

RTU studiju kurss "Datu apstrādes un datizraces pamati"

33000 Datorzinātnes, informācijas tehnoloģijas un enerģētikas fakultāte

Vispārējā informācija

Kods	DMI738
Nosaukums	Datu apstrādes un datizraces pamati
Studiju kursa statuss programmā	Obligātais/Ierobežotās izvēles
Atbildīgais mācītspēks	Sergejs Paršutins - Doktors, Asociētais profesors
Mācītspēks	Arnis Kiršners - Doktors, Docents, Laboratorijas darbi Inese Poļaka - Doktors, Asociētais profesors, Lekciju lasīšana, konsultācijas Darja Plinere - Doktors, Lektors, Laboratorijas darbi
Apjoms daļās un kredītpunktos	1 daļa, 4.5 kredītpunkti
Studiju kursa īstenošanas valodas	LV, EN
Anotācija	Datu apstrāde un datizrace nodrošina metožu un tehnoloģiju kopumu, kas ļauj pārvērst uzņēmumā pieejamos datus noderīgā informācijā un zināšanās. Datu apstrāde risina datu izgūšanas, datu transformēšanas un datu strukturēšanas problēmas. Tās rezultātā dati tiek sagatavoti tālākai apstrādei. Datizrace risina datu analīzes problēmas, lai atklātu iepriekš nezināmas attiecības starp datiem. Datizraces rezultāti ļauj uzņēmumiem un sistēmām pieņemt pareizus un pamatotus lēmumus. Kursā tiek aptverta datu apstrāde un analīze, sākot no datu izgūšanas no datu avotiem līdz zināšanu iegūšanai un to prezentācijai lietotājam saprotamā formā. Kursa tēmas ietver datu glabāšanas veidu un tehnoloģiju apskatu, datu analīzes metodes, ietverot gan klasiskās statistikas, gan modernas mašīnāpmācības un datizraces metodes.
Mērķis un uzdevumi, izteikti kompetencēs un prasmēs	Sniegt pamata teorētiskās un praktiskās zināšanas datu savākšanas, uzglabāšanas, apstrādes, analīzes un prezentācijas metožu un algoritmu izmantošanā dažādās problēmsfērās un nozarēs.
Patstāvīgais darbs, tā organizācija un uzdevumi	Patstāvīgais darbs sastāv no praktiskiem uzdevumiem. Praktiskie uzdevumi ietver šādas tēmas: 1. Datu izgūšana, strukturēšana un sagatavošana analīzei; 2. Datu statistiskā analīze; 3. Klasifikācijas modeļu izveidošana; 4. Klasterizācijas modeļu izveidošana; 5. Mākslīgo neironu tīklu pielietošana prognozēšanas uzdevumos; 6. Datu analīzes rezultātu pasniegšana.
Literatūra	1. Datu ieguve: Pamati. / A.Sukovs, L.Aleksejeva, K.Makejeva, A.Borisovs. – Rīga: RTU, SIA „Drukātava”, 2007. – 130 lpp. ISBN: 978-9984-798-19-6. 2. Datu ieguve: Programmlīdzekļi. / S.Paršutins, A.Sukovs, G.Kolešova, A.Borisovs, L.Aleksejeva. – Rīga: RTU, SIA „Latgales Druka”, 2008. – 111 lpp. ISBN: 978-9984-798-68-4 3. Krishnan, K. Data Warehousing in the Age of Big Data - Morgan Kaufmann, 2013. – 370 p. ISBN: 978-0124058910. 4. Liu, H., Motoda, H. (Eds). Computational Methods of Feature Selection. – Chapman & Hall/CRC, 2008. – 419 p. ISBN: 978-1-58488-878-9. 5. Maimon, O., Rokach, L. (Eds.). Data Mining and Knowledge Discovery Handbook. – Springer, 2010, 2nd edition. – 1305 p. ISBN: 978-0-387-09822-7. 6. Gan, G., Ma, C., Wu, J. Data Clustering: Theory, Algorithms and Applications. – ASA SIAM, 2007. – 466 p. ISBN: 978-0-898716-23-8. 7. Turley, P. Professional Microsoft SQL Server 2016 Reporting Services and Mobile Dashboards – Wrox, 2016. – 528 p. ISBN: 978-1119258353
Nepieciešamās priekšzināšanas	Matemātika, matemātiskā statistika, datu bāzu vadības sistēmas, programmēšana bāzes līmenī

Studiju kursa saturs

Saturs	Pilna un nepilna laika klātienes studijas		Nepilna laika neklātienes studijas	
	Kontakt stundas	Patstāv. darbs	Kontakt stundas	Patstāv. darbs
Ievads priekšmetā un darba organizācija	2	0	0	0
Datu izgūšana un strukturēšana	4	4	0	0
Datu transformēšana un datu kvalitāte	4	4	0	0
Datu sagatavošana analīzei	4	4	0	0
Datu statistiskā analīze	4	4	0	0
Lineārās regresijas pamati	8	8	0	0
Klasifikācijas algoritmi, rezultātu interpretācija	8	8	0	0
Klasterizācijas algoritmi, rezultātu interpretācija	8	8	0	0
Ievads mākslīgajos neironu tīklos	4	4	0	0
Prognozēšana ar mākslīgajiem neironu tīkliem	6	8	0	0
Datu analīzes rezultātu pasniegšana	8	8	0	0
Kopā:	60	60	0	0

Sasniedzamie studiju rezultāti un to vērtēšana

Sasniedzamie studiju rezultāti	Rezultātu vērtēšanas metodes
Spēj izgūt datus no datu avotiem un strukturēt tos dažādos datu glabāšanas veidos	Teorētiskā daļa: eksāmens; Praktiskā daļa: darbs par tēmu „Datu izgūšana, strukturēšana un sagatavošana analīzei”, 1. daļa.
Spēj realizēt datu apstrādes darbplūsmas, transformēt datus, novērtēt datu kvalitāti un sagatavot datus analīzei	Teorētiskā daļa: eksāmens; Praktiskā daļa: darbs par tēmu „Datu izgūšana, strukturēšana un sagatavošana analīzei”, 2. daļa.
Spēj atklāt sakarības datus ar klasiskajām statistiskās analīzes metodēm (regresija, korelācija, X2 tests)	Teorētiskā daļa: eksāmens; Praktiskā daļa: darbs par tēmu „Datu statistiskā analīze”
Spēj izveidot datu klasifikācijas modeli un pielietot to jaunu ierakstu klasifikācijai	Teorētiskā daļa: eksāmens; Praktiskā daļa: darbs par tēmu „Klasifikācijas uzdevums un algoritmi”
Spēj veikt datu analīzi, pielietojot datu klasterizācijas algoritmus	Teorētiskā daļa: eksāmens; Praktiskā daļa: darbs par tēmu „Klasterizācijas uzdevums un algoritmi”
Spēj pielietot mākslīgus neironu tīklus prognozēšanas uzdevumu risināšanai	Teorētiskā daļa: eksāmens; Praktiskā daļa: darbs par tēmu "Prognozēšana ar mākslīgajiem neironu tīkliem"
Spēj pasniegt un vizualizēt datu apstrādes un analīzes rezultātus	Teorētiskā daļa: eksāmens; Praktiskā daļa: darbs par tēmu "Datu analīzes rezultātu pasniegšana"

Studiju rezultātu vērtēšanas kritēriji

Kritērijs	% no kopējā vērtējuma
Laboratorijas darbu atskaite	30
Gala pārbaudījums teorētisko un praktisko zināšanu novērtēšanai	50
Eksāmens	20
Kopā:	100

Studiju kursa plānojums

Daļa	KP	Stundas			Pārbaudījumi		
		Lekcijas	Prakt d.	Laborat	Ieskaite	Eksām.	Darbs
1.	4.5	2.0	1.0	0.0		*	