



STUDIJŲ DALYKO (MODULIO) APRAŠAS

Dalyko (modulio) pavadinimas	Kodas
Didžiųjų duomenų analitika	

Dėstytojas (-ai)	Padalinys (-iai)
Koordinuojantis: dr. Laura Ringienė Kitas (-i):	Matematikos ir informatikos fakultetas Duomenų mokslo ir skaitmeninių technologijų institutas

Studijų pakopa	Dalyko (modulio) tipas
Pirmoji	Pasirenkamas

Igyvendinimo forma	Vykdymo laikotarpis	Vykdymo kalba (-os)
Auditorinė	6 semestras	Lietuvių

Reikalavimai studijuojančiajam	
Išankstiniai reikalavimai: Statistiniai duomenų analizės metodai, Kompiuteriniai tinklai, Duomenų tyryba ir mašininis mokymas.	Gretutiniai reikalavimai (jei yra):

Dalyko (modulio) apimtis kreditais	Visas studento darbo krūvis	Kontaktinio darbo valandos	Savarankiško darbo valandos
5	133	48	85

Dalyko (modulio) tikslas: studijų programos ugdomos kompetencijos		
Dalyko tikslas – siekiama, kad studentas įgytų žinių apie didžiųjų duomenų analitiką bei ugdytų praktinius gebėjimus taikyti duomenų apdorojimo ir tyrybos metodus analizuojant didžiuosius duomenis.		
Dalyko (modulio) studijų siekiniai	Studijų metodai	Vertinimo metodai
Gebės bendrauti su ne inžinerinės srities ekspertais valstybine bei užsienio kalbomis, diskutuojant ir perteikiant didžiųjų duomenų analitikos žinias ne specialistams.	Paskaita-diskusija, sąvokų žemėlapis, pavyzdžių demonstravimas ir analizė, informacijos paieška ir analizė, savarankiškas darbas, grupinis darbas, laboratoriniai darbai, rašto darbas, pateikčių rengimas ir darbų pristatymas.	Laboratorinių darbų atlikimas savarankiškai arba grupėse, aktyvumas paskaitose, testai, egzaminas.
Gebės laikytis profesinės etikos normų, susijusių su duomenų kaupimu, jų viešinimu.		
Gebės atlikti su didžiais duomenimis ir jų analitika susijusios tinkamos literatūros paiešką ir jos analizę, įsisavinti žinias apie didžiųjų duomenų analitiką ir jas pritaikyti sprendžiant praktinėje veikloje kylančias užduotis.		
Gebės atlikti duomenų analitiką išnaudojant debesų kompiuterijos teikiamas analitikos priemones.		
Gebės planuoti ir atlikti didžiųjų duomenų analizės eksperimentus, kelti hipotezes, vertinti skaičiavimų rezultatus.		
Gebės apdoroti didžiuosius duomenis, atlikti jų statistinę analizę, spręsti didžiųjų duomenų tyrybos uždavinius.		

Temos	Kontaktinio darbo valandos							Savarankiškų studijų laikas ir užduotys
	Paskaitos	Konsultacijos	Seminarai	Pratybos	Laboratoriniai darbai	Praktika	Visas kontaktinis darbas	
1. Įvadinė dalis. Didžiųjų duomenų apibrėžtys bei jų vystymosi istorija. Pagrindiniai didžiųjų duomenų šaltiniai, charakteristikos. Didžiųjų duomenų tipai: struktūriniai, pusiau struktūriniai ir nestrukūriniai duomenys. Didžiųjų duomenų tyrybos procesas. Realūs pavyzdžiai.	6				2		8	7
								Pateiktos informacijos analizavimas ir nevertinamų užduočių atlikimas <i>Praktinė užduotis:</i> Susipažinimas su sistemomis ir duomenų aprašymo pateikimas

2. Didžiųjų duomenų analitikos pagrindai. Duomenų tyryba žinių gavimo procese. Didžiųjų duomenų pirminio apdorojimo būdai. Didžiųjų duomenų tyrybos metodai bei jais sprendžiami uždaviniai. Didžiųjų duomenų vizualizavimo iššūkiai. Praktinių taikymų pavyzdžiai sprendžiant uždavinius.	12				14		26	24	Pateiktos informacijos analizavimas ir nevertinamų užduočių atlikimas <i>Praktinės užduotys:</i> • Klausimų kėlimas. • Duomenų rinkimas. • Duomenų valymas. • Duomenų analizavimas. • Duomenų vizualizavimas.
3. Didžiųjų duomenų analitikos metodai ir technologijos. MapReduce paradigma. Apache Hadoop. Srautiniai duomenys. Duomenų bazių valdymo sistemos.	4				4		8	16	Pateiktos informacijos analizavimas ir nevertinamų užduočių atlikimas <i>Praktinės užduotys:</i> • Hadoop. • MapReduce. • Srautiniai duomenys.
4. Debesų kompiuterijos sprendimai, taikomi didžiųjų duomenų analitikoje. Sistemos tinkančios didelių duomenų apdorojimui (Kntime, Neo4j ir k. t.)	2				4		6	18	Pateiktos informacijos analizavimas ir nevertinamų užduočių atlikimas <i>Praktinė užduotis:</i> Darbas su debesų kompiuterijos sistemomis.
5. Egzaminas								20	Literatūros kartojimas
Iš viso	24				24		48	85	

Vertinimo strategija	Svoris proc.	Atsiskaitymo laikas	Vertinimo kriterijai
			Kurso vertinimui taikoma kaupiamojo balo sistema, kuri susideda iš 11 laboratorinių darbų atsiskaitymų ir egzamino. Visa kurso informacija talpinama VMA https://emokymai.vu.lt
Praktinės užduotys	60	Semestro metu	Studentai privalo atlikti 11 laboratorinių darbų ir juos atsiskaityti. Prie kiekvienos atsiskaitymo užduoties yra pateikiamas jos vertinimas balais ir atsiskaitymo tipas. Atsiskaitymas gali būti dvejopas: savarankiškas darbo atlikimas ir apsigynimas arba savarankiško darbo atlikimas ir kitų besimokančiųjų įvertinimas. Atskiras atsiskaitymas vertinamas 10 balų sistemoje. Semestro gale suskaičiuojamas laboratorinių darbų įvertinimo vidurkis. Darbai atsiskaitomi nurodytu laiku. Vėluojant atsiskaityti įvertinimas mažinamas per pusę. Daugiausia vėluoti atsiskaityti galima savaitę. Atsiskaicius vieną laboratorinį darbą suteikiama galimybė matyti ir atlikti kitą laboratorinį darbą. Laboratorinių darbų perlaikymų semestro metu nėra.
Egzaminas	40	Egzaminų sesijos metu	Egzaminą gali laikyti studentai, kurie atsiskaitė 11 laboratorinių darbų. Egzaminas yra testo tipo, kurį sudaro atvirojo ir uždarojo tipo klausimai. Teste prie kiekvieno klausimo nurodomas jo vertinimas balais. Egzaminas vertinamas 10 balų sistemoje. Egzamino vertinimo strategija yra aptariama paskaitų metu ir pateikiama VMA https://emokymai.vu.lt . Neatvykus į egzaminą, kaupiamasis balas nėra skaičiuojamas ir egzamino žiniaraštyje yra žymimas neatvykimas. Egzaminas išlaikytas surinkus nemažiau 4 balų. Atsiskaityti eksternu galimybės nėra, nes kurso vertinimui yra taikoma kaupiamojo balo sistema, kuri susideda iš 11 laboratorinių darbų atsiskaitymų ir egzamino.

Autorius	Leidimo metai	Pavadinimas	Periodinio leidinio Nr. ar leidinio tomas	Leidimo vieta ir leidykla ar internetinė nuoroda
Privaloma literatūra				
EMC Education Services	2015	Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data		John Wiley & Sons
Erl, T., Khattak, W., & Buhler, P.	2016	Big data fundamentals: concepts, drivers & techniques		Prentice Hall Press, https://www.informit.com/articles/article.aspx?p=2473128&seqNum=11&ranMID=24808
Soraya Sedkaoui	2018	Data Analytics and Big Data		https://ebookcentral.proquest.com/lib/viluniv-ebooks/detail.action?docID=5401178



COURSE UNIT (MODULE) DESCRIPTION

Course unit (module) title	Code
Big Data Analytics	

Lecturer(s)	Department(s) where the course unit (module) is delivered
Coordinator: Dr. Laura Ringienė Other(s):	Faculty of Mathematics and Informatics Institute of Data Science and Digital Technologies

Study cycle	Type of the course unit (module)
First	Optional

Mode of delivery	The period when the course unit (module) is delivered	Language(s) of instruction
face-to-face	6 th semester	Lithuanian

Requirements for students	
Prerequisites: Students must have completed the following courses: Statistical Data Analysis Methods, Computer Networks, Data Mining and Machine Learning.	Additional requirements (if any):

Course (module) volume in credits	Total student workload	Contact hours	Self-study hours
5	133	48	85

Purpose of the course unit (module): programme competencies to be developed
The purpose of this course unit is to provide students with knowledge on Big Data analytics and to develop the practical abilities to apply the data processing and mining methods for analyzing Big Data.

Learning outcomes of the course unit (module)	Teaching and learning methods	Assessment methods
Will be able to discuss big data topics with both: experts and non-experts.	Lecture-discussion; Concept mapping; Example demonstration and analysis; Literature search and analysis; Self-dependent work; Teamwork; Practical tasks; Report preparation and presentation.	Laboratory work on your own or in groups; Active participation in lectures; Tests; Exam
Will preserve the ethics and norms that are related to data storage.		
Ability to conduct a search and analysis of the proper literature related to Big Data and their analytics, to absorb new knowledge on Big Data analytics, and to apply those while analyzing Big Data.		
Ability to design and implement Big Data processing and mining algorithms for multiprocessor systems; understanding applications of high-performance computing when processing Big Data		
Ability to plan and perform research experiments, investigate Big Data, evaluate the obtained results, and draw conclusions.		
Ability to handle Big Data, perform statistical analysis, to solve data mining tasks.		

Content: breakdown of the topics	Contact hours							Self-study work: time and assignments	
	Lectures	Tutorials	Seminars	Exercises	Laboratory work	Internship/work placement	Contact hours	Self-study hours	Assignments
1. Introduction. Definitions of Big Data and their development history. The main sources and characteristics of Big Data. Types of Big Data: structured, semi-structured, and unstructured data. The process of Big Data exploration. Real-life examples.	6				2		8	7	Analyzing provided information and completing non-assessable tasks <i>Practical task:</i> Familiarising yourself with the systems and providing a description of the data

2. The basics of Big Data analytics. Data exploration in the knowledge discovery process. Big Data pre-processing techniques. Big Data mining methods and the problems they solve. Challenges of visualizing Big Data. Examples of practical applications in problem-solving.	12				14		26	24	Analyzing provided information and completing non-assessable tasks <i>Practical task:</i> Raising questions; Collecting data; Data cleaning; Analysing data; Visualising data.
3. Big Data analytics techniques and technologies. MapReduce paradigm. Apache Hadoop. Streaming data. Database management systems.	4				4		8	16	Analyzing provided information and completing non-assessable tasks <i>Practical task:</i> Hadoop; MapReduce; Streaming data.
4. Cloud computing solutions for Big Data analytics. Systems suitable for Big Data processing (Knime, Neo4j, etc.)	2				4		6	18	Analyzing provided information and completing non-assessable tasks <i>Practical task:</i> Working with cloud computing systems.
5. Exam								20	Literature review
Total	24				24		48	85	

Assessment strategy	Weight, %	Deadline	Assessment criteria
			The course is assessed by a cumulative marking system consisting of 11 laboratory assignments and an exam. All course information is hosted on the VMA https://emokymai.vu.lt
Practical tasks	60	During the semester	Students are required to complete and submit 11 laboratory assignments. Laboratory work is computer-based. Each assignment is graded and the type of assignment is indicated. The reporting can take two forms: independent work and defence or independent work and peer assessment. The individual report shall be assessed on a 10-point system. At the end of the semester, the average of the marks for the laboratory work is calculated. The assignments are due at the specified time. In case of late submission, the mark will be halved. The maximum number of late submissions is one week. After one laboratory paper has been submitted, the next laboratory paper will be allowed to be seen and completed. There are no retakes of laboratory work during the semester.
Exam	40	Exam session	Students who have completed 11 of their laboratory work can sit the exam. The exam is a test-type examination consisting of open and closed questions. Each question in the test shall be accompanied by a score. The examination shall be graded on a 10-point system. The marking strategy for the examination is discussed during lectures and made available on the VMA https://emokymai.vu.lt Failure to attend an examination will not result in a cumulative mark and will be marked as an absence on the examination paper. The examination is passed with a score of at least 4 points. There is no possibility of an external examination, as the course is assessed by a cumulative marking system consisting of 11 laboratory assignments and an examination.

Author	Year of publication	Title	Issue of a periodical or volume of a publication	Publishing place and house or weblink
Compulsory reading				
EMC Education Services	2015	Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data		John Wiley & Sons
Erl, T., Khattak, W., & Buhler, P.	2016	Big data fundamentals: concepts, drivers & techniques		Prentice Hall Press, https://www.informit.com/articles/article.aspx?p=2473128&seqNum=11&ranMID=24808
Soraya Sedkaoui	2018	Data Analytics and Big Data		https://ebookcentral.proquest.com/lib/viluniv-ebooks/detail.action?docID=5401178