

**Reading list for the course**  
*Izbrana poglavja iz umetne inteligence 2b*  
*(Predictive Analytics for Structured Data) – prof. dr. Sašo Džeroski*

a) General framework for data mining

1. Džeroski, S. (2007). Towards a general framework for data mining. In Proc. Fifth International Workshop on Knowledge Discovery in Inductive Databases: Revised, Selected and Invited Papers, LNCS 4747: 259-300. (2007\_KDID\_TGFDm.pdf)

b) Predictive clustering for structured output prediction (Trees and rules for various tasks)

2. Vens, C., Struyf, J., Schietgat, L., Džeroski, S., and Blockeel, H. (2008). Decision trees for hierarchical multi-label classification. *Machine Learning*, 73(2): 185-214 (2008-VensEtAl-ML.pdf)
3. Levatić, J., Kocev, D., and Džeroski, S. (2014). The importance of the label hierarchy in hierarchical multi-label classification. *Journal of Intelligent Information Systems*, pp. 25 (2015jiis\_hierarchy.pdf)
4. Slavkov, I., Gjorgjioski, V., Struyf, J., and Džeroski, S. (2010). Finding explained groups of time-course gene expression profiles with predictive clustering trees. *Molecular BioSystems*, 6(4): 729-740 (2010-Slavkovetal-MBS.pdf)
5. Struyf, J., and Džeroski, S. (2006). Constraint based induction of multi-objective regression trees. In Proc. Fourth International Workshop on Knowledge Discovery in Inductive Databases: Revised, Selected and Invited Papers, LNCS 3933: 222-233. (2006KDID\_CBIPCT.pdf)
6. Stojanova, D., Ceci, M., Appice, A., and Džeroski, S. (2012). Network regression with predictive clustering trees. *Data Mining and Knowledge Discovery*, 25(2): 378-413 (2012-StojanovaEtAl-DAMI.pdf)
7. Kocev, D., Struyf, J., and Džeroski, S. (2007). Beam search induction and similarity constraints for predictive clustering trees. In Proc. Fifth International Workshop on Knowledge Discovery in Inductive Databases: Revised, Selected and Invited Papers, LNCS 4747: 134-151. (2007kdid\_bs\_similarity\_constraints\_pcts.pdf)
8. Aho, T., Ženko, B., Džeroski, S., and Elomaa, T. (2012). Multi-target regression with rule ensembles. *Journal of Machine Learning Research*, 13: 2367-2407. (2012-AhoEtAl-JMLR.pdf)
9. Ženko, B., and Džeroski, S. (2008). Learning classification rules for multiple target attributes. In Proc. Twelfth Pacific-Asia Conference on Knowledge Discovery and Data Mining, LNCS 5012: 454-465. (2008-PCR-Zenko-PAKDD08.pdf)
10. Mileski, V., Džeroski, S., and Kocev, D. (2017) Predictive Clustering Trees for Hierarchical Multi-Target Regression. In Proc. Sixteenth International Symposium on Intelligent Data Analysis, LNCS 10584: 223-234. (2017ida\_hmtr.pdf)
11. Osojnik, A., Džeroski, S., and Kocev, D. (2016). Option Predictive Clustering Trees for Multi-target Regression. In Proc. Nineteenth International Conference on Discovery Science, LNCS 9956: 118-133. (2016ds\_opcts\_mtr.pdf)
12. Stepišnik Perdiš, T., Osojnik, A., Džeroski, S., and Kocev, D. (2017). Option Predictive Clustering Trees for Hierarchical Multi-label Classification. In Proc. 20th International Conference on Discovery Science, LNCS 10558: 116-123. (2017ds\_opcts\_hmlc.pdf)
13. Breskvar, M., Kocev, D., and Džeroski, S. (2017). Multi-label Classification Using Random Label Subset Selections. In Proc. 20th International Conference on Discovery Science, LNCS 10558: 108-115. (2017ds\_ros\_mlc.pdf)
14. Breskvar, M., Kocev, D., and Džeroski, S. (2018). Ensembles for multi-target regression with random output selections. *Machine Learning* 107(11): 1673-1709. (2018mlj\_ros\_mtr.pdf)

c) Ensemble methods for structured output prediction

15. Kocev, D., Vens, C., Struyf, J., and Džeroski, S. (2013). Tree ensembles for predicting structured outputs. *Pattern Recognition*, 46(3): 817-833. (2013pr\_ensembles\_sp.pdf)
16. Madjarov, G., Kocev, D., Gjorgjevikj, D., and Džeroski, S. (2012). An extensive experimental comparison of methods for multi-label learning. *Pattern Recognition*, 45(9): 3084-3104 (2012pr\_ml\_comparison.pdf)
17. Panov, P., and Džeroski, S. (2007) Combining bagging and random subspaces to create better ensembles. In Proc. Seventh International Symposium on Intelligent Data Analysis, LNCS 4723: 118-129. (2008\_IDA\_BaggedSubspaces.pdf)
18. Kocev, D., and Ceci, M. (2015). Ensembles of extremely randomized trees for multi-target regression, *Discovery science DS-2015*, LNCS 9356: 86-100. (2015ds\_extra\_pcts.pdf)

d) Feature ranking for structured output prediction

19. Slavkov, I., Karcheska, J., Kocev, D., and Džeroski, S. (2018). HMC-ReliefF: Feature Ranking for Hierarchical Multi-label Classification. *Computer Science and Information Systems* 15(1):187–209. (2018comsis-reliefhmc.pdf)
20. Kocev, D., Slavkov, I., and Džeroski, S. (2013). Feature ranking for multi-label classification using predictive clustering trees. In *Proc. Workshop on Solving Complex Machine Learning Problems with Ensemble Methods* (held at ECML-PKDD 2013), pp. 56-68. Prague, Czech Republic. (2013copem\_frank\_rfpct.pdf)
21. Petković, M., Džeroski, S. and Kocev, D. (2017). Feature Ranking for Multi-Target Regression with Tree Ensemble Methods. In *Proc. 20th International Conference on Discovery Science, LNCS 10558*: 171-185. (2017ds\_frank\_mtr.pdf)
22. Petković, M., Džeroski, S. and Kocev, D. (2018). Feature Ranking with Relief for Multi-label Classification: Does Distance Matter?. In *Proc. 21th International Conference on Discovery Science, LNCS 11198*: 51-65. (2018ds\_frank\_mlc.pdf)

e) Ontologies of data mining

23. Panov, P., Soldatova, L., Džeroski, S. (2015). Generic ontology of datatypes. *Information Sciences*, 7: 35. (2015-PanovEtAl-IS.pdf)
24. Panov, P., Soldatova, L., and Džeroski, S. (2014). Ontology of core data mining entities. *Data Mining and Knowledge Discovery*, 28(5/6): 1222-1265 (2014-PanovEtAl-DAMI.pdf)

f) Semi-supervised learning for structured output prediction

25. Levatić, J., Ceci, M., Kocev, D., and Džeroski, S. (2017). Semi-supervised classification trees. *Journal of Intelligent Information Systems*, DOI 10.1007/s10844-017-0457-4. (2017jiis\_ssl\_classification)
26. Levatić, J., Ceci, M., Kocev, D., and Džeroski, S. (2017). Self-training for multi-target regression with tree ensembles. *Knowledge Based Systems*. DOI 10.1016/j.knosys.2017.02.014. (2017knosys\_self\_mtr.pdf)
27. Levatić, J., Kocev, D., Ceci, M., and Džeroski, S. (2018). Self-training for multi-target regression with tree ensembles. *Information Sciences*. <https://doi.org/10.1016/j.ins.2018.03.033> (2018infsci\_ssl\_mtr.pdf)

g) Structured output prediction from data streams

28. Osojnik, A., Panov, P., and Džeroski, S. (2017). Tree-based methods for online multi-target regression. *Journal of Intelligent Information Systems*, DOI 10.1007/s10844-017-0462-7. (2017jiis\_tree\_streams.pdf)
29. Osojnik, A., Panov, P., and Džeroski, S. (2017). Multi-label classification via multi-target regression on data streams. *Machine Learning*, DOI 10.1007/s10994-016-5613-5. (2017mlj\_mlc\_mtr.pdf)

g) Applications of structured output prediction: Environmental sciences

30. Džeroski, S., Demšar, D., and Grbović, J. (2000). Predicting chemical parameters of river water quality from bioindicator data. *Applied Intelligence*, 13(1): 7-17 (2000-DžeroskiEtAl-AI.pdf)
31. Demšar, D., Džeroski, S., Larsen, T., Struyf, J., Axelsen, J., Bruns-Pedersen, M., and Henning Krogh, P. (2006). Using multi-objective classification to model communities of soil microarthropods. *Ecological Modelling*, 191: 131-143 (2006-DemsarEtAl-EM.pdf)
32. Jerina, K., Debeljak, M., Džeroski, S., Kobler, A., and Adamič, M. (2003). Modeling the brown bear population in Slovenia: a tool in the conservation management of a threatened species. *Ecological Modelling*, 170 (2/3): 453-469. (2003-JerinaEtAl-EM.pdf)
33. Kocev, D., Džeroski, S., White, M., Newell, G., and Griffioen, P. (2009). Using single- and multi-target regression trees and ensembles to model a compound index of vegetation condition. *Ecological Modelling*, 220(8): 1159-1168. (2009ecomod\_ensembles\_mtrts\_hhi.pdf)
34. Debeljak, M., Kocev, D., Towers, W., Jones, M., Griffiths, B., and Hallett, P. (2008). Potential of multi-objective models for risk-based mapping of the resilience characteristics of soils: demonstration at a national level, *Soil Use and Management*, 25(1):66-77. (2009soil\_potential\_mt.pdf)
35. Stojanova, D., Kobler, A., Ogrinc, P., Ženko, B., and Džeroski, S. (2012). Estimating the risk of fire outbreaks in the natural environment. *Data Mining and Knowledge Discovery*, 24(2): 411-442. (2012-StojanovaEtAl-DAMI\_2.pdf)
36. Stojanova, D., Panov, P., Gjorgjioski, V., Kobler, A., and Džeroski, S. (2010). Estimating vegetation height and canopy cover from remotely sensed data with machine learning. *Ecological Informatics*, 5(4): 256-266 (2010-Stojanovaetal-ECOINF.pdf)
37. Debeljak, M., Squire, G., Kocev, D., Hawes, C., Young, M., and Džeroski, S. (2011). Analysis of time series data on agroecosystem vegetation using predictive clustering trees. *Ecological Modelling*, 222(14): 2524-2529 (2011ecomod\_time\_series.pdf)
38. Everaert, G., Boets, P., Lock, K., Džeroski, S., and Goethals, P. (2011). Using classification trees to analyze

the impact of exotic species on the ecological assessment of polder lakes in Flanders, Belgium. *Ecological Modelling*, 222(14): 2202-2212 (2011-EM-polder\_lakes.pdf)

39. Kocev, D., and Džeroski, S. (2013). Habitat modeling with single- and multi-target trees and ensembles. *Ecological Informatics*, 18: 79-92 (2013ei\_habitatmodelling.pdf)
40. Levatić, J., Kocev, D., Debeljak, M., and Džeroski, S. (2014). Community structure models are improved by exploiting taxonomic rank with predictive clustering trees. *Ecological Modelling*, pp. 11. (2015ecomod\_taxonomic\_ranks.pdf)

h) Applications of structured output prediction: Life sciences

41. Schietgat, L., Vens, C., Struyf, J., Blockeel, H., Kocev, D., and Džeroski, S. (2010). Predicting gene function using hierarchical multi-label decision tree ensembles. *BMC Bioinformatics*, 11(2): 1-14. (2010bmc\_hmc\_ensembles.pdf)
42. Škunca, N., Bošnjak, M., Kriško, A., Panov, P., Džeroski, S., and Šmuc, T. (2013). Phyletic profiling with cliques of orthologs Is enhanced by signatures of paralogy relationships. *PLoS Computational Biology*, 9(1): 1002852-1-1002852-14. (2013-SkuncaEtAl-PLOS-CompBio.pdf)
43. Stojanova, D., Ceci, M., Malerba, D., and Džeroski, S. (2013). Using PPI network autocorrelation in hierarchical multi-label classification trees for gene function prediction. *BMC Bioinformatics*, 14: 285-1-285-18. (2013-StojanovaEtAl-BMCBioinformatics.pdf)
44. Vidulin, V., Šmuc, T., Džeroski, S., and Supek, F. (2018). The evolutionary signal in metagenome phyletic profiles predicts many gene functions. *Microbiome*, 6: 129. DOI: 10.1186/s40168-018-0506-4. (2018microbiome\_phylogenetic.pdf)
45. Korbee, C., Heemskerk, M., Kocev, D., Van Strijen, E., Rabiee, O., Franken, K., Wilson, L., Savage, N., Džeroski, S., Haks, M., and Ottenhoff, T. (2018). Combined chemical genetics and data-driven bioinformatics approach identifies receptor tyrosine kinase inhibitors as host-directed antimicrobials. *Nature Communications*, 9(358). DOI: 10.1038/s41467-017-02777-6 (2018natcom\_repositioning.pdf)

i) Applications of structured output prediction: Image analysis

46. Dimitrovski, I., Kocev, D., Loskovska, S., and Džeroski, S. (2015). Improving bag-of-visual-words image retrieval with predictive clustering trees. *Information Sciences*, pp. 15. (2015is\_pcts\_retrieval.pdf)
47. Dimitrovski, I., Kocev, D., Loskovska, S., and Džeroski, S. (2013). Fast and efficient visual codebook construction for multi-label annotation using predictive clustering trees. *Pattern Recognition Letters*, 38: 38-45 (2013prl\_codebook\_construction.pdf)
48. Dimitrovski, I., Kocev, D., Loskovska, S., and Džeroski, S. (2012). Hierarchical classification of diatom images using ensembles of predictive clustering trees. *Ecological Informatics*, 7(1): 19-29 (2012ei\_diatom\_images.pdf)
49. Dimitrovski, I., Kocev, D., Loskovska, S., and Džeroski, S. (2011). Hierarchical annotation of medical images. *Pattern Recognition*, 44(10/11): 2436-2449. (2011pr\_image\_annotation.pdf)

j) Applications of structured output prediction: Engineering

50. Breskvar, M., Kocev, D., Levatić, J., Osojnik, A., Petković, M., Simidjievski, N., Ženko, B., Boumghar, R., and Lucas, L. (2017). Predicting thermal power consumption of the Mars Express satellite with machine learning. In *Proc. Sixth IEEE International Conference on Space Mission Challenges for Information Technology*, pp. 88-93. Alcalá de Henares, Spain. (2017\_smc\_mex.pdf)