DESCRIPTION OF THE COURSE OF STUDY

Course code	**0719-2ID-F48-AED**	
Name of the course in	Polish	**Data exploration algortihms**
	English	**Algorytmy eksploracji danych**

1. LOCATION OF THE COURSE OF STUDY within the system of studies

1.1. Field of study	Data engineering
1.2. Mode of study	Undergraduate engineering study
1.3. Level of study	First-cycle engineering
1.4. Profile of study*	General academic
1.5. Person/s preparing the course description	Mgr Dariusz Pasieka
1.6. Contact	dariusz.pasieka@ujk.edu.pl

2. GENERAL CHARACTERISTICS OF THE course of study

2.1. Language of instruction	Polish
2.2. Prerequisites*	Discrete mathematics Statistics Probability theory

3. DETAILED CHARACTERISTICS OF THE COURSE OF STUDY

3.1. **Form of classes**		Lectures, laboratories, project
3.2. **Place of classes**		Courses in the UJK teaching rooms of the Faculty of Physics
3.3. **Form of assessment**		Lectures – graded credit Laboratories – graded credit Project – graded credit
3.4. **Teaching methods**		Verbal, visual, practical
3.5. **Bibliography**	**Required reading**	1. T. Morzy, *Eksploracja danych*, Wydawnictwo Naukowe PWN, Warsaw, 2013
	Further reading	1. J. Koronacki, J. Ćwik, Statystyczne systemy uczące się, Akademicka Oficyna Wydawnicza EXIT Andrzej Lang, Warsaw, 2015 2. T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer, 2009 3. A. Géron, Uczenie maszynowe z użyciem Scikit-Learn i TensorFlow, 3rd ed., Helion, Gliwice, 2023

4. OBJECTIVES, SYLLABUS CONTENT AND INTENDED LEARNING OUTCOMES

4.1. Course objectives *(including form of classes)*

Lectures

C1. To familiarize students with fundamental concepts, definitions, and objectives of data exploration, as well as the challenges related to the analysis of large datasets.

C2. To present the main tasks of data exploration, including classification, regression, clustering, rule discovery, and semantic analysis.

C3. To discuss principles of model construction, quality assessment, and validation, including testing and cross-validation methods.

C4. To introduce selected algorithms and techniques of data exploration, in particular Bayesian classifiers, decision trees, and association rule algorithms.

Laboratories

C1. To develop practical skills in working with data in Python using data analysis libraries.

C2. To master preprocessing and transformation techniques, such as encoding, normalization, standardization, and dimensionality reduction.

C3. To enhance skills in implementing and testing classifiers and regression methods, as well as in evaluating their performance.

C4. To gain experience in applying clustering, association rule mining, and semantic analysis methods to practical data exploration problems.

Project

C1. To develop the ability to independently analyze and interpret real-world data by carrying out a team project

using data exploration methods.

4.2. **Detailed syllabus** ***(including form of classes)***

Lectures

1. Introduction to data exploration: definitions, objectives, challenges
2. Types of data exploration tasks: distribution identification, preprocessing and transformation, classification, regression, rule and pattern discovery
3. Clustering and data segmentation
4. Assessing the quality of classifiers and regression models: testing, cross-validation
5. Bayesian classifiers: naïve Bayes assumption, optimal Bayesian classifier
6. Decision trees: construction, impurity measures, pruning
7. Association rule discovery: support, confidence, Apriori algorithm
8. Semantic analysis, natural language processing

Laboratories

1. Data analysis tools: working with Python, scikit-learn, and pandas; loading datasets; public repositories
2. Data preprocessing: one-hot encoding, normalization, standardization
3. Naïve Bayes classifier
4. Classifier testing: cross-validation, majority voting, confusion matrix
5. Classifiers: decision trees, random forests, k-nearest neighbors
6. Principal Component Analysis, Linear Discriminant Analysis – data visualization applications
7. Predicting continuous target variables using regression analysis
8. Association rule mining (implementation), Apriori variants
9. Text data processing with semantic analysis: Latent Dirichlet Allocation
10. Working with unlabeled data: clustering analysis

Project

1. Students, working in teams, develop their own data analysis using data exploration methods and provide interpretation of the results.

4.3 Intended learning outcomes

Code	A student, who passed the course	Relation to learning outcomes
	within the scope of **KNOWLEDGE:**	
W01	distinguishes fundamental concepts, objectives, and tasks of data exploration, and is familiar with selected IT technologies applied in data analysis	ID1A_W06
W02	knows selected algorithms for data analysis and exploration, including classification, regression, clustering, and rule discovery	ID1A_W08
	within the scope of **ABILITIES:**	
U01	is able to use software and IT tools for data collection, preprocessing, visualization, and interpretation .	ID1A_U06 ID1A_U07
U02	can design and implement database systems, selecting appropriate programming tools, including their structure and transactions.	ID1A_U07 ID1A_U08
U03	is able to implement and test selected data exploration algorithms using a high-level programming language .	ID1A_U07
U04	is able to analyze and solve complex problems related to data exploration and select appropriate methods under partially unpredictable conditions	ID1A_U08
	within the scope of **SOCIAL COMPETENCE:**	
K01	can apply business and social requirements in a project.	ID1A_K03 ID1A_K04
K02	is aware of the need for continuous modification of the implemented project in both substantive and social aspects.	ID1A_K04

4.4. **Methods of assessment of the intended learning outcomes**

Teaching outcomes ***(code)***	**Method of assessment (+/-)**																				
	Exam oral/written*			**Test***			**Project***			**Effort in class***			**Self-study***			**Group work***			**Others* e.g. standardized test used in e-learning**		
	Form of classes			***Form of classes***			***Form of classes***			***Form of classes***			***Form of classes***			***Form of classes***			***Form of classes***		
	L	*C*	*P*	*L*	*C*	*P*	*L*	*C*	*P*	*L*	*C*	*P*	*L*	*C*	*P*	*L*	*C*	*P*	*L*	*C*	*P*

																				.
W01									+	+										
W02									+	+										
U01								+					+				+			
U02								+					+				+			
U03								+					+				+			
U04								+					+				+			
K01								+									+			
K02								+									+			

**delete as appropriate*

4.5. **Criteria of assessment of the intended learning outcomes**		
Form of classes	**Grade**	**Criterion of assessment**
lecture (L) (including e-learning)	**3**	achievement of 50–60% of assessment requirements
	3,5	achievement of 61–70% of assessment requirements
	4	achievement of 71–80% of assessment requirements
	4,5	achievement of 81–90% of assessment requirements
	5	achievement of 91–100% of assessment requirements
classes (C)* (including e-learning)	**3**	achievement of 50–60% of assessment requirements
	3,5	achievement of 61–70% of assessment requirements
	4	achievement of 71–80% of assessment requirements
	4,5	achievement of 81–90% of assessment requirements
	5	achievement of 91–100% of assessment requirements
others (...)* (including e-learning)	**3**	achievement of 50–60% of assessment requirements
	3,5	achievement of 61–70% of assessment requirements
	4	achievement of 71–80% of assessment requirements
	4,5	achievement of 81–90% of assessment requirements
	5	achievement of 91–100% of assessment requirements

5. BALANCE OF ECTS CREDITS – STUDENT'S WORK INPUT

Category	**Student's workload**	
	Full-time studies	**Extramural studies**
NUMBER OF HOURS WITH THE DIRECT PARTICIPATION OF THE TEACHER /CONTACT HOURS/	***90***	
*Participation in lectures**	30	
*Participation in classes, seminars, laboratories**	30	
Projekt	30	
INDEPENDENT WORK OF THE STUDENT/NON-CONTACT HOURS/	***60***	
*Preparation for the lecture**	10	
*Preparation for the classes, seminars, laboratories**	20	
*Preparation for the exam/test**	15	
*Gathering materials for the project/Internet query**	15	
TOTAL NUMBER OF HOURS	**150**	
ECTS credits for the course of study	***6***	

**delete as appropriate*

Accepted for execution *(date and legible signatures of the teachers running the course in the given academic year)*

.....