

UČNI NAČRT PREDMETA / COURSE SYLLABUS

Predmet:	Statistično učenje in modeliranje
Course title:	Statistical learning and modeling

Študijski program in stopnja Study programme and level	Študijska smer Study field	Letnik Academic year	Semester Semester
Podatkovne znanosti, magistrski študijski program druge stopnje	-	Drugi	Tretji
The second cycle masters study programme Data Sciences	-	Second	Third

Vrsta predmeta / Course type Obvezni / Obligatory

Univerzitetna koda predmeta / University course code: 2-PZ-MAG-SUM-2020-06-30

Predavanja Lectures	Seminar Seminar	Vaje Tutorial	Klinične vaje work	Druge oblike študija	Samost. delo Individ. work	ECTS
35	-	45	-	-	130	7

Nosilec predmeta / Lecturer: Doc. dr. Nuša Erman

Jeziki / Languages:

Predavanja / Lectures:	Slovenski, angleški / Slovene, English
Vaje / Tutorial:	Slovenski, angleški / Slovene, English

Pogoji za vključitev v delo oz. za opravljanje študijskih obveznosti:

Pogoj za vključitev v delo predstavlja pridobljeno znanje pri predmetu Izbrana poglavja iz verjetnosti in statistike.

Pogoj za pristop k pisnemu izpitu je pravočasna oddaja in pozitivno ocenjena seminarska naloga.

Prerequisites:

Knowledge obtained at Selected topics in probability and statistics is required for the inclusion in the work.

Student has to submit seminar work within the due time. If the seminar work is positively graded, he/she is allowed to write the exam.

Vsebina:

Statistični modeli:

- opredelitev in namen,
- ocenjevanje statističnih modelov,
- ocenjevanje točnosti modela: kvaliteta prilaganja modela podatkom, kompromis med pristranostjo in razpršenostjo.

Content (Syllabus outline):

Introduction to statistical learning:

- definition and purpose,
- estimation of statistical models,
- assessing model accuracy: measuring the quality of fit, the bias-variance trade-off.

Linearne regresijske metode:

- enostavna linearna regresija,
- multipla linearna regresija
- ocenjevanje točnosti ocen koeficientov,
- ocenjevanje točnosti modela,
- kvalitativni prediktorji,
- razširitve linearnega modela: učinek interakcije,
- izbira in regularizacija linearnega modela: izbira podskupine, metode krčenja (Ridge in Lasso regresija), metode redukcije dimenzionalnosti podatkov (regresija glavnih komponent, metoda parcialnih najmanjših kvadratov).

Posplošeno linearno modeliranje:

- posplošeni linearni modeli; ocenjevanje, sklepanje,
- Logit in Probit modeli,
- Poissonova regresija

Linearne klasifikacijske metode:

- logistična regresija,
- multipla logistična regresija,
- linearna diskriminantna analiza,
- kvadratna diskriminantna analiza,
- k-najbližjih sosedov,
- primerjava klasifikacijskih modelov.

Samovzorčenje:

- križno preverjanje
- zankanje.

Nelinearno modeliranje:

- polinomska regresija,
- regresija zlepkov,
- glajenje zlepkov,
- lokalna regresija,
- posplošeni aditivni modeli.

Drevesne metode:

- regresijska in klasifikacijska drevesa,
- bagging, slučajni gozdovi, boosting.

Nenadzorovano učenje:

- analiza glavnih komponent,
- razvrščanje v skupine (metoda voditeljev, hierarhično razvrščanje).

Uporaba sodobnih računalniških programov za napredno statistično analizo (R in RStudio).

Linear regression methods:

- simple linear regression,
- multiple linear regression,
- assessing the accuracy of the coefficient estimates,
- assessing the accuracy of the model,
- qualitative predictors,
- extensions of the linear model: interaction effect,
- linear model selection and regularization: subset selection, shrinkage methods (Ridge and Lasso regression), dimension reduction methods (principal components regression, partial least squares method).

Generalized linear modeling:

- generalized linear models: estimation, inference,
- Logit in Probit models,
- Poisson regression.

Linear classification methods:

- logistic regression,
- multiple logistic regression,
- linear discriminant analysis,
- quadratic discriminant analysis,
- k-nearest neighbors,
- comparison of classification methods.

Resampling methods:

- cross-validation,
- bootstrap

Non-linear modeling:

- polynomial regression,
- regression splines,
- smoothing splines,
- local regression,
- generalized additive models.

Tree-based methods:

- regression and classification trees,
- bagging, random forests, boosting.

Unsupervised learning:

- principal component analysis,
- clustering methods (k-means clustering, hierarchical clustering).

Application of state-of-the-art software for applied statistical analysis (R and RStudio).

Temeljna literatura in viri / Readings:

- Hastie, T., Tibshirani, R. in Friedman, J.H. (2016): *The elements of statistical learning: data mining, inference, and prediction, 2nd Edition*. New York, Springer.
- Gareth, J., Witten, D., Hastie, R. in Tibshirani R. (2017): *An Introduction to Statistical Learning with Applications in R, Corr. 7th ed.* New York: Springer.
- Dunn, P.K. in Smyth, G.F. (2018): *Generalized Linear Models with Examples in R*. New York, Springer.
- Erman, N.: Prosojnice iz predavanj in gradiva z vaj **pri predmetu** Statistično učenje in modeliranje. Moodle, FIŠ.

Cilji in kompetence:

Učna enota prispeva k razvoju naslednjih splošnih in predmetno specifičnih kompetenc:

Splošne kompetence:

- sposobnost skrbeti za kakovost strokovnega dela skozi avtonomnost, samoiniciativnost, (samo)kritičnost, (samo)refleksivnost in (samo)evalviranje
- sposobnost fleksibilne uporabe znanja v praksi
- zmožnost artikulacije raziskovalnega problema in na tej podlagi sposobnost pridobivanja, selekcije, ocenjevanja in umeščanja novih informacij;
- uporaba ustreznih metodoloških pristopov za izvajanje, koordiniranje in organiziranje raziskav

Predmetno-specifične kompetence:

- poznavanje osnovnih in naprednih metod analize podatkov in poizvedovanja v podatkih
- obvladanje raziskovalnih metod, postopkov in procesov
- sposobnost izvedbe kvantitativne raziskave in analize podatkov z uporabo ustreznih statističnih metod in modelov s pomočjo primerne programske opreme

Objectives and competences:

The instructional unit contributes to the development of the following general and subject-specific competences:

General competences:

- the ability to manage quality of professional work through autonomy, initiative, as well as (self-)criticism, (self-)reflection and (self-)evaluation
- the ability of flexible usage of knowledge in practice
- the ability to articulate the research problem and correspondingly, obtain, select, evaluate and embed the new information;
- utilization of adequate methodological approaches to conduct, coordination and organization of research

Subject-specific competences:

- familiarity with the basic and applied data analysis and data inquiry methods
- competence in research methods, procedures and processes
- ability to perform quantitative research and data analysis using appropriate statistical methods and models and suitable software

Predvideni študijski rezultati:

Znanje in razumevanje:

Sposobnost študenta/študentke bo:

- v povezavi z drugimi predmeti bo poznal in razumel relevantna poglavja iz družboslovnega raziskovanja,

Intended learning outcomes:

Knowledge and understanding:

The ability of the student:

- to realize and understand the relevant chapters from the research in social science, data bases and data analytics.

<p>podatkovnih baz in podatkovne analitike.</p> <ul style="list-style-type: none"> • sposoben zavzeti stališče do ključnih etičnih vprašanj v raziskovalnem procesu in kritično vrednotiti konkreten primer • poznal in bil sposoben uporabiti izbrane metode in tehnike kvantitativnega raziskovanja na višjem nivoju • sposoben uporabe osnovne programske opreme za kvantitativno analizo • sposoben pripraviti in izvesti načrt kvantitativne raziskave: raziskovalno vprašanje, hipoteze, načrt zbiranja in obdelave podatkov, zbiranje in obdelava podatkov, diskusija o rezultatih • sposoben refleksije in kritičnega vrednotenja primernosti določene raziskovalne metode za analizo konkretnega problema 	<ul style="list-style-type: none"> • to take a position on key ethical issues in the research process and to be critical in evaluating concrete examples; • to apply methods and techniques of quantitative research on higher level; • to use of basic software for quantitative analysis; • to prepare and implement a quantitative research plan: research questions, hypotheses, data collection and processing plan, collection and processing of data, discussion about the results; • reflection and critical evaluation of the appropriateness of certain research methods for the analysis of concrete problems
---	---

Metode poučevanja in učenja:

<ul style="list-style-type: none"> • <i>predavanja</i> z aktivno udeležbo študentov (razlaga, diskusija, vprašanja, primeri, reševanje problemov) • <i>vaje</i>, kjer študentje na enostavnih primerih ponovijo temeljne koncepte in metode, predstavljene na predavanjih • <i>laboratorijske vaje</i>, kjer se študenti seznanijo s programskimi orodji za zbiranje in analiziranje podatkov
--

Learning and teaching methods:

<ul style="list-style-type: none"> • <i>lectures</i> with active students participation (explanations, discussion, questions, examples, problem solving); • <i>tutorials</i> (students will recall, reinforce, and shed light on the concepts and methods taught on lectures); • <i>lab work</i> (students will learn state of the art software for data collection and analysis).

Načini ocenjevanja:

Način:	Delež (v %) / Weight (in %)	Assessment: Type:
<ul style="list-style-type: none"> • pisni izpit • seminarska naloga 	<p>60 %</p> <p>40 %</p>	<ul style="list-style-type: none"> • written exam • seminar

Reference nosilca / Lecturer's references:

<ul style="list-style-type: none"> • ERMAN, Nuša, GOLOB, Tea, JELOVAC, Dejan, RAKOVEC, Primož. The impact of internal dialogue on aggressive driving. <i>The social sciences</i>, ISSN 1993-6125. [Online ed.], 2020, vol. 15, iss. 3, str. 119-127
--

- ERMAN, Nuša, TODOROVSKI, Ljupčo. The effects of measurement error in case of scientific network analysis. *Scientometrics*, aug. 2015, vol. 104, iss. 2, str. 453-473.
- ERMAN, Nuša. Izbrani vidiki proučevanja znanstvenih omrežij : teorija in praksa. 1. izd. Ljubljana: Vega, 2015. 103 str.
- ERMAN, Nuša, KOROŠEC, Aleš, SUKLAN, Jana. Performance of selected agglomerative hierarchical clustering methods. *Innovative issues and approaches in social sciences*, Jan. 2015, vol. 8, no. 1, str. 180-204.
- ERMAN, Nuša, TODOROVSKI, Ljupčo, JEREB, Berta. Late somatic sequelae after treatment of childhood cancer in Slovenia. *BMC research notes*, May 2012, vol. 5, no. 254, str. [1-19].