Syllabus

1. Programme information

1.1. Institution	THE BUCHAREST UNIVERSITY OF ECONOMIC STUDIES
1.2. Faculty	Economic Cybernetics, Statistics and Informatics
1.3. Departments	Department of Statistics and Econometrics
1.4. Field of study	Cybernetics and statistics
1.5. Cycle of studies	Master Studies
1.6. Education type	Full-time
1.7. Study programme	Applied data analytics
1.8. Language of study	English
1.9. Academic year	2025-2026

2. Information on the discipline

2.1. Name	**Text Mining**								
2.2. Code	**25.0318IF1.2-0003**								
2.3. Year of study	**1**	2.4. Semester	**2**	2.5. Type of assessment	**Exam**	2.6. Status of the discipline	**O**	2.7. Number of ECTS credits	**5**
2.8. Leaders	C(C)	**Cadru asoc Conf.dr. FANTOZZI PAOLO**							
	S(S)	**Cadru asoc Conf.dr. FANTOZZI PAOLO**							

3. Estimated Total Time

3.1. Number of weeks	14.00		
3.2. Number of hours per week	2.00	of which	
		S(S)	1.00
		C(C)	1.00
3.3. Total hours from curriculum	28.00	of which	
		S(S)	14.00
		C(C)	14.00
3.4. Total hours of study per semester (ECTS*25)	125.00		
3.5. Total hours of individual study	97.00		
Distribution of time for individual study			
Study by the textbook, lecture notes, bibliography and student's own notes	60.00		
Additional documentation in the library, on specialized online platforms and in the field	14.00		
Preparation of seminars, labs, assignments, portfolios and essays	14.00		
Tutorials	7.00		
Examinations	2.00		
Other activities	0.00		

4. Prerequisites

4.1. of curriculum	● Basic Statistics ● Programming logic
4.2. of competences	Python and R

5. Conditions

for the S(S)	The seminars will be conducted in rooms with computers which have appropriate software: Python, R, Tableau etc
for the C(C)	The lectures will be held in rooms with Internet access and multimedia teaching equipment

6. Acquired specific competences

PREFESSIONAL	CC1	Digital skills – advanced use of software tools and platforms, development of software applications on different platforms.
PREFESSIONAL	CP5	Identifies of statistical patterns
PREFESSIONAL	CT1	Demonstrates an understanding of mathematical terms and concepts and applies basic mathematical principles and processes for interpreting data and facts.
PREFESSIONAL	CT4	Evaluates and analyses information and its sources. Demostrates ability to access and gain critical understanding of both traditional and new media, and their role and function in democratic societies.
PREFESSIONAL	CT5	Assumes the need for continuous training to create the premises for career progression and adapts one's own professional and managerial skills to the dynamics of the economic environment.

7. Objectives of the discipline

7.1. General objective	Providing the necessary tools for the analysis of corpora through quantitative methods
7.2. Specific objectives	Knowledge: C1: Acquisition of advanced knowledge of applied statistical analysis concepts, methods, and techniques, including inference, stochastic modeling, econometrics, and Bayesian methods. C2: Deep understanding of the processes of data collection, processing, analysis and interpretation in economic, social and industrial contexts, as well as the integration of interdisciplinary knowledge (data science, machine learning, advanced visualization). C3: Acquisition of knowledge about accessing data from different sources and handling large volumes of data. C5: Acquisition of natural language processing and social network analysis techniques. Skills: A1: Development and implementation of statistical models and artificial intelligence algorithms for the analysis of complex data, using programming environments such as R, Python and SQL. A3: Ability to use AI techniques such as ML, Neural Networks/Deep Learning to solve problems in different fields, including GIS-based methods A4: Ability to develop software modules for natural language processing, integrate them into applications, and implement analysis modules using social networks data Responsibility and autonomy: RA1: Ability to lead complex analytical projects and make autonomous statistical decisions under uncertainty, in compliance with ethical and data protection principles. RA3: Continuous development of professional skills in accordance with technology trends and business environment

8. Contents

8.1. C(C)		Teaching/Work methods	Recommendations for students
1	Introduction to the module and to text mining [C1,C2,C3,C5,A1,A3,A4] [2h]	Lecture based on multimedia presentations and interaction with students.	Lecture 1. Class notes online.ase.ro
2	Documents as vectors, cosine similarity, tf-idf [C1,C2,C3,C5,A1,A3,A4] [2h]	Lecture based on multimedia presentations and interaction with students.	Lecture 2. Class notes online.ase.ro
3	Part-Of-Speech tagging [C1,C2,C3,C5,A1,A3,A4] [2h]	Lecture based on multimedia presentations and interaction with students.	Lecture 3. Class notes online.ase.ro
4	Text classification and sentiment analysis [C1,C2,C3,C5,A1,A3,A4] [2h]	Lecture based on multimedia presentations and interaction with students.	Lecture 4. Class notes online.ase.ro
5	Text classification and sentiment analysis [C1,C2,C3,C5,A1,A3,A4] [2h]	Lecture based on multimedia presentations and interaction with students.	Lecture 5. Class notes online.ase.ro
6	Topic modelling and embeddings [C1,C2,C3,C5,A1,A3,A4] [2h]	Lecture based on multimedia presentations and interaction with students.	Lecture 6. Class notes online.ase.ro
7	Evaluation	Lecture based on multimedia presentations and interaction with students.	

Bibliography

- Manning, Schutze , Foundations of Statistical Natural Language Processing, The MIT Press , 1999
- Jurafsky, Martin , Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, Pearson, 2024
- Thanaki, Jalaj , Python natural language processing , Packt Publishing Ltd , 2017
- Bird, Klein, Loper , Natural language processing with Python: analyzing text with the natural language toolkit , https://www.nltk.org/book/
- Chopra, Deepti, Nisheeth Joshi, and Iti Mathur, Mastering natural language processing with python, Packt Publishing, 2016
- Fantozzi Paolo, Note de curs, online.ase.ro

8.2. S(S)		Teaching/Work methods	Recommendations for students
1	Working with strings in Python [C1,C2,C3,C5,A1,A3,A4,RA1,RA3] [2h]	Aplications / Workshop	Lecture 1. Class notes online.ase.ro
2	NLTK, tokenization, normalization, co-occurrences, n-gram models [C1,C2,C3,C5,A1,A3,A4,RA1,RA3] [2h]	Aplications / Workshop	Lecture 2. Class notes online.ase.ro
3	Scikit-learn, frequencies, vectors, similarities, POS [C1,C2,C3,C5,A1,A3,A4,RA1,RA3] [2h]	Aplications / Workshop	Lecture 3. Class notes online.ase.ro
4	Classification, sentiment analysis and topic modelling [C1,C2,C3,C5,A1,A3,A4,RA1,RA3] [2h]	Aplications / Workshop	Lecture 4. Class notes online.ase.ro
5	Transformers, SBert, Embeddings [C1,C2,C3,C5,A1,A3,A4,RA1,RA3] [2h]	Aplications / Workshop	Lecture 5. Class notes online.ase.ro
6	Project Presentation	Project Presentation	

Bibliography

- Chopra, Deepti, Nisheeth Joshi, and Iti Mathur, Mastering natural language processing with python, Packt Publishing, 2016
- Thanaki, Jalaj , Python natural language processing , Packt Publishing Ltd , 2017
- Bird, Klein, Loper , Natural language processing with Python: analyzing text with the natural language toolkit , https://www.nltk.org/book/
- Fantozzi Paolo, Curses notes, online.ase.ro

9. Corroboration of the contents of the discipline with the expectations of the representatives of the epistemic community, of the professional associations and representative employers in the field associated with the programme

The proposed course and seminar topics are consistent with national and international literature, as well as employers' requirements both in terms of theoretical knowledge and software packages used.

10. Assessment

Type of activity	Assessment criteria	Assessment methods	Percentage in the final grade
10.1. C(C)	Attendance at the course and seminars and involvement in discussing issues	Number of courses and seminar attendance and interventions Teste	10.00
10.2. S(S)	Project	Written project and presentation	40.00
10.3. Final assessment	Written exam	Written exam	50.00
10.4. Modality of grading	Whole notes 1-10		
10.5. Minimum standard of performance	Project presentation at the end of the module. The project can be done alone or in groups of up to 4 master’s students. To pass the exam, the student must achieve at least 50% on the written exam		

Date of listing,
04/28/2026

Signature of the discipline leaders,

Date of approval in the department

Signature of the Department Director,