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 2020	Academic year of realisation of subject	2022/2023
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	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

Business Data Analytics-2022 - Moodle ID: 23765
https://enauczanie.pg.edu.pl/moodle/course/view.php?id=23765

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 the basic algorithms of data mining in context discovering knowledge from unstructured data of public and private organizations.

Learning outcomes

Course outcome	Subject outcome	Method of verification
[K6_K05] understands the need for self-improvement through systematic acquisition of knowledge and skills.	The student understands the need for self-improvement by updating knowledge on data mining	[SK2] Assessment of progress of work [SK1] Assessment of group work skills
[K6_W08] Knows the models and structure of the data mining process and their multidimensional analysis and can assess the results of such analyses	The student is able to assess the accuracy of the output data, select effective mining methods to solve a specific business problem, and can interpret and present the results of data mining	[SW2] Assessment of knowledge contained in presentation [SW3] Assessment of knowledge contained in written work and projects
[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 a program code for simulation, analysis and presentation of intelligent data mining	[SU2] Assessment of ability to analyse information [SU4] Assessment of ability to use methods and tools [SU5] Assessment of ability to present the results of task

<table>
<tr><td>Subject contents</td><td>1. What is Business Data Analytics? Data can help us solve specific problems. KDD Process. What techniques do people apply on data? Data Mining: the Dark Side

2. What is Business Text Analytics (BTA)? Why is BTA relevant? Why do we study it? Application domains. The complexity of unstructured text (the origin of BTA challenges). Bag-of-words representation of text. Vector Space Model. Methods/techniques for text pre-processing. Zipfs Law

3. Term frequency and weighting. TF-IDF weighting. Clustering: Theoretical background. Euclidian Distance. Hierarchical Clustering. K-means Clustering

4. Vector Space Model. Cosine Similarity. Similarity Measure for Text Mining. Tf-IDF and Cosine similarity Example. Adjacency Matrix. Cosine Similarity as a Weight of Graph Edges. Community Detection Algorithms Within the Cosine Similarity Graph

5. Latent Semantic Analysis: Motivation. Singular Value Decomposition (SVD). LSA-based Similarity Search. Latent Dirichlet allocation. Topic modeling

6. Introduction to Sentiment Analysis

7. Structural Topic Modeling

8. Social Network Analysis history. Networks in the Real World. Network Study. Basic Concepts of Network Theory. Representing Relations as Networks. Walks, Paths and Trails. Centrality in Social Network</td></tr>
<tr><td>Prerequisites and co-requisites</td><td>Knowledge of statistical methods</td></tr>
<tr><td>Assessment methods and criteria</td><td>
<table>
<tr><th>Subject passing criteria</th><th>Passing threshold</th><th>Percentage of the final grade</th></tr>
<tr><td>Project</td><td>60.0%</td><td>40.0%</td></tr>
<tr><td>Final test</td><td>60.0%</td><td>40.0%</td></tr>
<tr><td>Excercises classes</td><td>60.0%</td><td>20.0%</td></tr>
</table>
</td></tr>
<tr><td>Recommended reading</td><td>
<table>
<tr><td>Basic literature</td><td>1. Arkadiusz Januszewski; Funkcjonalność Informatycznych systemów zarządzania - Zintegrowane systemy transakcyjne; PWN W-wa 2008 2. pod red. Stanisław Wrycza; Informatyka ekonomiczna; PWE Warszawa 2010 3.

2. Pollak, B. (Ed.): Ultra-Large-Scale Systems, 150 pages, ISBN: 0-9786956-0-7, June 2006.

3. Report of a Workshop on The Scope and Nature of Computational Thinking, Committee for the Workshops on Computational Thinking; National Research Council 126 pages, ISBN-10: 0-309-14957-6, 2010.

4. http://books.nap.edu/openbook.php?record_id=12840&page=2</td></tr>
<tr><td>Supplementary literature</td><td>1. Zander, J., Mosterman,P.J., et al.: On the Structure of Time in Computational Semantics of a Variable-Step Solver for Hybrid Behavior Analysis, 18th World Congress of the International Federation of Automatic Control (IFAC), Milano, Italy, 2011.

2. Kelly, K., On the next 5,000 days of the web, TED Talk, 2007.</td></tr>
<tr><td>eResources addresses</td><td></td></tr>
</table>
</td></tr>
</table>

Example issues/ example questions/ tasks being completed	*Examples of Test Questions:* - Choose the main properties of Euclidian distance - Choose the right value of the Cosine similarity measure between Document 1 and Document 2 - K-means algorithm allows ... - What's the difference between Euclidian distance and Cosine similarity measure? - Person with high Eigenvector centrality is ... *Examples of Open Questions*: - Please calculate the IDF for the terms from the following Corpus - Build the Zipfslaw chart for the following Corpus - Give the general characteristics of the actors of following Social network
Work placement	Not applicable