

## DOKTORANTŪROS STUDIJŲ DALYKO SANDAS

Dalyko pavadinimas	Mokslo kryptis, kodas	Fakultetas	Katedra
Didžiųjų duomenų analitika	Informatikos inžinerija (T 007)	MIF	Duomenų mokslo ir skaitmeninių technologijų institutas

Studijų būdas	Kreditų skaičius ECTS	Studijų būdas	Kreditų skaičius
paskaitos	1 (rudens sem.)	konsultacijos	1
individualus	4	seminarai	1

Dalyko anotacija
<p><b>Dalyko tikslas</b> – siekiama, kad doktorantas įgytų esminių įgūdžių apie didžiųjų duomenų analitiką bei ugdytų praktinius gebėjimus taikyti didžiųjų duomenų saugojimo ir apdorojimo būdus ir technologijas bei tyrybos metodus.</p> <p><b>Įvadinė dalis:</b> Didžiųjų duomenų apibrėžtys bei jų vystymosi istorija. Pagrindiniai didžiųjų duomenų šaltiniai. Didžiųjų duomenų tipai: struktūriniai ir nestrukūriniai duomenys. Didžiųjų duomenų saugojimo būdai ir technologijos. Didžiųjų duomenų tyryba: taikymo uždavinių pavyzdžiai.</p> <p><b>Didžiųjų duomenų analitika:</b>                  Duomenų tyryba žinių gavimo procese, žinių gavimo proceso etapai.                  Didžiųjų duomenų pirminio apdorojimo būdai.                  Didžiųjų duomenų analitikos uždaviniai                  Didžiųjų duomenų tyryba:</p> <ul style="list-style-type: none"> <li>• Panašumų ir dažnų sekų paieška</li> <li>• Prognozavimo ir klasifikavimo metodai, paremti didžiųjų duomenų analize</li> <li>• Srautinių duomenų analizė</li> <li>• Klasterizavimo metodai</li> <li>• Dimensijos mažinimo būdai</li> <li>• Mašininis mokymasis didelių apimčių duomenims: neuroniniai tinklai ir gilusis mokymas</li> <li>• Teksto tyrybos metodai.</li> <li>• Didžiųjų duomenų vizualizavimo iššūkliai.</li> </ul> <p><b>Didžiųjų duomenų technologijos:</b> MapReduce paradigma. Apache Hadoop ir Spark platformos. Spark MLlib biblioteka mašiniam mokymuisi, Didžiųjų duomenų analitika taikant R paketą.</p> <p><b>Debesų kompiuterijos sprendimai, taikomi didžiųjų duomenų analitikoje.</b> Microsoft Azure Machine Learning platforma. Amazon Web Service sprendimai. KNIME Analytics Platform praplėtimai didiesiems duomenims.</p> <p><b>Didžiųjų duomenų tyrybos praktiniai aspektai.</b> Praktinių taikymų pavyzdžiai sprendžiant uždavinius, kylančius versle, finansuose, medicinoje ir kt.</p> <p><b>Praktinė užduotis:</b> išspręsti nurodytą duomenų analizės uždavinį, taikant didžiųjų duomenų technologijas.</p>
Pagrindinė literatūra
Marr, Bernard (2015). Big data: using smart big data, analytics and metrics to make better decisions and improve performance, ISBN 978-1-118-96583-2.
Simon Walkowiak (2016). Big Data Analytics with R, Publisher: Packt Publishing, ISBN: 9781786466457.
Favaretto, M., De Clercq, E., Schneble, C. O., & Elger, B. S. (2020). What is your definition of Big Data? Researchers' understanding of the phenomenon of the decade. <i>PloS one</i> , 15(2), e0228987.

Reddy, G.T., Reddy, M.P.K., Lakshmana, K., Kaluri, R., Rajput, D.S., Srivastava, G. and Baker, T. (2020). Analysis of dimensionality reduction techniques on big data. <i>IEEE Access</i> , 8, pp.54776-54788.
Parteek Bhatia (2019). Data Mining and Data Warehousing Principles and Practical Techniques. Cambridge University Press. Prieiga internete <a href="https://doi.org/10.1017/9781108635592">https://doi.org/10.1017/9781108635592</a>
Prajapati, V. Big data analytics with R and Hadoop. Packt Publishing Ltd, 2013.
Wolfgang Pietsch (2021), Big Data, Cambridge University Press. Prieiga internete <a href="https://doi.org/10.1017/9781108588676">https://doi.org/10.1017/9781108588676</a>
Mining Massive Data Sets, SOE-YCS0007, Stanford School of Engineering Prieiga internete <a href="https://online.stanford.edu/courses/soe-ycs0007-mining-massive-data-sets">https://online.stanford.edu/courses/soe-ycs0007-mining-massive-data-sets</a>
Leskovec, Jure, Anand Rajaraman, and Jeffrey David Ullman. Mining of massive data sets. Cambridge university press, 2020. Prieiga internete <a href="https://doi.org/10.1017/9781108684163">https://doi.org/10.1017/9781108684163</a> , <a href="http://www.mmds.org/">http://www.mmds.org/</a>
Spark and Python for Big Data with PySpark, Udemy, Prieiga internete <a href="https://www.udemy.com/course/spark-and-python-for-big-data-with-pyspark/">https://www.udemy.com/course/spark-and-python-for-big-data-with-pyspark/</a>

Konsultuojančiųjų dėstytojų vardas, pavardė	Mokslo laipsnis	Svarbiausieji darbai mokslo kryptyje (šakoje) paskelbti per pastaruosius 5 metus
Viktor Medvedev	dr.	<a href="http://www.elaba.mb.vu.lt/dmsti/?aut=Viktor+Medvedev">http://www.elaba.mb.vu.lt/dmsti/?aut=Viktor+Medvedev</a>
Jolita Bernatavičienė	dr.	<a href="http://www.elaba.mb.vu.lt/dmsti/?aut=Jolita+Bernatavičienė">http://www.elaba.mb.vu.lt/dmsti/?aut=Jolita+Bernatavičienė</a>
Gintautas Dzemyda	habil. dr.	<a href="http://www.elaba.mb.vu.lt/dmsti/?aut=Gintautas+Dzemyda">http://www.elaba.mb.vu.lt/dmsti/?aut=Gintautas+Dzemyda</a>