

RTU studiju kurss "Digitālā sentimentu analīze"

33000 Datorzinātnes, informācijas tehnoloģijas un enerģētikas fakultāte

Vispārējā informācija

Kods	ETH729
Nosaukums	Digitālā sentimentu analīze
Studiju kursa statuss programmā	Brīvās izvēles
Atbildīgais mācībspēks	Tatjana Kelebeka - Doktors, Asociētais profesors
Mācībspēks	Tatjana Smirnova - Doktors, Asociētais profesors Zane Seņko - Docents (praktiskais) Oksana Ivanova - Doktors, Docents (praktiskais) Alīna Vagele-Kricina - Lektors Tatjana Menise - Habilitētais doktors, Docents Marina Platonova - Doktors, Profesors
Apjoms daļās un kredītpunktos	1 daļa, 3.0 kredītpunkti
Studiju kursa īstenošanas valodas	EN
Anotācija	<p>Studiju kurss galvenokārt ir vērsts uz padziļinātu un augsti specializētu zināšanu līmeņa kompetenču un prasmi attīstīšanu studentiem, kuri apgūst gan humanitāro zinātņu, gan starpdisciplināro STEM+ balstīto, gan informācijas tehnoloģiju studiju programmas. Studiju kurss paredzēts augstākā līmeņa studiju studentiem ar pamatzināšanām dabiskās valodas apstrādē, kuri vēlas pilnveidot savas kompetences noskaņojuma analīzē un tekstuālu datu apstrādē dažādu ar nozari saistītu lietišķo uzdevumu risināšanai. Studenti mācīsies klasificēt nestrukturētus un daļēji strukturētus datus, lai noteiktu noskaņojuma polaritāti (piemēram, pozitīvs, negatīvs vai neitrāls), izmantojot bezmaksas un/vai maksas rīkus, ko viņi pielāgos savām pētniecības, mācību un profesionālajām vajadzībām. Studenti mācīsies izgūt un atlasīt ar noskaņojumu saistītus raksturojumus un kontekstuālās iezīmes, lietojot atbilstošus modeļus, un novērtēt ar emocijām saistīto vārdu nozīmi tekstā un kopējo analizētā teksta noskaņojumu. Studenti attīstīs prasmes izgūt teksta semantisko informāciju, to analizēt, klasificēt un novērtēt noskaņojuma kontekstā, lai uzlabotu klientu pieredzi, kā arī kvalitātes nodrošināšanas nolīkos. Studenti apgūs runas marķēšanu, lietvārdu salikņu iegūšanu, emociju identificēšanu un noskaņojuma analīzi, pievērsīsies tādiem jēdzieniem kā polaritāte, nodomi un subjektivitāte, praktiski strādājot ar Python programmēšanas valodu un tajā pieejamām bibliotēkām noskaņojuma analīzei, piemēram, NLTK un TextBlob. Praktisko uzdevumu ietvaros studenti realizēs leksikonā balstītu noskaņojuma analīzi (piemēram, izmantojot VADER (Valence Aware Dictionary and sEntiment Reasoner) leksikonu), izmantos iepriekš apmācītus modeļus dažādiem klasifikācijas uzdevumiem, ieskaitot noskaņojuma identificēšanu (piemēram, RoBERTa modeli vai citus pieejamos transformera tipa modeļus, kas pieejami apkopotā veidā Hugging Face platformā) un citus risinājumus, piemēram, Matplotlib bibliotēku rezultātu vizualizācijai un novērtēšanai.</p> <p>Nemot vērā noskaņojuma analīzes pievienoto vērtību mērķtieku datu izpētē, e-komercijā, sabiedriskās domas izpētē, komercizpētē, politikas plānošanā, izglītības efektivitātes novērtēšanā un citās jomās, zināšanas, kompetences un prasmes, ko studenti apgūs studiju kursā, veicinās viņu profesionālās spējas un nodarbinātību un palīdzēs risināt aktuālas nozares kontekstuzdevumus.</p>
Mērķis un uzdevumi, izteikti kompetencēs un prasmēs	<p>Studiju kursa mērķi:</p> <ul style="list-style-type: none"> - attīstīt studentu zināšanas par noskaņojuma analīzes pamatjēdzieniem, piemēram, polaritāti, nodomiem un subjektivitāti, bi-gramiem, tri-gramiem un n-gramiem; - attīstīt studentu kompetenci normalizēt un attēlot tekstu, izmantojot dažādas metodes (Bag-of-Words, TF-IDF u.c. metodes); - stiprināt studentu kompetenci izstrādāt, pielāgot un testēt apmācītus klasifikācijas modeļus noskaņojuma analīzei, izmantojot atbilstošu programmatūru. <p>Studiju kursa uzdevumi:</p> <ul style="list-style-type: none"> - padziļināt studentu prasmes veikt datu priekšapstrādi un izpētīti noskaņojuma analīzei; - attīstīt studentu prasmes praktiski veikt teksta tokenizāciju, lematizāciju, datu attīrīšanu; - attīstīt studentu prasmes izgūt un analizēt n-gramus; - attīstīt studentu prasmes sasaitēt vārdu vektorus ar noskaņojumu klasifikācijas modeļiem; - attīstīt studentu prasmes līdzdarboties, lai piedāvātu un radītu risinājumus starpdisciplināriem, dažāda sarežģītības līmeņa, ar valodu tehnoloģiju un ierobežotu definīciju saistītiem semiotikas izaicinājumiem, kas vērsti uz pārdomātu cilvēkiem un videi risku minimizējošu digitālo tehnoloģiju lietošanu, tostarp to lietošanu sociālai labklājībai un sociālajai integrācijai; - attīstīt studentu izpratni par noskaņojuma analīzes veikšanas ētiskajiem aspektiem un tās ilgtspējīgu un ētisku piemērošanu.
Patstāvīgais darbs, tā organizācija un uzdevumi	Studenti iepazīstas ar plašu teorētiskās literatūras klāstu, gūst praktisku pieredzi noskaņojuma analīzē, lietojot bezmaksas un/vai maksas rīkus, modeļus un bibliotēkas, īsteno situācijas izpēti par noskaņojuma analīzes pielietojumu specifiskam mērķim.

Literatūra	<p>Obligātā / Obligatory : Brink, Henrik. Real-world machine learning / Henrik Brink, Joseph W. Richards, Mark Fetherolf., xxii, 242 lpp. : diagrammas, ilustrācijas, tabulas ; 24 cm.</p> <p>Farhadloo, Mohsen & Rolland, Erik. Fundamentals of Sentiment Analysis and Its Applications. 2016. http://dx.doi.org/10.1007/978-3-319-30319-2_1 Bing Liu . Sentiment Analysis: Mining Opinions, Sentiments, and Emotions CUP, 2020. Dhawale, Chitra & Chaudhari, Vandana. Sentiment Analysis Techniques, Tools, Applications, and Challenge 2020. DOI:10.4018/978-1-5225-8575-6.ch003 Papildu. // Additional : Sentiment Analysis for Social Media. Eds. Antonio Moreno, Carlos A. Iglesias MDPI AG, 2020. Brand, C.O, Acerbi, A. and Mesoudi, A. Cultural evolution of emotional expression in 50 years of song lyrics SocArXiv : 2019. Špats, Gatis,. Opinion Mining in Latvian Text Using Semantic Polarity Analysis and Machine Learning Approach / Gatis Špats, Ilze Birzniece</p>
Nepieciešamās priekšzināšanas	Pamatzināšanas Python programmēšanas valodā un dabiskās valodas apstrādē; DigComp 3.-4. Līmenī atbilstoši Eiropas Iedzīvotāju digitālās kompetences ietvaram, angļu valodas zināšanas vismaz B2 līmenī saskaņā ar EKPVA

Studiju kursa saturs

Saturs	Pilna un nepilna laika klātienes studijas		Nepilna laika neklātienes studijas	
	Kontakt stundas	Patstāv. darbs	Kontakt stundas	Patstāv. darbs
Ievads noskaņojuma analīzē.	3	4	0	0
Noskaņojums un emocijas.	3	4	0	0
Noskaņojuma analīze mārketinga kampaņās.	3	5	0	0
Noskaņojuma analīze kultūras diskursā.	3	5	0	0
Teksta priekšapstrāde noskaņojuma analīzei.	3	4	0	0
Noskaņojuma analīzes process: metodes un procedūras.	3	5	0	0
Konteksta nozīme noskaņojuma analīzē.	3	4	0	0
Nozares specifiskās valodas nozīme noskaņojuma analīzē.	3	4	0	0
Noskaņojuma leksikoni un vārdnīcas.	3	4	0	0
Iepriekš apmācīti valodu modeļi noskaņojuma analīzei.	3	5	0	0
Ētiskie apsvērumi noskaņojuma analīzē.	2	4	0	0
Kopā:	32	48	0	0

Sasniedzamie studiju rezultāti un to vērtēšana

Sasniedzamie studiju rezultāti	Rezultātu vērtēšanas metodes
Studenti spēj izstrādāt, pielāgot un testēt apmācītos modeļus noskaņojuma analīzei, lietojot atbilstošu programmatūru.	Metodes: praktiski uzdevumi, situācijas izpēte, eksāmens. Studenti demonstrē atbilstošu kompetences līmeni dabiskās valodas apstrādes jomā.
Studenti iegūst zināšanas par noskaņojuma analīzes pamatjēdzieniem, piemēram, polaritāti, nodomiem un subjektivitāti, bi-gramiem, tri-gramiem un n-gramiem.	Metodes: praktiski uzdevumi, situācijas izpēte, eksāmens. Studenti lieto atbilstošu terminoloģiju noskaņojuma analīzes uzdevumu izvirzīšanai un īstenošanai.
Studenti spēj demonstrēt zināšanas testa normalizēšanā un attēlošanā, izmantojot dažādas metodes (piemēram, Bag-of-Words, TF-IDF u.c. metodes).	Metodes: praktiski uzdevumi, situācijas izpēte, eksāmens. Studenti sekmīgi veic praktiskus uzdevumus teksta normalizācijā un attēlošanā ar dažādām metodēm.
Studenti spēj veikt teksta priekšapstrādi un izpēti noskaņojuma analīzei, izgūt n-gramus, kā arī sasaisīt izgūtos vārdu vektorus ar noskaņojuma klasifikāciju.	Metodes: praktiski uzdevumi, situācijas izpēte, eksāmens. Studenti sekmīgi veic praktiskus noskaņojuma analīzes uzdevumus, sasniedzot noteiktos izpildes rādītājus.
Studenti spēj līdzdarboties, lai piedāvātu un radītu risinājumus starpdisciplināriem, dažāda sarežģītības līmeņa, ar valodu tehnoloģiju un ierobežotu definīciju saistītiem noskaņojuma analīzes izaicinājumiem, kas vērsti uz pārdomātu cilvēkiem un videi risku minimizējošu digitālo tehnoloģiju lietošanu, tostarp to lietošanu sociāli labklājībai.	Metodes: praktiski uzdevumi, situācijas izpēte. Studenti demonstrē augsta līmeņa digitālās, sociālās un komunikācijas prasmes, risinot praktiskus kontekstuzdevumus un izaicinošas problēmas nozarē.

Studiju rezultātu vērtēšanas kritēriji

Kritērijs	% no kopējā vērtējuma
Praktiski uzdevumi	20
Situācijas izpēte	40
Eksāmens	40
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	1.0	1.0	0.0		*				