Faculty of Humanities
AGH University of Science and Technology
Academic Year 2016/2017

Academic field: Sociology

1. Subject title: Data Mining and Media Analysis

2. Subject duration: 30 hrs.

3. Lecturer: Prof. PhD Bohdan Yuskiv (yuskivbm@gmail.com)

4. Classes form: lecturing + laptop-based laboratory works

5. Objective:

introducing students to data mining analysis of media

6. Subject requirements:

– knowledge of the subject-related basics (statistics and related methods),
– PC user knowlegde,
– no more than 19 persons per group,
– Internet access is required.

7. Subject results:

– students will get knowledge in data mining analysis, characterizing basics of data processing, describing the data exploration as well as specific data mining methods;
– students will use new methods in the R language to perform deep data analysis in the media;
– students will be able to obtain additional information from the data using scientific approach.

8. Subject contents with specific classes:

CLASSES:

1. Introducing the Data Mining practice (2 hrs.)
2. Data Import and Export (2 hrs.)
3. Data exploration methods (2 hrs.)
4. Data and results visualization (2 hrs.)
5. Data clusters and factors analysis algorithms (2 hrs.)
6. Basic association analysis (2 hrs.)
7. Time series. The SARIMA model prognosis (3 hrs.)

LABORATORY WORKS (based on the media-knowledge examples)

1. Introduction of R basics as a statistics calculation environment. Program installation (3 hrs.)
2. Data import and export in R (2 hrs.)
3. Pre-analysis data exploration in R (2 hrs.)
4. The R visualization (2 hrs.)
5. Data clusters and factors analysis in R (2 hrs.)
6. Basic association analysis (2 hrs.)
7. Time series. The SARIMA model prognosis in R (2 hrs.)

PROGRAMS / SUBJECT TOOLS:

- R
- RStudio
- R packages

9. Literature

9.1. Recommended reading

1. Crawley Michael J. *The R Book*, John Wiley & Sons Ltd 2007
2. Ledolter Johannes *Data mining and business analytics with R*, University of Iowa, Wiley 2013
3. Maindonald John, Braun W. John *Data Analysis and Graphics. Using R – an Example-Based Approach*, Cambridge University Press 2003
4. Pang-Ning Tan, Michael Steinbach, Vipin Kumar, *Introduction to Data Mining*, Pearson 2005
5. Paradis Emmanuel *R for Beginners*, Institut des Sciences de l' Evolution 2005
6. Przemysław Biecek, *Przewodnik po pakiecie*, Oficyna Wydawnicza 2008
7. *Statystyczna analiza danych z wykorzystaniem programu R*, pod red. nauk. M.Walesiaka, E.Gatnara, Wydawnictwo Naukowe PWN, Warszawa 2013
8. Torgo Luis *Data Mining with R: learning with case studies*, LIACC-FEP, University of Porto 2003
9. Yanchang Zhao *R and Data Mining: Examples and Case Studies*, Elsevier 2012

9.2. Additional literature

10. Adler Joseph *R in a Nutshell*, O'Reilly 2010
11. Coghlan Avril *A Little Book of R For Multivariate Analysis*, https://github.com/avrilcoghlan/LittleBookofRMultivariateAnalysis/raw/master/_build/latex/MultivariateAnalysis.pdf
12. Gatnar E. *Podejście wielomodelowe w zagadnieniach dyskryminacji i regresji*, PWN, Warszawa 2008
13. Kabacoff Robert I. *R in Action. Data analysis and graphics with R*, Manning Publications Co 2011 (Кабаков Роберт И. *R в действии. Анализ и визуализация данных в программе R*, ДМК Пресс, Москва 2014)
14. Larose D.T. *Metody i modele eksploracji danych*, PWN, Warszawa 2008
15. Pałka Dariusz, Zaskórski Piotr, *Data mining w procesach decyzyjnych*, Zeszyty Naukowe Warszawskiej Wyższej Szkoły Informatyki, nr 7, Warszawa 2012, s. 143-161
16. Stanisz A. *Przystępny kurs statystyki. t III Analizy wielowymiarowe*, StatSoft, Kraków 2005
17. StatSoft: *Internetowy Podręcznik Statystyki*. *Techniki zgłębiania danych (data mining)*, http://www.statsoft.pl/textbook/glosfra.html
18. Venables W. N., Smith D. M. and the R Core Team *An Introduction to R. A Programming Environment for Data Analysis and Graphics*, 2015
19. Vries Andrie de, Meys Joris *R For Dummies*, John Wiley & Sons, Ltd 2012