Razlaga klasifikatorjev na podlagi podkonceptov

Nejc Mušič
Mentor: prof. dr. Marko Robnik Šikonja

Fakulteta za računalništvo in informatiko
Univerza v Ljubljani

nm4063@student.uni-lj.si

September 10, 2023

Pregled predstavitve

Motivacija za interpretacijo klasifikatorjev

- Klasifikatorji imajo dobro točnost pri napovedovanju
- Zaradi kompleksnosti in narave modelov ne razumemo vzrokov za sprejeto odločitev
- Kritične odločitve potrebujejo razlago
- Ljudem bo razlaga modelov omogočila zaupati v odločitve

Algoritmi za gručenje

- MDEC (Multidiversified Ensemble Clustering) se uporablja za gručenje visoko dimenzionalnih prostorov
- K-MEANS dobro gruči podatke sferičnih oblik
- DBSCAN (Density-Based Spatial Clustering of Applications with Noise) temelji na osnovi gostote podatkovnih točk v prostoru
- HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) izvaja algoritem DBSCAN pri različnih vrednostih epsilon in najde gručenje, ki zagotavlja najboljšo stabilnost

Hevristike za oceno gručenj

- Uporabimo za oceno gručenja in avtomatski izbor parametrov pri algoritmih gručenja
- Koeficient silhuete oceni, kako dobro so točke znotraj iste gruče povezane in kako dobro so razmejene od drugih gruč
- Indeks DBCV (Density-Based Clustering Validation) temelji na gostoti gruč in je primeren tudi za nepravilne oblike in postavitev v prostoru

Klasifikatorja

- XGBoost (Extreme Gradient Boosting) kombinira moč odločitvenih dreves in tehnike gradientnega spusta
- Logistična regresija napoveduje verjetnosti, da bo določen vhodni primer spadal v enega izmed razredov

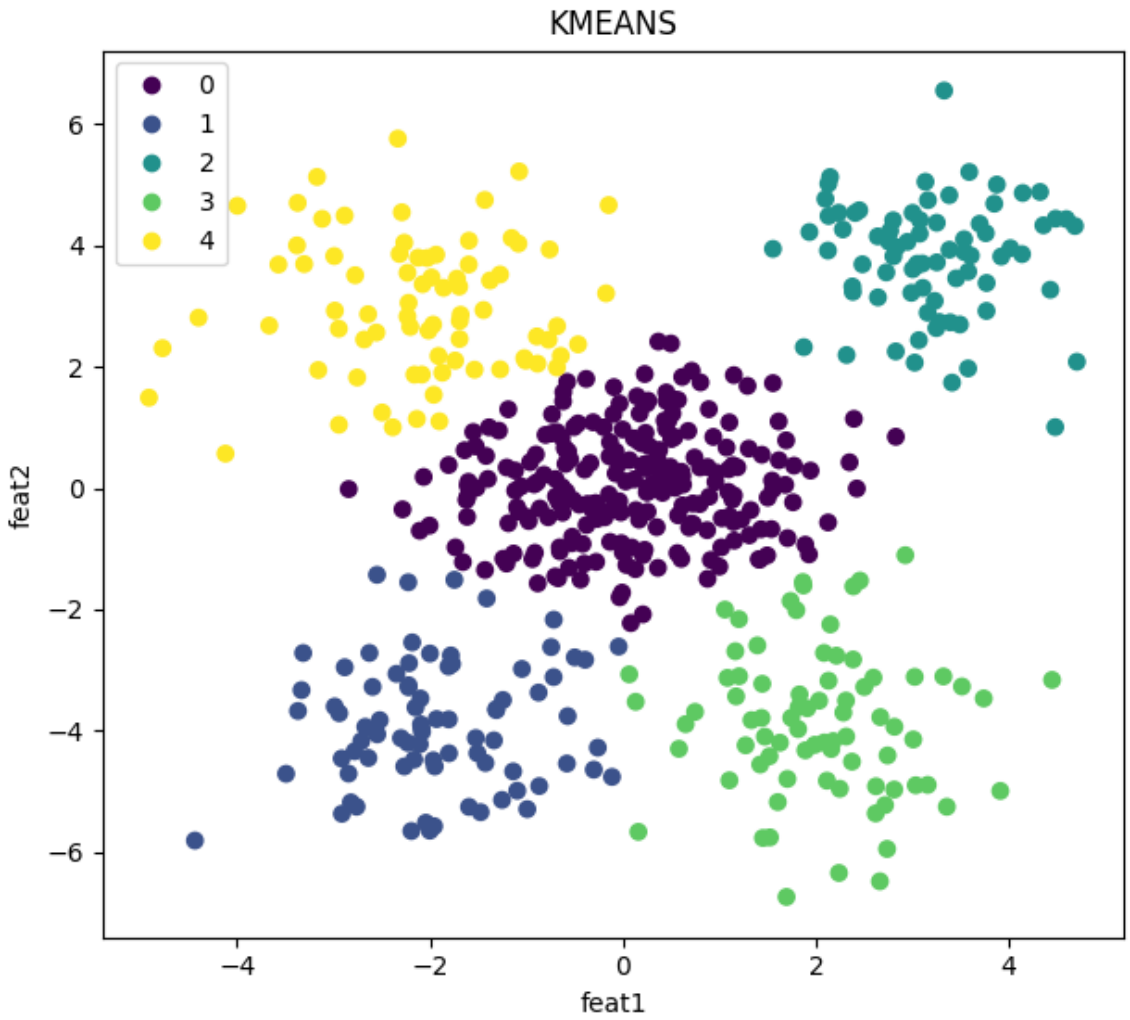
Metode razlage

- Odločitvena pravila uporabimo, saj so lahko razložljiva z obliko "IF conditions THEN response"
- Medoide uporabimo za opis homogenih gruč, saj gre za tipičen primer in ponuja enostaven opis
- Vrednosti SHAP uporabimo pri interpretaciji medoidov in odločitvenih pravil v visokih dimenzijah, saj razložijo pomembnost atributov

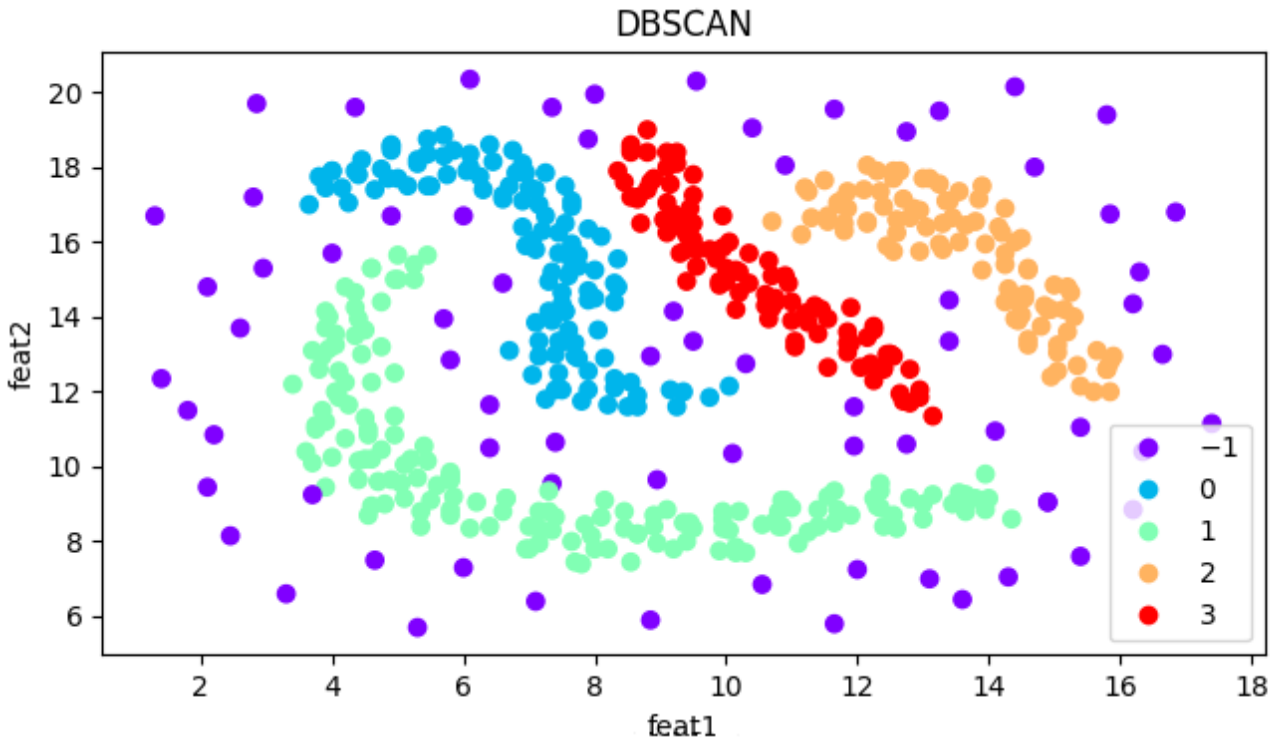
Zmanjšanje dimenzionalnosti

- Tehniko t-SNE uporabimo za vizualizacijo visoko dimenzionalnih podatkov v nižjih dimenzijah
- Nesistematično preizkusimo tudi tehniko UMAP (Uniform Manifold Approximation and Projection) in PCA (Principal Component Analysis), a t-SNE vrne bolje ločene gruče
- Tehniko t-SNE uporabimo tudi za predprocesiranje visoko dimenzionalnih podatkov pri gručenju z algoritmi K-MEANS, DBSCAN in HDBSCAN

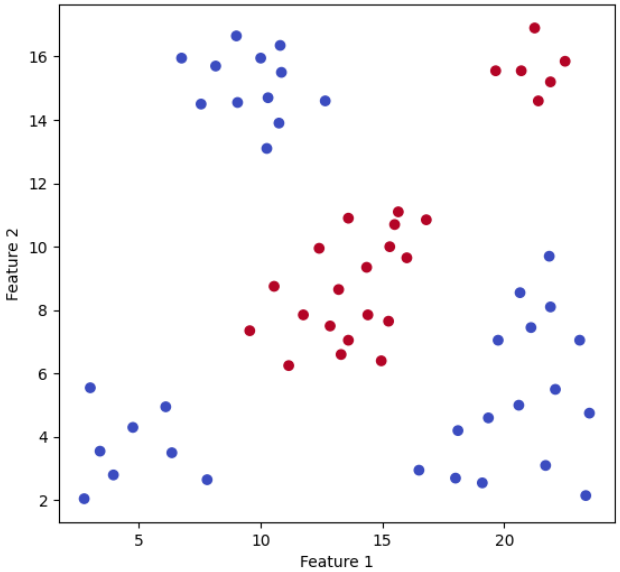
Podatkovna množica Krožne gruče

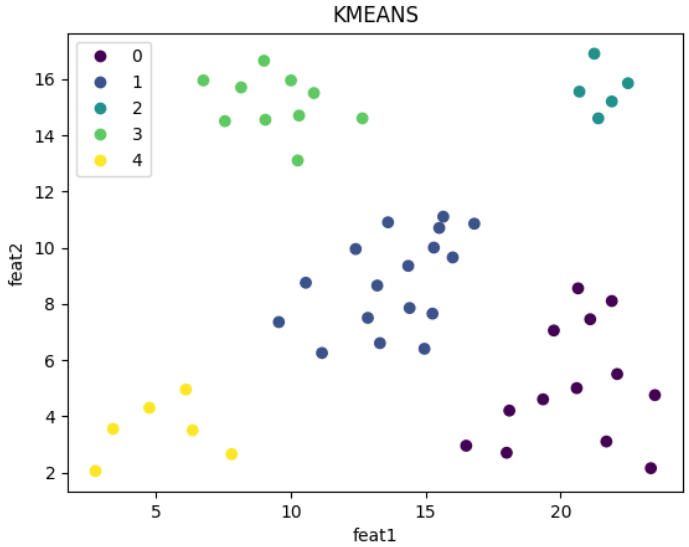


Podatkovna množica Trakovi 4-3

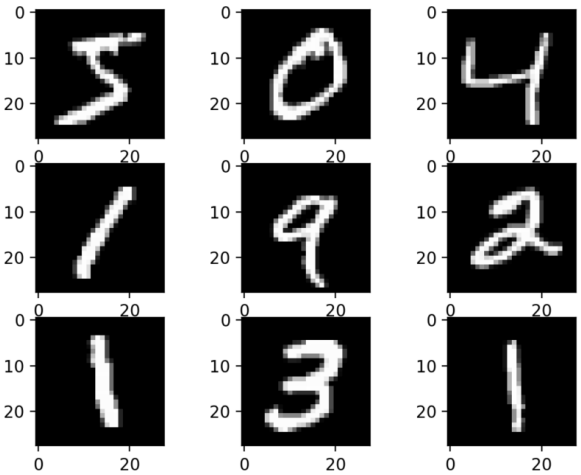


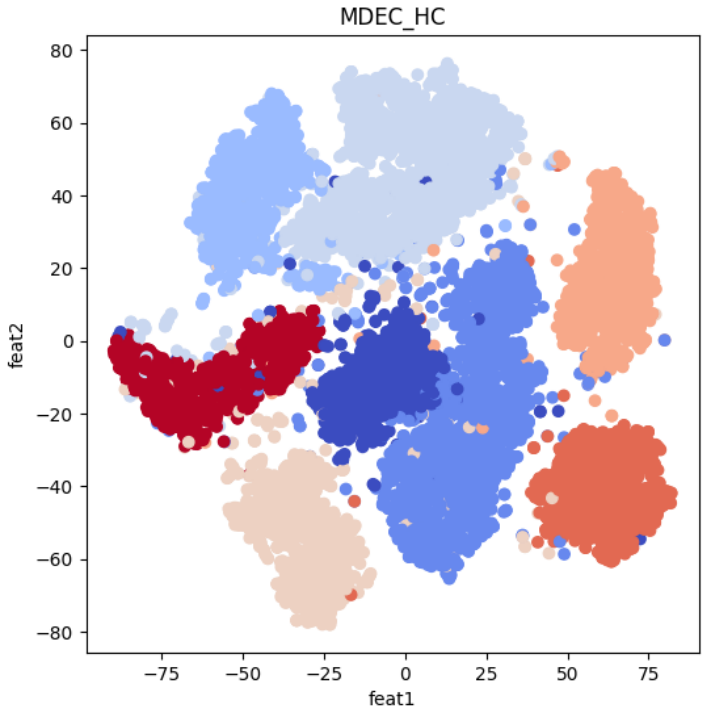
Podatkovna množica Homogene gruče



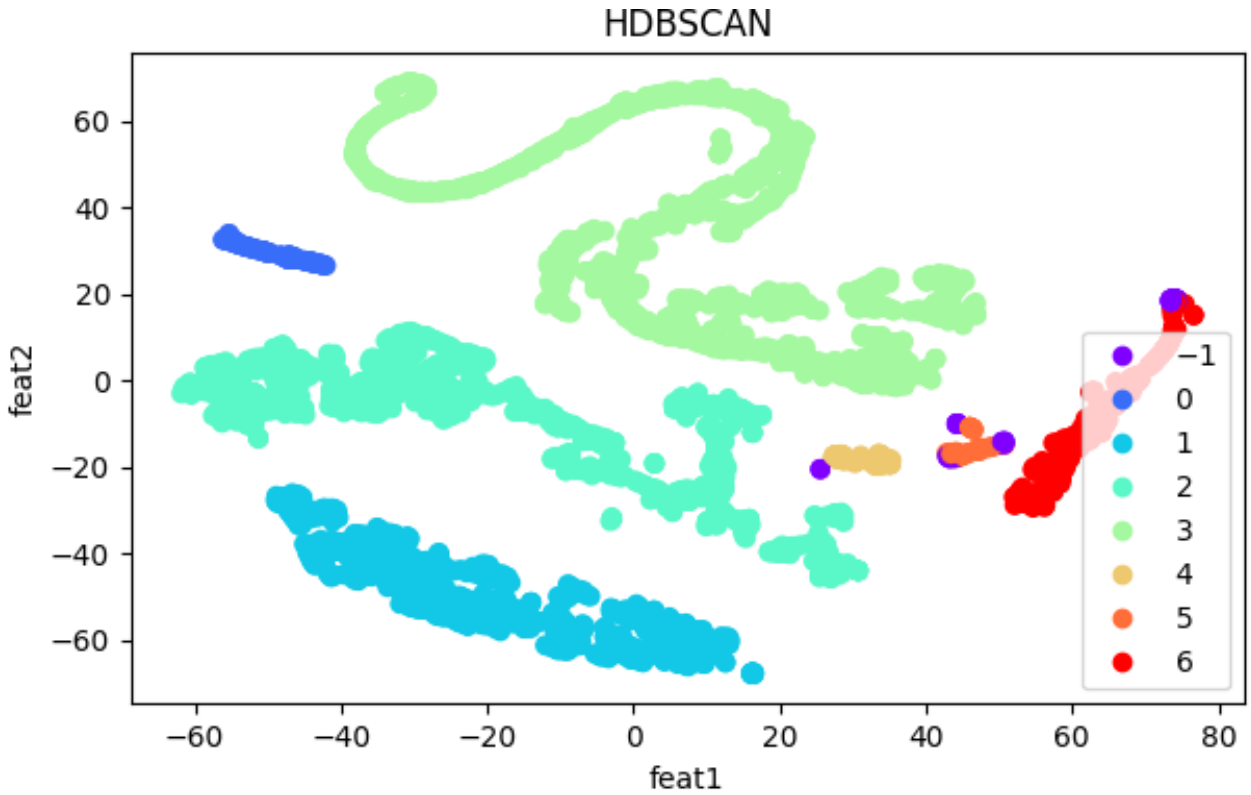


Podatkovna množica MNIST





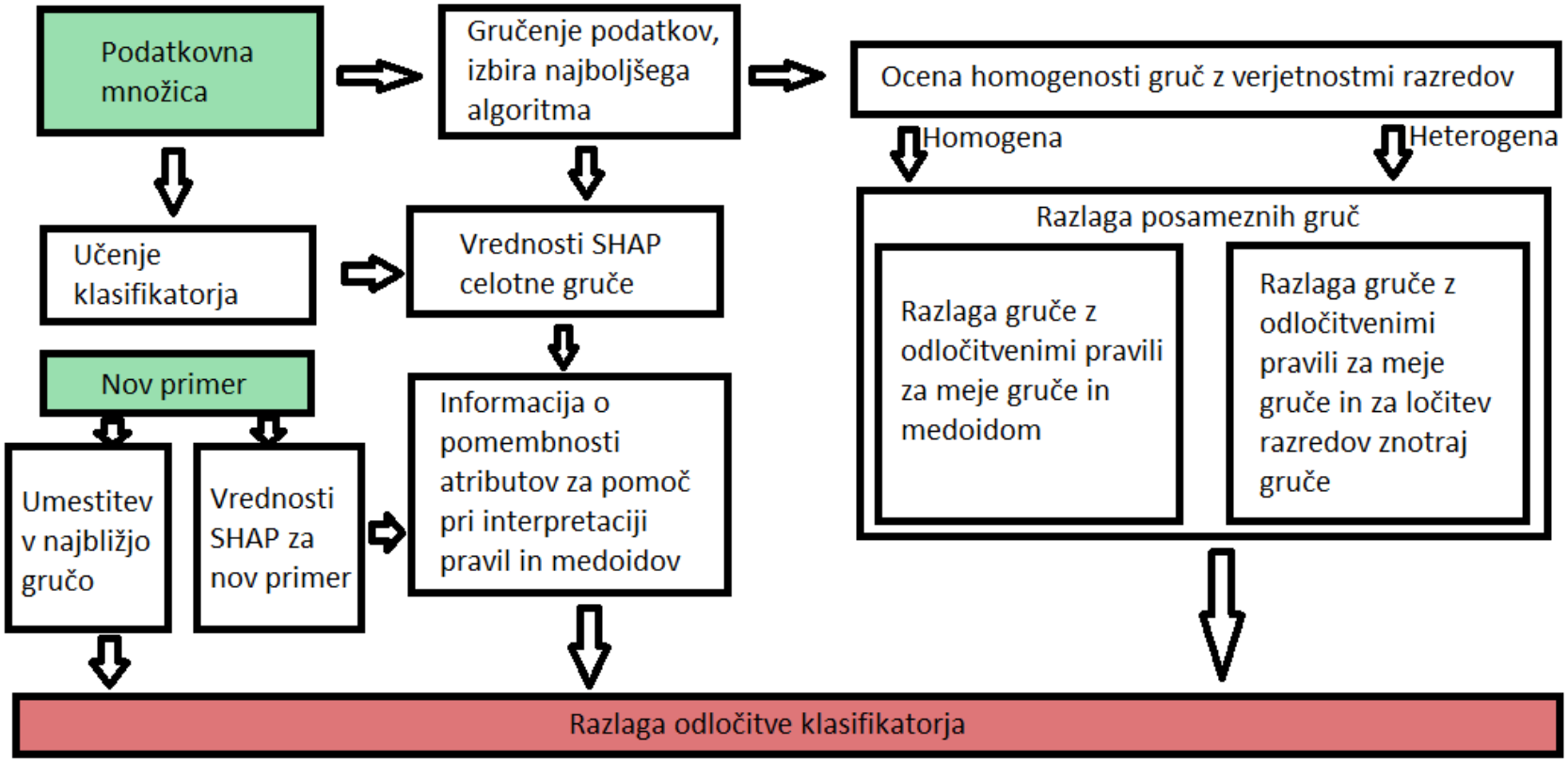
Podatkovna množica KDD99



Pregled rezultatov algoritmov gručenja

Podatkovna množica	Najboljši algoritem	Smiselni rezultati
Krožne gruče	K-MEANS	HDBSCAN
Trakovi 4-3	DBSCAN	HDBSCAN
MNIST	MDEC	HDBSCAN

Postopek razlage



Razlaga klasifikatorja na podatkovni množici KDD99

- Algoritem HDBSCAN je v kombinaciji s tehniko t-SNE našel sedem skupin (5 homogenih in 2 nehomogeni)
- Dve gruči imata večinske normalne mrežne povezave, ostale imajo večinske škodljive mrežne povezave
- Po izračunu vseh metod razlage, odločitev klasifikatorja interpretiramo na podlagi novega primera (normalna povezava)
- Opišemo, kako se medoida normalnih mrežnih povezav razlikujeta od medoidov škodljivih mrežnih povezav
- Ogledamo si vrednosti SHAP, ki pomagajo pri interpretaciji odločitvenih pravil in medoidov
- Podamo dodatno primerjavo med medoidoma, ki pripadata normalnim mrežnim povezavam

Zaključki

- Pomembnost števila gruč pri nekaterih algoritmih za gručenje in razlagi (preveč gruč: slabo razložljivi medoidi in pravila; premalo gruč: posplošitev)
- Medoidi smiselni pri homogenih gručah, v visokih dimenzijah nepregledni in neintuitivni (pomagamo si z vrednostmi SHAP)
- V visokih dimenzijah imamo več podmnožic odločitvenih pravil, ki dobro ločijo gruče. Nekatere značilke v pogojih niso pomembne.
- Postopek razlage težek za interpretacijo
- Struktura podatkov, ki jo pričakuje naša razlaga, je redka pri realnih podatkovnih bazah
- Postopek ni v celoti avtomatiziran (izbor algoritma gručenja, človeška interpretacija posameznih komponent razlage)
- Nepreglednost v visokih dimenzijah, težaven prikaz

Viri in literatura

- Dong Huang, Chang-Dong Wang, Jian-Huang Lai in Chee-Keong Kwoh. "Toward multidiversified ensemble clustering of high-dimensional data: From subspaces to metrics and beyond". V: IEEE Transactions on Cybernetics 52.11 (2021), str. 12231–12244.
- John A Hartigan in Manchek A Wong. "Algorithm AS 136: A k-means clustering algorithm". V: Journal of the royal statistical society. series c (applied statistics) 28.1 (1979), str. 100–108.
- Kamran Khan, Saif Ur Rehman, Kamran Aziz, Simon Fong in Sababady Sarasvady. "DBSCAN: Past, present and future". V: The fifth international conference on the applications of digital information and web technologies (ICADIWT 2014). IEEE. 2014, str. 232–238.
- Leland McInnes, John Healy in Steve Astels. "hdbscan: Hierarchical density based clustering." V: J. Open Source Softw. 2.11 (2017), str. 205. doi: 10.21105/joss.00205.
- Peter J. Rousseeuw. "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis". V: Journal of Computational and Applied Mathematics 20 (1987), str. 53–65. doi: https://doi.org/10.1016/0377-0427(87)90125-7.
- Davoud Moulavi, Pablo A Jaskowiak, Ricardo JGB Campello, Arthur Zimek in J¨org Sander. "Density-based clustering validation". V: Proceedings of the 2014 SIAM international conference on data mining. SIAM. 2014, str. 839–847.
- Ashutosh Nayak. "XGBoost: An Intuitive Explanation". V: Towards Data Science (2019). Datum dosega 15.8.2023. url: https://towardsdatascience.com/xgboost-an-intuitive-explanation-88eb32a48eff.
- Alfred DeMaris. "A tutorial in logistic regression". V: Journal of Marriage and the Family (1995), str. 956–968.
- Jerome H. Friedman in Bogdan E. Popescu. "Predictive learning via rule ensembles". V: The Annals of Applied Statistics 2.3 (2008), str. 916–954. doi: 10.1214/07-AOAS148.
- Scott M Lundberg in Su-In Lee. "A unified approach to interpreting model predictions". V: Advances in neural information processing systems 30 (2017).
- Laurens Van der Maaten in Geoffrey Hinton. "Visualizing data using t-SNE." V: Journal of machine learning research 9.11 (2008).
- Jason Brownlee. "How to Develop a CNN for MNIST Handwritten Digit Classification". V: Machine Learning Mastery (2019). Datum dosega 11.8.2023. url: https://machinelearningmastery.com/how-to-develop-a-convolutional-neural-network-from-scratch-for-mnist-handwritten-digit-classification/.