University of Gdańsk

Subject card

Subject name and code	Data Mining – Data Sets Exploration, PG_00178106						
Field of study	Informatics and Econometrics						
Date of commencement of studies	October 2026		Academic year of realisation of subject		2027/2028		
Education level	Bachelor's studies		Subject group		Obligatory subject group in the field of study 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	2		Language of instruction		Polish		
Semester of study	4		ECTS credits		7.0		
Learning profile	academic		Assessment form		exam		
Conducting unit	Department of Statistics -> Faculty of Management -> Rector						
Name and surname of lecturer (lecturers)	Subject supervisor		dr hab. Kamila Migdał-Najman				
	Teachers						
Lesson types	Lesson type	Lecture	Tutorial	Laboratory	Project	Seminar	SUM
	Number of study hours	30.0	30.0	15.0	0.0	0.0	75
	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	75		4.0		96.0	175
Subject objectives	The aim of the course is to familiarize students with the basic methods of Exploratory Data Analysis and its practical applications.						

Learning outcomes

Course outcome	Subject outcome	Method of verification
[IiEL3_U02] Can select or construct econometrics, informatics or statistics tools and apply them to describe and solve economic and social problems	The student is able to identify, select, and develop statistical, econometric, and IT tools used in exploratory data analysis. They can analyze structures, relationships, and trends in data related to economic and social phenomena. The student can formulate conclusions supporting analytical and decision-making processes and verifiy the effectiveness of the applied tools.	[SU2] presentation/project/paper/ report
[IiEL3_U03] Can obtain data from appropriately selected sources, use these data to solve economic and social problems, and process and interpret them using econometrics, informatics or statistics tools	The student is able to identifiy and verifiy appropriate data sources related to economic and social phenomena. They can analyze and interpret the acquired data using exploratory data analysis methods and statistical, IT, and econometric tools. The student can organize, process, and visualize data to support problem diagnosis and rational decision-making processes.	[SU2] presentation/project/paper/ report
[IiEL3_W06] Knows and understands, to an advanced degree, the processes and methods of creating, developing, and providing appropriate conditions for using informatics or statistics tools, particularly those that improve human and organizational functioning	The student knows and understands processes and techniques for creating, developing, and ensuring optimal conditions for using exploratory data analysis tools, including statistical and IT tools. They know and can verify their effectiveness in supporting data analysis and interpretation and assess the impact on the efficiency of human and organizational performance. The student understands principles for the optimal use of these tools.	[SW4] test/exam - oral or written [SW2] presentation/project/paper/ report

Subject contents

Basic concepts, objectives, and functions of data exploration. Exploratory Data Analysis (EDA), data mining, Knowledge Discovery in Databases (KDD), data exploration as a process, stages of the data exploration process. Multidimensional data sets, the concept of Big Data. Analytical software. Tasks of data exploration: description, estimation, prediction, classification, clustering, rule discovery. Examples of data exploration applications in socio-economic research. Data sources. Preliminary data processing: data preparation for analysis, the problem of selecting variables and cases, dimensionality reduction of the feature space, the problem of missing data, graphical methods for identifying outliers, analytical methods for identifying outliers. Procedures for variable and feature normalization. Basic tabular and graphical EDA methods. Partial and multiple correlation. Dependence matrices. The process of data clustering and classification. Object similarity and its measurement. Measurement of object diversity based on quantitative features distance measures. Measurement of similarity of objects based on quantitative features. Similarity of features and its measurement. Measurement of similarity of quantitative features based on correlation. Hierarchical clustering. The essence of agglomerative hierarchical clustering. Nearest neighbor method, furthest neighbor method, group average method, centroid method, median method, Wards method. LanceWilliams algorithm. Visualization of group structure. Evaluation of clustering based on dendrograms. Selected EDA methods: classification trees, regression trees, k-nearest neighbors algorithm.

Prerequisites and co-requisites

Descriptive Statistics, Mathematical Statistics, Mathematics

Assessment methods and criteria

Subject passing criteria	Passing threshold	Percentage of the final grade
Semester project	51.0%	50.0%
Written exam	51.0%	50.0%

Recommended reading

Basic literature	1. M. Lasek, *Data Mining*, Manager and Bankers Library, Warsaw, 2002 2. D. T. Larose, *Discovering Knowledge from Data: An Introduction to Data Mining*, PWN Scientific Publishing, Warsaw, 2006 3. E. Gatnar, *Nonparametric Methods of Discrimination and Regression* , PWN Scientific Publishing, Warsaw, 2001 4. A. Balicki, *Multivariate Statistical Analysis and Its Socio-Economic Applications*, University of Gdańsk Publishing, Gdańsk, 2009
Supplementary literature	Nong Ye, The Handbook of Data Mining, Lawrence Erlbaum Associates, New Jersey, 2003

	eResources addresses	
Example issues/ example questions/ tasks being completed		
Work placement	Not applicable	

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