

DATA MINING FOR BUSINESS INTELLIGENCE

Course code	<i>IT101</i>
Compulsory in the programmes	<i>Economics and Data Analytics</i>
Level of studies	<i>Undergraduate</i>
Number of credits and hours	<i>6 ECTS (48 contact hours + 6 consultation hours + 2 exam hours, 106 individual work hours)</i>
Course coordinator	<i>Dr. Simonas Čepėnas, Haroldas Mackevičius,</i>
Prerequisites	<i>Statistical Data Analysis, Mathematical Analysis, Computer Programming, Econometrics</i>
Language of instruction	<i>English</i>

THE AIM OF THE COURSE:

Modern technologies enable business and institutions to generate vast amounts of data, and the ability to put that data to good use yields a major competitive advantage. Data can be used to answer critical business questions, improve the efficiency of operations, or directly enhance products and services. To be utilized, the data needs to be collected, stored, pre-processed, analyzed and modelled using statistical techniques. Data scientists are specialists who analyze and model data.

The goal of this course is to provide the necessary technical expertise for extracting patterns from data. Specifically, the course focuses on 5 core topics: (1) exploratory data analysis; (2) building supervised predictive models; (3) unsupervised learning; (4) experimentation; and (5) model deployment. Having went through these topics, students will be able to apply data science techniques in their workplaces and individual projects. Moreover, having a good grasp of these topics will prove useful in any modern company, regardless of occupation.

The course heavily focuses on predictive modeling using Python. Therefore, foundational knowledge in statistics, mathematical analysis, and Python programming is assumed.

MAPPING OF COURSE-LEVEL LEARNING OUTCOMES (OBJECTIVES) WITH DEGREE LEVEL LEARNING OBJECTIVES (See Annex), ASSESSMENT AND TEACHING METHODS

Course level learning outcomes (objectives)	Degree level learning objectives (Number of LO)	Assessment methods	Teaching methods
CLO1. Understand the key elements of relational databases and data storage. Extract data from databases by writing SQL queries.	ELO1.1 ELO4.3	Final exam, Mid-term exam, group project	Lectures, seminars, independent work
CLO2. Perform exploratory data analysis on tabular data using relevant Python packages. Visualize the relationship between variables.	ELO1.1 ELO3.1 ELO3.2 ELO4.3	Final exam, Mid-term exam, group project	Lectures, seminars, independent work
CLO3. Understand the differences between supervised and unsupervised models, the bias-variance trade-off, and train-test	ELO1.1	Final exam, Mid-term exam,	Lectures, seminars,

splits. Critically evaluate which models are best suited for specific tasks.	ELO4.3	group project	independent work
CLO4. Understand and apply linear regression, including shrinkage methods for variable selection and addressing high-dimensional data problems. Understand how to measure the performance of a linear regression model.	ELO1.1 ELO4.3	Final exam, Mid-term exam, group project	Lectures, seminars, independent work
CLO5. Understand and apply classification models for modeling binary and non-binary response variables. Understand how to measure the performance of a classification model.	ELO1.1 ELO4.3	Final exam, Mid-term exam, group project	Lectures, seminars, independent work
CLO6. Understand and apply tree-based methods for both classification and regression problems.	ELO1.1 ELO4.3	Final exam, Mid-term exam, group project	Lectures, seminars, independent work
CLO7. Understand the importance of experimentation in modern organizations.	ELO1.1	Final exam, Mid-term exam, group project	Lectures, seminars, independent work
CLO8. Write clear, reproducible, and well-documented code in Python using the Jupyter Notebook environment. Be able to use the most relevant Python packages for data wrangling.	ELO1.1 ELO3.1 ELO3.2 ELO4.3	Final exam, Mid-term exam, group project	Lectures, seminars, independent work

ACADEMIC HONESTY AND INTEGRITY

The ISM University of Management and Economics Code of Ethics, including cheating and plagiarism, is fully applicable and will be strictly enforced in the course. Academic dishonesty and cheating can and will lead to a report to the ISM Committee of Ethics. With regard to remote learning, ISM reminds students that they are expected to adhere to and maintain the same academic honesty and integrity that they would in a classroom setting.

COURSE OUTLINE

Week	Topic	In-class hours	Readings
I. High-level overview			
1.	Introduction to data science.	4	James et al. (2023), Chapter 1
2.	Exploratory data analysis.	4	James et al. (2023), Chapter 2
II. Supervised Learning			
3.	Linear regression.	4	James et al. (2023) Chapters 2.2, 3, 6, 7.1
4.	Classification.	4	James et al. (2023), Chapter 4
5.	Model Evaluation.	4	James et al. (2023), Chapter 5

6.	Tree-based Methods.	4	James et al. (2023), Chapter 8
7.	<i>Midterm exam</i>	2	
8.	Other topics in supervised learning.	4	Will be provided during the lectures
III. Deep Learning and Unsupervised Learning			
9.	Introduction to deep learning.	4	James et al. (2023) Chapter 10.1
10.	Unsupervised learning.	4	James et al. 2023 Chapter 12
IV. Experimentation and Model Deployment			
11.	Experimentation.	4	Will be provided during the lectures
12.	Model deployment.	4	Will be provided during the lectures
13.	Course review.	2	
		Total: 48 hours	
	CONSULTATIONS	6	
	FINAL EXAM	2	

FINAL GRADE COMPOSITION

Type of assignment	%
<i>Group Components 20%</i>	
Group project	20%
<i>Individual Components, 80%</i>	
Homework 1	10%
Homework 2	10%
Homework 3	10%
Mid-term exam	25%
Final exam	25%
Total:	100%

DESCRIPTION AND GRADING CRITERIA OF EACH ASSIGNMENT

Group project. In the group project, students will have to prepare a report detailing the analysis they have performed. Details on the group project will be provided later in the course. Broadly, students can expect to perform any or all of the following: (i) data extraction from a database, (ii) predictive modeling; (iii) evaluation of the model(s), and (iv) presentation of the results through data visualization. The group sizes are expected to be between 3-4 people. There will be one such project worth 20% of the final grade.

Homework. Students will be assigned homework tasks to be completed with Python. The homework tasks will include applying the knowledge gained during the class in the Jupyter Notebook environment. This will include completing tasks such as performing exploratory data analysis, fitting supervised and unsupervised statistical models, fine-tuning the models and evaluating results. It counts towards 30% of the final grade. There will be 3 homework assignments.

Mid-term exam. The mid-term exam will be held during the midterm exam session. It counts towards 25% of the final grade. The midterm will be based on topics 1-6. The midterm will consist of theoretical and practical questions (both multiple-choice and open-ended).

Final exam. The final examination will take place during the final examination session. It counts towards 25% of the final grade. The exam will be primarily based on topics 7-11, although it might include questions from the previous chapters as well. The midterm will consist of theoretical and practical questions (both multiple-choice and open-ended).

Retake Exam. The retake comprises 50% of the final grade and will consist of all the study material assigned for the midterm as well as final exam. Homework as well as group project are not subject for retake.

REQUIRED READINGS

James, Gareth, et al. (2023). An Introduction to Statistical Learning with Applications in Python. 1st ed., Springer.
<https://www.statlearning.com/>

ADDITIONAL READINGS

McKinney, W. Python for Data Analysis. O'Reilly Media, Inc. <https://wesmckinney.com/book/>

ANNEX

DEGREE LEVEL LEARNING OBJECTIVES

Learning objectives for the Bachelor of Business Management

Programmes:

International Business and Communication,

Business Management and Marketing,

Finance,

Industrial Technology Management,

Entrepreneurship and Innovation

Learning Goals	Learning Objectives
Students will be critical thinkers	BLO1.1. Students will be able to understand core concepts and methods in the business disciplines
	BLO1.2. Students will be able to conduct a contextual analysis to identify a problem associated with their discipline, to generate managerial options and propose viable solutions
Students will be socially responsible in their related discipline	BLO2.1. Students will be knowledgeable about ethics and social responsibility
Students will be technology agile	BLO3.1. Students will demonstrate proficiency in common business software packages
	BLO3.2. Students will be able to make decisions using appropriate IT tools
Students will be effective communicators	BLO4.1. Students will be able to communicate reasonably in different settings according to target audience tasks and situations
	BLO4.2. Students will be able to convey their ideas effectively through an oral presentation
	BLO4.3. Students will be able to convey their ideas effectively in a written paper

Learning objectives for the Bachelor of Social Science

Programmes:

Economics and Data Analytics,

Economics and Politics

Learning Goals	Learning Objectives
Students will be critical thinkers	ELO1.1. Students will be able to understand core concepts and methods in the key economics disciplines
	ELO1.2. Students will be able to identify underlying assumptions and logical consistency of causal statements
Students will have skills to employ economic thought for the common good	ELO2.1. Students will have a keen sense of ethical criteria for practical problem-solving
Students will be technology agile	ELO3.1. Students will demonstrate proficiency in common business software packages
	ELO3.2. Students will be able to make decisions using appropriate IT tools
Students will be effective communicators	ELO4.1. Students will be able to communicate reasonably in different settings according to target audience tasks and situations
	ELO4.2. Students will be able to convey their ideas effectively through an oral presentation
	ELO4.3. Students will be able to convey their ideas effectively in a written paper