



DALYKO (MODULIO) APRAŠAS

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

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

Studijų pakopa	Dalyko (modulio) tipas
Pirmoji	Privalomas

Igyvendinimo forma	Vykdymo laikotarpis	Vykdymo kalba (-os)
Auditorinė	Penktas (rudens) semestras	Lietuvių/Anglų

Reikalavimai studijuojančiajam	
Išankstiniai reikalavimai: Statistikos įvadas, Praktinė duomenų analizė su R ir Python I	Gretutiniai reikalavimai (jei yra): nėra

Dalyko (modulio) apimtis kreditais	Visas studento darbo krūvis	Kontaktinio darbo valandos	Savarankiško darbo valandos
5	135	64	71

Dalyko (modulio) tikslas: studijų programos ugdomos kompetencijos		
Gebėjimas dirbti grupėje ir prisiimti atsakomybę už patikėtą užduotį. Gebėjimas modeliuoti reiškinius matematinėmis ir statistinėmis priemonėmis. Mokėjimas naudotis specializuota programine įranga.		
Dalyko (modulio) studijų siekiniai	Studijų metodai	Vertinimo metodai
Gebėti naudotis šiuolaikiniais duomenų analizės įrankiais (R, Python, Git,...).	Probleminis dėstymas, atvejų analizė, diskusija.	Statistinių programų įsisavinimo testavimas, kursinio darbo aprašymas ir pristatymas.
Gebėti pritaikyti mašininio mokymosi metodus imčių sudarymui, modelių validacijai ir skirtingų modelių prognozių apjungimui.		
Gebėti analizuoti tekstinę informaciją.		

Temos	Kontaktinio darbo valandos							Savarankiškų studijų laikas ir užduotys		
	Paskaitos	Konsultacijos	Seminarai	Pratybos	Laboratoriniai darbai	Praktika	E. mokymas(is)	Visas kontaktinis darbas	Savarankiškas darbas	Studijuojama literatūra
1. Šiuolaikiniai duomenų analizės įrankiai (git, kanban, docker, MLflow, draw.io ir kt.) ir MLOps paradigma.	5				4			9	6	[1]
2. Modelių validacijai skirti imčių metodai: Monte Karlo metodas, kryžminė validacija, saviranka.	4				2			6	4	[1] ir [2]
3. Dimensijos mažinimo metodai ir reguliarizacija.	4				3			7	4	[1] ir [2]
4. Ansamblio metodai: gradiento didinimas, atsitiktinis miškas, savirankos agregavimas, balsavimas, krovimas.	4				2			6	4	[1] ir [2]
5. Neprižiūrimas mokymas: klasterizavimo metodai.	4				3			7	5	[1] ir [2]
6. Natūralios kalbos apdorojimas: tekstinių duomenų analizė	5				2			7	6	[1] ir [3]
7. Natūralios kalbos apdorojimas: Sentimentų analizė.	4				2			6	5	[1] ir [3]
8. Natūralios kalbos apdorojimas: teksto klasifikavimas.	4				2			6	4	[1] ir [3]
9. Natūralios kalbos apdorojimas: teksto generavimas ir dideli kalbos modeliai (LLM).	4				2			6	4	[1] ir [3]
10. LLM: atvirai prieinamų modelių pavyzdžiai.	2				2			4	4	[1]
11. Projektinio darbo aprašymas, pristatymo parengimas ir pasiruošimas egzaminui.									25	
Iš viso	40				24			64	71	

Vertinimo strategija	Svoris proc.	Atsiskaitymo laikas	Vertinimo kriterijai
Semestro metu atliekamas projektinis darbas grupėse. Projektiniame darbe analizuojami duomenys ir taikomi statistiniai ir mašininio mokymosi metodai, kodo versijavimo, projekto eigos dokumentavimo, duomenų ir rezultatų vizualizavimo įrankiai. Aprašoma atlikta analizė, naudoti įrankiai, gauti rezultatai ir atliekamas projektinio darbo pristatymas. Sesijos metu rašomas egzaminas iš sąvokų, metodų, bei programavimo kalbų įvesties ir išvesties interpretavimo. Bendra vertinimo sistema: Teigiamam pažymiui gauti reikia surinkti bent 45 taškus (iš maksimalių 100 t.) Dalyko laikymas eksternu neleidžiamas.			
Projektinio darbo ataskaita	30	Gruodžio mėn. pradžia	Vertinamas taikytų metodų pagrįstumas, naudotų programinių paketų tinkamumas ir gautų rezultatų aprašymas ir pagrįstumas. Projektinio darbo ataskaitos apimtis – iki 20 psl.
Projektinio darbo pristatymas	30	Gruodžio mėn. vidurys	Studentų grupė pristato atliktą projektą. Vertinama: - komandos gebėjimas pristatyti naudotus metodus, sudarytus grafikus, gautus rezultatus ir projekto metu naudotą programinę įrangą, - komandos narių atsakymai į klausimus. Pristatymo apimtis – 15-20 skaidrių, iki 30 min.
Egzaminas	40	Sausio mėn.	Studentas atsako į 5 pateiktus klausimus, kuriuos sudaro 2 atviri ir 3 uždari klausimai. Klausimai susiję su mašininio mokymosi sąvokomis, apibrėžimais, metodų paaiškinimais ir programinio kodo įvesties ir išvesties interpretavimu.

Autorius	Leidimo metai	Pavadinimas	Periodinio leidinio Nr. ar leidinio tomas	Leidimo vieta ir leidykla	Prieiga internete ar VU bibliotekoje
Privalomoji literatūra					
Buteikis A.	2024	Paskaitų konspektai ir skaidrės			
James G., Witten D., Hastie T., Tibshirani R.	2023	An Introduction to Statistical Learning with Application in R, 2nd Ed,		Springer	https://www.statlearning.com/
Jurafsky D., Martin J. H.	2024	Speech and Language Processing, 3rd Ed.			https://web.stanford.edu/~jurafsky/slp3/
Papildoma literatūra					
Visengeriyeva L., Kammer A., Bär I., Kniesz A., Plöd M.	2024	Machine Learning Model Operationalization Management (MLOps)		INNOQ	https://ml-ops.org/
Abu-Mostafa Y. S., Magdon-Ismael M., Lin H.T.	2012	Learning From Data: A short Course			https://amlbook.com/