

RTU studiju kurss "Statistikā izpēte: teorija un pielietojumi finansēs"

33000 Datorzinātnes, informācijas tehnoloģijas un enerģētikas fakultāte

Vispārējā informācija

Kods	DE0854
Nosaukums	Statistikā izpēte: teorija un pielietojumi finansēs
Studiju kursa statuss programmā	Obligātais/Ierobežotās izvēles
Atbildīgais mācībspēks	Andrejs Matvejevs - Doktors, Profesors
Mācībspēks	Viktors Ajevskis - Doktors, Asociētais profesors
Apjoms daļās un kredītpunktos	1 daļa, 6.0 kredītpunkti
Studiju kursa īstenošanas valodas	LV, EN
Anotācija	Statistikās izpētes pamatā ir koncepcija, ka datorprogramma var iemācīties un pielāgoties jauniem datiem bez cilvēka iejaukšanās. Statistikā izpētē tiek izmantoti statistikas algoritmi, lai identificētu vēsturiskos paraugus datus, kuri var tikt izmantoti, lai izveidotu nākotnes tendenču prognozes. Izmantojot piemērus no ekonomikas un finansēm, studiju kurss nodrošina studentiem praktisku pieredzi ar tipisku statistikās izpētes darbplūsmu, kas ietver datu priekšapstrādi un modeļu parametru novērtējumu, kā arī modeļu prognožu veiktspējas analīzi. Pēc studiju kursa apgūšanas studenti varēs praktiski pielietot iegūtās zināšanas, izmantojot statistikas programmatūras paketes R vidē.
Mērķis un uzdevumi, izteikti kompetencēs un prasmēs	Studiju kursa mērķis ir iepazīstināt studentus ar mūsdienu statistikās izpētes rīkiem un to pielietojumu praksē. Papildus mērķis ir iepazīstināt studentus ar moderniem teksta analīzes metodēm un to pielietojumu finansēs. Studiju kursa uzdevums ir apmācīt studentus pielietot datu prognozēšanas metodes, kas ietver lineāru regresiju, datu sarūkšanas un datu dimensijas samazināšanas pieejas, kā arī koku regresijas modeļus.
Patstāvīgais darbs, tā organizācija un uzdevumi	Patstāvīgā darba mērķis ir apmācīt studentus pašiem definēt pētījuma jautājumu, atrast, sagatavot un izanalizēt datus, izmantojot kursā pasniegto materiālu. Studiju kursa ietvaros studentiem jā sagatavo un jāveic īsa kursa projekta publiskā prezentācija.
Literatūra	Obligātā. / Obligatory: Gareth, J., Witten, D., Hastie, T., Tibshirani, R. An Introduction to Statistical Learning: with Applications in R Springer Texts in Statistics. 2nd Edition. 2021 Hastie, T., Tibshirani, R. & Friedman, J. H. . The elements of statistical learning: data mining, inference, and prediction. New York: Springer. Open Access. 2009. Grolemund, G., & Wickham, H. . R for Data Science: Import, Tidy, Transform, Visualize, and Model Data O'Reilly media. Open Access. 2017 Papildu. / Additional: Silge, J., Robinson. D. . Text Mining with R: A Tidy Approach. O'Reilly media. Open Access. 2017 Gentzkow, M., Kelly, B., Taddy, M. . Text as Data. Journal of Economic Literature, 57(3): 535-74. 2019.
Nepieciešamās priekšzināšanas	Augstākā matemātika pamatkursa apjomā, Vairbūtību teorija un matemātiskā statistika pamatkursa apjomā.

Studiju kursa saturs

Saturs	Pilna un nepilna laika klātienes studijas		Nepilna laika neklātienes studijas	
	Kontakt stundas	Patstāv. darbs	Kontakt stundas	Patstāv. darbs
Ievads kursā un organizatoriskā informācija. Statistikā programmatūra R/Rstudio. Reproducējami pētījumi ar R Markdown. Datu apstrādes un vizualizācijas rīki. Pētniecisko datu pirmā analīze R vidē.	16	24	0	0
Ievads statistikās izpētes kursā. Ievads prognozēšanas modelēšanas procesā un tā galvenajos posmos, piemēram, savstarpējā validācijā un paraugu sadalē. Dispersijas un novirzes kompromiss.	8	12	0	0
Lineārā regresija un mainīgo inženierija. Mainīgo izvēles metodes lineārās modeļos, kuru mērķis ir atlasīt ierobežotu regresoru skaitu, ar vislabāko atkarīgā mainīgā prognozēs veiktspēju. Pielietojums	8	12	0	0
Sarūkšanas pieeja regresijas parametru novērtēšanai, kas balstās uz koeficienta vērtību samazināšanas nulles virzienā, salīdzinot ar OLS aplēsēm. Modeļu novērtēšana, kad regresoru skaits $p > n$.	3	5	0	0
Koeficienta sarūkšanas veikšana, izmantojot sodu regresijas modeļus kā Ridge regresijas un LASSO / Elastīgo tīklu (ET) modeļus. LASSO / ET modeļi salīdzināšana. Pielietojums finansēs.	3	4	0	0
Datu dimensijas samazināšanas paņēmieni. Alternatīvu pieeju informācijas izmantošana ļoti lielās datu kopās, veidojot k lineāras kombinācijas no sākotnējiem p regresoriem, ar $p \gg k$.	3	5	0	0
K lineārās kombinācijas vai projekcijas izmantošana, lai prognozētu atkarīgo mainīgo. Visplašāk izmantotas metodes ietver galveno komponentu analīzi (PCA), galveno komponentu regresiju (PCR).	3	4	0	0
Regresijas koka modeļi, kas balstīti uz atkarīgā mainīgā stratifikāciju vai segmentēšanu atsevišķos nepārklājošos blokos. Metodes, kurām ir augsta prognozes veiktspēja un meža regresija.	2	3	0	0

Koku regresijas modeļi un mainīgie, kas automātiski nodrošina datu segmentēšanas. Pielietojums finansēs.	2	3	0	0
Regresijas koka modeļi, kas tie var patvaļīgi labi tuvināt nelineāras attiecības starp atkarīgiem un skaidrojošajiem mainīgajiem.	2	3	0	0
Statistikās izpētes metodes, kuru veikspēja var būt kritiski atkarīga no hiperparametru izvēlētajām vērtībām. Hiperparametru pielāgošanas nozīme modeļa izvēles procesā.	3	4	0	0
R programmatūras sistēma, kas nodrošina, lai pārredzamā un reproducējamā veidā tika veiktas hiperparametru vērtību ietekmes analīzi uz tā paša modeļa prognozēšanas veikspēju. Pielietojums finansēs.	3	5	0	0
Teksta dati. Sentimentu analīze. Pielietojums finansēs.	8	12	0	0
Kopā:	64	96	0	0

Sasniedzamie studiju rezultāti un to vērtēšana

Sasniedzamie studiju rezultāti	Rezultātu vērtēšanas metodes
Prot analizēt datus ar lineārās regresijas palīdzību izmantojot R programmatūru un izveidot datu analīzes kopsavilkumu R Markdown vidē.	Laboratorijas darbs. Tests. Eksāmens.
Prot pielietot koeficientu sarūkšanas regresijas metodes/datu dimensijas samazināšanas metodes/koku regresijas modeļus datu analīzei.	Laboratorijas darbs. Tests. Eksāmens.
Prot pielietot teksta apstrādes un analīzes rīkus.	Laboratorijas darbs. Tests. Eksāmens.
Prot izveidot statistiskās izpētes darba plūsmu jaunu datu analīzei un prezentēt analīzes rezultātus publikas priekšā.	Kursa darbs. Tests. Eksāmens.

Studiju rezultātu vērtēšanas kritēriji

Kritērijs	% no kopējā vērtējuma
Laboratorijas darbs. 1. tests	10
Laboratorijas darbs. 2. tests	10
Laboratorijas darbs. 3. tests	10
Kursa darbs. 4. tests	35
Eksāmens	35
Kopā:	100

Studiju kursa plānojums

Daļa	KP	Stundas			Pārbaudījumi		
		Lekcijas	Prakt d.	Laborat	Ieskaite	Eksām.	Darbs
1.	6.0	32.0	0.0	32.0		*	