Syllabus on the course “Advanced Data Analysis&Big Data for Business Intelligence”

Approved by Programme Academic Council

Process-verbal 2 from April 10, 2018

|  |  |
| --- | --- |
| Author | Piotr Baranov |
| Credits | 5 |
| Academic Hours | 190 |
| Year of study | 1 |
| Mode of study | Full-time |

1. **Pre-requisites**

This program of an academic discipline establishes minimum requirements for knowledge  
and skills of the student and determines the content and types of studies and reports.

1. **Course Type**

Advanced Methods of Data Analysis and Big Data in Business Intelligence is an *compulsory* course for first year master students enrolled on the program “Big Data Systems”.

1. **Abstract**

Advanced Data Analysis and Big Data for Business Intelligence is the study of the techniques for analyzing big data and big data technologies. Big data is the term for a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. The challenges include capture,сleaning, storage, search, sharing, transfer, analysis and visualization.Course is focused on understanding the role of big data analysis for business intelligence. Course content includes techniques for analyzing big data and big data technologies.

1. **Learning Objectives**

* Formation of the theoretical knowledge and practical basic skills in the collection, storage, processing and analysis of large data.
* Develop skills and practical skills to analyze large data to tackle a wide range of applications, including analysis of corporate data, financial data from the data warehousing world markets, modeling data storage and processing, prediction of complex indicators.

1. **Learning Outcomes**

As a result, during the studying of the discipline a student should:

* Understand the theory and fundamentals of storage, processing and analysis of big data, advanced tools for collection, storage, transmission and visualization of big data.
* To be able to process and analyze big amounts of data using modern software IBM Bluemix and R scripts.
* Have the skills to use neural networks and fuzzy models for compression, processing and analysis of large data.

1. **Course Plan**
   * 1. Introduction to the analysis and management of large data
     2. Data Management. Model of distributed file systems and databases computing
     3. MapReduce. Search for similarities in data
     4. Analysis of streaming data
     5. Link analysis. Page Rank
     6. PageRank. Topic-sensitive PageRank. Link Spam
     7. TrustRank and Spam Mass. Hubs and authorities.
     8. Clustering algorithms and their applications
     9. Neural networks and their applications in Big Data
     10. Fuzzy Logic. Notions and Applications
     11. Analysis of social network graphs
     12. Reducing the dimension of data
2. **Reading List**

Required

* Mayer-Schonberger V. Big data : a revolution that will transform how we live, work and think John Murray,2013
* Minelli M., Big data, big analytics : emerging business intelligence and analytic trends for today's businesses John Wiley & Sons, 2013

Optional

* Eaton C., Deutsch T., Deroos D., Lapis G., Zikopoulos P. Understanding Big Data.  
  Analytics for Enterprise Class Hadoop and Streaming Data
* Leskovec J., Rajaraman A., Jeffrey D. Ullman. Mining of Massive Datasets. Stanford  
  University, 2010
* Heaton C. Introduction to the Math of Neural Networks. Heaton Research, 2010
* F. Martin McNeill. Fuzzy Logic: a Practical Approach. AP Professional. 1994

1. **Grading System**

The student should demonstrate the knowledge of sections of the discipline and the ability to  
present the results of homework and tests in accordance with the required competencies.  
Evaluation of all forms of monitoring are set on a 10-point scale.

|  |  |
| --- | --- |
| **Grade** | **10-point scale** |
| Excellent | 10 |
| 9 |
| 8 |
| Good | 7 |
| 6 |
| Satisfactorily | 5 |
| 4 |
| Bad | 3 |
| 2 |
| 1 |
| 0 |

1. **Guidelines for Knowledge Assessment**

On the intermediate evaluation (Oi) on a subject matter consists of ratings for:  
•practical work – *O1*•exam (including oral and practical parts) – *O2*according to the formula: *Oi = 0.6\*O1+ 0.4\* O2*On the final evaluation (Of) on a subject matter consists of ratings for:  
•practical work (including control test) – *O3*•exam (including oral and practical parts) – *O4*according to the formula: *Of=0,3\*Oi+0,4\*O3+0,4\*O4*

1. **Methods of Instruction**

Presentation, cross-subject practices, collective practice work (programming and  
projects)

1. **Special Equipment and Software Support (if required)**

Lectures require an internet-connected PC and a projector. Practice sessions are carried out in computer-equipped classes. Each student should be provided with an individual PC with the necessary software installed (see previous section).