COURSE UNIT DESCRIPTION

Course unit title	Code
DATA MINING TECHNOLOGIES	

Lecturer(s)	Department, Faculty
Coordinating: Lect. Paulius Danielius **Other:**	Kaunas Faculty Institute of Social Sciences and Applied Informatics

Study cycle	Type of the course unit
Bachelor	Compulsory

Mode of delivery	Semester or period when it is delivered	Language of instruction
Auditorium	6	EN

Requisites	
Prerequisites: Advanced Mathematics, Informatics	**Co-requisites (if relevant):**

Number of ECTS credits allocated	Student's workload (total)	Contact hours	Individual work
5	130	52	78

Purpose of the course unit: programme competences to be developed

To develop the ability to work with data mining and data analysis technologies. To provide students with an understanding of data mining technologies, data analysis technologies, trends, methods. To develop the ability to apply in practice the main data analysis tools and the methods and algorithms used in them. Develop the ability to process, interpret and analyse data mining results.

Learning outcomes of the course unit	Teaching and learning methods	Assessment methods
Will be able to communicate effectively in written and oral English and/or national languages, using appropriate subject-specific terminology, presenting proposed ideas and justifying innovative solutions to problems to an audience of both professionals and users of IT services.	Active learning methods (individual presentation and group discussion).	Presentation of an individual written work to the class.
Will be able to apply knowledge of informatics engineering to develop safe and secure informatics application solutions that meet other relevant criteria for solving societally relevant problems.	Lectures, case studies, practical tasks.	• Defense of practical tasks reports. • Colloquium: Test in e-learning system. • Examination
Will be able to carry out a skilled analysis of a problem in the subject area of informatics engineering, using relevant and well-chosen empirical data, adhering to theoretical approaches, and applying effective methods.	Lectures, practical tasks, Active learning methods (group discussion, comparative analysis of methods).	
Will be able to systematise, summarise, interpret and critically evaluate research results, formulate conclusions and make recommendations for practical application.	Practical tasks, case studies.	

Course content: breakdown of the topics	Contact hours							Individual work: time and assignments	
	Lectures	Consultation	Exam	Workshops	Laboratory work	Internship	**Contact hours, total**	**Individual work**	**Assignments**
1. Business intelligence systems	**2**				**2**		**4**	**8**	Individual assignment - written work (scientific review on a topic chosen from a list of data mining applications in different subject areas), analysis and synthesis of the scientific literature. Presentation of the written work. Preparing for the colloquium.
Introduction. Evolution of business intelligence (BI) systems and classification of BI systems. Data mining tools, Rapid Miner package. Data Visualisation.	2				2		4	8	
2. Data storage systems	**2**						**4**	**8**	
Data integration, data warehousing systems and technologies, architectures, ETL.	2						4	8	
3. Data mining	**4**				**6**		**12**	**24**	
Data mining lifecycle. Knowledge discovery in databases.	2				2		4	8	
Data mining applications in marketing, education and other fields.					2		2	8	
Data analysis standards: CRISP-DM. Examples of data mining tasks. Data preparation processes.	2				2		6	8	
Colloquium	**2**						**2**		
3. Data mining algorithms	**10**				**20**		**26**	**38**	Laboratory work. Preparing for the exam.
Types of data mining tasks. Decision tree algorithm.	2				4		6	8	
Association rules.	2				4		6	8	
Text and web data mining.	2				4		6	8	
Neural networks.	2				4		6	8	
Other data mining algorithms. Regression analysis. Nearest Neighbour Algorithm. Clustering algorithms. Genetic algorithms.	2						2	6	
Python programming language tools for data mining.					4				
Consultation		2					**2**		
Exam			2				**2**		
Total	**20**	**2**	**2**		**28**		**52**	**78**	

Assessment strategy	Weight %	Deadline	Assessment criteria *For more details on the evaluation strategy and the use of AI generative models, see the table below.*
Practical assignments	40	2-16th week	Practical assignments are to be completed throughout the semester, and assignment reports are to be uploaded to the VMA (in MS Word or PDF format) according to the provided timetable. Each assignment is evaluated on a 10-point scale according to the VU assessment criteria. The grade for the practical assignments (overall grade) is the average of the grades for all assignments.
Written work with presentation	10	6th week	Essay: Students must prepare a literary review and analysis and give an oral presentation. The topic must be agreed with the lecturer. NO late assignments will be accepted, consequently, a student will be assigned a zero if the work is turned in late without the lecturer's permission.
Colloquium	20	9th week	The colloquium includes the presented theoretical and practical material, evaluation on a 10-point scale according to VU evaluation criteria.
Exam	30	During the session	The exam covers the presented theoretical and practical material of the subject, evaluation on a 10-point scale according to VU evaluation criteria.

A strategy for evaluating practical assignments and guidelines for using AI (Artificial intelligence) generative models.

The individual task – essay is written after selecting a topic from the provided list and agreeing it with the lecturer. In preparing this work and the practical reports, the student must use credible internet sources and scholarly articles and may use AI generative models in accordance with the principles of academic integrity policy (*Copy-Paste* is considered plagiarism - citations must be used, **see below for more details**). If the written work does not use enough sources, including scholarly articles, the mark will be reduced accordingly.
For the presentation, students use slides as an aid, they are not meant to be read, so students aiming for the maximum mark must present the topic in a proper way (reading from the slides is assessed with the minimum mark).
More detailed instructions for the written work and presentation are given to students at the time of topic selection.
Examples of the use of AI generative models
The best way to use such tools is:

- When explaining the principles of methods, algorithms and techniques during the learning process,
- for explaining concepts,
- for the development of a structure,
- generating ideas,
- case studies,
- generating summaries (for further work),
- processing large texts (for further work).

All information generated by AI tools **must be verified** and **the work must be referenced to external sources to prove it**, ensuring proper citation (in any case, the Copy-Paste principle is considered plagiarism if no citation is given). It is also important to understand that AI generative models are not co-authors of the work.
More about academic integrity at VU (in particular, see point 19):
https://www.vu.lt/site_files/Studies/Study_regulations/Code_of_academic_ethics_VU.pdf.
When should AI generative models not be used in this course?
These tools cannot be used:

- In written work and presentations, for copy-paste submissions without appropriate citation.
- For text embellishment (this does not apply to machine translation tools such as Deepl).
- For assessment tests during the semester and exams.

If AI generative models have been used in the preparation of the work?
If AI generative models have been used to generate ideas for the work, it must start with a description of

- The strategy for using AI tools,
- what questions were asked,
- what result was obtained and what percentage of the result was modified and adapted for the work.

The annexes shall contain the queries (e.g. Chat GPT query: "...") and the results (e.g. Chat GPT generated response "...") and the name, version and date of use of the generative model. More on citation: https://apastyle.apa.org/blog/how-to-cite-chatgpt, https://guides.library.uq.edu.au/referencing/chatgpt-and-generative-ai-tools.
It must also describe the volume of text generated by the AI tools used in the work. If the text is copied from generative model systems, it must be cited, as must any source. The number of citations and AI-generated text in the thesis cannot exceed 20% (e.g. more https://plagiarismcheck.org/blog/what-is-the-acceptable-percentage-of-plagiarism/).
When using AI generative models, it is important for students to be critical of the answers given, to be ethical, to be accurate, and for each student to be transparent with the rest of the group.
In the case of academic dishonesty: the lecturer informs the administration if he/she notices signs of plagiarism or if he/she discovers that a piece of written work contains blocks of text generated by artificial intelligence tools (i.e. academic dishonesty is suspected). In this case, a process for assessing academic integrity will be initiated.

Author	Publishing year	Title	Publishing house or internet site
Required reading			
P. Danielius	2024	Moodle environment	https://emokymai.vu.lt/?lang=en
Mohammed J. Zaki, Wagner Meira	2014	Data mining and analysis: fundamental concepts and algorithms	Cambridge University Press
Ian H. Witten, Eibe Frank, Mark A. Hall	2011	Data mining: practical machine learning tools and techniques	Morgan Kaufmann
G. Dzemyda, O. Kurasova, J. Zilinskas	2013	Multidimensional Data Visualization Methods and Applications	Springer
Recommended reading			
Andreas C. Müller, Sarah Guido	2018	Introduction to machine learning with Python: a guide for data scientists	O'Reilly
Gabe Ignatow, Rada Mihalcea	2018	An introduction to text mining: research design, data collection, and analysis	SAGE