

RTU studiju kurss "Ievads datu ieguvē"

33000 Datorzinātnes, informācijas tehnoloģijas un enerģētikas fakultāte

Vispārējā informācija

Kods	DID308
Nosaukums	Ievads datu ieguvē
Studiju kursa statuss programmā	Obligātais/Ierobežotās izvēles
Atbildīgais mācībspēks	Sergejs Paršutins - Doktors, Asociētais profesors
Mācībspēks	Arnis Kiršners - Doktors, Docents
Apjoms daļās un kredītpunktos	1 daļa, 3.0 kredītpunkti
Studiju kursa īstenošanas valodas	LV
Anotācija	Zināšanu iegūšana no datiem. Klasifikācijas, klasterizācijas un asociatīvās analīzes pamati. Datu sagatavošana analīzei. Datu līdzības un atšķirības mēri. Uz piemēriem balstīta apmācība. Induktīvā apmācība. Baijesa „Naivais” klasifikators. K-tuvāko kaimiņu klasifikators. Klasifikācijas lēmumu koki. Klasifikācijas koka apgriešanas metodes. Klasifikatoru precizitātes mērīšana. Šķēršvalidācija. Klasifikācijas likumu inducēšana. Hierarhiskā klasterizācija. Sadalošā klasterizācija. Klasterizācijas rezultātu novērtēšana. Asociatīvo likumu meklēšana. Apriori algoritms.
Mērķis un uzdevumi, izteikti kompetencēs un prasmēs	Sniegt pamatzināšanas par tādām intelektuālās datu analīzes tehnoloģijām, kā „Datu ieguve” un „Zināšanu atklāšana no datiem”. Iegūt praktiskas iemaņas zināšanu (noderīgas informācijas) iegūšanai no datiem, pielietojot klasifikācijas, klasterizācijas un asociatīvās analīzes metodes. Veicināt studenta spējas un kompetences veikt analizējamiem datiem nepieciešamo pirmsapstrādi un izvēlēties datu ieguves metodes, kas piemērotas konkrētam uzdevumam.
Patstāvīgais darbs, tā organizācija un uzdevumi	Studentu patstāvīgais darbs izpaužas šādās aktivitātēs: praktisko darbu teorētisko pamatojumu sagatavošana, eksperimentāli iegūto rezultātu apkopošana un analīze.
Literatūra	1. Sukovs A., Aleksejeva L., Makejeva K., Borisovs A. (2007). Datu ieguve: Pamati. Rīga, 130 lpp. 2. Paršutins S., Sukovs A., Kuļešova G., Borisovs A., Aleksejeva L. (2008). Datu ieguve: Programmlīdzekļi. Rīga, 111 lpp. 3. Tan P.N., Steinbach M., Kumar V. (2006). Introduction to Data Mining. Addison Wesley, 769 p. 4. Gan G., Ma C., Wu J. Data Clustering: Theory, Algorithms and Applications. (2007). ASA and SIAM, 466 p. 5. Quinlan J.R. (1993, 2006). C4.5 – Programms for Machine Learning. Morgan Kaufmann, 302 p. 6. Zhang C., Zhang S. (2002). Association Rule Mining: Models and Algorithms. Springer, 238 p.
Nepieciešamās priekšzināšanas	Matemātika, datu bāzes

Studiju kursa saturs

Saturs	Pilna un nepilna laika klātienēs studijas		Nepilna laika neklātienēs studijas	
	Kontakt stundas	Patstāv. darbs	Kontakt stundas	Patstāv. darbs
Datu ieguves tehnoloģiju apskats	2	0	0	0
Dati kā analīzes objekts (datu tipi; datu objektu līdzība un atšķirība)	2	0	0	0
Klasifikācijas pamati	2	0	0	0
Uz piemēriem balstīta apmācība	4	0	0	0
Klasifikācijas koki	4	0	0	0
Klasterizācijas pamati	2	0	0	0
Hierarhiskā klasterizācija	2	0	0	0
Sadalošā klasterizācija	2	0	0	0
Datu asociatīvā analīze	4	0	0	0
Praktiskās nodarbības	8	0	0	0
Kopā:	32	0	0	0

Sasniedzamie studiju rezultāti un to vērtēšana

Sasniedzamie studiju rezultāti	Rezultātu vērtēšanas metodes
Students spēj sagatavot datus analīzei, pielietojot datu pirmsapstrādes pieejas.	Datus nepieciešams sagatavot analīzei katrā no patstāvīgajiem darbiem. Ceturtajā darbā ir nepieciešams veikt datu kopas ar reāliem datiem apstrādi.
Students spēj atlasīt un apstrādāt nepieciešamo teorētisko materiālu.	Katrā no patstāvīgajiem darbiem ir nepieciešams sagatavot teorētisko daļu par izmantotajām metodēm.

Students spēj izpildīt nepieciešamos aprēķinus un pielietot „Uz piemēriem balstītas apmācības” metodes.	Pirmajā patstāvīgajā darbā students veic sagatavoto datu klasifikāciju ar „Uz piemēriem balstītas apmācības” metodēm.
Students spēj izpildīt nepieciešamos aprēķinus, pielietot klasifikācijas koku konstruēšanas metodes un savstarpēji salīdzināt klasifikācijas metodes.	klasifikācijas kokiem. Apkopojot rezultātus, veic analīzi, salīdzinot pirmā un otrā darba klasifikācijas metodes pēc to precizitātes.
Students spēj izpildīt nepieciešamos aprēķinus, pielietot hierarhiskas un sadalošas klasterizācijas metodes un savstarpēji tās salīdzināt.	Trešajā patstāvīgajā darbā students veic sagatavoto datu klasterizāciju ar hierarhiskas un sadalošas klasterizācijas metodēm. Apkopojot rezultātus, veic metožu salīdzināšanu pēc to precizitātes.

Studiju kursa plānojums

Daļa	KP	Stundas			Pārbaudījumi		
		Lekcijas	Prakt d.	Laborat	Ieskaite	Eksām.	Darbs
1.	3.0	1.0	0.0	1.0	*		