**Nearest neighbor search** (**NNS**), also known as **proximity search**, [**similarity search**](https://en.wikipedia.org/wiki/Similarity_search) or [**closest point search**](https://en.wikipedia.org/wiki/Closest_pair_of_points_problem), is an [optimization problem](https://en.wikipedia.org/wiki/Optimization_problem) for finding closest (or most similar) points. Closeness is typically expressed in terms of a dissimilarity function: the less similar the objects, the larger the function values.

[*k*-nearest neighbor search](https://en.wikipedia.org/wiki/K-nearest_neighbor_algorithm) identifies the top *k* nearest neighbors to the query. This technique is commonly used in predictive analytics to estimate or classify a point based on the consensus of its neighbors. *k*-nearest neighbor graphs are graphs in which every point is connected to its *k* nearest neighbors.

The nearest neighbor search problem arises in numerous fields of application, including:

* [Pattern recognition](https://en.wikipedia.org/wiki/Pattern_recognition) - in particular for [optical character recognition](https://en.wikipedia.org/wiki/Optical_character_recognition)
* [Statistical classification](https://en.wikipedia.org/wiki/Statistical_classification)- see [k-nearest neighbor algorithm](https://en.wikipedia.org/wiki/K-nearest_neighbor_algorithm)
* [Computer vision](https://en.wikipedia.org/wiki/Computer_vision)
* [Computational geometry](https://en.wikipedia.org/wiki/Computational_geometry) - see [Closest pair of points problem](https://en.wikipedia.org/wiki/Closest_pair_of_points_problem)
* [Databases](https://en.wikipedia.org/wiki/Database) - e.g. [content-based image retrieval](https://en.wikipedia.org/wiki/Content-based_image_retrieval)
* [Coding theory](https://en.wikipedia.org/wiki/Coding_theory) - see [maximum likelihood decoding](https://en.wikipedia.org/wiki/Decoding_methods)
* [Data compression](https://en.wikipedia.org/wiki/Data_compression) - see [MPEG-2](https://en.wikipedia.org/wiki/MPEG-2) standard
* [Robotic](https://en.wikipedia.org/wiki/Robotic) sensing
* [Recommendation systems](https://en.wikipedia.org/wiki/Recommender_system), e.g. see [Collaborative filtering](https://en.wikipedia.org/wiki/Collaborative_filtering)
* [Internet marketing](https://en.wikipedia.org/wiki/Internet_marketing) - see [contextual advertising](https://en.wikipedia.org/wiki/Contextual_advertising) and [behavioral targeting](https://en.wikipedia.org/wiki/Behavioral_targeting)
* [DNA sequencing](https://en.wikipedia.org/wiki/DNA_sequencing)
* [Spell checking](https://en.wikipedia.org/wiki/Spell_checking) - suggesting correct spelling
* [Plagiarism detection](https://en.wikipedia.org/wiki/Plagiarism_detection)
* [Contact searching algorithms in FEA](https://en.wikipedia.org/w/index.php?title=Contact_searching_algorithms_in_FEA&action=edit&redlink=1)
* [Similarity scores](https://en.wikipedia.org/wiki/Similarity_score) for predicting career paths of professional athletes.
* [Cluster analysis](https://en.wikipedia.org/wiki/Cluster_analysis) - assignment of a set of observations into subsets (called clusters) so that observations in the same cluster are similar in some sense, usually based on [Euclidean distance](https://en.wikipedia.org/wiki/Euclidean_distance)
* [Chemical similarity](https://en.wikipedia.org/wiki/Chemical_similarity)
* [Sampling-Based Motion Planning](https://en.wikipedia.org/wiki/Motion_planning#Sampling-Based_Algorithms)

The traveling salesman problem will create ‘clusters’ with cities that are not too distant from each other. At each stage, we visit an unvisited city nearest to the current city.

['TYC', 'JCI'] – Johnson Controls (Consumer Discretionary) and Tyro International (Industrials)

['GOOG', 'GOOGL'] – Alphabet Inc Class A and C (Information technology)

['FOX', 'FOXA'] – Twenty-First Century Fox Class A and B (Consumer Discretionary)

['NWS', 'NWSA'] – News Corp. Class A and B (Consumer Discretionary)

['DISCK', 'DISCA'] – Discovery Communications A and C (Consumer Discretionary)

K = 5

{'JCI', 'TYC'}

{'GOOG', 'GOOGL'}

{'FOX', 'FOXA'}

{'NWS', 'NWSA'}

{'DISCA', 'DISCK'}

K = 10

{'JCI', 'TYC'}

{'GOOG', 'GOOGL'}

{'FOX', 'FOXA'}

{'NWS', 'NWSA'}

{'DISCA', 'DISCK'}

{'BBT', 'HBAN', 'PNC', 'STI'}

{'CMS', 'DTE'}

{'LNC', 'MET'}

K = 30

{'JCI', 'TYC'} - Tyco International (Industrials), Johnson Controls (Consumer Discretionary)

{'GOOG', 'GOOGL'}

{'FOX', 'FOXA'}

{'NWS', 'NWSA'}

{'DISCA', 'DISCK'}

{'BAC', 'C', 'JPM'}

{'LNC', 'MET', 'PRU'}

{'CMS', 'DTE', 'XEL'}

{'BHI', 'HAL'}

{'AVB', 'EQR', 'ESS', 'UDR'}

{'BBT', 'HBAN', 'KEY', 'PNC', 'STI', 'USB', 'WFC'}

{'CMA', 'ZION'}

K = 50

{'JCI', 'TYC'} - Tyco International (Industrials), Johnson Controls (Consumer Discretionary)

{'GOOG', 'GOOGL'}

{'FOX', 'FOXA'}

{'NWS', 'NWSA'}

{'DISCA', 'DISCK'}

{'LNC', 'MET', 'PRU'}

{'BHI', 'HAL'}

{'AVB', 'EQR', 'ESS', 'UDR'}

{'CMS', 'DTE', 'ES', 'PNW', 'WEC', 'XEL'}

{'BAC', 'BBT', 'C', 'CMA’, ‘HBAN', 'JPM', 'KEY', 'PNC','RF', 'STI', 'USB', 'WFC', 'ZION'}

{'AEE', 'LNT'}

{'GS', 'MS'}

{'SLG', 'VNO'}

K = 100

[{'JCI', 'TYC'} - Tyco International (Industrials), Johnson Controls (Consumer Discretionary)

{'GOOG', 'GOOGL'}

{'FOX', 'FOXA'}

{'NWS', 'NWSA',

{'DISCA', 'DISCK'}

{'BHI', 'HAL'}, - Baker Hughes (Energy), Halliburton (Energy)

{'AVB', 'EQR', 'ESS', 'UDR'} - AvalonBay Communities, Equity Residential, Essex Property Trust, United Dominion Realty Trust

{'BXP', 'FRT', 'KIM', 'SLG', 'VNO'} - Boston Properties, Federal Realty Investment Trust, Kimberly Clark, SL Green Realty, Vornado Realty Trust {AIV, AVB, EQR, ESS, GGP, SPG, UDR} – Apartment Investment & Mgmt, AvalonBay Communities, Equity Residential, Essex Property Trust, General Growth Properties Inc, Simon Property Group Inc, United Dominion Realty Trust, Inc

{'AEE','AEP','CMS', 'D’, ‘DTE’, ‘DUK’, ‘ED’, ‘ES’, ‘LNT', 'PNW', 'SCG', 'SO', 'WEC', 'XEL'} – Ameren Corp, American Electric Power, CMS Energy, Dominion Resources, DTE Energy Co., Duke Energy, Consolidated Edison, Eversource Energy, Alliant Energy Corp, Pinnacle West Capital, SCANA Corp, Southern Co., Wisconsin Energy Corporation, Xcel Energy Inc {EIX, PCG} – Edison Intl, PG&E Corp

{'BK', 'NTRS', 'STT'} - The Bank of New York Mellon Corp (banking and financial services corporation), Northern Trust Corp (wealth and asset management), State Street Corp (investment management)

{'LNC', 'MET', 'PRU', 'UNM'} – Lincoln National (insurance and investment management), MetLife Inc, Prudential Financial, Unum Group

{'BAC’, ‘BBT’, ‘C’, ‘CMA’, ‘FITB’, ‘GS’, ‘HBAN’, ‘JPM', 'KEY', 'MS', 'PNC', 'RF', 'STI', 'USB', 'WFC', 'ZION'} - Bank of America Corp, BB&T Corporation (Financial Services holding company), Citigroup Inc, Comerica Inc (financial services company), Fifth Third Bancorp, Goldman Sachs Group, Huntington Bancshares (American bank holding company), JPMorgan Chase & Co., KeyCorp (American regional bank), Morgan Stanley, PNC Financial Services, Regions Financial Corp., SunTrust Banks, U.S. Bancorp, Wells Fargo (American international banking and financial services holding company), Zions Bancorp {BK, LNC, MET, MTB, NTRS, PBCT, PRU, STT, UNM} – The Bank of New York Mellon Corp., Lincoln National, MetLife Inc, M&T Bank Corp, Northern Trust Corp, People’s United Financial, Prudential Financial, State Street Corp., Unum Group

{'DHI', 'LEN'} - D.R. Horton (home construction company), Lennar Corp (homebuilder)

K = 200

{'JCI', 'TYC'}

{'GOOG', 'GOOGL'}

{'FOX', 'FOXA'}

{'NWS', 'NWSA'}

{'DISCA', 'DISCK'}

{'BHI', 'HAL'}

{'BAC', 'BBT','BK', 'C', 'CMA', 'FITB', 'GS', 'HBAN', 'JPM', 'KEY', 'LNC' 'MET', 'MS' 'MTB', 'NTRS', 'PBCT', 'PNC', 'PRU', 'RF', 'STI', 'STT', 'UNM', 'USB', 'WFC', 'ZION'}

{'AON', 'MMC'} - Aon plc, Marsh & McLennan

{'CB', 'TRV'} - Chubb Limited (publicly traded property and casualty insurer) , The Travelers Companies Inc ( second largest writer of U.S. commercial property casualty insurance)

{'AIV', 'AVB', 'BXP', 'EQR', 'ESS','FRT','GGP', 'KIM', 'SLG', 'SPG', 'UDR', 'VNO'}

{'DHI', 'LEN', 'PHM'} - D.R. Horton (home construction company), Lennar Corp (homebuilder), Pulte Homes Inc.

{'HCN', 'HCP'} - Welltower Inc (real estate investment trust), HCP Inc (Real estate investment trust)

{'COP', 'CVX', 'XOM'} - ConocoPhillips (Energy), Chevron Corp, Exxon Mobil Corp

{'BEN', 'TROW'} - Franklin Resources (American holding company/global investment firm), T. Rowe Price Group (American publicly owned investment firm)

{'AEE', 'AEP', 'CMS', 'D', 'DTE', 'DUK', 'ED', 'EIX', 'ES', 'LNT', 'PCG', 'PNW', 'SCG', 'SO', 'WEC', 'XEL'}

With an increase in k, we see more elements being added to existing clusters and more new clusters being formed. After a certain value of k, we also see clusters becoming more ‘broad’ – for example, what was initially two clusters – insurance and banking – becomes a general ‘finance’ one. Firms within a certain sub-industry, like insurance, are highly likely to have strong correlations with others engaging in similar activities, and slightly lower correlations with firms in a different sub-industry.