



STUDIJŲ DALYKO (MODULIO) APRAŠAS

Dalyko (modulio) pavadinimas	Kodas
Tikimybiniai mašininio mokymo algoritmai I	

Dėstytojas (-ai)	Padalinys (-iai)
Koordinuojantis: dr. Saulius Jokubaitis	Statistinės analizės katedra
Kitas (-i): dr. Viktor Skorniakov	

Studijų pakopa	Dalyko lygmuo	Dalyko (modulio) tipas
Antroji	Mokančiųjų	Privalomasis

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

Reikalavimai studijuojančiajam	
Išankstiniai reikalavimai: R ir Python pagrindai; anglų kalba B1 (arba aukštesniu) lygiu pagal įprastą Europoje galiojančią lygių sistemą; parametrinio statistinio modeliavimo pagrindai	Gretutiniai reikalavimai (jei yra):

Dalyko (modulio) apimtis kreditais	Visas studento darbo krūvis	Kontaktinio darbo valandos	Savarankiško darbo valandos
5	125	48	77

Dalyko (modulio) tikslas: studijų programos ugdomos kompetencijos		
Studentai turėtų išugdyti šias kompetencijas (B – bendrosios, D – dalykinės kompetencijos): - gebėjimas analizuoti, sisteminti, mokytis ir taikyti įgytas žinias praktikoje (B1); - gebėjimas naudotis matematine kalba, spręsti analitines problemas panaudojant matematinius įrankius (D4); - gebėjimas rinktis tinkamą analizės metodologiją bei jai reikalingus įrankius (D6); - gebėjimas interpretuoti ir reprezentuoti analizės rezultatus (D7).		
Dalyko (modulio) studijų siekiniai (išklause modulį studentai turėtų):	Studijų metodai	Vertinimo metodai
- susipažinti su mašininio mokymosi paradigma; - teoriškai ir praktiškai susipažinti su bendrais prižiūrimo mašininio mokymo modeliavimo metodais ir keliais modeliais; - susipažinti su programine įranga, skirta mašininio mokymosi modelių analizei.	Paskaitos, pratybos naudojant programinę įrangą, savarankiškas užduočių sprendimas ir teorinės medžiagos studijavimas.	Kontroliniai darbai

Temos	Kontaktinio darbo valandos	Savarankiškų studijų laikas ir užduotys
-------	----------------------------	---

	Paskaitos	Konsultacijos	Seminarai	Pratybos	Laboratoriniai darbai	Praktika	Visas kontaktinis darbas	Savarankiškas darbas	Užduotys ¹
1. Įvadas. Dirbtinis intelektas ir mašininis mokymasis. Mašininio mokymosi terminologija, uždaviniai ir jų klasifikavimas.	1						1	5	Perskaityti įvadinis [3] ir [4] knygų poskyrius.
2. Teoriniai pagrindai. Statistinio mokymosi modelis. Empirinės rizikos minimizavimo taisyklė. Apytikriai teisingo mokymosi modelis.	2				6		8	10	Išspręsti dėstytojo suformuluotas užduotis. Perskaityti nurodytus [3] knygos 2, 3, 4 ir 6 skyrių poskyrius.
3. Optimizavimo metodų pagrindai. Pagrindinės sąvokos. Iškilos funkcijos ir aibės. Iškilų funkcijų savybės. Dualumo teorijos pagrindai. Skaitiniai algoritmai: determinuotas ir stochastinis gradientinio nusileidimo metodai, teorinės garantijos. Taikymai mašininio mokymosi kontekste.	5				8		13	12	Išspręsti dėstytojo suformuluotas užduotis. Perskaityti nurodytus [5] ir [6] knygų poskyrius.
4. Bendri metodai. Hiperparametrai, kryžminė validacija, reguliarizacija, validavimo ir mokymosi kreivės.	3				2		5	10	Išspręsti dėstytojo suformuluotas užduotis. Perskaityti nurodytus [3] ir [4] poskyrius.
5. Programinė įranga. Darbas sklearn aplinkoje: duomenų paruošimas; modelių klasės, jų metodai ir atributai; hiperparametrų parinkimas.	2				3		5	11	Išspręsti dėstytojo suformuluotas užduotis. Susipažinti su nurodytais [2] dokumentacijos poskyriais ir pavyzdžiais.
6. 0-1 klasifikavimo modeliai – pirma pažintis. Logistinis klasifikatorius, atraminių vektorių mašinos, artimiausių kaimynų metodas, klasifikavimo tikslumo matai.	3				3		6	9	Susipažinti su nurodytais [2] dokumentacijos poskyriais ir pavyzdžiais.
7. Kontroliniai darbai ir egzaminas.							10	24	Pasiruošti atsiskaitymams.
Iš viso							48	77	

Vertinimo strategija	Svoris proc.	Atsiskaitymo laikas	Vertinimo kriterijai
1 kontrolinis darbas	25	ketvirta studijų savaitė	Kontrolinį sudaro ne daugiau 5 užduočių, skirtų patikrinti žinių lygiui. Bendra užduočių vertė – 2,5 balo. Kiekvienos užduoties vertė svyruoja nuo 0,1 iki 1 balo. Užduotys atliekamos raštu arba prie kompiuterio. Atskiros užduoties vertinimo principai: a) išskiriamos dalys, už kurias skiriama dalis visos užduoties taškų; b) atlikus
2 kontrolinis darbas	25	aštunta studijų savaitė	
3 kontrolinis darbas	25	dvylikta studijų savaitė	

¹ dėstytojo nurodytos užduotys talpinamos virtualioje mokymosi aplinkoje (toliau VMA)

4 kontrolinis darbas	25	šešiolikta studijų savaitė	atitinkamą dalį be klaidų už ją skiriamas maksimalus taškų skaičius, priešingu atveju taškų skaičius mažinamas atsižvelgiant į padarytas klaidas; c) klaidingas kažkurios dalies atlikimas neturi įtakos kitų dalių vertinimui.
Egzaminas		Sesijos metu	Egzaminas skirtas pakelti semestro metu sukauptam balui. Dėstytojo sprendimu norintiems pasikelti balą gali būti taikoma viena iš šių strategijų: a) raštu ir/arba žodžiu pateikiamos individualios užduotys; b) pasirinktinai leidžiama perrašyti vieną kontrolinį darbą. Gavus prastesnį balą paliekamas ankstesnis įvertinimas.
Egzamino perlaikymas eksternu yra negalimas.			

Autorius	Leidimo metai	Pavadinimas	Periodinio leidinio Nr. ar leidinio tomas	Leidimo vieta ir leidykla ar internetinė nuoroda
Privaloma literatūra				
1. V. Skorniakov	2023	MM paskaitų skaidrės, I dalis		Pateikiamos virtualioje mokymosi aplinkoje
2. scikit-learn developers	2023	Documentation of scikit-learn		https://scikit-learn.org/stable/documentation.html
3. Shai Shalev-Shwartz and Shai Ben-David	2014	Understanding Machine Learning: From Theory to Algorithms		Cambridge University Press (nuo spausdintos nesmarkiai besiskirianti versija prieinama internete)
4. Andreas Lindholm, Niklas Wahlström, Fredrik Lindsten, and Thomas B. Schön	2022	Machine Learning - A First Course for Engineers and Scientists		Cambridge University Press (prieiga internete)
5. Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar	2018	Foundations of Machine Learning		MIT Press, Second Edition, 2018 (prieiga internete)
6. Hui Jiang	2021	Machine Learning Fundamentals: A Concise Introduction		York University, Toronto, Cambridge University Press, 2021 (prieiga galima per virtualią VU biblioteką ²)
Papildoma literatūra				
1. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani	2013	An Introduction to Statistical Learning: with Applications in R		Springer (prieiga internete)
2. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani	2023, summer	An Introduction to Statistical Learning: with Applications in Python		Springer (prieiga internete)
3. Trevor Hastie, Robert Tibshirani, Jerome Friedman	2016	The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition		Springer (prieiga internete)

² norint parsisiųsti knygą reikia būti prisijungus prie VU tinklo per VPN



COURSE UNIT (MODULE) DESCRIPTION

Course unit (module) title	Code
Probabilistic machine learning I	

Lecturer(s)	Department(s) where the course unit (module) is delivered
Coordinator: assistant prof. S. Jokubaitis	Department of Statistical Analysis
Other(s): associate prof. V. Skorniakov	

Study cycle	Level of course	Type of the course unit (module)
First	Advanced	Compulsory

Mode of delivery	Period when the course unit (module) is delivered	Language(s) of instruction
Face-to-face	Sixth (spring) semester	Lithuanian

Requirements for students	
Prerequisites: basics of R and Python; ability to understand English at the level of independent user (B1 according to CEFR classification), familiarity with simplest parametric statistical models.	Additional requirements (if any):

Course (module) volume in credits	Total student's workload	Contact hours	Self-study hours
5	125	48	77

Purpose of the course unit (module): programme competences to be developed (the number in the brackets coincides with that given in the official description of the programme)		
- Ability to analyse, systematize, learn and apply the obtained knowledge in practice (1); - Ability to use mathematical language and solve analytical problems by making use of mathematical tools (4); - Ability to choose appropriate methodology and tools (6); - Ability to interpret and represent the results obtained (7);		
Learning outcomes of the course unit (module); after completing the course students should:	Teaching and learning methods	Assessment methods
- become familiar with machine learning paradigm; - become familiar with several general Machine Learning modelling principles on the level required in applications; - become familiar with software tools designed for machine learning modelling.	Lectures, problem solving and reading, assignments	Tests

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. Artificial intelligence and Machine Learning (ML). Terminology, tasks, and their classification.	1						1	5	To read introductory chapters of [3] and [4].
2. Theoretical foundations. Statistical learning. Empirical risk minimization. Approximately correct learning.	2				6		8	10	To solve assigned problems. To read assigned subsections of sections 2, 3, 4, and 6 in [3].
3. Optimization. Main concepts. Convex functions and sets: main properties. Duality. Numerical algorithms: (non-)stochastic gradient descent, theoretical guaranties, applications for ML problems.	5				8		13	12	To solve assigned problems. To read assigned (sub)sections of [5] and [6].
4. General methods. Hyper-parameters, cross validation, regularization, validation and learning curves.	3				2		5	10	To solve assigned problems. To read assigned (sub)sections of [3] and [4].
5. Software. sklearn basics: data preparation; modelling; selection of hyper-parameters.	2				3		5	11	To solve assigned problems. To get acquainted with prescribed subsections and examples of [2].
6. Binary classification: first touch. Logistic classifier, support vector machines, k nearest neighbours, measuring of classification accuracy.	3				3		6	9	To get acquainted with prescribed subsections and examples of [2].
7. Tests and exam.							10	24	To prepare for assessments.
Total							48	77	

Assessment strategy	Weight, %	Deadline	Assessment criteria
Test 1	25	4th study week	The test consists of several practical tasks intended to check the level of knowledge obtained. The total weight of these tasks equals to 2.5 points. The weight of each task ranges from 0.1 to 1 point. Tasks are designed to be solved by making use of computer and appropriate software, or by hand. Each task is evaluated as follows: a) the task is divided into parts and each part is assigned an appropriate amount of points; b) if student accomplishes the part without mistakes, the whole amount of that part is attained; otherwise, the amount is reduced considering the mistakes made; c) the parts are evaluated independently.
Test 2	25	8th study week	
Test 3	25	12th study week	
Test 4	25	16th study week	
Exam		The final examination session	Exam is devoted to increase the grade compiled during the semester. For this, the lecturer may choose one of the following strategies: (a) to assign individual tasks; (b) to allow students to rewrite one of the four tests (chosen individually) taken during the regular semester. In case of worse outcome, the grade remains unchanged.
External exam is not allowed.			

Author	Year	Title	Issue of a	Publishing place and house
--------	------	-------	------------	----------------------------

	of public ation		periodical or volume of a publication	or web link
Compulsory reading				
1. V. Skorniakov	2023	ML lecture notes, part I		Available via Virtual Learning environment
2. scikit-learn developers	2023	Documentation of scikit-learn		https://scikit-learn.org/stable/documentation.html
3. Shai Shalev-Shwartz and Shai Ben-David	2014	Understanding Machine Learning: From Theory to Algorithms		Cambridge University Press (a version very slightly differing from printed is available online)
4. Andreas Lindholm, Niklas Wahlström, Fredrik Lindsten, and Thomas B. Schön	2022	Machine Learning - A First Course for Engineers and Scientists		Cambridge University Press (available online for personal use)
5. Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar	2018	Foundations of Machine Learning		MIT Press, Second Edition, 2018 (available online)
6. Hui Jiang	2021	Machine Learning Fundamentals: A Concise Introduction		York University, Toronto, Cambridge University Press, 2021 (available via virtual VU library ¹)
Optional reading				
1. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani	2013	An Introduction to Statistical Learning: with Applications in R		Springer (available online)
2. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani	2023, summer	An Introduction to Statistical Learning: with Applications in Python		Springer (available online)
3. Trevor Hastie, Robert Tibshirani, Jerome Friedman	2016	The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition		Springer (available online)

¹ to make use of virtual library, one needs to use VPN of Vilnius University