



I E G U L D Ī J U M S T A V Ā N Ā K O T N Ē

Datizrace (LV)
Data Mining (ENG)

Kursa īss nosaukums: DM2019

Autors	PhD cand., Mg.sc.comp., vieslektors Juris Kļonovs
LAIS kods	
Pārbaudes forma	Eksāmens
Kredītpunkti (ECTS kredītpunkti)	2 KP (3 ECTS)
Kopējais kontaktnodarību skaits	16
Lekciju skaits	8
Praktisko nodarbību skaits	8
Nepieciešamās zināšanas kursa uzsākšanai	MATLAB / Python programmēšanas pamatzināšanas
Studiju programmas daļa	Nozares obligātās izvēles kursi

Studiju kursa mērķis

Šis ir praktisks kurss, kas paredzēts, lai iepazīstinātu studentus ar datizraci (Data Mining) un zināšanu izgūšanas mehānismiem no datubāzēm (KDD, no angļu val. Knowledge Discovery from Databases), kā arī datizraces pielietošanu mūsdienu pasaules tehnoloģijās un lietojumprogrammās. Kurss iepazīstinās studentus arī ar esošajiem datizraces programmēšanas modeļiem, ko varēs pielietot uz reālas pasaules datiem.

Šis kurss ietver POPBL (uz projektiem orientētu problēmu risināšanas) apguves modeli, kura mērķis ir saskaņot studentu prasmes ar industrijas vajadzībām, kas attiecīgi nodrošina, ka students iegūst nepieciešamo praktisko pieredzi un zināšanas reālistiska projekta realizācijas laikā.

Studiju rezultāti

Studiju kursa apgūšanas rezultātā students:

- Spēj izprast pamatprincipus datizraces teorijā (**p1**)
- Spēj identificēt un formalizēt datizraces problēmas (**z1, p1**)

- Spēj pielietot mašīnmācīšanās metodes dažādu problēmu risināšanai datizracē (**z2, p3, a2**)
- Spēj realizēt dažādas datizraces metodes, pielietojot mūsdienu programmatūras rīkus un arhitektūras (**p3, p7, a2**)
- Prot implementēt gan klasiskos, gan dziļās apmācības algoritmus, izmantojot MATLAB un/vai Python programmēšanas valodu (**a4,p8**).

Studējošo patstāvīgā darba organizācijas veids

Studentu patstāvīgais darbs ietver:

- regulāru studiju kursa vielas apgūšanu, izmantojot lekciju materiālus, mācību literatūru, interneta resursus,
- patstāvīgu darbu ar zinātnisko literatūru un datizraces piemēriem,
- vidēja apjoma projekta realizēšanu ar datizraces pielietojumu,
- gatavošanos eksāmenam.

Studiju rezultātu vērtēšana

Gala rezultātu veido:

- Laboratorijas darbi 40%
- Kursa projekts / eksāmens 60%

Studiju kursa saturs

N.p.k.	Tēmas nosaukums
1.	Kursa ievads. Vispārīgas definīcijas. Kādi aspekti iekļaujas Datu ieguves un KDD teorijā un neatbilst tiem. Atšķirības starp datu ieguvu un mašīnu apguvi. Datu zinātnieku pamatprasmes.
2.	KDD pamati. Uzraudzīta un nepārraudzīta mācīšanās. Priekšapstrāde un datu tīrīšana.
3.	Statistikas secinājums attiecībā uz datu ieguvu.
4.	Datu ieguves klasifikācijas algoritmi. Grafiskie modeļi.
5.	Datu ieguves regresijas algoritmi. Pastāvīgo mainīgo noteikšana.
6.	Datu ieguves algoritmi. Gausa metodes.
7.	Hibrida metodes datu ieguves procesā.
8.	Kursa projekta izstrāde un praktiskie uzdevumi.

Studiju kursa kalendārais plāns

Nodarbības numurs	Tēmas nosaukums	Nodarbības veids (lekcijas, semināri, praktiskās nodarbības, laboratorijas darbi), akadēmisko stundu skaits
1.- 2.	Kursa ievads. Vispārīgas definīcijas. Kādi aspekti iekļaujas Datu ieguves un KDD teorijā un neatbilst tiem. Atšķirības starp datu ieguvu un mašīnu apguvi. Datu zinātnieku pamatprasmes.	lekcijas, 2 semināri, 2
3.-4.	KDD pamati. Uzraudzīta un nepārraudzīta mācīšanās. Priekšapstrāde un datu tīrīšana.	lekcijas, 2 semināri, 2
5.-6.	Statistikas secinājums attiecībā uz datu ieguvu.	lekcijas, 2 semināri, 2

7.-8.	Datu ieguves klasifikācijas algoritmi. Grafiskie modeļi.	lekcijas, 2 semināri, 2
9.-10.	Datu ieguves regresijas algoritmi. Pastāvīgo mainīgo noteikšana.	lekcijas, 2 semināri, 2
11.-12.	Datu ieguves algoritmi. Gausa metodes.	lekcijas, 2 semināri, 2
13.-14.	Hibrīda metodes datu ieguves procesā.	lekcijas, 2 semināri, 2
15.-16.	Kursa projekta izstrāde un praktiskie uzdevumi.	lekcijas, 2 semināri, 2

Pamatliteratūra

- Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems) – 2013.
- Data Mining and Analysis Fundamental Concepts and Algorithms, Zaki & Meira – 2014.
- Yaser S. Abu-Mostafa, Learning from Data - A Short Course – 2017, ISBN 978-1600490064

Papildliteratūra

- “The Elements of Statistical Learning” by Freidman et al – 2009.

Citi informācijas avoti

- SciKit Learn: <https://scikit-learn.org/stable/>
- Weka 3: <https://www.cs.waikato.ac.nz/ml/weka/>
- Tensor Flow: <https://www.tensorflow.org/>
- Tiešsaistes kursi MOOC:
 - Udemy: <https://www.udemy.com/datamining>
 - edX: <https://www.edx.org/learn/data-mining>

AL. 56.¹ pants. Studiju kurss

(1) Augstskolas un koledžas nosaka kārtību, kādā izstrādājami un iekļaujami studiju programmās studiju kursi, lai nodrošinātu studiju programmas kopējo studiju rezultātu sasniegšanu. Studiju kursa aprakstu sagatavo un apstiprina augstskolas un koledžas noteiktajā kārtībā.

(2) Studiju kursa apraksts:

- 1) definē prasības studiju kursa apguves uzsākšanai;
- 2) nosaka studiju kursa īstenošanas mērķi un plānotos studiju rezultātus;
- 3) izklāsta studiju rezultātu sasniegšanai nepieciešamo studiju kursa saturu, ietver studiju kursa kalendāru, obligāto literatūru un papildliteratūru, norāda citus informācijas avotus;
- 4) raksturo studējošo patstāvīgā darba organizāciju un uzdevumus;
- 5) nosaka studiju rezultātu vērtēšanas kritērijus.

Šie studiju materiāli ir sagatavoti projekta Nr. 8.2.2.0/18/A/009

"Stiprināt Ventpils Augstskolas akadēmisko personālu stratēģiskās specializācijas jomās".

NACIONĀLAIS
ATTĪSTĪBAS
PLĀNS 2020



EIROPAS SAVIENĪBA
Eiropas Sociālais
fonds