

APPROVED BY

Director of Institute of Cybernetics
 / D.M. Sonkin

Data analysis methods

Field of Study: 09.04.04 Software Engineering

Programme name: Big data technologies

Level of Study: Master Degree Programme

Year of admission: 2020

Semester, year: Summer

ECTS: 3

Total Hours: 108

Contact Hours: 48

- **Lectures:** 16
- **Labs:** 16
- **Practical experience:** 16


Assessment: Exam

Department: Software Engineering

Head of Department


_____/ V.S. Sherstnev

Instructor(s)


_____/ O.V. Marukhina

Data analysis methods

Course Overview

Course Objectives	The goal of the course is to provide students with the theoretical knowledge and practical skills necessary for the data analysis. At the end of the course the students should be able to identify and apply the key methods of data analysis, carry out the analysis using specialized software, and to interpret the results.
Learning Outcomes	Acquiring knowledge of foundations and application of methods in data mining and data analysis. The course will prepare students to use the methods and tools of data science in their own research, whether focused on methods themselves, or more on applications.
Course Outline	Considerable attention is directed toward understanding the procedures, especially the relevant assumptions and appropriate interpretations. The course will employ a variety of teaching methods, including lecture, class discussion, and individual homework assignments. The multivariate techniques will be described as well as demonstrated on a computer. Students' questions and comments, as well as participation in class discussions, are an integral part of the course and are heartily encouraged. Data-drive homework assignments—one for each technique—will be assigned. These assignments should be well-researched, carefully thought out, meticulously organized, and, especially, well written. Data sets are provided for all techniques, but additionally, students are encouraged to use their own data sets, as these are often more meaningful to a given individual.
Prerequisites (if available)	-
Course Structure	<p>Topic 1. DATA ANALYSIS FOUNDATIONS: Categorical Attributes, Graph Data, Kernel Methods, High-dimensional Data, Dimensionality Reduction.</p> <p>Topic 2. CLUSTERING: Representative-based Clustering, Hierarchical Clustering, Density-based Clustering, K-Means Cluster Analysis</p> <p>Topic 3. CLASSIFICATION: Probabilistic Classification, Decision Tree Classifier, Linear Discriminant Analysis, Support Vector Machines, Classification Assessment</p> <p>Topic 4. Principal Components and Factor Analysis.</p>
Facilities and Equipment	Room with 10 computers: Core i5-750(MB S-1156 Asus P7P55 LX, iP55 2.66GHz/8Mb, 2 Dimm 2048 Mb, HDD 1500 Gb, Sparkle PCI-E 1024MB GeForce DVI)
Grading Policy	<p>In accordance with TPU rating system we use:</p> <ul style="list-style-type: none"> - Current assessment which is performed on a regular basis during the semester by scoring the quality of mastering of theoretical material and the results of practical activities (performance tests, perform tasks, problem solving). Max score for current assessment is 60 points, min – 40 points. - Course final assessment (exam/ credit test) is performed at the end of the semester. Max score for course final assessment is 40 points, min – 22 points. <p>The final rating is determined by summing the points of the current assessment during the semester and exam (credit test) scores at the end of the semester. Maximum</p>

	overall rating corresponds to 100 points, min pass score is 60.
Course Policy	Students are expected to make and present the results of all tasks provided in practice and lab works.
Teaching Aids and Resources	<p>Compulsory Readings:</p> <ol style="list-style-type: none"> 1. Pocket Data Mining electronic resource : Big Data on Small Devices / by Mohamed Medhat Gaber, Frederic Stahl, João Bártolo Gomes. - Cham : : Springer International Publishing : : Imprint: Springer, , 2014. - 108 p. – URL: http://dx.doi.org/10.1007/978-3-319-02711-1. - 1 2. 3 Principles of Data Mining electronic resource /by Max Bramer. - London : Springer London : Imprint: Springer, 2013. - 440 p. – URL: http://dx.doi.org/10.1007/978-1-4471-4884-5 <p>Additional Readings:</p> <ol style="list-style-type: none"> 3. Everitt, B.S., Hothorn, T., 2011. An introduction to Applied Multivariate Analysis with R, Springer. 4. Johnson, R.A., Wichern, D.W., 2007. Applied Multivariate Statistical Analysis (6th edition), Pearson Prentice Hall. 5. Manly, B.F.J, 2005. Multivariate Statistical Methods: A Primer (3rd edition), Chapman & Hall/CRC. 6. Everitt, B.S., Dunn, G., 2010. Applied Multivariate Data Analysis (2nd edition), Wiley.
Instructor (-s)	Marukhina Olga, http://portal.tpu.ru/SHARED/m/MARUKHINA/eng , marukhina@tpu.ru