

Osnove strojnega učenja in podatkovnega rudarjenja

Vsebina predmeta & “pravila igre”

0. (UVODNO) PREDAVANJE

izr. prof. Branko Kavšek

O predmetu

Naziv:

Osnove strojnega učenja in podatkovnega rudarjenja

Predavatelj: Branko Kavšek

Asistent: Domen Vake

Tip predmeta: izbirni

Slušatelji:

- dodiplomski študenti programov ***Računalništvo in informatika*** in ***Bioinformatika***
(1. stopnja) UP FAMNIT – 2. in 3. letnik ++

Predznanje: ni potrebno

koristno: osnove statistike, osnove programiranja (v Pythonu)

e-učilnica: <https://e.famnit.upr.si/course/view.php?id=6196>

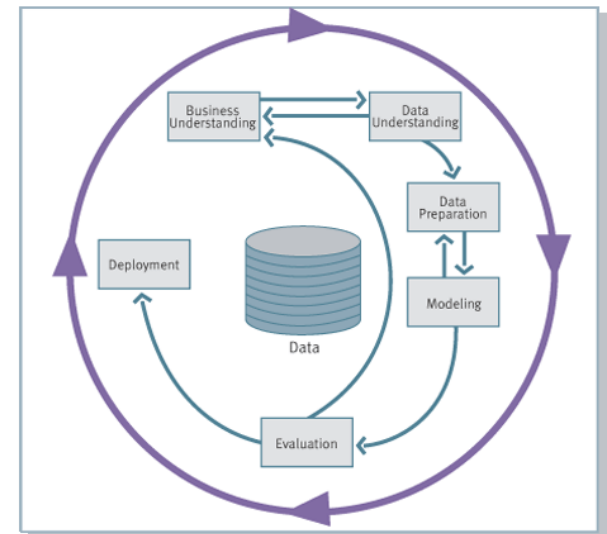
Vsebina predmeta – teorija

Odkrivanje zakonitosti (iskanje vzorcev) v podatkih po CRISP-DM metodologiji:

- Razumevanje problema
- Razumevanje podatkov
- Priprava podatkov
- Modeliranje – odkrivanje zakonitosti
- Evalvacija
- Uporaba

Vir:

https://en.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining



Vsebina predmeta – na “vajah”

Uporaba tehnik podatkovnega rudarjenja na podatkih:

- “ročno”
- z uporabo (računalniških) orodij:
 - Naprednejši urejevalniki besedil (gedit, Notepad++ ...)
 - Urejevalniki preglednic (Calc, Excel ...)
 - Namenski odprtokodni programi (**WEKA**, Python ...)

Pravila igre

Ocena predmeta je sestavljena iz:

- Ocene **pisnega izpita** – ali 2 kolokvijev (**50%**)
 - Pogoj: pisni izpit mora biti “pozitiven” ($\geq 50/100$ točk), kolokvija morata biti oba “pozitivna” ($\geq 40/100$ točk);
- Ocene (kratkih) **domačih nalog** ~ 1 sklop/teden (**15%**)
 - Pogoj: doseženih vsaj 50% vseh točk;
- Ocene **programerskega projekta** na vajah (**30%**)
 - Pogoj: v roku oddan in pozitivno ocenjen projekt;
- Ocene **ustnega izpita** (**5%**):
 - Pogoj: vsaj “zadovoljiv” odgovor na vsakega izmed 3 vprašanj;

Podrobnejše informacije v **izvedbenem predmetniku**:

<https://e.famnit.upr.si/mod/page/view.php?id=118036>

Kako torej do končne ocene?

Aktivnost	Del ocene	Pogoj
Pisni izpit / kolokvija	50 %	Doseženih vsaj polovica točk, ne manj kot 40% pri vsakem kolokviju.
Programerski projekt	30 %	Delujoč program + dokumentacija, oddano v roku = 100 % zamuda pri oddaji = – %.
Domače naloge	15 %	Doseženih vsaj 50% točk, oddano v roku = 100 % 1 teden po roku = 50 % > 1 teden po roku = 0 %.
Ustni izpit	5 %	Prikazano vsaj osnovno znanje pri odgovarjanju na vsakega izmed 3 zastavljenih vprašanj.
Skupaj:	100 %	Izpolnjeni vsi pogoji pri posameznih aktivnostih

95 do 100 točk =
odlično (**10**),

85 do 94,99 točk =
prav-dobro (**9**),

75 do 84,99 točk =
prav-dobro (**8**),

65 do 74,99 točk =
dobro (**7**),

50 do 64,99 točk =
zadostno (**6**),

manj kot 50 točk =
nezadostno (**ni opravl/a**).

Literatura in viri

1. Ian H. Witten, Eibe Frank, Mark A. Hall, and Christopher J. Pal. *Data Mining: Practical Machine Learning Tools and Techniques*, 4th Edition, Morgan Kaufmann, 2016.
-
2. Mohammed J. Zaki, Wagner Meira, Jr. *Data Mining and Analysis: Fundamental Concepts and Algorithms*, Cambridge University Press, 2014.
3. David J. Hand, Heikki Mannila and Padhraic Smyth, *Principles of Data Mining*, MIT Press, Fall 2000.
4. Trevor Hastie, Robert Tibshirani, Jerome Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Verlag, 2001.
5. Tom Mitchell: *Machine Learning*, McGraw Hill, 1997.
-
6. UCI ML repozitorij: <http://archive.ics.uci.edu/ml/>
7. Kaggle: <https://www.kaggle.com/>
-
8. WEKA softver: <http://www.cs.waikato.ac.nz/ml/weka/>