



DALYKO (MODULIO) APRAŠAS

Dalyko (modulio) pavadinimas lietuvių kalba	Dalyko (modulio) pavadinimas anglų kalba	Kodas
Praktinė duomenų analizė su R ir Python	Practical Data Analysis with R and Python	

Dėstytojas (-ai)	Padalinys (-iai)
Koordinuojantis: Andrius Buteikis Kitas (-i):	Matematikos ir informatikos fakultetas Taikomosios matematikos institutas Statistinės analizės katedra

Studijų pakopa	Dalyko (modulio) tipas
Pirmoji	Privalomas

Igyvendinimo forma	Vykdyto laikotarpis	Vykdyto kalba (-os)
Auditorinė	Ketvirtas (pavasario) semestras	Lietuvių/Anglų

Reikalavimai studijuojančiajam	
Išankstiniai reikalavimai: Statistika ir tikimybių teorija	Gretutiniai reikalavimai (jei yra): Mikroekonomika

Dalyko (modulio) apimtis kreditais	Visas studento darbo krūvis	Kontaktinio darbo valandos	Savarankiško darbo valandos
5	140	64	76

Dalyko (modulio) tikslas: studijų programos ugdomos kompetencijos		
<p>Gebėjimas modeliuoti reiškinius matematinėmis, statistinėmis, bei mašininio mokymo priemonėmis. Mokėjimas naudotis specializuota statistine ir mašininio mokymo programine įranga. Gebėjimas savarankiškai analizuoti empirinius duomenis, pasirinkti tinkamus mašininio mokymo metodus ir juos pritaikyti. Gebėjimas analizuoti ir interpretuoti atliktos analizės rezultatus.</p>		
Dalyko (modulio) studijų siekiniai	Studijų metodai	Vertinimo metodai
Susipažinti su prižiūrimo ir neprižiūrimo mašininio mokymo modelių teoriniais ir praktiniais aspektais.	Paskaitos, pratybos naudojant programinę įrangą, savarankiškas užduočių sprendimas ir teorinės medžiagos studijavimas, darbas kompiuterinėje laboratorijoje	Statistinių programų įsisavinimo testavimas, vidurio semestro ir galutinis egzaminai kompiuteriu, tiriamasis projektas.
Gebėti pritaikyti mašininio mokymosi metodus imčių sudarymui, modelių validacijai ir prognozavimui.		
Gebėti analizuoti tekstinę informaciją.		
Gebėti atlikti modelio adekvatumo analizę.		

Temos	Kontaktinio darbo valandos							Visas kontaktinis darbas	Savarankiškų studijų laikas ir užduotys	
	Paskaitos	Konsultacijos	Seminariai	Pratimai	Laboratoriniai darbai	Praktika	E. mokymas(is)		Savarankiškas darbas	Užduotys
1. Statistinių duomenų tipai ir jų modeliai. R ir Python programavimo kalbos, šiuolaikiniai statistinio modeliavimo įrankiai (RStudio, JupyterLab, VSCode, Docker).	2				2			4	6	[1]
2. Tiesinė regresija su Python: modelio funkcinės formos specifikacija, parametrų vertinimas, koeficientų interpretacija, standartizuoti koeficientai, pasikliautiniai intervalai, instrumentiniai kintamieji. Modelio adekvatumo analizė - autokoreliacijos ir heteroskedastiškumo statistiniai testai, multikolinearumo tikrinimas, modelio specifikacijos testai, determinacijos koeficientas ir informaciniai kriterijai .	4				2			6	8	[1] ir [2]
3. Praleisti kintamieji, pervaldinimas, išskirčių nustatymo metodai.	2				1			3	2	[1] ir [2]
4. Klasifikavimas, klasterizavimas ir diskrečiojo atsako modeliai.	6				2			8	10	[1] ir [2]
5. Modelių validacijai skirti imčių metodai: Monte Karlo metodas, kryžminė validacija, saviranka.	3				2			5	6	[1] ir [2]
6. Dimensijos mažinimo metodai ir reguliarizacija.	3				2			5	6	[1] ir [2]
7. Ansamblio metodai: gradiento didinimas, atsitiktinis miškas, savirankos agregavimas, balsavimas, krovimas.	4				2			6	6	[1] ir [2]
8. Gilus mokymasis: neuroniniai tinklai, kaštų funkcija, svorių (parametrų) vertinimas, hiperparametrai, aktyvacijos funkcijos, atgalinio sklidimo algoritmas.	6				3			9	14	[1] ir [2]
9. Natūralios kalbos apdorojimas: tekstinių duomenų paruošimas, analizė ir modeliavimas.	8				4			12	12	[3]
10. Neuroniniais tinklais: MLP (daugiasluksniai perceptronai) ir KAN (Kolmogorov-Arnold tinklai).	2				1			3	3	[1] ir [4]
11. Įvadas į generatyvius (vaizdo, teksto ir garso) dirbtinio intelekto modelius.	2				1			3	3	[1]
Iš viso	42				22			64	76	

Vertinimo strategija	Svoris proc.	Atsiskaitymo laikas	Vertinimo kriterijai
Bendra vertinimo sistema: Teigiamam pažymiui gauti reikia surinkti bent 45 t. (iš maksimalių 100 t.).			
Dalyko laikymas eksternu neleidžiamas.			
Kontrolinis darbas	35	Semestro viduryje	Kontrolinis kompiuteriu iš praeitos medžiagos. Iš viso 10 užduočių. Studentas analizuoja konkrečius duomenis ir pritaiko atitinkamus metodus, statistinius testus, bei aprašo gautus rezultatus.
Tiriamasis projektas	30	Semestro metu	Studentas iki nustatyto laiko turi atlikti ir aprašyti individualiai priskirtą projekcinį darbą. Nepateikus aprašymo iki nurodytos datos, balai neskiriami. Vertinant atsižvelgiama į pasirinktų metodų tinkamumą, atliktos analizės nuoseklumą, parašyto programinio kodo tinkamumą ir išvadų formulavimą.
Sesijinis egzaminas	35	Egzaminų sesijos metu	Egzaminas kompiuteriu. Iš viso 10 užduočių. Studentas analizuoja konkrečius duomenis ir pritaiko atitinkamus metodus, statistinius testus, bei aprašo gautus rezultatus.

Autorius	Leidimo metai	Pavadinimas	Periodinio leidinio Nr. ar leidinio tomas	Leidimo vieta ir leidykla	Prieiga internete ar VU bibliotekoje
Privalomoji literatūra					
1. Buteikis A.		Paskaitų konspektai ir skaidrės			Pateikiamos virtualioje mokymosi aplinkoje
2. James, G., Witten, D., Hastie, T. and Tibshirani, R.	2021	An Introduction to Statistical Learning, 2nd Edition			Prieiga internete: [su R] ir [su Python]
3. Jurafsky D., Martin J. H.	2025	Speech and Language Processing, 3rd Ed.			Prieiga internete: https://web.stanford.edu/~jurafsky/slp3/
Papildoma literatūra					
4. Liu Z., Wang Y., Vaidya S., Ruehle F., Halverson J., Soljačić M., Hou T. Y., Tegmark M.	2024	KAN: Kolmogorov-Arnold Networks			https://arxiv.org/abs/2404.19756 ir https://github.com/KindXiaoming/pykan
5. Abu-Mostafa Y. S., Magdon-Ismail M., Lin H.T.	2012	Learning From Data: A short Course			https://amlbook.com/