| Z1-PU7 | WYDANIE N1 |
|---|---|

(*faculty stamp*)

**COURSE DESCRIPTION**

| | |
|---|---|
| **1. Course title: DATA MINING, Knowledge discovery** | **2. Course code DM_KD** |
| **3. Validity of course description:** 2018/2019 | |
| **4. Level of studies:** MSc programme | |
| **5. Mode of studies:** intramural studies | |
| **6. Field of study**:<br>CONTROL, ELECTRONIC AND INFORMATION ENGINEERING (MACRO) | (FACULTY SYMBOL)<br>RAU-2 |
| **7. Profile of studies:** ACADEMIC | |
| **8. Programme:** DATA SCIENCE | |
| **9. Semester: 2** | |
| **10. Faculty teaching the course:** Faculty of Automatic Control, Electronics and Computer Science | |
| **11. Course instructor**: Dr hab. Marek Sikora prof. nzw. | |
| **12. Course classification:** common courses | |
| **13. Course status:** compulsory ~~/elective~~ | |
| **14. Language of instruction:** English | |
| **15. Pre-requisite qualifications:** Discrete Mathematics, Algorithms and data structures, Machine Learning, Soft Computing, Statistical Learning, | |
| **16. Course objectives:** The aim of the course is to make the students familiar with the methods of knowledge discovery in data (particularly in databases). The methods for building tree and rule based classification, regression, survival (survival analysis data models will be presented. The foundations of the rough set theory will be discussed along with its application in knowledge discovery. | |

**17. Description of learning outcomes:**

| Nr | Learning outcomes description | Method of assessment | Teaching methods | Learning outcomes reference code |
|---|---|---|---|---|
| 1. | Student knows methods of knowledge discovery based on rough set theory. Knows definitions of the rough set, reduct. | Exam | Lecture | K2A_W01, K2A_W03 |
| 2. | Student knows method of rule induction and rules for classification, regression and survival analyses tasks. | Exam | Lecture | K2A_W02, K2A_W03 |
| 3. | Student knows methods of induction of association rules. | Exam | Lecture | K2A_W02, K2A_W03 |
| 4. | Student knows methods of verification and selection of rule based data models. | Exam | Lecture | K2A_W02, K2A_W03 |
| 5. | Student is able to use the algorithm for induction trees and rules to tasks of description and classification of data. | Laboratory tasks | Laboratory | K2A_U14, K2A_U15 |

| | | | | |
|---|---|---|---|---|
| 6. | Student can perform selection of the most interesting rules. Can define subjective performance measures. | Laboratory tasks | Laboratory | K2A_U18, K2A_U19, K2A_U23 |
| 7. | Student can realize tasks of analysis based on association rules induction. | Laboratory tasks | Laboratory | K2A_U14, K2A_U23, K2A_K01 |
| 8. | Student knows libraries (R, Java) of algorithms of tree and rule induction. Can modify their contents to obtain specialized algorithms. | Laboratory tasks | Laboratory | K2A_U17, K2A_U18, K2A_K01 |
| 9. | | | | |

**18. Teaching modes and hours**

**Lecture 15 / ~~BA /MA Seminar / Class / Project~~ / Laboratory 15**

**19. Syllabus description:**

**Lecture:**

1. Rough Set Theory in Knowledge Discovery (rough sets, data reduction, decision rules and algorithm, exact and approximate reduct).
2. Classification tree and classification rule induction (divide-and-conquer approaches, separate-and-conquer approaches, splitting criteria, rule quality measures, pruning). Subgroup discovery.
3. Regression tree and regression rule induction (divide-and-conquer approaches, separate-and-conquer approaches, splitting criteria, rule quality measures, pruning)
4. Survival tree and survival rule induction (divide-and-conquer approaches, separate-and-conquer approaches, splitting criteria, rule quality measures, pruning, log-rank statistics)
5. Action rule induction (rule based action planning, meta-actions, actionability of data mining models)
6. Association rule induction (apriori, fp-growth, rule selection)
7. Rule interestingness measures (objective measures, subjective measures)

**Laboratory:**

1. Knowledge discovery – rough sets, classification and action rule induction – uses cases (industry
2. – seismic hazard assessment, good candidates for fighter pilots description and selection,
3. analysis of benchmark data sets from UCI Repository)
4. Knowledge discovery – regression and survival analysis – uses cases (medical data – bone
5. marrow transplantation, industry – methane forecasting in coal mines, retail – sales forecasting)
6. Association rule induction – use cases (rule selection and evaluation, market
7. basket analysis, analysis of benchmark data sets form UCI Repository)

**20. Examination:** semester 2

**21. Primary sources:**

Witten I. H., Frank E., Hail M.A: Data Mining. Practical Machine Learning Tools and Techniques. Fourth Edition. heory and its applications, Morgan Kaufmann / Elsevier 2017.
Pawlak Z.: Rough Sets – Theoretical Aspect of Reasoning about Data. Kluwer Academic Press 1991.

**22. Secondary sources:**

Cichosz P.: Data Mining Algorithms: Explained Using R, Wiley 2015.

**23. Total workload required to achieve learning outcomes**

| Lp. | Teaching mode : | Contact hours / Student workload hours |
|---|---|---|
| 1 | Lecture | 15/30 |
| 2 | Classes | / |
| 3 | Laboratory | 15/30 |
| 4 | Project | / |
| 5 | BA/ MA Seminar | / |
| 6 | Other | / |
| | Total number of hours | 30/60 |

**24. Total hours:** 90

**25. Number of ECTS credits:** 3

**26. Number of ECTS credits allocated for contact hours: 1**

**27. Number of ECTS credits allocated for in-practice hours (laboratory classes, projects): 2**

**26. Comments:**

Approved:

…….
*(date, Instructor's signature)*

……
(*date , the Director of the Faculty Unit signature)*