

RTU studiju kurss "Mašīnmācīšanās tekstuālo datu apstrādei"

33000 Datorzinātnes, informācijas tehnoloģijas un enerģētikas fakultāte

Vispārējā informācija

Kods	DE1023
Nosaukums	Mašīnmācīšanās tekstuālo datu apstrādei
Studiju kursa statuss programmā	Brīvās izvēles
Atbildīgais mācībspēks	Zane Seņko - Docents (praktiskais)
Mācībspēks	Tatjana Smirnova - Doktors, Asociētais profesors Marina Platonova - Doktors, Profesors Tatjana Menise - Habilitētais doktors, Docents Tatjana Kelebeka - Doktors, Asociētais profesors Oksana Ivanova - Doktors, Docents (praktiskais) Sintija Petroviča-Kļaviņa - Doktors, Lektors
Apjoms daļās un kredītpunktos	1 daļa, 3.0 kredītpunkti
Studiju kursa īstenošanas valodas	EN
Anotācija	Studiju kurss piedāvā pirmā cikla augstākās izglītības bakalaura studiju programmu studējošajiem iespēju attīstīt zināšanas, kompetenci un prasmes dažādu mašīnmācīšanās rīku lietošanā un to pielāgošanā tekstuālo datu apstrādei, lai risinātu virkni kontekstuzdevumu nozarē un zinātnē ieskaitot korpusa un tekstuālo datu analīzi, datu pirmāpstrādi un attēlojumu, sentimenta analīzi un mašīntulkšanas rīku lietošanu. Studējošajiem jāattīsta vispusīga izpratne par mūsdienu multimodālā digitālā teksta būtību, cita starpā ņemot vērā tekstuālo datu vākšanas, apstrādes un attēlošanas ētiskos, drošības un ilgtspējības aspektus. Studējošie gūs pieredzi datu ieguves un tekstizraces pieeju, datu strukturēšanas un datu vizualizācijas metožu praktiskā pielietošanā. Viņi spēs pārbaudīt, segmentēt un atkārtoti izmantot tekstuālo datu analīzes rezultātus, pielietojot atbilstošas mašīnmācīšanās metodes, kā arī attīstīs prasmes kvalitatīvo un kvantitatīvo datu analīzes metožu lietošanā. Pēc studiju kursa pabeigšanas studējošie spēs risināt ierobežotas definīcijas problēmas, demonstrējot digitālo kompetenci un prasmes, kas atbilst vismaz 5.-6. līmenim saskaņā ar DIGICOMP 2.2 (t.sk., dabiskās valodas apstrāde (DVA), dabiskās valodas ģenerēšana (DVĢ), uzraudzīta un neuzraudzīta mašīnmācīšanās tekstuālo datu apstrādei, mašīntulkšanas lietojumprogrammas, progresīvas teksta apstrādes lietojumprogrammas).
Mērķis un uzdevumi, izteikti kompetencēs un prasmēs	Studiju kursa mērķi: - attīstīt studējošo izpratni par automatizētas tekstuālo datu apstrādes iespējām, ierobežojumiem, lietojumprogrammām un to iespējamiem paplašinājumiem; - padziļināt studējošo zināšanas par mašīnmācīšanās rīku piedāvātajām iespējām dažādu tekstuālo datu apstrādes uzdevumu risināšanā; - nodrošināt studējošajiem iespēju apgūt un attīstīt nozarē pieprasītas prasmes tekstuālo datu apstrādē; Studiju kursa uzdevumi: - attīstīt studējošo prasmes tekstuālo datu apstrādē, strukturēšanā un attēlošanā, ieskaitot datu un datu analīzes rezultātu vizualizāciju; - attīstīt studējošo iemaņas, lai meistarīgi lietotu kvantitatīvas un kvalitatīvas tekstuālo datu apstrādes metodes nozares kontekstuzdevumu efektīvai risināšanai; - iepazīstināt studējošos ar DVA, DVĢ, sentimentu analīzes un digitālo stāstījumu veidošanas principiem; - iepazīstināt studējošos ar mašīnmācīšanās lietojumprogrammu paplašinājumiem tekstuālo datu apstrādē (runas atpazīšana, teksta rezumēšana, sarunvalodas mākslīgais intelekts, sarunboti u.c.).
Patstāvīgais darbs, tā organizācija un uzdevumi	Studējošie pēta un testē brīvpieejas un komerciālās mašīnmācīšanās programmatūras un rīku funkcionalitāti tekstuālo datu apstrādē, veic tekstuālo datu pirmāpstrādes, strukturēšanas, attēlošanas un vizualizācijas uzdevumus, profesionāli pilnveidojot savu kā valodas pakalpojumu sniedzēju iemaņu klāstu.

Literatūra	<p>Obligātā. / Obligatory: Tsourakis N. Machine Learning Techniques for Text: Apply modern techniques with Python for text processing, dimensionality reduction, classification, and evaluation. Packt Publishing, 2022. Online Machine Learning : a practical guide with examples in Python /Eva Bartz, Thomas Bartz-Beielstein, editors., xiii, 155 lpp. : ilustrācijas ; 24 cm.</p> <p>Rowse, Samuel. Python For Data Science : what every Python developer needs to know, step by step, learn the basics of machine learning and Big Data analysis : (2022 guide for beginners) /Samuel Rowse., 133 lpp. : ilustrācijas ; 23 cm</p> <p>George A. Python Text Mining: Perform Text Processing, Word Embedding, Text Classification and Machine Translation. BPB Publications, 2022. Aggarwal C.C. Machine Learning for Text (2nd ed.). Springer, 2022.</p> <p>Papildu. / Additional: Bell, Jason. Machine Learning : hands-on for developers and technical professionals /Jason Bell., xxxi, 400 lpp. : ilustrācijas ; 24 cm</p> <p>Stevenson, Halbert. Python for data science : a crash course for data science and analysis, Python machine learning and big data /[Halbert Stevenson]., 112, [1] lpp. : ilustrācijas ; 26 cm</p> <p>Zong C., Xia R., Zhang J. Text Data Mining (1st ed.). Springer, 2022 Žižka J., Dařena F., Svoboda A. Text Mining with Machine Learning: Principles and Techniques. CRC Press, 2019</p>
Nepieciešamās priekšzināšanas	Angļu valodas zināšanas B2 līmenī, saskaņā ar Eiropas kopīgajām pamatnostādņēm valodu apguvei, programmēšanas pamatprasmes.

Studiju kursa saturs

Saturs	Pilna un nepilna laika klātienē studijas		Nepilna laika neklātienē studijas	
	Kontakt stundas	Patstāv. darbs	Kontakt stundas	Patstāv. darbs
Teksta būtība.	4	4	0	0
Teksts kā datu kopa: Tekstuālo datu būtība, organizācija un principi.	4	4	0	0
Tekstuālo datu analīzes ierobežojumi.	3	4	0	0
Teksta pirmapstrāde un attēlošana.	4	4	0	0
Tekstuālo datu analīzes kvantitatīvās metodes.	3	4	0	0
Neuzraudzīta mašīnmācīšanās tekstuālo datu apstrādei.	4	4	0	0
Tekstuālo datu analīzes kvalitatīvās metodes.	4	4	0	0
Uzraudzīta mašīnmācīšanās tekstuālo datu apstrādei.	4	4	0	0
Transformatoru modeļi teksta apstrādē.	4	4	0	0
Progresīvas teksta apstrādes lietojumprogrammas.	4	4	0	0
Pārbaudes darbi.	2	0	0	0
Kopā:	40	40	0	0

Sasniedzamie studiju rezultāti un to vērtēšana

Sasniedzamie studiju rezultāti	Rezultātu vērtēšanas metodes
Spēj apzināties automatizētas tekstuālo datu apstrādes iespējas, ierobežojumus, lietojumprogrammas un to iespējamās paplašinājumus.	Praktiskie uzdevumi, pārbaudes darbi. Studējošie spēj izvēlēties, pielāgot un izmantot tekstuālo datu apstrādes rīkus ar mašīnmācīšanās funkcionalitāti kontekstuzdevumu risināšanai.
Spēj apstrādāt, strukturēt un attēlot tekstuālus datus dažādu nozares un pētniecības kontekstuzdevumu risināšanai.	Praktiskie uzdevumi, pārbaudes darbi, eksāmens. Studējošie prasmīgi izmanto teksta pirmapstrādes, vārdšķiras marķēšanas, celmošanas un lemmatizācijas metodes un attēlo tekstuālos datus atbilstošā formātā (vektORIZĀCIJA, vienizcēluma kodējums, Bag-of-Words, N-grammas, TF-IDF).
Spēj meistarīgi pielietot kvantitatīvās un kvalitatīvās tekstuālo datu apstrādes metodes kontekstuzdevumu efektīvai risināšanai nozarē.	Praktiskie uzdevumi, pārbaudes darbi, eksāmens. Studējošie demonstrē spēju piemērot uz biežumu balstītu pieeju, veikt tematu modelēšanu, viedokļu apkopošanu un informācijas meklēšanu, kā arī citus saistītus uzdevumus.

Pārzina DVA, DVĢ, sentimentu analīzes un digitālo stāstījumu veidošanas pamatprincipus un spēj tos izmantot, lai risinātu uzdevumus ar ierobežotu definīciju.	Praktiskie uzdevumi, pārbaudes darbi, eksāmens. Studējošie demonstrē spēju veikt tematisko analīzi, satura analīzi un stāstījuma analīzi un pielietot stāstījuma veidošanas paņēmienus.
Izprot mašīnmācīšanās lietojumprogrammu iespējamās paplašinājumus tekstuālo datu apstrādē un iespējamās scenārijus turpmākai attiecīgo prasmju un iemaņu pilnveidošanai.	Praktiskie uzdevumi, pārbaudes darbi. Studējošie demonstrē prasmi izmantot progresīvus mašīnmācīšanās rīkus (runas atpazīšana, teksta rezumēšana, sarunvalodas mākslīgais intelekts, sarunboti u.c.).

Studiju rezultātu vērtēšanas kritēriji

Kritērijs	% no kopējā vērtējuma
Praktiskie uzdevumi	40
Pārbaudes darbi	30
Eksāmens	30
Kopā:	100

Studiju kursa plānojums

Daļa	KP	Stundas			Pārbaudījumi			Brīvās izvēles pārbaudījumi		
		Lekcijas	Prakt d.	Laborat	Ieskaite	Eksām.	Darbs	Ieskaite	Eksām.	Darbs
1.	3.0	20.0	20.0	0.0		*			*	