtener una idea de las situaciones en las que un algoritmo funciona mejor que otro, pero esto no implica que un algoritmo produzca resultados más válidos que otro. [[4] La](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-estivill-4) validez medida por dicho índice depende de la afirmación de que este tipo de estructura existe en el conjunto de datos. Un algoritmo diseñado para algún tipo de modelos no tiene ninguna posibilidad si el conjunto de datos contiene un conjunto radicalmente diferente de modelos, o si la evaluación mide un criterio radicalmente diferente. [[4]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-estivill-4) Por ejemplo, el agrupamiento de k-means sólo puede encontrar agrupaciones convexas, y muchos índices de evaluación asumen grupos convexos. En un conjunto de datos con grupos no convexos ni el uso de k-medios, ni de un criterio de evaluación que asume la convexidad, es el sonido.

Los siguientes métodos se pueden utilizar para evaluar la calidad de los algoritmos de agrupación basados ​​en criterios internos:

[**Davies-Bouldin índice**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Davies%25E2%2580%2593Bouldin_index&usg=ALkJrhieDZ_EH9r8lfaZq7hUrX9ooJCWHg)

El [índice de Davies-Bouldin](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Davies%25E2%2580%2593Bouldin_index&usg=ALkJrhieDZ_EH9r8lfaZq7hUrX9ooJCWHg) puede calcularse mediante la siguiente fórmula:

 {\ Displaystyle DB = {\ frac {1} {n}} \ sum \_ {i = 1} ^ {n} \ max \_ {j \ neq i} \ Sigma \_ {j}} {d (c\_ {i}, c\_ {j}}}}}}

Donde n es el número de racimos,  {\ Displaystyle c\_ {x}} Es el [centroide](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Centroid&usg=ALkJrhhoTy8KS-TBXQBWnUsZehakbA21kQ) de cluster  {\ Displaystyle x} ,  {\ Displaystyle \ sigma \_ {x}} Es la distancia media de todos los elementos del cluster  {\ Displaystyle x} Al centroide  {\ Displaystyle c\_ {x}} , Y  {\ Displaystyle d (c\_ {i}, c\_ {j})} Es la distancia entre los centroides  {\ Displaystyle c\_ {i}} y  {\ Displaystyle c\_ {j}} . Debido a que los algoritmos que producen clusters con distancias intra-clúster bajas (alta similitud intra-cluster) y altas distancias inter-cluster (baja similitud inter-cluster) tendrán un bajo índice de Davies-Bouldin, el algoritmo de agrupación que produce una colección de clusters con El [índice Davies-Bouldin](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Davies%25E2%2580%2593Bouldin_index&usg=ALkJrhieDZ_EH9r8lfaZq7hUrX9ooJCWHg) más pequeño se considera el mejor algoritmo basado en este criterio.

[**Índice de Dunn**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Dunn_index&usg=ALkJrhh_DlnwMXIjESl2oY7tqD7S6Hcgfg)

El índice Dunn tiene como objetivo identificar grupos densos y bien separados. Se define como la relación entre la distancia mínima inter-clúster a la máxima distancia intra-clúster. Para cada partición de clúster, el índice de Dunn se puede calcular mediante la siguiente fórmula: [[33]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-33)

 {\ Displaystyle D = {\ frac {\ min \_ {1 \ leq i <j \ leq n} d (i, j)} {\ max \_ {1 \ leq k \ leq n} K

Donde *d* ( *i* , *j* ) representa la distancia entre los clusters *iy* *j* , y *d* '( *k* ) mide la distancia intra-clúster del grupo *k* . La distancia inter-clúster *d* ( *i* , *j* ) entre dos grupos puede ser [cualquier](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Centroids&usg=ALkJrhivtHGYkJnGeC3bycfyExnN8yIajw) número de medidas de [distancia](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Centroids&usg=ALkJrhivtHGYkJnGeC3bycfyExnN8yIajw) , tales como la [distancia](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Centroids&usg=ALkJrhivtHGYkJnGeC3bycfyExnN8yIajw) entre los [centroides](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Centroids&usg=ALkJrhivtHGYkJnGeC3bycfyExnN8yIajw) de los [conglomerados](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Centroids&usg=ALkJrhivtHGYkJnGeC3bycfyExnN8yIajw) . De manera similar, la distancia *d* '( *k* ) intra-agrupación puede medirse de diversas maneras, tal como la distancia máxima entre cualquier par de elementos en el grupo *k* . Dado que los criterios internos buscan clusters con alta similitud intra-cluster y baja similitud inter-cluster, los algoritmos que producen clusters con alto índice Dunn son más deseables.

[Coeficiente de silueta](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Silhouette_(clustering)&usg=ALkJrhiufLNcgFoI_KNYs_ALCcOE1DqxEg)

El coeficiente de silueta contrasta la distancia media a elementos en el mismo grupo con la distancia media a elementos en otros grupos. Los objetos con un alto valor de silueta se consideran bien agrupados, los objetos con un valor bajo pueden ser valores atípicos. Este índice funciona bien con el agrupamiento de k-means y también se utiliza para determinar el número óptimo de clusters.

**Evaluación externa**[ [editar](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/w/index.php%3Ftitle%3DCluster_analysis%26action%3Dedit%26section%3D11&usg=ALkJrhj5NBqKVGl_uJOry5aCCO6ip0UZ7w" \o "Editar sección: Evaluación externa) ]

En la evaluación externa, los resultados de la agrupación se evalúan en base a datos que no se utilizaron para la agrupación, como las etiquetas de clase conocidas y los puntos de referencia externos. Dichos puntos de referencia consisten en un conjunto de elementos pre-clasificados, y estos conjuntos suelen ser creados por humanos (expertos). Por lo tanto, los conjuntos de referencia pueden considerarse como un [patrón oro](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Gold_standard_(test)&usg=ALkJrhgHePf0J6qo8jM8Nm6Wi2CysW20Pw) para la evaluación. Estos tipos de métodos de evaluación miden cuán cerca está el clustering de las clases de referencia predeterminadas. Sin embargo, se ha discutido recientemente si esto es adecuado para datos reales, o sólo en conjuntos de datos sintéticos con una verdad real, ya que las clases pueden contener estructura interna, los atributos presentes pueden no permitir la separación de clusters o las clases pueden contener [anomalías](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Anomaly_detection&usg=ALkJrhhQJ0-GweI0Xv6IEQwqj4P8y7Ar4g) . [[34]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-Faerberetal2010-34) Además, desde el punto de vista del descubrimiento del conocimiento, la reproducción del conocimiento conocido puede no ser necesariamente el resultado deseado. [[34]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-Faerberetal2010-34) En el [escenario](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Constrained_clustering&usg=ALkJrhgX9gupA8yPMneBBX91umRKSmHB8g) especial de [agrupación restringida](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Constrained_clustering&usg=ALkJrhgX9gupA8yPMneBBX91umRKSmHB8g) , donde la información meta (como las etiquetas de clase) se utiliza ya en el proceso de agrupación, la retención de información para fines de evaluación no es trivial. [[35]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-pourrajabi-35)

Se han adaptado una serie de medidas a partir de las variantes utilizadas para evaluar las tareas de clasificación. En lugar de contar el número de veces que una clase se asignó correctamente a un único punto de datos ( [conocidos](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/True_positive&usg=ALkJrhhTxDZ8EDZjTP9HMe9ekjoL3PDlKg) como [verdaderos positivos](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/True_positive&usg=ALkJrhhTxDZ8EDZjTP9HMe9ekjoL3PDlKg) ), tales métricas de *conteo de pares* evalúan si cada par de puntos de datos que están verdaderamente en el mismo grupo se predice que estarán en el mismo racimo.

Algunas de las medidas de calidad de un algoritmo de clúster que utilizan criterios externos incluyen:

[**Medida de Rand**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Rand_measure&usg=ALkJrhgo6bDbJmYjQbkiMPXFhKv9ITCEYA) (William M. Rand 1971) [[36]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-36)

El índice Rand calcula la similitud de los clústeres (devueltos por el algoritmo de agrupación) a las clasificaciones de referencia. También se puede ver el índice Rand como una medida del porcentaje de decisiones correctas tomadas por el algoritmo. Se puede calcular utilizando la siguiente fórmula:

 {\ Displaystyle RI = {\ frac {TP + TN} {TP + FP + FN + TN}}}

dónde  {\ Displaystyle TP} Es el número de verdaderos positivos,  {\ Displaystyle TN} Es el número de [negativos verdaderos](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/True_negative&usg=ALkJrhhQCtwrqdBMBUHgP8qEBy9Oj31W_A) ,  {\ Displaystyle FP} Es el número de [falsos positivos](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/False_positives&usg=ALkJrhg9brTiyaZnYz3CLVLbqRxX9enQ_Q) , y  {\ Displaystyle FN} Es el número de [falsos negativos](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/False_negatives&usg=ALkJrhjQW4FK7desffk-2n7ukO8WZDL3HA) . Un problema con el [índice Rand](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Rand_index&usg=ALkJrhjZuXoMdkQx-67zfcMcfRApT56_DA) es que los [falsos positivos](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/False_positive&usg=ALkJrhhsPIoAF1rRmhTebVxZROxC1xm_zg) y los [falsos negativos](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/False_negative&usg=ALkJrhhxU93j95bL4YRxjxhWOm5vqaAVug) están [igualmente ponderados](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/False_negative&usg=ALkJrhhxU93j95bL4YRxjxhWOm5vqaAVug) . Esto puede ser una característica indeseable para algunas aplicaciones de agrupación. La medida F aborda esta preocupación, al igual que el [índice Rand ajustado](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Adjusted_Rand_index&usg=ALkJrhhr-afFCz3OVsptrnj4U0PIdfMqMQ) corregido por la probabilidad.

[**Medida F**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/F-measure&usg=ALkJrhj6NWZ_KyfSUn9QhYCd-Db4h39BAw)

La medida F puede utilizarse para equilibrar la contribución de los [falsos negativos](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/False_negative&usg=ALkJrhhxU93j95bL4YRxjxhWOm5vqaAVug) ponderando la [memoria](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Recall_(information_retrieval)&usg=ALkJrhgSe7z0VRVfoKvhC0KB9xQuboq8Bg) a través de un parámetro  {\ Displaystyle \ beta \ geq 0} . Que la [precisión](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Precision_(information_retrieval)&usg=ALkJrhil9aP6OJ0_V-7fBUeLvcY16Ka1MQ) y el [recuerdo](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Recall_(information_retrieval)&usg=ALkJrhgSe7z0VRVfoKvhC0KB9xQuboq8Bg) se definan como sigue:

 {\ Displaystyle P = {\ frac {TP} {TP + FP}}}

 {\ Displaystyle R = {\ frac {TP} {TP + FN}}}

dónde  {\ Displaystyle P} Es la tasa de [precisión](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Precision_(information_retrieval)&usg=ALkJrhil9aP6OJ0_V-7fBUeLvcY16Ka1MQ) y  {\ Displaystyle R} Es la tasa de [recuperación](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Recall_(information_retrieval)&usg=ALkJrhgSe7z0VRVfoKvhC0KB9xQuboq8Bg) . Podemos calcular la medida F mediante la siguiente fórmula: [[32]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-Christopher_D._Manning.2C_Prabhakar_Raghavan_.26_Hinrich_Schutze-32)

 {\ Displaystyle F \_ {\ beta} = {\ frac {(\ beta ^ {2} +1) \ cdot P \ cdot R} {\ beta ^ {2} \ cdot P + R}}

Observe que cuando  {\ Displaystyle \ beta = 0} ,  {\ Displaystyle F\_ {0} = P} . En otras palabras, [recordar](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Recall_(information_retrieval)&usg=ALkJrhgSe7z0VRVfoKvhC0KB9xQuboq8Bg) no tiene impacto en la medida-F cuando  {\ Displaystyle \ beta = 0} , Y aumentando  {\ Displaystyle \ beta} Asigna una cantidad creciente de peso para recordar en la medida F final.

[**Índice de Jaccard**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Jaccard_coefficient&usg=ALkJrhir3riD2Pd_9DVy-98Xo2uAD-q0mQ)

El índice Jaccard se utiliza para cuantificar la similitud entre dos conjuntos de datos. El [índice de Jaccard](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Jaccard_coefficient&usg=ALkJrhir3riD2Pd_9DVy-98Xo2uAD-q0mQ) toma un valor entre 0 y 1. Un índice de 1 significa que los dos conjuntos de datos son idénticos, y un índice de 0 indica que los conjuntos de datos no tienen elementos comunes. El índice de Jaccard se define por la siguiente fórmula:

 {\ Displaystyle J (A, B) = {\ frac {| A \ cap B |} {| A \ cup B |}} = {\ frac {TP} {TP + FP + FN}}}

Esto es simplemente el número de elementos únicos comunes a ambos conjuntos divididos por el número total de elementos únicos en ambos conjuntos.

[**Índice de Fowlkes-Mallows**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Fowlkes%25E2%2580%2593Mallows_Index&usg=ALkJrhgh4uZRV5S3pkHrPcIw6Qt-QYcrIg) (EB Fowlkes y CL Mallows 1983) [[37]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-37)

El índice Fowlkes-Mallows calcula la similitud entre los clústeres devueltos por el algoritmo de agrupación y las clasificaciones de referencia. Cuanto mayor sea el valor del índice Fowlkes-Mallows, más similares serán los clusters y las clasificaciones de referencia. Se puede calcular utilizando la siguiente fórmula:

 {\ Displaystyle FM = {\ frac {TP} {TP} FP}} \ cdot {\ frac {TP} {TP + FN}}}}}

dónde  {\ Displaystyle TP} Es el número de [verdaderos positivos](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/True_positive&usg=ALkJrhhTxDZ8EDZjTP9HMe9ekjoL3PDlKg) ,  {\ Displaystyle FP} Es el número de [falsos positivos](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/False_positives&usg=ALkJrhg9brTiyaZnYz3CLVLbqRxX9enQ_Q) , y  {\ Displaystyle FN} Es el número de [falsos negativos](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/False_negatives&usg=ALkJrhjQW4FK7desffk-2n7ukO8WZDL3HA) . los  {\ Displaystyle FM} Índice es la media geométrica de la [precisión](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Precision_(information_retrieval)&usg=ALkJrhil9aP6OJ0_V-7fBUeLvcY16Ka1MQ) y [memoria](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Recall_(information_retrieval)&usg=ALkJrhgSe7z0VRVfoKvhC0KB9xQuboq8Bg)  {\ Displaystyle P} y  {\ Displaystyle R} , Mientras que la medida F es su media armónica. [[38]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-38) Por otra parte,[la precisión](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Precision_(information_retrieval)&usg=ALkJrhil9aP6OJ0_V-7fBUeLvcY16Ka1MQ) y[el recuerdo](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Recall_(information_retrieval)&usg=ALkJrhgSe7z0VRVfoKvhC0KB9xQuboq8Bg) están también conocidos como índices de Wallace{\ Displaystyle B ^ {I}}  y {\ Displaystyle B ^ {II}} . [[39]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-39)

La [**información mutua**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Mutual_Information&usg=ALkJrhjcxJOd6yfGSdX_V2h5VP_MPjVRiw) es una [información teórica](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Information_theory&usg=ALkJrhgWKwbhlj1tjSmBmEaaLVHADDvmvA) medida de la cantidad de información que se comparte entre una agrupación y una clasificación realidad sobre el terreno que puede detectar una similitud no lineal entre dos agrupamientos. [Información mutua ajustada](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Adjusted_mutual_information&usg=ALkJrhh6aqsnbQVqVBHZ4m4TWWyrb2Wilg) es la variante corregida para la posibilidad de que esto que tiene un sesgo reducida por un número variable de racimo.

[**Matriz de confusión**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Confusion_matrix&usg=ALkJrhjJ5SNH57cBQO2U6kFWjcZv8n8A6Q)

Una matriz de confusión se puede utilizar para visualizar rápidamente los resultados de un algoritmo de clasificación (o agrupamiento). Se muestra cómo los diferentes clúster es un grupo de la regla de oro.

Aplicaciones [ [editar](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/w/index.php%3Ftitle%3DCluster_analysis%26action%3Dedit%26section%3D12&usg=ALkJrhgp8pnwVv3N66IP1JTF2j2VM28KwA" \o "Editar sección: Aplicaciones) ]

|  |  |
| --- | --- |
| [https://upload.wikimedia.org/wikipedia/en/thumb/9/99/Question_book-new.svg/50px-Question_book-new.svg.png](https://en.wikipedia.org/wiki/File:Question_book-new.svg) | Esta sección **necesita citas adicionales para la**[**verificación**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Wikipedia:Verifiability&usg=ALkJrhgPY6Kny0wctVUlQy0BcIT2RhgZAQ) . Por favor ayude a [mejorar este artículo](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/w/index.php%3Ftitle%3DCluster_analysis%26action%3Dedit&usg=ALkJrhiNtR0ec-mbVfvfZ3UM103fZ_mZYQ) [agregando citas a fuentes confiables](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Help:Introduction_to_referencing_with_Wiki_Markup/1&usg=ALkJrhjSTpzabezUQ6pM-io26VO19mo0hw) . El material sin suministro puede ser desafiado y eliminado. *(Noviembre de 2016)* *(*[*Aprenda cómo y cuándo quitar este mensaje de plantilla*](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Help:Maintenance_template_removal&usg=ALkJrhgyLcHEZ4mWw84KkqIKwg9T10h3gA)*)* |

[**Biología**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Biology&usg=ALkJrhi6IttldtjzB3WEQhGXQZbZbdSumg)**, biología computacional y**[**bioinformática**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Bioinformatics&usg=ALkJrhjsGDYOyj92RtpwhLgggBiR7PVEPg)

[**Planta**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Plant&usg=ALkJrhjvc5qEccSWMHvs8TIJbVPRxDJl4g)**y**[**animal de**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Animal&usg=ALkJrhhdEsXjL4hK6HlxWzgZTQVUR9SZ5A)[**la ecología**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Ecology&usg=ALkJrhhyAgHt_3555eKBn9QZDRzBZA11-Q)

análisis de conglomerados se utiliza para describir y hacer comparaciones espaciales y temporales de las comunidades (ensamblajes) de organismos en ambientes heterogéneos; también se utiliza en la [sistemática de plantas](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Systematics&usg=ALkJrhgH7ED3PvdTHLvFY1YU-QGINqbm3g) para generar artificiales [filogenias](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Phylogeny&usg=ALkJrhh_rOIN26H3zgOIZHTP6XvIw8x0BQ) o grupos de organismos (personas) en la especie, género o nivel más alto que comparten una serie de atributos

[**transcriptómica**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Transcriptome&usg=ALkJrhid_w3VLGNOcFZCezJaVI0tFvVdOw)

la agrupación se utiliza para construir grupos de [genes](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Genes&usg=ALkJrhj17-n5vn-zayvikeerfuFi_KDvrw) con patrones de expresión relacionados (también conocido como coexpressed genes) como en [algoritmo de agrupamiento HCS](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/HCS_clustering_algorithm&usg=ALkJrhhKGthz-OmNB-bfDdPzqs31IlBEFw) . A menudo, estos grupos contienen proteínas funcionalmente relacionadas, tales como [enzimas](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Enzyme&usg=ALkJrhii47iWuyHrI9p8RKvyu-sjSIYaKA) para una determinada [vía](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Metabolic_pathway&usg=ALkJrhhUSBJMW1MCRLqnUHp3K2alPRvarg) , o genes que son co-regulados. Experimentos de alto rendimiento utilizando [las etiquetas de secuencias expresadas](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Expressed_sequence_tag&usg=ALkJrhjaOkiqMzCmvwMcCw3Ypr7cF0Pmhw) (EST) o [micromatrices de ADN](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/DNA_microarray&usg=ALkJrhgnKfNox4BWjrq-vWpLO_sfO4N-jQ) pueden ser una poderosa herramienta para [la anotación del genoma](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Genome_annotation&usg=ALkJrhhrRihJqMsyvT4Frm91tcoQR5msYA) , un aspecto general de [la genómica](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Genomics&usg=ALkJrhgjOl84Agt2ZirMztLU_7zpoj6dHA) .

[**El análisis de secuencia**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Sequence_analysis&usg=ALkJrhgckGUE5k1btzL7ROGRoY_52dh9_Q)

la agrupación se utiliza para secuencias homólogas de grupo en [familias de genes](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/List_of_gene_families&usg=ALkJrhhUk9PKj05GRQaNbWQvvHopEADpYw) . Este es un concepto muy importante en la bioinformática y [la biología evolutiva](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Evolutionary_biology&usg=ALkJrhgW-pBQ2cLLnSC2bHmEJxlavG59Mg) en general. Ver la evolución por [duplicación de genes](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Gene_duplication&usg=ALkJrhiH9UME5xkAC-aBqhvdDOImedWR5g) .

**De alto rendimiento**[**de genotipos**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Genotype&usg=ALkJrhi1S-kTciSFWh1qShveylBEAyWMwQ)**plataformas**

algoritmos de agrupamiento se utilizan para asignar automáticamente los genotipos.

[**agrupación genética humana**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Human_genetic_clustering&usg=ALkJrhihvS6Fpy2TRsTi4embo0aEZgVdCA)

La similitud de los datos genéticos se utiliza en la agrupación para inferir las estructuras de población.

[**Medicina**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Medicine&usg=ALkJrhhVVxiu1k2FpfaXkvTULg8s7FYvTw)

[**Imagenes medicas**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Medical_imaging&usg=ALkJrhg4RQG0X3P3qXzA80PTJTXUZMRU3A)

En [la PET](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/PET_scan&usg=ALkJrhgQQ3fodU_ZM4tyvO9NGdlVHREX9g) , el análisis de conglomerados se puede utilizar para diferenciar entre diferentes tipos de [tejido](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Tissue_(biology)&usg=ALkJrhhkcUhWRNJM4Fp52_pbOQCd9WBwzA) en una imagen tridimensional para muchos propósitos diferentes. [[40]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-40)

**Negocios y**[**comercialización**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Marketing&usg=ALkJrhi2I0kIu0_TRdxcjfxlWRmedEpPCw)

[**Investigación de mercado**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Market_research&usg=ALkJrhgZsI6DfKnVR-tKevypy-Sczk5pKg)

El análisis de conglomerados es ampliamente utilizado en la investigación de mercado cuando se trabaja con datos multivariantes a partir de [encuestas](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Statistical_survey&usg=ALkJrhj-me_qzs2HejBgKn3q7xWka5lfoQ) y paneles de prueba. Los investigadores de mercado utilizan análisis de conglomerados para particionar el general de [la población](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Population&usg=ALkJrhjcZ45cPBTTsGzsoD6Yt7n810AprQ) de [consumidores](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Consumer&usg=ALkJrhiEtO3wPXb6iO4oWaUK87-IBtcCog) en segmentos de mercado y para comprender mejor las relaciones entre los diferentes grupos de consumidores / potenciales [clientes](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Customers&usg=ALkJrhj29JAH2G21Mv1SaiB_6VMqyY6Hog) , y para su uso en [la segmentación del mercado](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Market_segmentation&usg=ALkJrhgYyTSNZ3E-X7S8rFYrCKRJ_7yBHg) , [posicionamiento del producto](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Positioning_(marketing)&usg=ALkJrhhk_bE3a3iUt0sj57mX5ZmCp6BlzQ) , [desarrollo de nuevos productos](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/New_product_development&usg=ALkJrhhutYWZQWJkqjObXAH2PNYghTFncQ) y selección de mercados de prueba.

**Agrupación de artículos de compra**

La agrupación se puede utilizar para agrupar todos los artículos de compra disponible en la web en un conjunto de productos únicos. Por ejemplo, todos los artículos en eBay se pueden agrupar en productos únicos. (eBay no tiene el concepto de una [SKU](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Stock-keeping_unit&usg=ALkJrhhn9YYf8GuN1Gwg76EczJ78yHE_9g) )

[**Red mundial**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/World_wide_web&usg=ALkJrhioC8NI6NwZS2GUqAajuHhNhNMnCA)

**Análisis de redes sociales**

En el estudio de [las redes sociales](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Social_network&usg=ALkJrhj471VqIftdQzLJ0HLwEezS9SE5Dw) , la agrupación puede ser utilizado para reconocer [las comunidades](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Communities&usg=ALkJrhiWndaNIt9UlcjX2j4eiSTniGBIHg) dentro de grandes grupos de personas.

**Resultado de la búsqueda agrupación**

En el proceso de la agrupación inteligente de los archivos y sitios web, la agrupación puede ser usado para crear un conjunto más relevante de los resultados de búsqueda en comparación con los motores de búsqueda normales como [Google](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Google&usg=ALkJrhikq3ZC6VVTD0gR6I1jtbasVOHwtg) . En este momento hay una serie de herramientas basadas en la web de agrupamiento como [Clusty](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Clusty&usg=ALkJrhj5qYMmfhQUT3x8E8EuGDOGGsd0Lg) .

**optimización mapa resbaladizo**

[Flickr](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Flickr&usg=ALkJrhhNMYVnOG9chamdOoRX4vq1g6eP6w) mapa de fotos y otros sitios de mapas 's utilizar agrupación para reducir el número de marcadores en un mapa. Esto hace que sea más rápido y reduce la cantidad de desorden visual.

[**Ciencias de la Computación**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Computer_science&usg=ALkJrhgOOlzvBxPQMJF_jf-bZ9pd4ydptg)

[**La evolución del software**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Software_evolution&usg=ALkJrhgIfyG9PpKzC_7RL0AhV8yH_Roi0Q)

La agrupación es útil en la evolución del software, ya que ayuda a reducir las propiedades heredadas en código mediante la reforma de la funcionalidad que se ha convertido dispersa. Es una forma de reestructuración y, por tanto, es una forma de mantenimiento preventivo directa.

[**Segmentación de imagen**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Image_segmentation&usg=ALkJrhj7SOTThMC0uP-2pXQDbQUvoozvkg)

La agrupación se puede utilizar para dividir una [digital de](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Digital_data&usg=ALkJrhjwYowz6XmtY1IHCx9Ude4pkZm4Aw)[la imagen](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Image&usg=ALkJrhiz4bSWa8o2iGVWjyhOxKeckiwRpw) en regiones distintas de [detección de bordes](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Border_detection&usg=ALkJrhh5H0bYVg2wvjAjEE-qQW-sZSp4LA) o [de reconocimiento de objetos](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Object_recognition&usg=ALkJrhi-h2WMA1OW1AeVDilILnbOdHucjA) . [[41]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-panSearch-41)

[**Los algoritmos evolutivos**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Evolutionary_algorithms&usg=ALkJrhjAxKR1YfTks2noyq4K05TnU-nsug)

La agrupación puede ser usado para identificar diferentes nichos dentro de la población de un algoritmo evolutivo para que oportunidad reproductiva se puede distribuir de manera más uniforme entre las especies o subespecies en evolución.

[**Sistemas de recomendación**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Recommender_systems&usg=ALkJrhjfz9EgFRRBdMncx5fLCemNwXODew)

Los sistemas de recomendación están diseñados para recomendar nuevos elementos en función de los gustos del usuario. A veces utilizan algoritmos de agrupamiento para predecir las preferencias del usuario en función de las preferencias de otros usuarios en el conglomerado del usuario.

[**la cadena de Markov métodos de Monte Carlo**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Markov_chain_Monte_Carlo&usg=ALkJrhiqCjSDgASmR42ln-B8pkn1dISn3Q)

La agrupación se utiliza a menudo para localizar y caracterizar los extremos en la distribución de destino.

[**Detección de anomalías**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Anomaly_detection&usg=ALkJrhhQJ0-GweI0Xv6IEQwqj4P8y7Ar4g)

Anomalías / valores atípicos son típicamente - ya sea explícita o implícitamente - definido con respecto a la estructura de la agrupación de los datos.

**Ciencias Sociales**

**análisis de la delincuencia**

El análisis de conglomerados se puede utilizar para identificar las áreas en las que hay una mayor incidencia de determinados tipos de delitos. Al identificar estas áreas distintas o "puntos calientes", donde un crimen similar ha ocurrido en un período de tiempo, es posible gestionar los recursos policiales con mayor eficacia.

[**minería de datos para la Educación**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Educational_data_mining&usg=ALkJrhiJHo8d7CHdZCM371hAzxI3h8rVfA)

El análisis de conglomerados es por ejemplo utilizado para identificar grupos de escuelas o estudiantes con propiedades similares.

**tipologías**

A partir de los datos de la encuesta, proyectos como las llevadas a cabo por el análisis de uso del clúster Centro de Investigación Pew para discernir tipologías de opiniones, hábitos y demográficos que pueden ser útiles en la política y la comercialización.

**Otros**

**robótica de campo**

agrupación de los algoritmos se utilizan para el conocimiento de la situación robótico para realizar un seguimiento de objetos y detectar anomalías en los datos del sensor. [[42]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-42)

[**química matemática**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Mathematical_chemistry&usg=ALkJrhjnHODIEYjqQqhS_5yn7ON-39JtKQ)

Para encontrar la semejanza estructural, etc., por ejemplo, 3000 compuestos químicos se agruparon en el espacio de 90 [índices topológicos](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Topological_index&usg=ALkJrhhlE4QBG_1YMy2r-RDJvRyDRfEQoA) . [[43]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-43)

[**Climatología**](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Climatology&usg=ALkJrhh_3xhv41CTK9m7CcLeFWXHqZIQ_A)

Para encontrar los regímenes meteorológicos o patrones atmosféricos de presión del nivel del mar preferido. [[44]](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster_analysis&usg=ALkJrhi7-6wS68hdwOXh71CuuIjKaUcHgg#cite_note-44)

**geología del petróleo**

El análisis de conglomerados se utiliza para reconstruir los datos que faltan núcleo del fondo del pozo o desaparecidos curvas de registro con el fin de evaluar las propiedades del yacimiento.

**Geografía física**

El agrupamiento de las propiedades químicas de diferentes lugares de las muestras.

Ver también [ [editar](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/w/index.php%3Ftitle%3DCluster_analysis%26action%3Dedit%26section%3D13&usg=ALkJrhi9fkW5u_s3GFN0DXyGmSVkX854VA" \o "Editar sección: Ver también) ]

|  |  |
| --- | --- |
| [https://upload.wikimedia.org/wikipedia/en/thumb/4/4a/Commons-logo.svg/30px-Commons-logo.svg.png](https://en.wikipedia.org/wiki/File:Commons-logo.svg) | Wikimedia Commons alberga contenido multimedia sobre [***análisis de conglomerados***](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://commons.wikimedia.org/wiki/Category:Cluster_analysis&usg=ALkJrhj8Md_nQeMIf7QL6Y0wH_Ls5ybs_A) . |

**Tipos especializados de análisis de agrupamiento**[ [editar](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/w/index.php%3Ftitle%3DCluster_analysis%26action%3Dedit%26section%3D14&usg=ALkJrhipR2Ql1gkFXpZLEYZlfPHchPOmyQ" \o "Editar sección: tipos especializados de análisis de conglomerados)]

[agrupación equilibrada](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Balanced_clustering&usg=ALkJrhiPKXX-N6ary2oItaOw7Hjeaj0GEw)

[La agrupación de datos de alta dimensión](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Clustering_high-dimensional_data&usg=ALkJrhi1KfryEhiBeqKxVOANr41oKW7Mqg)

[agrupamiento conceptual](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Conceptual_clustering&usg=ALkJrhhG5OlOcpzX9R-T1-bcLubR4sQFcQ)

[la agrupación consenso](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Consensus_clustering&usg=ALkJrhj0Ujwvg_fBNok3JCqTZunbaFgQuA)

[agrupación restringida](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Constrained_clustering&usg=ALkJrhgX9gupA8yPMneBBX91umRKSmHB8g)

[agrupación flujo de datos](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Data_stream_clustering&usg=ALkJrhgxPfmqSsHx066oB0_hfM89fXNB_A)

[agrupación HCS](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/HCS_clustering_algorithm&usg=ALkJrhhKGthz-OmNB-bfDdPzqs31IlBEFw)

[secuencia de la agrupación](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Sequence_clustering&usg=ALkJrhjxPR1MujLw9UZqSMlfCFhIMyK6Jg)

[La agrupación espectral](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Spectral_clustering&usg=ALkJrhjeTjkY9ZjUqLW6AU4HtAxVSqhf9w)

**Las técnicas utilizadas en el análisis de agrupamiento**[ [editar](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/w/index.php%3Ftitle%3DCluster_analysis%26action%3Dedit%26section%3D15&usg=ALkJrhjuUiNPINHJclV-x9wPSV2Z86kxHw" \o "Editar sección: Las técnicas utilizadas en el análisis de conglomerados)]

[Redes neuronales artificiales](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Artificial_neural_network&usg=ALkJrhjrru950dvqm72cdwXCx2H2IS_vLg) (ANN)

[La búsqueda del vecino más cercano](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Nearest_neighbor_search&usg=ALkJrhiczWIckqaSdrlyJ_G2WSx9TTEmOA)

[El análisis de componentes Barrio](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Neighbourhood_components_analysis&usg=ALkJrhg0qNhzpYouLhyWJvd_TUEXss1p7w)

[El análisis de clases latentes](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Latent_class_model&usg=ALkJrhjNaTrWHw5IbHVFZ74xkbZQrtPO1g)

[la propagación de afinidad](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Affinity_propagation&usg=ALkJrhjRs8WVj877S1mtBBkEyE164-ejFQ)

**Proyección de datos y procesamiento previo**[ [editar](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/w/index.php%3Ftitle%3DCluster_analysis%26action%3Dedit%26section%3D16&usg=ALkJrhiIDVaZKJ_03ARMTaDZSNKJa-bg0Q" \o "Editar sección: proyección de datos y preprocesamiento)]

[reducción de la dimensión](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Dimension_reduction&usg=ALkJrhjUL9Pd_sw1KAluOGF6phpg0-qPvA)

[Análisis de componentes principales](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Principal_component_analysis&usg=ALkJrhiHNcosX0QB8uyyLew1EyMoaP77DQ)

[Escalamiento multidimensional](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Multidimensional_scaling&usg=ALkJrhhwUH95P1T1FdCDHvf41eMmiTnUNA)

**Otros**[ [editar](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/w/index.php%3Ftitle%3DCluster_analysis%26action%3Dedit%26section%3D17&usg=ALkJrhik1n_MHt0kmjj5IPCOOHIVOJaeVA" \o "Editar sección: Otros) ]

[modelado](https://translate.googleusercontent.com/translate_c?depth=1&hl=es&prev=search&rurl=translate.google.com.mx&sl=en&sp=nmt4&u=https://en.wikipedia.org/wiki/Cluster-weighted_modeling&usg=ALkJrhhHzLXQ7RCPNiznFZByJNjfR_5FhQ)