

RTU studiju kurss "Datu vizualizācijas pamati"

33000 Datorzinātnes, informācijas tehnoloģijas un enerģētikas fakultāte

Vispārējā informācija

Kods	DE0601
Nosaukums	Datu vizualizācijas pamati
Studiju kursa statuss programmā	Brīvās izvēles
Atbildīgais mācītspēks	Vita Šakele - Doktors, Docents
Apjoms daļās un kredītpunktos	1 daļa, 3.0 kredītpunkti
Studiju kursa īstenošanas valodas	LV, EN
Anotācija	Datu vizualizācija ir neatņemama datu analīzes un datu zinātnes daļa. Ar tās palīdzību tiek veikta gan datu pētnieciskā analīze, gan datu un to analīzes rezultātu prezentēšana dažādām mērķauditorijām. Studiju kursa ietvaros studenti apgūst datu vizualizāciju ar programmēšanas valodām Python un R, kas ir nozīmīgākās programmēšanas valodas datu zinātnē, kā arī iepazīstas ar datu vizualizācijas rīkiem Power BI un Tableau. Datu vizualizāciju valodā R ir iespējams veikt ar tās bāzes pakotnē ietilptošajiem līdzekļiem. Tomēr papildus pakotņu lietošana nodrošina bagātīgākas vizualizācijas iespējas. Studiju kursa ietvaros studenti apgūst valodas R pakotni ggplot2 (un tās analoģu - valodas Python bibliotēku plotnine), ar kuru ir iespējams izveidot grafikus un diagrammas gan viena mainīgā, gan divu, gan vairāku mainīgo attēlošanai. Arī valodas Python bibliotēku matplotlib un seaborn apgūšana ir iekļauta kursā. Studentus iepazīstina arī ar specifiskām pakotnēm interaktīvajai un telpisko datu vizualizācijai, kā arī ar datu vizualizācijas rīkiem Power BI un Tableau. Paralēli datu vizualizācijai studenti apgūst arī vienkāršas manipulācijas ar datiem. Izstrādājot patstāvīgo darbu, studenti gūst iemaņas darbā ar pakotnēm RMarkdown, knitr un tīmekļa lietotni Jupyter Notebook, kas atbalsta atskaites dokumenta veidošanu un programmatūras koda dokumentēšanu jau vizualizācijas izstrādes laikā, kā arī dokumenta eksportu dažādos formātos, tādā veidā nodrošinot atkārtoti reproducējamu datu analīzi. Visas kursa nodarbības notiek datorklasē, tādā veidā sniedzot studentiem iespēju iegūtās zināšanas uzreiz izmantot praksē, un lekciju teorētisko materiālu papildina nodarbības laikā risināmie uzdevumi.
Mērķis un uzdevumi, izteikti kompetencēs un prasmēs	Studentu mērķis - iegūt zināšanas un prasmes datu vizualizēšanā ar programmēšanas valodām Python un R un datu vizualizācijas rīkiem Power BI un Tableau. Kursa noslēgumā studenti prātīs izvēlēties datu dabai piemērotāko diagrammu un izveidot to, izmantojot dažādas pakotnes un bibliotēkas, kā arī dokumentēt veikto datu analīzi.
Patstāvīgais darbs, tā organizācija un uzdevumi	Studenta patstāvīgo darbu veido nodarbību laikā izsniegto uzdevumu risināšana un ieskaite darba izstrādāšanā. Ieskaite darbā students patstāvīgi izstrādā datu kopas vizualizāciju, ko prezentē pēdējā nodarbībā. Nodvēumu veido ieskaite darba apraksts, kas ir iesniegts kā *.pdf dokuments, kurš ir iegūts ar Jupyter Notebook vai RMarkdown palīdzību, *.rmd fails vai *.ipynb, kas ir avots uzģenerētajam *.pdf dokumentam, un datu kopas fails.
Literatūra	H. Wickham. ggplot2: Elegant Graphics for Data Analysis (Use R!). Springer, 2016 R. I. Kabacoff. R in Action: Data analysis and graphics with R. Manning, 2011 W. McKinney. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. O'Reilly, 2018 C.O. Wilke. Fundamentals of Data Visualization. O'Reilly, 2019. Internet resources: 1. Reference to ggplot2. http://ggplot2.tidyverse.org/reference/index.html 2. Contributed R packages. https://cran.r-project.org/web/packages/index.html 3. Matplotlib. https://matplotlib.org/ Downloads: 1. R. https://cran.r-project.org/ 2. RStudio. https://www.rstudio.com 3. Jupyter Notebooks/Spyder. https://www.anaconda.com/download/ 4. Power BI. https://powerbi.microsoft.com/en-us/downloads/ 5. Tableau. https://public.tableau.com/s/download?source=cta
Nepieciešamās priekšzināšanas	Matemātiskās statistikas pamatjēdzieni un minimālas prasmes programmēšanā.

Studiju kursa saturs

Saturs	Pilna un nepilna laika klātienēs studijas		Nepilna laika neklātienēs studijas	
	Kontakt stundas	Patstāv. darbs	Kontakt stundas	Patstāv. darbs
Ievads datu vizualizācijā.	8	0	0	0
Valodās Python un R pieejamie datu vizualizācijas līdzekļi. Programmatūras izstrādes vides RStudio un Anaconda.	2	6	0	0
Grafikas gramatika.	2	0	0	0
Valodu R un Python sintakses pamati. Pakotnes ggplot2 un plotnine.	8	8	0	0
Citas valodas R pakotnes.	4	6	0	0
Valodas Python bibliotēkas matplotlib un seaborn.	6	6	0	0

Interaktīva datu vizualizācija ar valodas Python bibliotēku Bokeh.	2	4	0	0
Datu vizualizācija ar rīkiem Power BI un Tableau.	2	0	0	0
Labā prakse datu vizualizācijā.	2	4	0	0
Atkārtojamas datu analīzes iespējas ar Jupyter Notebook un RMarkdown. Patstāvīgā darba noformēšana un prezentācijas sagatavošana.	2	6	0	0
Kursa kopsavilkums un ieskaite darba prezentācija.	2	0	0	0
Kopā:	40	40	0	0

Sasniedzamie studiju rezultāti un to vērtēšana

Sasniedzamie studiju rezultāti	Rezultātu vērtēšanas metodes
Studenti spēj izvēlēties vizualizāciju, kas atbilst datu dabai un vislabāk tos atklāj, un pamatot savu izvēli	Uzdevumi lekciju laikā. Ieskaite darbs un tā aizstāvēšana.
Studenti prot izveidot grafikus un diagrammas valodās Python un R, kas atklāj viena, divu un vairāku mainīgo datus	Uzdevumi lekciju laikā. Ieskaite darbā izveidotie grafiki.
Studenti prot veidot datu apstrādes, analīzes un vizualizācijas atskaites ar Jupyter Notebook un RMarkdown	Ieskaite darba noformējums.

Studiju rezultātu vērtēšanas kritēriji

Kritērijs	% no kopējā vērtējuma
Uzdevumi par grafiku novērtēšanu	10
Programmēšanas uzdevumi	10
Ieskaite darbs	80
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	*			*		