## 582448   TIEDON LOUHINNAN MENETELMÄTKurssin keskeisten käsitteiden englanti-suomi-sanasto

* [Aineiston kuvailu](https://www.cs.helsinki.fi/u/ronkaine/tilome/sanasto.html#kuvailu)
* [Assosiaatiosäännöt](https://www.cs.helsinki.fi/u/ronkaine/tilome/sanasto.html#assosiaatiot)
* [Klusterointi](https://www.cs.helsinki.fi/u/ronkaine/tilome/sanasto.html#klusterointi)
* [Luokittelu](https://www.cs.helsinki.fi/u/ronkaine/tilome/sanasto.html#luokittelu)
* [Tiedon louhinta ja sen prosessi](https://www.cs.helsinki.fi/u/ronkaine/tilome/sanasto.html#prosessi)

#### Aineiston kuvailu

* analytical characterization = analyyttinen luonnehdinta
* analytical comparison, analytical discrimination = analyyttinen vertailu
* attribute-oriented induction = attribuuttiyleistysmenetelmä
* attribute relevance analysis = attribuuttien merkittävyysanalyysi, relevanssianalyysi
* data aggregation = datan koostaminen
* data characterization = datan luonnehdinta, aineiston luonnehdinta
* data comparison, data discrimination = data vertailu, aineiston vertailu
* data cube = tietokuutio
* data generalization = datan yleistys
* data focusing = datan kohdentaminen
* concept description = käsitteen tai käsitteiden kuvailu, aineiston kuvailu
* contingency table = kontingenssitaulu
* cross-tabulation = ristiintaulukointi
* data comparison = datan vertailu
* data discrimination = datan vertailu, datan erottelu
* entropy = entropia
* gain ratio = hyötysuhde
* generalized relation = yleistetty relaatio
* information = informaatio
* information gain = informaatiohyöty
* quantitative characteristic rule = määrällinen luonnehtiva sääntö, määrällinen ominaissääntö
* quantitative description rule = määrällinen kuvaileva sääntö
* quantitative discriminant rule = määrällinen erotteleva sääntö

#### Assosiaatiosäännöt

* antecedent of a rule = säännön edeltäjäosa, säännön vasen puoli
* Apriori algorithm = Apriori-algoritmi
* Apriori principle = Apriori-periaate
* association rule = assosiaatiosääntö
* body of a rule = säännön vartalo, säännön vasen puoli
* Boolean association rule = assosiaatiosääntö, totuusarvoinen assosiaatiosääntö
* candidate itemset = kandidaattijoukko
* confidence = luottamus
* consequent of a rule = säännön seuraajaosa, säännön oikea puoli
* constraint-based association rule = rajoitepohjainen assosiaatiosääntö
* conviction = tuomitsevuus
* frequency = tuki, kattavuus
* frequent = kattava
* frequent itemset = kattava joukko
* frequent set = kattava joukko
* head of a rule = säännön pää, säännön oikea puoli
* item = tietoalkio, tapahtuma
* itemset = tietoalkiojoukko, tapahtumajoukko
* k-itemset = k:n kokoinen alkiojoukko, k:n alkion joukko
* left-hand side of a rule = säännön vasen puoli
* level-wise search = tasoittainen haku
* leverage = nostovoima
* lift = noste
* minimum support = minimituki
* minimum confidence = minimiluottamus
* multi-dimensional association rule = moniulotteinen assosiaatiosääntö
* multi-level association rule = monitasoinen assosiaatiosääntö
* right-hand side of a rule = säännön oikea puoli
* single-dimensional association rule = yksiulotteinen assosiaatiosääntö
* single-level association rule = yksitasoinen assosiaatiosääntö
* strength = luottamus, vahvuus
* support = tuki
* quantitative association rule = määrällinen assosiaatiosääntö

#### Klusterointi

* agglomerative hierarchical clustering = kokoava hierarkkinen klusterointi
* average linkage method = keskiarvosidosryvästys, naapurikeskiarvomenetelmä
* cluster = ryhmä, ryväs, klusteri
* cluster analysis = klusterianalyysi
* clustering = klusterointi, ryvästys, ryhmittely
* clustering tree = klusterointipuu
* complete linkage clustering = täydellinen sidosklusterointimenetelmä, kauimmaisen naapurin menetelmä
* conceptual clustering = käsitteellinen klusterointi
* dendrogram = dendrogrammi, klusterointipuu
* density-based clustering = tiheyspohjainen klusterointi
* dissimilarity = erilaisuus
* distance matrix = etäisyysmatriisi
* distance measure = etäisyysmitta
* divisive hierarchical clustering = hajottava hierarkkinen klusterointi
* Euclidean distance = euklidinen etäisyys
* furthest neighbor method = kauimman naapurin menetelmä
* grid-based clustering = ristikkopohjainen klusterointi
* hierarchical clustering = hierarkkinen klusterointi
* inter-class similarity = klusterien välinen samankaltaisuus
* intra-class similarity = klusterin sisäinen samankaltaisuus
* inter-cluster distance = klustereiden välinen etäisyys
* Jaccard coefficient = Jaccardin kerroin
* k-means clustering method = k:n keskiarvon klusterointimenetelmä
* k-medoids clustering method = k:n medoidin klusterointimenetelmä
* Manhattan distance = Manhattan-etäisyys
* Minkowski distance = Minkowskin etäisyys
* model-based clustering = mallipohjainen klusterointi
* nearest-neighbor clustering = lähimmän naapurin klusterointimenetelmä
* partitioning clustering = osittava klusterointi
* similarity = samankaltaisuus
* similarity measure = samankaltaisuusmitta
* simple matching coefficient = yksinkertainen yhteensopivuuskerroin
* single linkage clustering = yksinkertaisen kytkennän klusterointimenetelmä, lähimmän naapurin klusterointimenetelmä

#### Luokittelu

* Bayes network = Bayes-verkko
* Bayes theorem = Bayesin teoreema
* Bayesian = bayesläinen, bayeslainen, bayesiläinen, bayesilainen
* Bayesian classification = bayesläinen luokittelu
* class = luokka
* class label = luokkatunnus
* classification = luokittelu
* classification accuracy = luokittelutarkkuus
* classification attribute = luokitteluattribuutti
* classification process = luokitteluprosessi
* classifier = luokittelija
* condition for stopping = pysähtymisehto
* cross-validation = ristiinvalidointi
* decision tree = päätöspuu
* decision tree classifier = päätöspuuluokittelija
* decision tree induction = päätöspuupäättely
* independence hypothesis = riippumattomuusoletus
* labeling rule = nimeämissääntö
* model construction = mallin rakentaminen
* model usage = mallin käyttö
* naive Bayes classifier = naiivi Bayes-luokittelija
* overfitting = ylisovittuminen
* posterior probability = posteriori todennäköisyys
* prediction = ennustaminen
* pruning = karsiminen, karsinta
* training set = opetusjoukko
* test set = testausjoukko

#### Tiedon louhinta ja sen prosessi

* association rule = assosiaatiosääntö
* binning = lokerointi
* classification = luokittelu
* clustering = klusterointi, ryvästys, ryhmittely
* concept description = käsitteen tai käsitteiden kuvailu, aineiston kuvailu
* concept hierarchy = käsitehierarkia
* categorical data = kategorinen, luokitteleva data
* data = data, (havainto)aineisto
* data cleaning = datan puhdistaminen, aineiston puhdistaminen
* data integration = datan yhdistäminen, aineistojen yhdistäminen
* data mining, DM = tiedon louhinta
* data mining methods = tiedon louhinnan menetelmät
* data preprocessing = datan esikäsittely, aineiston esikäsittely
* data post-processing = datan esikäsittely, aineiston jälkikäsittely
* data reduction = datan tiivistys, aineiston tiivistys
* descriptive data mining = kuvaileva tiedon louhinta
* dimensionality reduction = datan ulottuvuuksien (siis attribuuttien) vähentäminen
* interestingness = mielenkiintoisuus
* KDD process = tiedon louhintaprosessi
* knowledge discovery, KDD = tietämyksen muodostus, tiedon louhinta
* missing value = puuttuva arvo
* noise = häly, häiriö
* numeric data = numeerinen data
* numerosity reduction = datan/aineiston monilukuisuuden vähentäminen
* outlier = poikkeva havainto, vieras havainto
* predictive data mining = ennustava tiedon louhinta
* prediction = ennustaminen, ennuste
* regression = regressio
* supervised learning = ohjattu oppiminen
* transformation = muunnos
* unsupervised learning = ohjaamaton oppiminen

Muita mahdollisesti hyödyllisiä sanakirjoja tai sanastoja löytyy sekä kirjastosta että verkosta, mm. osoitteista

* Tietotekniikkaliiton [ATK-sanakirja](http://www.tt-tori.fi/atk-sanakirja/index.htm),
* Petri Koistisen keräämä [Tilastotieteen ja todennäköisyyslaskennan sanasto](http://www.rni.helsinki.fi/~pek/sanasto/tilastosanasto.html) ja
* Jukka Korpelan sivusto [Datatekniikan sanastoista](http://www.cs.tut.fi/~jkorpela/dt.html8).

Kerännyt: [*Pirjo Moen*](http://www.cs.helsinki.fi/pirjo.moen/)