Opis **zajęć (sylabus)**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Title: | | Advanced data exploration techniques for big data | | | | | | | | **ECTS** | **4** |
| Translation if applicable: | | Advanced data exploration techniques for big data | | | | | | | | | |
| Field of study: | | **Informatics and econometrics** | | | | | | | | | |
|  | |  | | | | | | | | | |
| Language: | | English | | | | Level: | | II (master) | | | |
| Form: | 🗷 regular  🞎 weekend | Status: | 🞎 primary  🗷 specialisation | 🞎 mandatory  🗷 elective | | Semester: ……3….. | | 🗷 winter semester 🞎 summer semester | | | |
|  |  | Didactic year: | | | | 2019/2020 | Numer katalogowy: | **ZIM-IE-BDA-2S-03Z-23\_1** | | | |
|  | | | | | | | | | | | |
| Coordinator: | |  | | | | | | | | | |
| Lecturer: | |  | | | | | | | | | |
| Department: | |  | | | | | | | | | |
| Contracting department: | |  | | | | | | | | | |
| Assumptions, the goal and the description | | The aim of the course is to present advanced methods of data analysis for solving practical problems covering classification, regression and forecasting in the context of large and diverse data sets. The focus will be on data processing and analysis of complex datasets so the audience can tackle the idea of new knowledge and patterns extraction from massive data. The subject will discuss in detail techniques such as decision trees, artificial neural networks and advanced machine learning techniques including bagging, boosting and ensemble methods.Lectures:  * Introduction, basic concepts (3V, 4V), problems of analysis of large data volumes; * Data processing, discovery of inconsistencies, peculiarities and dependencies in large data repositories, methods of dimension reduction; * SAS technology and specificity of SAS 4GL, SAS SQL, SAS MACRO syntax * Introduction to problems of classification, regression and forecasting in business practice * Data mining techniques   - classification trees - basic algorithms, C4.5, C5.0, ID.3, CART, learning classification trees, tree division criteria, comparison of model validity;  - neural networks - history, neuron construction, construction of artificial neural networks of the RBF and MLP types, selected algorithms of learning artificial neural networks, optimization of neural network architecture, assessment of network performance;   * Advanced machine learning techniques – ensemble classifiers; * Application of data mining methods in computer science and telecommunications, finance, industry, medicine, marketing;   Computer labs:   * SAS technology and the solutions supporting the processing of large data volumes; * Data processing in SAS 4GL and SAS SQL language; * Automation of data extraction processes in SAS Macro language; * Data mining techniques and implementation in the SAS environment: decision trees and neural networks for classification and regression; * Advanced techniques of machine learning: groups of classifiers and predictors; * Case study on large data volumes (in terms of the number of observations and variables) | | | | | | | | | |
| Forms of didactics, hours: | | 1. Lectures: 15h; 2. Computer labs: 30h; | | | | | | | | | |
| Didactic methods: | | Presentations, discussion, practical examples, numerical experiments, case study | | | | | | | | | |
| Assumptions: | | Basic skills in data processing, information systems, databases, statistics and data analysis. | | | | | | | | | |
| Effects: | | Knowledge:  2 – Has knowledge about the process of collecting, processing, extracting data from database systems and data warehouses.  3 – Has advanced knowledge of basic algorithms of data mining (decision trees, neural networks, algorithms) and advanced algorithms (teams of classifiers) used to support decisions in business practice (in IT, industry, marketing)  2 – Has knowledge on mass data specifics and extracting it from source systems using IT technology and efficiently prepare it depending on the problem and the requirements. | | | Skills:  3 – Can use literature to gain required skills on analytical methods (data mining techniques)and to formulate and solve tasks at the interface between IT and business applications (in IT, industry, medicine, marketing).  3 – Is able to implement selected analytical techniques in information systems (including generating scoring lists, generating decision rules) using the right methods and IT tools.  2 – Has substantial skills in analysis of market phenomena and business processes, including the specificity of collecting, processing, extracting data from database systems and data warehouses. | | | | Competence: | | |
| The way of verifying the effects of education: | | Written test, project/case study | | | | | | | | | |
| Form of documentation of the learning outcomes achieved: | | Test to verify ability to process data and knowledge of data mining techniques, project realization | | | | | | | | | |
| Elements and weights of the final grade: | | **Written test – 50% (minimum 50% points to pass), Project/case study – 50%** | | | | | | | | | |
| Type of the classes | | Lectures + computer labs | | | | | | | | | |
| Literature:   1. Cukier K., Mayer-Schonberger V. (2014) Big data. A Revolution That Will Transform How We Live, Work, and Think, Houghton Mifflin Harcourt, New York. 2. Jared D. (2014) Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners, Wiley 3. Larose T.D. (2006) Discovering Knowledge in Data: An Introduction to Data Mining, Wiley-IEEE Press 4. Larose T.D. (2008) Data Mining Methods and Models, Wiley-IEEE Press | | | | | | | | | | | |
| Comments | | | | | | | | | | | |

Quantitative indicators characterizing the module:

|  |  |
| --- | --- |
| Estimated total number of student work hours (contact and own work) necessary to achieve the assumed learning outcomes | **100 h** |
| The total number of ECTS points that a student receives in classes requiring direct participation of academic teachers: | **2,5 ECTS** |

Effects to program mapping table:

|  |  |  |  |
| --- | --- | --- | --- |
| Effect category | Effects: | Effects to program mapping | Impact\*) |
| Knowledge 1 | Has knowledge about the process of collecting, processing, extracting data from database systems and data warehouses. | K\_W06 / P7S\_WG | 2 |
| Knowledge 2 | Has advanced knowledge of basic algorithms of data mining (decision trees, neural networks, algorithms) and advanced algorithms (teams of classifiers) used to support decisions in business practice (in IT, industry, marketing) | K\_W07 / P7S\_WG | 3 |
| Knowledge 3 | Has knowledge on mass data specifics and extracting it from source systems using IT technology and efficiently prepare it depending on the problem and the requirements. | K\_W10 / P7S\_WG | 2 |
| Skill 1 | Can use literature to gain required skills on analytical methods (data mining techniques)and to formulate and solve tasks at the interface between IT and business applications (in IT, industry, medicine, marketing). | K\_U01 / P7S\_UW | 3 |
| Skill 2 | Is able to implement selected analytical techniques in information systems (including generating scoring lists, generating decision rules) using the right methods and IT tools. | K\_U13 / P7S\_UW | 3 |
| Skill 3 | Has substantial skills in analysis of market phenomena and business processes, including the specificity of collecting, processing, extracting data from database systems and data warehouses. | K\_U15/ P7S\_UW | 2 |
|  |  |  |  |

\*)

3 – advanced,

2 – substantial,

1 – basic,