

RTU studiju kurss "Mašīnmācīšanās tekstuālo datu apstrādei"**33000 Datorzinātnes, informācijas tehnoloģijas un enerģētikas fakultāte****Vispārējā informācija**

Kods	DE1086
Nosaukums	Mašīnmācīšanās tekstuālo datu apstrādei
Studiju kursa statuss programmā	Obligātais/Ierobežotās izvēles; Brīvās izvēles
Atbildīgais mācītspēks	Oksana Ivanova - Doktors, Docents (praktiskais)
Mācītspēks	Sintija Petroviča-Kļaviņa - Doktors, Docents
Apjoms daļās un kredītpunktos	1 daļa, 3.0 kredītpunkti
Studiju kursa īstenošanas valodas	LV, EN
Anotācija	Studiju kurss sniedz ieskatu jēdzienos, izaicinājumos un iespējās, kas saistītas ar tekstuālo datu apstrādi, izmantojot mašīnmācīšanās rīkus. Studiju kursa laikā studējošie attīsta kompetenci lietot un pielāgot mašīnmācīšanās metodes tekstuālo datu apstrādei. Studējošie apgūst, kā validēt, segmentēt un atkārtoti izmantot tekstuālo datu analīzes rezultātus, izmantojot piemērotas mašīnmācīšanās metodes. Studiju kursa pamataaktivitātes ietver teksta/korpusa pirmapstrādi un attēlojumu, sentimenta analīzi, tekstizraci, vizualizāciju, kā arī ievadprocesus mašīntulkošanā, dabiskās valodas ģenerēšanā (DVG) un dabiskās valodas apstrādē (DVA). Studējošie strādā ar mūsdienīgiem digitāliem tekstiem (tostarp multimodālos kontekstos) un risina ētikas, drošības un ilgtspējas jautājumus datu vākšanā, apstrādē un attēlošanā.
Mērķis un uzdevumi, izteikti kompetencēs un prasmēs	Studiju kursa mērķi: - veidot studējošo konceptuālu izpratni par galvenajiem jēdzieniem, izaicinājumiem un iespējām tekstuālo datu apstrādē, izmantojot mašīnmācīšanās metodes; - attīstīt studējošo praktisku kompetenci izvēlēties, pielietot un pielāgot mašīnmācīšanās metodes digitālajiem tekstiem (tostarp multimodālos kontekstos); - veicināt studējošo atbildīgu praksi, visā darba plūsmā integrējot validāciju, atkārtotu izmantojamību, kā arī ētikas, drošības un ilgtspējas apsvērumus. Studiju kursa uzdevumi: - veidot studējošo kompetenci tekstuālo datu apstrādē un strukturēšanā, ar uzsvaru uz datu attēlošanu un rezultātu vizualizāciju; - eksperimentēt ar plašu mašīnmācīšanās metožu klāstu un pielāgot tās problēmas kontekstam; - nodrošināt studējošajiem pamatzināšanas sentimenta analīzē, tekstizracē, kā arī ievadprocesos mašīntulkošanā/DVG/DVA; - iepazīstināt studējošos ar progresīvām un jaunradītām mašīnmācīšanās lietojumprogrammām tekstuālo datu apstrādē.
Patstāvīgais darbs, tā organizācija un uzdevumi	Studējošie izvērtē gan brīvpieejas, gan komerciālu komerciālās mašīnmācīšanās programmatūras tekstuālo datu apstrādei un veic pirmapstrādes, strukturēšanas, attēlošanas un vizualizācijas darba plūsmas. Studējošie eksperimentē ar uzraudzītu un neuzraudzītu mašīnmācīšanos (piem., sentimenta analīze, klasifikācija, klasterizācija) un pēta ievadprocesus DVA/DVG/mašīntulkošanas darba plūsmās.
Literatūra	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 Aggarwal, C.C. Machine Learning for Text. 2022. Springer Cham. https://doi.org/10.1007/978-3-030-96623-2 Sharma, N.A., Ali, A.B.M.S., & Kabir, M.A. A Review of Sentiment Analysis: Tasks, Applications, and Deep Learning Techniques. International Journal of Data Science and Analytics, 2024. https://doi.org/10.1007/s41060-024-00594-x George, A. Python Text Mining: Perform Text Processing, Word Embedding, Text Classification and Machine Translation. BPB Publications, 2022 Liu, B. Sentiment Analysis. Mining Opinions, Sentiments and Emotions. Cambridge: CUP, 2015 Papildu. // Additional: Albrecht, J., Ramachandran, S., & Winkler, C. Blueprints for Text Analytics Using Python: Machine Learning-Based Solutions for Common Real World (NLP) Applications. O'Reilly Media, 2021 Li, Q., Peng, H., Li, J., Xia, C., Yang, R., Sun, L., ... & He, L. A Survey on Text Classification: From Traditional to Deep Learning. ACM Transactions on Intelligent Systems and Technology, 13 (2), 1–41, 2022. https://dl.acm.org/doi/full/10.1145/3495162 Žižka, J., Dařena, F., & Svoboda, A. Text Mining with Machine Learning: Principles and Techniques. CRC Press, 2019
Nepieciešamās priekšzināšanas	Angļu valodas zināšanas B2 līmenī, saskaņā ar Eiropas kopīgajām pamatnostādnēm valodu apguvei, ja studiju kurss tiek apgūts angļu valodā.

Studiju kursa saturs

Saturs	Pilna un nepilna laika klātienēs studijas		Nepilna laika neklātienēs studijas	
	Kontakt stundas	Patstāv. darbs	Kontakt stundas	Patstāv. darbs
Ievads tekstuālo datu jomā.	3	4	0	0
Teksts kā dati un ētikas normas (tekstu tipi, korpusi, metadati, licencēšana, privātums, neobjektivitāte un atkārtojamība).	2	4	0	0

Teksta pirmapstrādes paņēmieni.	3	5	0	0
Dabiskās valodas apstrādes izaicinājumi.	4	6	0	0
Kvantitatīvās pieejas tekstuālo datu analīzei.	3	5	0	0
Neuzraudzīta mašīnmācīšanās: metodes tekstuālo datu apstrādei.	3	4	0	0
Kvalitatīvās pieejas tekstuālo datu analīzei.	3	4	0	0
Uzraudzīta mašīnmācīšanās: no iezīmēm līdz transformatoriem.	3	4	0	0
Transformatori un lielie valodas modeļi.	3	4	0	0
Progresīvas teksta apstrādes lietojumprogrammas: multimodāla dabiskās valodas apstrāde.	3	5	0	0
Pārbaudes darbi.	2	4	0	0
Kopā:	32	49	0	0

Sasniedzamie studiju rezultāti un to vērtēšana

Sasniedzamie studiju rezultāti	Rezultātu vērtēšanas metodes
Spēj apliecināt izpratni par tekstuālo datu apstrādes izaicinājumiem un iespējām, izmantojot mašīnmācīšanās rīkus.	Praktiskie uzdevumi, pārbaudes darbi. Tiek novērtēta studējošo spēja apstrādāt tekstuālus datus, izmantot mašīnmācīšanās rīkus.
Spēj veikt teksta korpusu apstrādi un analīzi (tīrīšanu, tokenizāciju, normalizāciju, marķēšanu/anotēšanu).	Praktiskie uzdevumi, pārbaudes darbi, eksāmens. Tiek novērtēta studējošo spēja apstrādāt un analizēt teksta korpusu.
Spēj apliecināt izpratni par uzraudzītas un neuzraudzītas mašīnmācīšanās principiem.	Praktiskie uzdevumi, eksāmens. Tiek novērtēta studējošo izpratne par dažādiem mašīnmācīšanās principiem.
Spēj pielietot DVA, DVĢ un sentimenta analīzes pamatprincipus, lai risinātu dažādus ar datu analītiku saistītus uzdevumus.	Praktiskie uzdevumi, eksāmens. Tiek novērtēta studējošo spēja risināt dažādus ar datu analītiku saistītus kontekstuzdevumus.
Spēj apliecināt izpratni par progresīviem mašīnmācīšanās pielietojumiem tekstuālo datu apstrādē.	Praktiskie uzdevumi, pārbaudes darbi. Tiek novērtēta studējošo spēja izmantot jaunākus mašīnmācīšanās rīkus tekstuālo datu apstrādē.

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	16.0	16.0	0.0		*			*	