

UČNI NAČRT PREDMETA / COURSE SYLLABUS

| | |
|----------------------|--|
| Predmet: | Metodologija in evalvacija v tehnologijah znanja |
| Course title: | Methodology and Evaluation in Knowledge Technologies |

| Študijski program in stopnja Study programme and level | Modul Module | Letnik Academic year | Semester Semester |
|---|------------------------|-------------------------|----------------------|
| Informacijske in komunikacijske tehnologije, 3. stopnja | Tehnologije znanja | 1 | 1 |
| Information and Communication Technologies, 3 rd cycle | Knowledge Technologies | 1 | 1 |

Vrsta predmeta / Course type Izbirni / Elective

Univerzitetna koda predmeta / University course code: IKT3-720

| Predavanja Lectures | Seminar Seminar | Sem. vaje Tutorial | Lab. vaje Laboratory work | Druge oblike | Samost. delo Individ. work | ECTS |
|------------------------|--------------------|-----------------------|------------------------------|--------------|-------------------------------|------|
| 15 | 15 | | | 15 | 105 | 5 |

**Navedena porazdelitev ur velja, če je vpisanih vsaj 15 študentov. Drugače se obseg izvedbe kontaktnih ur sorazmerno zmanjša in prenese v samostojno delo. / This distribution of hours is valid if at least 15 students are enrolled. Otherwise the contact hours are linearly reduced and transferred to individual work.*

Nosilec predmeta / Lecturer: Prof. dr. Marko Robnik-Šikonja

Jeziki / Languages: Predavanja /Lectures: Slovenščina, angleščina / Slovenian, English
Vaje / Tutorial: -

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

Zaključen študij druge stopnje s področja informacijskih ali komunikacijskih tehnologij ali zaključen študij druge stopnje na drugih področjih z znanjem osnov s področja predmeta. Potrebna so tudi osnovna znanja matematike, računalništva in informatike.

Prerequisites:

Completed second cycle studies in information or communication technologies or completed second cycle studies in other fields with knowledge of fundamentals in the field of this course. Basic knowledge of mathematics, computer science and informatics is also requested.

Vsebina:

Učenje kot modeliranje in optimizacija:
cilji učenja, različne naloge učenja, klasifikacija, regresija, ocenjevanje verjetnosti in gostote verjetnosti, rangiranje, razvrščanje, posplošitve osnovnih nalog, učenje kot optimizacija
Uspešnost učenja:
mere uspešnosti učenja za različne naloge, intervali zaupanja, strmenske in permutacijske metode, kalibracija verjetnosti
Ocene napake:
pretirana prilagoditev učnim podatkom, regularizacija, dekompozicija napake na pristranost

Content (Syllabus outline):

Learning as modeling and optimization:
learning goals, learning tasks, classification, regression, probability estimation, density estimation, ranking, clustering, generalizations of basic tasks, learning as optimization
Validation of learning;
validation measures for different tasks, confidence intervals, bootstrap and permutation approaches, probability calibration
Error estimation:
data overfitting, regularization, bias-variance error, decomposition, margin, cross-validation, VC

in razpršenost, rob, prečno preverjanje, VC dimenzija, princip najkrajše dolžine opisa
Primerjava modelov:
neobstoj zastonskega kosila, statistični testi za primerjavo modelov
Kombiniranje modelov:
šibko učenje in principi kombiniranja metod, napake in raznolikost kombiniranih metod
Vizualizacija napovednih modelov:
aditivni modeli, razlage odločitev, vizualizacijske tehnike za nekatere kompleksne modele, orodje VIPER
Primeri uporabe za področje jezikovnih tehnologij

dimension, minimum description length principle
Model comparison:
no free lunch theorem, statistical test for model comparison
Combining models:
weak learning and principles of ensemble learning, error and diversity in ensembles
Visualization of predictive models:
additive models, explaining decisions, visualization techniques for some complex models, VIPER toolbox
Use cases from natural language processing

Temeljna literatura in viri / Readings:

Izbrana poglavja iz naslednjih knjig: / Selected chapters from the following books:

- G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning with Applications in R*. Springer, 2013. ISBN 978-1-4614-7137-0
- T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning*, 2nd edition. Springer, 2009. ISBN 978-0-387-84857-0
- P. K. Janert, *Data analysis with open source tools*. O'Reilly Media. 2010. ISBN 978-0-596-80235-6
- S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python*, O'Reilly Media. 2009. ISBN 978-0-596-51649-9

Cilji in kompetence:

Cilj predmeta je pregledno predstaviti metodologijo in evalvacijo statističnega učenja na primerih s področja procesiranja naravnega jezika. Poudarek je na povezavi poznavanja teoretičnih konceptov s praktičnimi znanji iz podatkovne analitike, povezanimi z rabo orodij, ki omogočajo analizo podatkov, evalvacijo in izbor modelov, njihovo vizualizacijo in interpretacijo.

Predvidene kompetence:

- poznavanje osnovnih metodoloških pristopov k statističnemu učenju,
- poznavanje mer uspešnosti učenja in njihovih lastnosti,
- poznavanje načinov za ocenjevanje posplošitvene napake učenja,
- praktična uporaba statističnih testov za primerjavo učnih modelov,
- poznavanje načinov za kombiniranje učnih metod,
- napredna vizualizacija napovednih modelov in odločitev,
- poznavanje odprtokodnih orodij za statistično učenje na področju naravnega jezika.

Objectives and competences:

The goal of the course is to present a comprehensive overview of statistical learning methodology and evaluation of learned models on natural language processing use cases. The emphasis is to integrate theoretical knowledge of methodology and evaluation with practical skills from data analytics, i.e., the use of analytical tools for statistical learning, evaluation, model selection, visualization and interpretation of models.

The competencies of the students completing this course would include:

- understanding basic methodological approaches for statistical learning,
- knowledge of measures for success of learning and their properties,
- knowledge of generalization error estimation,
- practical use of statistical tests for comparison of learned models,
- knowledge of combining different models,
- competences in visualization of predictive models and their decisions,
- capability to apply open-source statistical learning tools to natural language processing tasks.

Predvideni študijski rezultati:

Študenti bodo z uspešno opravljenimi obveznostmi tega predmeta pridobili:

- pregled nalog statističnega učenja in njihove praktične rabe na področju obdelave naravnega jezika,
- pregled mer za ocenjevanje napake učenja in razumevanje njihovih lastnosti,
- sposobnost uporabe tehnik za dekompozicijo napake na pristranost in razpršenost,
- sposobnost ugotavljanja primernosti učnih modelov in izbora najprimernejšega,
- veščine vizualizacije učnih modelov in njihovih rezultatov.

Intended learning outcomes:

Students successfully completing this course will acquire:

- overview of statistical learning tasks and their practical use in the natural language processing,
- overview of error estimation measures and comprehension of their properties,
- the ability to apply bias variance decomposition analysis,
- the ability to select learned models,
- the ability to visualize learning models and their predictions.

Metode poučevanja in učenja:

Predavanja, seminar, konzultacije, individualno delo

Learning and teaching methods:

Lectures, seminar, consultancy, individual work

Delež (v %) /

Načini ocenjevanja:

Weight (in %)

Assessment:

| | | |
|--------------------------------------|------|---|
| Pisni ali ustni izpit | 50 % | Written or oral exam |
| Seminarska naloga | 25 % | Seminar work |
| Javna predstavitev seminarske naloge | 25 % | Public presentation of the seminar work |

Reference nosilca / Lecturer's references:

- M. Robnik Šikonja, and K. Vanhoof, Evaluation of ordinal attributes at value level. Data mining and knowledge discovery, 2007, vol. 14, no. 2, str. 225-243.
- M. Robnik Šikonja, and I. Kononeko, Theoretical and empirical analysis of ReliefF and RReliefF. Machine learning, 2003, 53:23-69.
- M. Robnik Šikonja, and I. Kononeko, Explaining classifications for individual instances. IEEE Transactions on Knowledge and Data Engineering, 2008, 20(5):589-600.
- M. Pičulin, and M. Robnik Šikonja, Handling numeric attributes with ant colony based classifier for medical decision making. Expert systems with applications, 2014, 41(16):7524-7535.
- M. Robnik Šikonja, I. Kononeko, and E. Štrumbelj: Quality of Classification Explanations with PRBF. Neurocomputing, 96:37-46, 2012