GDAŃSK UNIVERSITY OF TECHNOLOGY

Subject card

Subject name and code	Business Data Analytics, PG_00053096		
Field of study	Data Engineering		
Date of commencement of studies	October 2022	Academic year of realisation of subject	2024/2025
Education level	first-cycle studies	Subject group	Optional subject group Subject group related to scientific research in the field of study
Mode of study	Full-time studies	Mode of delivery	at the university
Year of study	3	Language of instruction	English
Semester of study	5	ECTS credits	4.0
Learning profile	general academic profile	Assessment form	assessment
Conducting unit	Department of Informatics in Management -> Faculty of Management and Economics		
Name and surname of lecturer (lecturers)	Subject supervisor	dr Nina Rizun	
	Teachers	mgr Jaromir Durkiewicz dr Nina Rizun	

Lesson types and methods of instruction

Lesson type	Lecture	Tutorial	Laboratory	Project	Seminar	SUM
Number of study hours	15.0	0.0	30.0	0.0	0.0	45

E-learning hours included: 0.0

Learning activity and number of study hours

Learning activity	Participation in didactic classes included in study plan	Participation in consultation hours	Self-study	SUM
Number of study hours	45	5.0	50.0	100

Subject objectives

The aim of the course is to familiarize students with basic data mining algorithms in the context of discovering knowledge from unstructured data of business organizations.

Learning outcomes

Course outcome	Subject outcome	Method of verification
[K6_U01] programs in procedural, object, functional and logic programming languages, codes programs at the processor instruction level, runs and tests programs.	The student is able to develop the code program to be carried out simulation, analysis and presentation intelligent data mining	[SU1] Assessment of task fulfilment [SU2] Assessment of ability to analyse information [SU3] Assessment of ability to use knowledge gained from the subject [SU4] Assessment of ability to use methods and tools
[K6_W08] Knows the models and structure of the data mining process and their multidimensional analysis and can assess the results of such analyses	Is able to assess the accuracy of data outputs, choose effective ones exploration methods to solve specific problem business, he can interpret and present data exploitation results	[SW2] Assessment of knowledge contained in presentation [SW3] Assessment of knowledge contained in written work and projects
[K6_K05] understands the need for self-improvement through systematic acquisition of knowledge and skills.	The student knows the notation modeling of business processes and their application in the functioning of enterprises	[SK1] Assessment of group work skills [SK2] Assessment of progress of work

Subject contents

1. What is Business Data Analytics? CRISP-DM. Data preparation. Data mining methods
2. What is Text Analytics (BTA)? Text Analytics & NLP. Text Analytics tasks. Search evaluation metrics. Keyword based search. Search based on vector representation. Techniques for pre-processing text documents. Zipf's law
3. Corpus. Properties of vector representation. Determining the frequency matrix (Calculating weights). Binary representation. Inverse-document frequency (IDF). TFIDF weighting scheme.Vector representation of a document. Distance/similarity measurement. Distance calculation methods. Similarity calculation methods. Cluster analysis grouping. Hierarchical agglomerative methods. Example of the Hierarchical Method. K-optimization methods.
4. Cosine similarity. Euclidean distance vs. Cosine similarity. Adjacency matrix and visualization. Similarities using graphs. Adjacency matrix. The degree of a graph node. Network community structure.
5. Multidimensional Scaling (MDS): Motivation. Goals. Formal model. Stress (Goodness-of-fit). Latent Semantic Group Analysis (LSA): Motivation. Vector model: Limitations. Singular Value Decomposition. LSA dimension reduction. Similarity of documents. Similarity of words. Automatic topic modeling of text data. Latent Dirichlet Allocation (LDA).
6. Introduction to Sentiment Analysis
7. Structural Topic Modeling
8. SNA - introduction. History of social network analysis theory. Small world theory. Milgram experiment. Relationships as a network. Directed graph. Undirected graph. Edge attributes. Graph density.Reachability and distance. Centrality measures

Prerequisites and co-requisites

Assessment methods and criteria

Subject passing criteria	Passing threshold	Percentage of the final grade
Project	60.0%	40.0%
Final test	60.0%	40.0%
Excercises classes	60.0%	20.0%

Recommended reading

Basic literature

1. Provost, Foster, Fawcett, Tom. Data Science for Business: What you need to know about data mining and data-analytic thinking. " O'Reilly Media, Inc.", 2019
2. Sharda, R., Delen, D., & Turban, E. Business intelligence: a managerial perspective on analytics. Pearson. 2018
3. Siegel, E. Predictive analytics: The power to predict who will click, buy, lie, or die. John Wiley & Sons, 2016. 338 pp.
4. James, G., Witten, D., Hastie, T., Tibshirani, R. An introduction to statistical learning with applications in R. Springer, 2018.

Supplementary literature

1. Han, J., Kamber, M., Pei, J. Data mining concepts and techniques. Morgan Kaufmann, 2021.
2. Hastie, T., Tibshirani, R., Friedman, J. The Elements of Statistical Learning, 2nd edition. Springer, 2022.
3. James, G. et al. An introduction to statistical learning. Springer, 2013. 426 pp.
4. Murphy, K. Machine Learning: A Probabilistic Perspective. MIT Press, 2019

eResources addresses

Podstawowe

https://enauczanie.pg.edu.pl/moodle/course/view.php?id=40079 - The Business Data Analytics e-course contains all the necessary didactic materials (lecture notes) and additional literature for independent reading.

Adresy na platformie eNauczanie:

Business Data Analytics-2024 /2025 - Moodle ID: 40079
https://enauczanie.pg.edu.pl/moodle/course/view.php?id=40079

Example issues/ example questions/ tasks being completed

- Select the main properties of Euclidean distance
- Select the appropriate cosine similarity measure value between document 1 and document 2
- K-means algorithm allows...- What is the difference between Euclidean distance and cosine similarity measure?
- A person with high eigenvector centrality is..
- Please calculate the IDF for terms from the following Corpus
- Construct a Zipf's law graph for the following Corpus
- Give a general description of the actors of the following social network

Work placement	Not applicable

Document generated electronically. Does not require a seal or signature.