



## DALYKO (MODULIO) APRAŠAS

Dalyko (modulio) pavadinimas	Kodas
Duomenų tyrimas	7BIODT

Dėstytojas (-ai)	Padalinys (-iai)
<b>Koordinuojantis:</b> Algirdas Mačiulis	Matematinės informatikos katedra Matematikos ir informatikos fakultetas Vilniaus universitetas
<b>Kitas (-i):</b>	

Studijų pakopa	Dalyko (modulio) tipas
Pirmoji	Privalomas

Igyvendinimo forma	Vykdymo laikotarpis	Vykdymo kalba (-os)
Auditorinė	7 semestras	Lietuvių

Reikalavimai studijuojančiajam
<b>Išankstiniai reikalavimai:</b> turi būti išklaustyti pradiniai matematinės analizės, algebros ir tikimybių teorijos kursai; pageidautini programavimo pagrindai

Dalyko (modulio) apimtis kreditais	Visas studento darbo krūvis	Kontaktinio darbo valandos	Savarankiško darbo valandos
5	126	68	58

Dalyko (modulio) tikslas: studijų programos ugdomos kompetencijos		
Siekiami ugdyti gebėjimus tinkamai parinkti ir pritaikyti matematinius duomenų tyrimo modelius, leidžiančius atskleisti objektyviai egzistuojančius dėsningumus įvairios prigimties duomenų aibėse.		
<b>Bendrosios kompetencijos:</b> <ul style="list-style-type: none"><li>Gebėjimas ieškoti duomenų informacijos šaltiniuose, analizuoti, vaizduoti ir sisteminti gautus duomenis. (BK1).</li></ul>		
<b>Dalykinės kompetencijos:</b> <ul style="list-style-type: none"><li>Bioinformatikos duomenų gavybos, vaizdavimo ir analizės (DK11).</li></ul>		
Dalyko (modulio) studijų siekiniai	Studijų metodai	Vertinimo metodai
- Gebės suvokti, skaičiuoti ir analizuoti duomenų aibės atributų entropiją, jų tarpusavio informaciją.	Paskaitos, praktiniai užsiėmimai, savarankiškas literatūros studijavimas, individualios praktinės užduotys.	Savarankiškų užduočių rezultatų ir sprendimų analizė, egzamino klausimų atsakymų peržiūra
- Gebės teisingai formuluoti klasifikavimo ir skaitinės prognozės uždavinius, tinkamai parinkti jų sprendimo metodus.		
- Gebės teisingai formuluoti asociacijų paieškos ir klasterinės analizės uždavinius, tinkamai parinkti jų sprendimo metodus.		
- Suvoks modelio kokybės charakteristikas, gebės jas vertinti ir interpretuoti.		

Temos	Kontaktinio darbo valandos							Savarankiškų studijų laikas ir užduotys	
	Paskaitos	Konsultacijos	Seminarai	Pratybos	Laboratoriniai darbai	Praktika	Visas kontaktinis darbas	Savarankiškas darbas	Užduotys
1. Pagrindinės tikimybių teorijos ir informacijos teorijos sąvokos ir teiginiai.	4				2		6	6	Literatūros studijavimas, uždavinių sprendimas
2. Duomenų aibės ir jų atributai. Pradinė duomenų analizė ir jų transformacijos.	2				4		6	6	Literatūros studijavimas, praktinių užduočių sprendimas, naudojant programų paketą WEKA arba kitą specializuotą programinę įrangą.
3. Klasifikavimo uždavinių sprendimas: sprendimų medžiai, klasifikavimo taisyklės, artimiausių kaimynų metodas, Bajeso klasifikatoriai.	8				8		16	10	
4. Regresijos modeliai	2				2		4	4	
5. Modelio klaidos įverčiai, kryžminis patikrinimas, pakartotinių imčių metodas.	4				4		8	6	
6. Asociacijos taisyklės, pirkėjo krepšelio uždavinys.	6				6		12	8	
7. Asociacijos taisyklių vertinimas	2				2		4	2	
8. Klasterinės analizės metodai, K-vidurkių metodas.	4				4		8	6	
Egzaminas		4					4	10	
Iš viso	32	4			32		68	58	

Vertinimo strategija	Svoris proc.	Atsiskaitymo laikas	Vertinimo kriterijai
Praktinės užduotys	50	Semestro metu	Semestro metu reikės atlikti 5 individualias praktines užduotis. Kiekviena užduotis vertinama 2 balais.
Egzaminas (raštu)	50	Sesijos metu	Egzaminą sudaro teoriniai klausimai ir uždaviniai. Egzaminas vertinamas 10 balų.

Autorius	Leidimo metai	Pavadinimas	Periodinio leidinio Nr. ar leidinio tomas	Leidimo vieta ir leidykla ar internetinė nuoroda
<b>Privalomoji literatūra</b>				
A. Mačiulis	2010	Duomenų tyrimas		<a href="http://www.mif.vu.lt/katedros/matinf/asm/ma/files/dama.pdf">http://www.mif.vu.lt/katedros/matinf/asm/ma/files/dama.pdf</a>
I. Witten, E. Frank	2005, 2011	Data Mining: Practical machine learning tools and techniques		Morgan Kaufmann Publishers
<b>Papildoma literatūra</b>				
Pang-Ning Tan, M. Steinbach, V. Kumar	2005	Introduction to Data Mining		Addison Wesley
T.Hastie, R.Tibshirani, J.Friedman	2001, 2009	The Elements of Statistical Learning: Data Mining, Inference, and Prediction		Springer
V.Čekanavičius, G.Murauskas	2000, 2002	Statistika ir jos taikymai, I,II		Vilnius, TEV