

RTU studiju kurss "Ievads datu analīzē ar mašīnmācīšanos"**33000 Datorzinātnes, informācijas tehnoloģijas un enerģētikas fakultāte****Vispārējā informācija**

Kods	DE1069
Nosaukums	Ievads datu analīzē ar mašīnmācīšanos
Studiju kursa statuss programmā	Obligātais/Ierobežotās izvēles
Atbildīgais mācītspēks	Ilze Birzniece - Doktors, Docents
Apjoms daļās un kredītpunktos	1 daļa, 6.0 kredītpunkti
Studiju kursa īstenošanas valodas	LV, EN
Anotācija	Datu apjoms pasaulē pieaug ik dienu, un šajos datos slēpjas potenciāla vērtība. Jaunu sakarību izpēti un atklāšana dātos mūsdienās ir būtiska zināšanu attīstības sastāvdaļa dažādās zinātnes nozarēs. Datu izpēti, vizualizācija, statistiskā analīze un informācijas paneļi veido datu analīzes pamatu. Zināšanu atklāšana dātos ir process, kas ietver datu izgūšanu, datu priekšapstrādi, atbilstošu analīzes metožu izvēli un pielietošanu, kā arī rezultātu interpretēšanu. Datizraze (data mining) ir statistikas un mašīnmācīšanos metožu lietojums vēsturiskajiem vai eksperimentālajiem datiem ar mērķi iegūt skaidrojumus, atklāt likumsakarības vai veidot prognozes. Studiju kursā tiek apskatītas galvenās datizrazes pieejas pārraudzītajā un nepārraudzītajā apmācībā – regresija, klasifikācija, klasterēšana – iepazīstoties ar populārākajām metodēm katrā no tām. Kā aktuāla joma tiek aplūkots teksta datu (nestrukturētu un daļēji strukturētu) analīze, kā arī ģeneratīvā mākslīgā intelekta rīku izmantošana datu analītikā. Studiju kursa uzsvārs tiek likts uz analītiskās izpratnes veidošanu un praktisku darbošanos, ar bezkoda rīkiem un Python programmēšanas valodu veicot datu analīzi reālām datu kopām un interpretējot iegūtās sakarības. Studenti strādā komandās un pielieto apgūtās zināšanas un prasmes datu analīzē, izstrādājot integrētu kursa projektu. Studiju kursa apguvei nav nepieciešamas iepriekšējās zināšanas datizracē vai programmēšanā.
Mērķis un uzdevumi, izteikti kompetencēs un prasmēs	Studiju kursa mērķis ir attīstīt studējošo izpratni par datu analītiskās iespējām un prasmes izvēlēties un lietot atbilstošas pieejas konkrētām datu apstrādes vajadzībām. Studiju kursa uzdevumi: - iepazīstināt studējošos ar datu analīzes vajadzībām un iespējām dažādās jomās; - veidot izpratni par datu ieguves un apstrādes procesiem, lai iegūtu dātos balstītas zināšanas; - attīstīt prasmes darbā ar datizrazes metodēm un rīkiem biznesa lēmumu atbalstam; - veicināt analītiskās spējas, kritisko domāšanu un akadēmiskās rakstīšanas prasmes.
Patstāvīgais darbs, tā organizācija un uzdevumi	Patstāvīgajos darbos studenti, ievērojot akadēmiskā godīguma principus, veic gan individuālus, gan grupu darbus, apgūstot studiju kursa vielu, praktiski vingrinot kontaktstundās apgūtās prasmes un padziļināti izpētot atsevišķus tematus. Patstāvīgais darbs iekļauj mājasdarbus – praktisko darbu izpildi, ar studiju kursa tematiku saistītās literatūras analīzi, projekta darba izstrādi, kā arī gatavošanos rezultātu demonstrēšanai prezentācijās un eksāmenā.
Literatūra	Obligātā / Obligatory: 1. Data Mining: Practical Machine Learning Tools and Techniques, 5th ed., Ian H. Witten, Eibe Frank, Mark A. Hall, Christopher J. Pal, James R. Foulds, Morgan Kaufmann, 2026 2. Data Mining: Concepts and Techniques, 4th ed., Jiawei Han, Micheline Kamber, Jian Pei, Morgan Kaufmann, 2022 (RTU bibliotēkā ir 2. izdevums - Data Mining: Concepts and Techniques, 2nd Ed., Jiawei Han, Micheline Kamber, Jian Pei, Morgan Kaufmann, 2006) Papildu / Additional: 1. Data Mining and Machine Learning: Fundamental Concepts and Algorithms, 2nd Ed., Mohammed J. Zaki and Wagner Meira, Jr, 2020 (online book: https://dataminingbook.info/book_html/) 2. Machine Learning for Business Analytics: Concepts, Techniques, and Applications in Python, 2nd Ed., Galit Shmueli, Peter C. Bruce, Peter Gedeck, Nitin R. Patel, Wiley, 2025 3. Ethics of Data and Analytics: Concepts and Cases, Kirsten Martin, Auerbach Publishers (CRC Press), 2022 4. Business Analytics: Data Analysis and Decision Making, 7th Ed., S. Christian Albright, Wayne L. Winston, 2020 5. Business analytics: combining data, analysis & judgement to inform decisions, Mary Ellen Gordon. London: SAGE Publications Ltd, 2023. 6. Python Data Analytics: With Pandas, NumPy, and Matplotlib, 3rd Ed., Fabio Nelli, Apress, 2023 (RTU bibliotēkā ir 2. izdevums Python Data Analytics: With Pandas, NumPy, and Matplotlib, 2nd Ed., Fabio Nelli, Apress, 2018) 7. Taming Text: How to Find, Organize, and Manipulate It, 1st ed., Gregory S. Ingersoll, Thomas S. Morton, Andrew L. Farris, Manning Publications, 2013 8. Data Analysis with Open Source Tools, Philipp K. Janert, O'Reilly Media, 2011 9. Principles of Data Mining, 4th ed., Max Bramer, Springer, 2020 (online book: https://content.e-bookshelf.de/media/reading/L-13831469-195d030d66.pdf)
Nepieciešamās priekšzināšanas	Pamatzināšanas par datu glabāšanu un apstrādi ar lietojumprogrammatūru, statistikas pamati.

Studiju kursa saturs

Saturs	Pilna un nepilna laika klātienes studijas	Nepilna laika neklātienes studijas
--------	---	------------------------------------

	Kontakt stundas	Patstāv. darbs	Kontakt stundas	Patstāv. darbs
Ievads datu analīzē	4	4	0	0
Datu apstrāde un izpēte: datu tipi, priekšapstrāde, datu tīrīšana un normalizēšana	8	8	0	0
Statistika eksperimentu datu analīzei: aprakstošā statistika, secinošā statistika (hipotēžu pārbaude ar parametriskajiem un neparametriskajiem testiem), korelāciju analīze, netipisku vērtību noteikšana	16	16	0	0
Datu atspoguļošana un vizualizācija	4	4	0	0
Datizrace un sakarību atklāšana datos	4	4	0	0
Regresijas analīze	4	4	0	0
Klasifikācija, izmantojot lēmumu kokus un likumus, gadījuma mežus un mākslīgos neironu tīklus	4	4	0	0
Prognožu novērtēšana, metrikas	4	4	0	0
Klasterēšana, izmantojot K-vidējo algoritmu, hierarhiskās pieejas un DBSCAN algoritmu	4	4	0	0
Tekstizrace un nestrukturētu datu analīze izmantojot vektoru telpas modeli ar TF-IDF un lielos valodu modeļus	4	4	0	0
Datu analīzes rīki – bezkoda un koda	2	2	0	0
Datu un to analīzes ētiskie aspekti	4	4	0	0
Datu analītikas projekta izstrāde un prezentēšana	16	16	0	0
Kopā:	78	78	0	0

Sasniedzamie studiju rezultāti un to vērtēšana

Sasniedzamie studiju rezultāti	Rezultātu vērtēšanas metodes
Pārzina datu priekšapstrādes uzdevumus un atbilstošās statistiskās metodes.	Vērtēšanas metodes: projekts, eksāmens. Kritēriji: spēj identificēt datu priekšapstrādes uzdevumus un izvēlēties atbilstošās statistiskās metodes konkrētos gadījumos.
Spēj raksturot galvenās datizraces pieejas, prot izvēlēties un pielietot atbilstošās metodes konkrētiem datiem.	Vērtēšanas metodes: izpildīti praktiskie darbi, projekts, eksāmens. Kritēriji: spēj raksturot un atšķirt galvenās datizraces pieejas, izvēlēlas un prot pielietot metodes atbilstoši konkrētiem datiem.
Spēj analizēt jomas vajadzības un sasaistīt tās ar datu analītikas iespējām.	Vērtēšanas metodes: projekts, eksāmens. Kritēriji: spēj demonstrēt izpratni par dažādu uzdevumu vajadzībām saistībā ar datu analītikas pielietošanu.
Izmantojot zināšanas datu analīzē un prasmes lietot datizraces rīkus, spēj radīt risinājumus zināšanu iegūšanai no datiem un to reprezentēšanai.	Vērtēšanas metodes: izpildīti praktiskie darbi, projekts. Kritēriji: spēj izstrādāt praktiski pielietojamu datorizētu risinājumu datu vizualizācijai, priekšapstrādei un mašīnmācīšanās metožu izmantošanai.
Prot veikt dažāda tipa uzdevumus, ievērojot akadēmiskā godīguma principus un jomas ētiku.	Vērtēšanas metodes: izpildīti praktiskie darbi, projekts, eksāmens. Kritēriji: uzdotos uzdevumus veic, nepārkāpjot akadēmiskā godīguma principus un lēmumu pieņemšanā izmanto profesionālo ētiku.

Studiju rezultātu vērtēšanas kritēriji

Kritērijs	% no kopējā vērtējuma
Praktiskie darbi	40
Projekta darbs	30
Eksāmens	30
Kopā:	100

Studiju kursa plānojums

Daļa	KP	Stundas			Pārbaudījumi		
		Lekcijas	Prakt d.	Laborat	Ieskaite	Eksām.	Darbs
1.	6.0	39.0	39.0	0.0		*	