

Dalyko sando aprašas

Dalyko sando kodas <i>(Course unit code)</i>	BDTY2114
Dalyko sando pavadinimas <i>(Course unit title)</i>	Duomenų tyrimas
Dėstytojo (-jų) pedagoginis vardas, vardas ir pavardė <i>(Name and title of lecturer)</i>	Doc. dr. Marius Radavičius Doc. dr. Gediminas Stepanauskas
Katedra, centras	Ekonometrinės analizės katedra Matematinės informatikos katedra
Fakultetas, padalinys	Matematikos ir informatikos fakultetas
Dalyko sando lygis <i>(Level of course)</i>	Pirmosios pakopos (First cycle)
Semestras <i>(Semester)</i>	Rudens (7)
ECTS kreditai <i>(ECTS credits)</i>	4,5
VU kreditai <i>(VU credits)</i>	3
Auditorinės valandos	Viso dalyko 64 Paskaitų 32 Seminarų Laboratorinių darbų, 32 Konsultacijų
Reikalavimai <i>(Prerequisites)</i>	Biometriją, Matematinį modeliavimą, tikimybes ir statistiką, tiesinę algebrą ir geometriją, diskrečias struktūras ir matematinę analizę I, II
Dėstomoji kalba <i>(Language of instruction)</i>	Lietuvių, anglų
Dalyko sando tikslai ir numatomi gebėjimai <i>(Objectives and learning outcomes)</i>	Tikslas: įsisavinti statistinės analizės principus ir išsiaiškinti jų sąsajas su informacinėmis technologijomis. Seminarai bus skirti konkrečių duomenų analizės algoritmų pristatymui: esmė, taikymų sritis ir konkretūs pavyzdžiai, galimybių ribos.
Dalyko sando turinys <i>(Course unit content)</i>	Turinys <i>Duomenys:</i> generavimo mechanizmas, stebėjimo procesas, matavimo skalės, transformacijos. Neapibrėžtumas ir informacija. Realių uždavinių pavyzdžiai. <i>Pradinė tiriamaoji analizė.</i> Empiriniai funkcionalai. Duomenų glodinimo metodai: slenkančio vidurkio, medianinis, branduolinis, artimiausių kaimynų. Projektavimas ir reguliarizacija. Standartinės bazinių funkcijų klasės. <i>Apibendrintas tiesinis modelis.</i> Pavyzdžiai: Gauso ir logistinė regresija.

	<p><i>Sprendimų priėmimas neapibrėžtumo sąlygomis.</i> Nuostoliai ir rizika. Didžiausio tikėtimumo principas ir informacija. Bajeso metodologija. Robastinės procedūros. Modelio parinkimo problema. AIC ir kiti informaciniai kriterijai. Kryžminio patikrinimo bei pakartotinių imčių (“bootstrap”) metodai.</p> <p><i>Informatyvių požymių paieška</i> ir duomenų dimensijos sumažinimo problema. Pagrindinių komponentų metodas, faktorinė analizė. Tikslinis projektavimas.</p> <p><i>Klasifikavimas.</i> Mišinių modeliai ir Bajeso klasifikatorius. Diskriminantinė analizė. Neparаметrinės procedūros, artimiausių kaimynų metodas. Tiesiniai klasifikatoriai ir jų apmokymas. Klasterinė analizė. EM ir k-centrų (prototipų) algoritmai. Hierarchinė klasterizacija.</p> <p><i>Kokybinių požymių analizė.</i> Logtiesiniai ir grafiniai modeliai. Bajeso tinklai.</p> <p>Algoritmai CART, MARS, AdaBoost, SOM, MCMC, sprendimų medžiai (DT) ir kiti.</p>
Pagrindinės literatūros sąrašas (<i>Reading list</i>)	<ol style="list-style-type: none"> 1. T.Hastie, R.Tibshirani, J.Friedman. (2001). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York. 2. M.Berthold, D.J.Hand (1999) Intelligent Data Analysis: An Introduction. Springer, New York.
Papildomos literatūros sąrašas	<ol style="list-style-type: none"> 1. I.H.Witten, E.Franke (2000), Data mining: Practical Machine Learning Tools and Techniques with JAVA Implementations (2nd ed.). San Francisco, CA:Morgan Kaufmann. 2. J.Han, M.Kambre (2000), Data Mining: Concepts and Techniques. San Francisco, CA:Morgan Kaufmann. 3. L.Devroye, L.Gyorfi, and L.Lugosi (1996), A Probabilistic Theory of Pattern Recognition. Springer, New York.
Mokymo metodai (<i>Teaching methods</i>)	Teorinė kurso dalis pateikiama paskaitose, kur formuluojami ir uždaviniai. Uždaviniai sprendžiami laboratorinių darbų metu. Numatomi studentų grupių projektai konkrečioms problemoms nagrinėti.
Lankomumo reikalavimai (<i>Attendance requirements</i>)	Privaloma lankyti ne mažiau kaip 80% užsiėmimų, atsiskaityti už laboratorinius darbus
Atsiskaitymo reikalavimai (<i>Assessment requirements</i>)	Egzaminas. Testas ir atsakymas į klausimus raštu
Vertinimo būdas (<i>Assessment methods</i>)	Galutinį įvertinimą sudaro 70% egzamino pažymio ir 30% laboratorinių darbų įvertinimo.
Aprobuota katedros	2004-08-30
Patvirtinta Studijų programos komiteto	2004-09-14