GDAŃSK UNIVERSITY OF TECHNOLOGY

Subject card

Subject name and code	Business Data Analytics, PG_00053096		
Field of study	Data Engineering, Data Engineering		
Date of commencement of studies	October 2026	Academic year of realisation of subject	2028/2029
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 -> Faculties of Gdańsk University of Technology		
Name and surname of lecturer (lecturers)	Subject supervisor	dr Nina Rizun	
	Teachers	dr Nina Rizun mgr Jaromir Durkiewicz	

Lesson types

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 analytics algorithms in the context of discovering knowledge from unstructured data of business organizations.

Learning outcomes

Course outcome	Subject outcome	Method of verification
[K6_K03] demonstrates the ability to think critically and analytically and integrates knowledge from many disciplines in order to make effective decisions	The student demonstrates the ability to critically interpret data analysis results, integrates knowledge from data analytics, economics, and management, and applies it to formulate conclusions and support business decision-making.	[SK2] Assessment of progress of work [SK4] Assessment of communication skills, including language correctness [SK3] Assessment of ability to organize work
[K6_W03] identifies veracious sources of information relevant to the analyzed issues	The student identifies, acquires, and critically evaluates data from various sources (e.g., databases, business reports, open datasets) and applies them to analyze and interpret business problems using basic data analytics tools.	[SW3] Assessment of knowledge contained in written work and projects [SW2] Assessment of knowledge contained in presentation
[K6_U06] acquires new knowledge, planning its own development in aiming at achieving defined goals	The student independently expands their knowledge and competencies in data analytics by searching for specialized materials, analytical tool documentation, and business application examples, and uses them to plan their professional development in data analysis and data-driven decision-making.	[SU2] Assessment of ability to analyse information [SU1] Assessment of task fulfilment [SU3] Assessment of ability to use knowledge gained from the subject [SU4] Assessment of ability to use methods and tools [SU5] Assessment of ability to present the results of task

Subject contents

Course content – lecture
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

Course content – laboratory
- Basic Text Analytics in RFile
- TF-IDF transformation. Cluster analysis using RFile
- Cosine similarity. Network graphs using R
- Topic Modeling. Multidimentional Scaling and Latent Semantic Analysis
- Opinion Mining and Sentiment Analysis
- Structural Topic Modelling
- Social Network Analysis. Basic Network Measures

Prerequisites and co-requisites

Assessment methods and criteria

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

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

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

Practical activites within the subject	Not applicable

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