

DALYKO (MODULIO) APRAŠAS

Dalyko (modulio) pavadinimas	Kodas
DIDELIŲ DUOMENŲ ANALITIKA	

Dėstytojas (-ai)	Padalinys (-iai)
Koordinuojantis: prof. dr. Dalia Krikščiūnienė Kitas (-i):	Kauno fakultetas Socialinių mokslų ir taikomosios informatikos institutas Muitinės g. 8, LT-44280 Kaunas

Studijų pakopa	Dalyko (modulio) tipas
Pirmoji	Privalomasis

Įgyvendinimo forma	Vykdymo laikotarpis	Vykdymo kalba (-os)
Auditorinė	5 semestras	Lietuvių ir anglų

Reikalavimai studijuojančiajam	
Išankstiniai reikalavimai: Statistika, Informacijos sistemos ir duomenų bazės	Gretutiniai reikalavimai (jei yra):

Dalyko (modulio) apimtis kreditais	Visas studento darbo krūvis	Kontaktinio darbo valandos	Savarankiško darbo valandos
5	130	48	82

Dalyko (modulio) tikslas: studijų programos ugdomos kompetencijos		
Dalyko tikslas yra suteikti studentams žinių apie didelių duomenų rinkimą, tvarkymą, vizualizavimą bei analizę, siekiant gauti vertingus rezultatus ir priimti teisingus situacijos valdymo sprendimus.		
Dalyko (modulio) studijų siekiniai	Studijų metodai	Vertinimo metodai
Studentas gebės: <ul style="list-style-type: none"> - taikyti informacines technologijas analizuojant grafinę bei skaitinę informaciją, susijusią su dideliais duomenimis; - įvertinti didelių duomenų naudą priimant sprendimus; - analizuoti tiek struktūrizuotus tiek nestruktūrizuotus duomenis; - taikyti intelektines sprendimo paramos sistemas didelių duomenų analizei; 	Paskaitos; pratybos; savarankiškas darbas; grupinis darbas.	Egzaminas, tarpiniai praktinių darbų atsiskaitymai, koliokviumas.
Gebės naudotis didelių duomenų analizei skirtais populiariausiais kompiuteriniais sprendimais.	Savarankiškas darbas; grupinis darbas.	Tarpiniai praktinių darbų atsiskaitymai

Temos	Kontaktinio darbo valandos						Savarankiškų studijų laikas ir užduotys		
	Paskaitos	Konsultacijos	Egzaminas	Pratybos	Laboratoriniai darbai	Praktika	Visas kontaktinis darbas	Savarankiškas darbas	Užduotys
Įvadinė paskaita. Supažindinimas su kursu, vertinimo kriterijais.	1						1		
Didelių duomenų samprata ir panaudojimo galimybės	2			4			6	10	Mokslinės literatūros analizė, informacijos

Struktūrizuoti ir nestruktūrizuoti duomenys. Didelių duomenų tipai ir grupės. Atsiradimo šaltiniai. Analizė ir taikymai. Paskirstyti skaičiavimai (Distributed computing).									paieška, jos sisteminimas ir apibendrinimas. Praktinių namų darbų atlikimas. [1] p.7-45; [2] p. 3-19; [5] p. 1-24
Didelių duomenų analizės technologijos Didelių duomenų surinkimas. Duomenų servisas, apdorojimo įrankiai ir virtualizavimo galimybės. Debesų technologijos taikymas dideliems duomenims ir Microsoft AZURE įrankiai, Google Cloud ir BIG DATA produktai	2			4			6	10	Mokslinės literatūros analizė, informacijos paieška, jos sisteminimas ir apibendrinimas. Praktinių namų darbų atlikimas. [1] p. 46-81; [2] p. 20-41; [5] p. 25-63
Didelių duomenų valdymas Reliacinės ir nereliacinės duomenų bazės. MapReduce pagrindai. Hadoop samprata. Hadoop paskirstytoms failų sistemoms. Hadoop MapReduce. Didelių duomenų sandėliai (warehouse). Hortonworks Sandbox technologijos panaudojimas didelių duomenų tvarkymui ir analizei.	2			4			6	10	Mokslinės literatūros analizė, informacijos paieška, jos sisteminimas ir apibendrinimas. Praktinių namų darbų atlikimas. Darbas su Hortonworks Sandbox [1] p.182-139; [5] p. 63-115; [3] p. 44-59
Didelių duomenų analitika Didelių duomenų panaudojimas sprendimų priėmimui. Intelektiniai analizės metodai, rezultatų modeliavimas, vizualizavimas, elgesio modeliavimas, anomalijų prognozavimas. Didelių duomenų analizės pavyzdžiai: Orbitz, Nokia, NASA. Didelių duomenų analitikos įrankiai Tableau, SAP analytic cloud, Google Big Query, ir Google Data Studio aplinkose. Socialinių tinklų analitika.	2			8			10	14	Mokslinės literatūros analizė, informacijos paieška, jos sisteminimas ir apibendrinimas. Praktinių namų darbų atlikimas. Darbas su analizės įrankiais [1] p. 140-179; [5] p. 117-137; [4] p. 197-221
Didelių duomenų sprendimų diegimas Identifikavimas ir atrinkimas tinkamų analizei duomenų. Rezultatų tikslumo reikalavimai. Srautiniai duomenys. Sudėtingų procesų valdymas. Sprendimų diegimo plano sudarymas. Duomenų patikimumo, prasmingumo ir laikinumo (validity, veracity, volatility) reikalavimai. Sprendimų diegimo kelias su SAP HANA.	2			4			6	10	Mokslinės literatūros analizė, informacijos paieška, jos sisteminimas ir apibendrinimas. Praktinių namų darbų atlikimas. Darbas su SAP HANA [1] p.179-233; [2] p. 41-55
Realūs didelių duomenų verslo sprendimų pavyzdžiai	2			8			10	14	Mokslinės literatūros analizė, informacijos

Didelių duomenų gavimas panaudojant sensorius. Realaus laiko duomenys. Didelių duomenų taikymai viešosios nuomonės analizei, aplinkos apsaugai, sveikatos apsaugos ir energetikos sektoriuose. Nestruktūrizuotos informacijos analizės pavyzdžiai. Verslo problemų sprendimų pavyzdžiai su Google Big Query įrankiu ir SAP analytic Cloud.									paieška, jos sisteminimas ir apibendrinimas. Praktinių namų darbų atlikimas. Darbas su Google Big Query ir SAP analytic Cloud. [1] p.235-263; [2] p. 55-83
Konsultacija		2						2	
Savarankiškas žinių sisteminimas ir ruošimasis egzaminui								14	Pasirošimas egzaminui (paskaitų konspektų analizė, mokslinės literatūros analizė).
Egzaminas	1							1	
Iš viso	14	2		32				48	82

Vertinimo strategija	Svoris proc.	Atsiskaitymo laikas	Vertinimo kriterijai
Du teoriniai/praktiniai testai (koliokviumai)	20%	Kas 6 savaites	Testą sudaro iki 20 įvairių teorinių klausimų. Reikia parinkti teisingą iš kelių pateiktų variantų. Vertinama 1-10 pažymių vertinimo skalėje: 10-9: Puikios žinios ir gebėjimai. Vertinimo lygmuo. 90-100 % teisingų atsakymų. 8-7: Geros žinios ir gebėjimai, gali būti neesminių klaidų. Sintezės lygmuo. 70-89 % teisingų atsakymų. 6-5: Vidutinės žinios ir gebėjimai, yra klaidų. Analizės lygmuo. 50-69 % teisingų atsakymų. 4-3: Žinios ir gebėjimai nesiekia vidutinių, yra (esminių) klaidų. Žinių taikymo lygmuo. 20-49 % teisingų atsakymų. 2-1: Netenkinami minimalūs reikalavimai. 0-19 % teisingų atsakymų.
Du praktiniai darbai ginami semestro metu (gali būti atliekami grupėje po 2 studentus)	30%	Nustatytu laiku	Praktiniai darbai suskirstyti į dvi dalis pagal teorines temas. Pirmas darbas - Didelių duomenų rinkimas, tvarkymas ir jų vizualizavimas. Antras- <i>Didelių duomenų analizė</i> . Vertinama 1-10 pažymių vertinimo skalėje: 10-9: Puikios žinios ir gebėjimai. Vertinimo lygmuo. 90-100 % teisingų atsakymų. 8-7: Geros žinios ir gebėjimai, gali būti neesminių klaidų. Sintezės lygmuo. 70-89 % teisingų atsakymų. 6-5: Vidutinės žinios ir gebėjimai, yra klaidų. Analizės lygmuo. 50-69 % teisingų atsakymų. 4-3: Žinios ir gebėjimai nesiekia vidutinių, yra (esminių) klaidų. Žinių taikymo lygmuo. 20-49 % teisingų atsakymų. 2-1: Netenkinami minimalūs reikalavimai. 0-19 % teisingų atsakymų.
Egzaminas	50 %	Nustatytu sesijos laiku	Egzaminas susideda iš teorinės ir praktinės dalies. Teorinės dalies egzamine yra pateikiamas 1 klausimas iš paskaitų metu išdėstytų temų. Praktinė dalis vyksta naudojant SAP HANA programinį paketą. Studentas turi pademonstruoti savo įgūdžius sprendžiant konkretų didelių duomenų tvarkymo ir analizės uždavinį.

			<p>Vertinama 1-10 pažymių vertinimo skalėje:</p> <p>10-9: Puikios žinios ir gebėjimai. Vertinimo lygmuo. 90-100 % teisingų atsakymų.</p> <p>8-7: Geros žinios ir gebėjimai, gali būti neesminių klaidų. Sintezės lygmuo. 70-89 % teisingų atsakymų.</p> <p>6-5: Vidutinės žinios ir gebėjimai, yra klaidų. Analizės lygmuo. 50-69 % teisingų atsakymų.</p> <p>4-3: Žinios ir gebėjimai nesiekia vidutinių, yra (esminių) klaidų. Žinių taikymo lygmuo. 20-49 % teisingų atsakymų.</p> <p>2-1: Netenkinami minimalūs reikalavimai. 0-19 % teisingų atsakymų.</p>
--	--	--	---

Autorius	Leidimo metai	Pavadinimas	Periodinio leidinio Nr. ar leidinio tomas	Leidimo vieta ir leidykla ar internetinė nuoroda
Privalomoji literatūra				
Judith Hurwitz, Alan Nugent, Dr. Fern Halper, and Marcia Kaufman	2013	Big Data For Dummies®		John Wiley & Sons, Inc.
O'Reilly Media	2012	Big Data Now: 2012 Edition		O'Reilly Media, Inc.
Vignesh Prajapati	2013	Big Data Analytics with R and Hadoop		Packt Publishing Ltd.
Soumendra Mohanty, Madhu Jagadeesh, Harsha Srivatsa.	2013	Big Data Imperatives, (Enterprise Big Data Warehouse, BI Implementations and Analytics)		APress
EMC Education Services	2015	Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data		John Wiley & Sons, Inc., Indianapolis, Indiana
Papildoma literatūra				
Francesco Corea	2017	Big Data Analytics: A Management Perspective		Springer International Publishing AG
Fujitsu	2016	Big Data. The definitive guide to the revolution in business analytics		Fujitsu Services Ltd.

