

| | | | | | | | | |
|---|---|----------------|------------------|-------------------|----------------|----------------|--------------|--|
| Course code | | | | | | | | |
| Type and description | EC | | | | | | | |
| ECTS credit | 1 | | | | | | | |
| Course name | Data exploration | | | | | | | |
| Course name in Polish | Eksploracja danych | | | | | | | |
| Language of instruction | English | | | | | | | |
| Course level | 8 PRK | | | | | | | |
| Course coordinator | Adam Pelikant | | | | | | | |
| Course instructors | Adam Pelikant, Dorota Kamińska | | | | | | | |
| Delivery methods and course duration | | Lecture | Tutorials | Laboratory | Project | Seminar | Other | Total of teaching hours during semester |
| | Contact hours | | | | 15 | | | 15 |
| | E-learning | No | No | No | No | No | No | |
| | Assessment criteria (weightage) | | | | 100% | | | |
| Course objective | The course aims are data exploration algorithms and their practical implementation and use in the analysis of data from transactional systems and data warehouses. | | | | | | | |
| Learning outcomes | 1) describe algorithms used in the data exploration (W4, U4, K1) 2) classify the algorithms used in the data mining (W4, U4, K1) 3) explain the principles of software that implements these algorithms (W4, U4, K1) 4) operate data mining software to solve the problem (W4, U4, K1) | | | | | | | |
| Assessment methods | Evaluation of algorithms implemented in the final project | | | | | | | |
| Prerequisites | Basis of computer science, basis of discrete mathematics, basis of programming, basis of databases | | | | | | | |
| Course content with deliv- | Mathematical foundations of exploration algorithms. | | | | | | | |

| | |
|--|--|
| <p>ery methods</p> | <p>Binary decision trees in data mining and their ordering</p> <p>Decision trees based on information gain (Shanon, Gini, Bayes)</p> <p>Linear, hierarchical, non-linear classification - implementation of the naive Bayes algorithm</p> <p>Clustering - sharp and fuzzy algorithms of clustering K-means, C-means.</p> <p>Creating clusters without defining their number - mountain algorithm, nearest neighbors, density algorithms.</p> <p>Linear and non-linear regression methods - extrapolation, logistic regression. Sequence clustering.</p> <p>Testing the data mining model and methods for estimating the quality of the classification.</p> <p>Basic concepts in the field of data mining at the level of relational structures</p> <p>Review of algorithm implementation in commercial products in terms of data mining requirements</p> <p>Use of the SQL extension for data mining (DMX SQL).</p> <p>The use of script languages in data mining (Python, R, Weka...)</p> <p>Creating and testing the model. Estimation of quality and inference from the model</p> |
| <p>Basic reference materials</p> | <p>Szeliga Marcin, Data Science i uczenie maszynowe, Wydawnictwo Naukowe PWN 2017</p> <p>Rutkowski L., Metody i techniki sztucznej inteligencji, Wydawnictwo Naukowe PWN, Warszawa, 2019</p> <p>Paweł Cichosz: Systemy uczące się, Wydawnictwa Naukowo-Techniczne, Warszawa, 2000</p> <p>Claude Seidman: Zgłębianie i analiza danych z Microsoft SQL Server 2000, PROMISE 2002</p> <p>Hand D, Mannila H, Smyth P: Eksploracja danych WNT 2005</p> <p>Trueblood R. P., Lovett J. N.: Zastosowanie języka SQL do analizy statystycznej i eksploracji danych, Mikom, Warszawa 2002.</p> <p>Kwiatkowska A. M.: Systemy wspomaganie decyzji, Jak korzystać z wiedzy i informacji, PWN, 2007</p> |
| <p>Other reference materials</p> | <p>Włodzisław Duch - materiały internetowe http://www.is.umk.pl/~duch/Wyklady/index.html</p> <p>Sebastian Raschka, Python. Uczenie maszynowe, Helion 2017</p> <p>A. Pelikant: Hurtownie danych. Od przetwarzania analitycznego do raportowania, Helion 2011</p> <p>Inteligentne systemy w zarządzaniu, teoria i praktyka, pod red. J. S. Zielińskiego, PWN, Warszawa, 2000</p> |
| <p>Average student workload outside classroom</p> | <p>10 h</p> |

| | |
|-----------------|------------|
| Comments | |
| | 25.04.2023 |