

RTU studiju kurss "Datizrace un zināšanu atklāšana"**33000 Datorzinātnes, informācijas tehnoloģijas un enerģētikas fakultāte****Vispārējā informācija**

Kods	DE0642
Nosaukums	Datizrace un zināšanu atklāšana
Studiju kursa statuss programmā	Obligātais/Ierobežotās izvēles
Atbildīgais mācītspēks	Sergejs Paršutins - Doktors, Asociētais profesors
Mācītspēks	Henrihs Gorskis - Doktors, Docents, Kursa vadīšana Kiberdrošības programmā
Apjoms daļās un kredītpunktos	1 daļa, 6.0 kredītpunkti
Studiju kursa īstenošanas valodas	LV, EN
Anotācija	Zināšanu atklāšana ir process, kurā noderīga informācija (zināšanas) tiek automātiski iegūta no datiem. Šo procesu var arī definēt kā zināšanu iegūšanu no lieliem datu apjomiem. Datizrace ir zināšanu atklāšanas procesa algoritmiskā daļa, posms, kurā tiek pielietoti algoritmi jau sagatavoto datu analīzei. Datizrace un zināšanu atklāšana ir starpdisciplināra tehnoloģija lielo datu apjomu apstrādei, kas apvieno tradicionālās datu analīzes metodes ar moderniem algoritmiem no tādiem zinātniskiem virzieniem kā statistika, datu bāzu tehnoloģijas, mašīnāpmācība, tēlu atpazīšana, mākslīgie neironu tīkli, nestingrā skaitļošana (Soft Computing) u.c. Tas paver plašas iespējas jaunu, kā arī jau esošo datu tipu pētīšanai un analīzei, izmantojot jaunas metodes. Kursā ietvaros tiek apskatītas datu pirmāpstrādes tehnoloģijas; klasifikācijas un klasteru analīzes metožu un algoritmu pielietošana zināšanu atklāšanai; īsu laika rindu un to raksturojošo parametru analīzes un apstrādes pieejas; izplūduši logika un zināšanu atklāšana izplūdušajā vidē; uz mākslīgiem neironu tīkliem balstītas datu analīzes metodes un pieejas. Kā arī tiek apskatītas datizraces un zināšanu atklāšanas praktiskās pielietošanas jomas
Mērķis un uzdevumi, izteikti kompetencēs un prasmēs	Sniegt padziļinātas teorētiskās un praktiskās zināšanas datizraces un zināšanu atklāšanas metožu un algoritmu izmantošanā dažādās jomās
Patstāvīgais darbs, tā organizācija un uzdevumi	Katrai izskatītai tēmai ir paredzēta prakse. Tā satur kopējo uzdevumu, kuru studenti pildīs kopā ar pasniedzēju, un patstāvīgo uzdevumu, kuru katrs students pildīs individuāli. Katrai praksei ir jā sagatavo atskaite par izpildītiem uzdevumiem.
Literatūra	Obligātā literatūra \ Obligatory resources. Wes McKinney. Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter. O'Reilly Media, 2022 Aurelien Geron. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 3rd Edition. O'Reilly Media, 2022 Chip Huyen. Designing Machine Learning Systems. O'Reilly Media, 2022 Francois Chollet. Deep Learning with Python, Second Edition. Manning, 2021. François Voron . Building Data Science Applications with FastAPI: Develop, manage, and deploy efficient machine learning applications with Python. https://github.com/PacktPublishing/Building-Data-Science-Applications-with-FastAPI . Packt Publishing, 2021. Tyler Richards. Getting Started with Streamlit for Data Science. https://github.com/PacktPublishing/Getting-started-with-Streamlit-for-Data-Science . Packt Publishing, 2021 Papildu literatūra \ Additional resources. Mark Treveil, Nicolas Omont, Clément Stenac, Kenji Lefevre, Du Phan, Joachim Zentici, Adrien Lavoillotte, Makoto Miyazak. Introducing MLOps. O'Reilly Media, 2020 Natu Lauchande. Machine Learning Engineering with MLflow. Packt Publishing, 2021 Jake VanderPlas. Python Data Science Handbook O'Reilly Media, 2016. https://jakevdp.github.io/PythonDataScienceHandbook/ Aston Zhang, Zack C. Lipton, Mu Li, Alex J. Smola. Dive into Deep Learning. https://d2l.ai/index.html . Interactive e-book MLFlow dokumentācija \ MLFlow documentation. https://mlflow.org/docs/latest/index.html . Elektronisks resurss \ E-resource Tensorflow. https://www.tensorflow.org/ . Elektronisks resurss / E-resource PyTorch. https://pytorch.org/ . Elektronisks resurss / E-resource Scikit-learn. https://scikit-learn.org/stable/ . Elektronisks resurss / E-resource FastAPI. https://fastapi.tiangolo.com/ . Elektronisks resurss / E-resource Streamlit. https://docs.streamlit.io/ . Elektronisks resurss / E-resource
Nepieciešamās priekšzināšanas	Matemātika, objektorientētā programmēšana, Python programmēšana pamata līmenī

Studiju kursa saturs

Saturš	Pilna un nepilna laika klātienē studijas		Nepilna laika neklātienē studijas	
	Kontakt stundas	Patstāv. darbs	Kontakt stundas	Patstāv. darbs
Ievads priekšmetā	2	0	0	0
Datu priekšapstrādes tehnoloģijas. Datu sagatavošana analīzei	8	10	0	0
Klasifikācijas modeļi un tehnoloģijas zināšanu atklāšanai	8	10	0	0
Klasteru analīzes modeļi un tehnoloģijas zināšanu atklāšanai	8	10	0	0
Īsu laika rindu un to raksturojošo parametru analīzes metodes	6	10	0	0

Zināšanu atklāšana, pielietojot mākslīgo neironu tīklu modeļus un tehnoloģijas	10	20	0	0
Ievads izplūdušajā loģikā. Zināšanu atklāšana izplūdušajā vidē	8	8	0	0
Datizraces modeļu kombinēšana	10	20	0	0
Datu asociatīvā analīze	4	8	0	0
Kopā:	64	96	0	0

Sasniedzamie studiju rezultāti un to vērtēšana

Sasniedzamie studiju rezultāti	Rezultātu vērtēšanas metodes
Spēj definēt datu pirmapstrādes soļus, izvēlēties metodes un pielietot tās datu pirmapstrādei	Teorētiskā daļa: eksāmens; Praktiskā daļa: darbs par tēmu „Datu sagatavošana analīzei”
Spēj izveidot un pielietot datu klasifikācijas un klasteru analīzes modeļus zināšanu atklāšanai	Teorētiskā daļa: eksāmens; Praktiskā daļa: darbs par tēmu „Klasifikācijas un klasteru analīzes modeļi”
Spēj analizēt īsas laika rindas un to raksturojošos parametrus ar datizraces un zināšanu atklāšanas metodēm un algoritmiem	Teorētiskā daļa: eksāmens; Praktiskā daļa: darbs par tēmu „Datizraces metodes īsu laika rindu un to raksturojošo parametru analīzei”
Spēj izveidot, apmācīt un pielietot mākslīgo neironu tīklu datizraces un zināšanu atklāšanas uzdevumu risināšanai	Teorētiskā daļa: eksāmens; Praktiskā daļa: darbs par tēmu „Datu analīze ar mākslīgiem neironu tīkliem”
Spēj integrēt izplūdušo loģiku ar datizraces metodēm un algoritmiem un pielietot to klasifikācijas, klasteru analīzes un citu datizraces un zināšanu atklāšanas uzdevumu risināšanai	Teorētiskā daļa: eksāmens; Praktiskā daļa: darbs par tēmu „Klasifikācija un klasteru analīze izplūdušajā vidē”

Studiju rezultātu vērtēšanas kritēriji

Kritērijs	% no kopējā vērtējuma
Teorētiskās zināšanas tiek vērtētas ar testveida eksāmenu	30
Prasmes pielietot teorētiskās zināšanas praksē tiks vērtētas ar patstāvīgiem praktiskiem darbiem. Kopējā vērtējumā tiks ņemta patstāvīgo darbu vidējā atzīme	70
Kopā:	100

Studiju kursa plānojums

Daļa	KP	Stundas			Pārbaudījumi		
		Lekcijas	Prakt d.	Laborat	Ieskaite	Eksām.	Darbs
1.	6.0	32.0	0.0	32.0		*	