

DOCTORAL (PHD) STUDIES
COURSE UNIT DESCRIPTION

| Course unit title | Scientific areas | Faculty | Institute, department |
|--------------------|---------------------------------|--|--|
| Big Data Analytics | Informatics engineering (T 007) | Faculty of Mathematics and Informatics | Institute of Data Science and Digital technologies |

| Study method | Number of credits | Study method | Number of credits |
|------------------|-------------------|---------------|-------------------|
| Lectures | 1 (autumn) | Consultations | 1 |
| Individual works | 4 | Seminars | 1 |

| |
|---|
| <p>Summary</p> <p>Course objective. The aim is to enable the doctoral student to acquire essential skills in analyzing Big data and develop practical abilities to apply methods and technologies for Big data storage and processing, as well as methods of research.</p> <p>Introductory part: Definition and evolution of Big data. The main sources of Big data. Types of Big data: structured and unstructured. Big data storage technologies and techniques. Big data mining: practical applications.</p> <p>Big data analytics: Knowledge discovery in databases, data mining in knowledge discovery process, Methods for Big data preprocessing Big data analysis tasks Big data mining:</p> <ul style="list-style-type: none"> • Frequent pattern analysis • Forecasting and classification methods based on the analysis of Big data • Streaming data analysis • Clustering methods • Dimension reduction methods • Machine learning for Big data processing: neural networks and deep learning • Text mining methods • Big data visualization challenges <p>Big data technologies: MapReduce, Apache Hadoop and Spark. Spark MLlib Library for Machine Learning, Big data analytics using the R package. Cloud computing solutions for Big data analytics. Microsoft Azure Machine Learning. Amazon Web Service. KNIME Analytics platform extensions for Big data. Practical aspects of Big data processing. Practical applications in addressing the challenges arising from the business, finance, medicine and so on.</p> <p>Practical task: to solve the task of data analysis using Big data technologies.</p> |
| <p>Main literature</p> <p>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.</p> <p>Simon Walkowiak (2016). Big Data Analytics with R, Publisher: Packt Publishing, ISBN: 9781786466457.</p> <p>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.</p> |

| |
|--|
| 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. https://doi.org/10.1017/9781108635592 |
| Prajapati, V. Big data analytics with R and Hadoop. Packt Publishing Ltd, 2013. |
| Wolfgang Pietsch (2021), Big Data, Cambridge University Press. https://doi.org/10.1017/9781108588676 |
| Mining Massive Data Sets, SOE-YCS0007, Stanford School of Engineering https://online.stanford.edu/courses/soe-ycs0007-mining-massive-data-sets |
| Leskovec, Jure, Anand Rajaraman, and Jeffrey David Ullman. Mining of massive data sets. Cambridge university press, 2020. https://doi.org/10.1017/9781108684163 , http://www.mmds.org/ |
| Spark and Python for Big Data with PySpark, Udemy https://www.udemy.com/course/spark-and-python-for-big-data-with-pyspark/ |

| Lecturer(s) (name, surname) | Science degree | Main publications |
|-----------------------------|----------------|---|
| Viktor Medvedev | dr. | http://www.elaba.mb.vu.lt/dmsti/?aut=Viktor+Medvedev |
| Jolita Bernatavičienė | dr. | http://www.elaba.mb.vu.lt/dmsti/?aut=Jolita+Bernatavičienė |
| Gintautas Dzemyda | habil. dr. | http://www.elaba.mb.vu.lt/dmsti/?aut=Gintautas+Dzemyda |