

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 Economic Informatics and Cybernetics
1.4. Field of study	Economic Cybernetics, Statistics and Informatics
1.5. Cycle of studies	Licence
1.6. Education type	Full-time
1.7. Study programme	Economic informatics (in english)
1.8. Language of study	English
1.9. Academic year	2019-2020

2. Information on the discipline

2.1. Name	Data Analysis								
2.2. Code	19.0233IF3.1-0005								
2.3. Year of study	3	2.4. Semester	1	2.5. Type of assessment	Exam	2.6. Status of the discipline	O	2.7. Number of ECTS credits	4
2.8. Leaders	C(C)	lect.univ.dr. VINȚE C CLAUDIU					claudiu.vinte@ie.ase.ro		
	S(S)	lect.univ.dr. VINȚE C CLAUDIU					claudiu.vinte@ie.ase.ro		

3. Estimated Total Time

3.1. Number of weeks	14.00		
3.2. Number of hours per week	3.00	of which	
		C(C)	2.00
		S(S)	1.00
3.3. Total hours from curriculum	42.00	of which	
		C(C)	28.00
		S(S)	14.00
3.4. Total hours of study per semester (ECTS*25)	100.00		
3.5. Total hours of individual study	58.00		
<i>Distribution of time for individual study</i>			
Study by the textbook, lecture notes, bibliography and student's own notes	16.00		
Additional documentation in the library, on specialized online platforms and in the field	16.00		
Preparation of seminars, labs, assignments, portfolios and essays	16.00		
Tutorials	1.00		
Examinations	3.00		
Other activities	6.00		

4. Prerequisites

4.1. of curriculum	Statistics, Algebra, Basics of Programming, Probability and Mathematical Statistics
4.2. of competences	

5. Conditions

for the C(C)	The courses are held in rooms with video projectors.
for the S(S)	The seminars are held in laboratorie with work stations with Internet acces and where specialized software packages are installed

6. Acquired specific competences

	C1	Using the concepts, theories, principles and methods for investigating the economic phenomena and processes
	CT3	Identify opportunities for long-life training and efficient use of resources and learning techniques for personal future development

7. Objectives of the discipline

7.1. General objective	Developing skills and knowledge necessary for an efficient use and implement of data analysis methods in order to detect and capture the essence of information hidden in the data and obtain a simplified representation, clear and easy to read of this essence.
7.2. Specific objectives	<ul style="list-style-type: none"> - intelegerea conceptelor, metodelor si tehnicilor de analiza a datelor, precum si a posibilitatilor de utilizare a acestora in domeniul economico-financiar; - asigurarea cunostintelor necesare pentru utilizarea eficientă a metodelor și tehnicilor probabilistice și statistico-matematice în scopul fundamentării deciziilor din domeniul afacerilor și finanelor; - abordarea interdisciplinara a masurarii, cuantificarii, analizei si predictiei economico-financiare prin utilizarea unor concepte, metode, instrumente și proceduri specifice teoriei și practicii economice, modelării statistico-matematice și informaticii; - formarea modului de gândire cantitativă a specialistului economist, dezvoltarea și aprofundarea cunostintelor acestuia în domeniul măsurării, cuantificării, analizei, evaluării, interpretării și predicției economico-financiare; - formarea de abilități privind analiza și interpretarea datelor și rezultatelor metodelor și tehnicilor de analiza a datelor; - formarea de abilități privind înțelegerea și utilizarea instrumentelor software dedicate analizei datelor.

8. Contents

8.1. C(C)		Teaching/Work methods	Recommendations for students
1	Preliminary Data Analysis. Fundamentals. Statistical Distributions. Goodness of fit tests for distributions. Types of data. measures of similarity and dissimilarity in data analysis.	Oral presentation and interaction with students during class	
2	Principal component analysis: concept definition, principal components computation, mathematical model.	Oral presentation and interaction with students during class	
3	Principal component analysis: model evaluation and data visualization.	Oral presentation and interaction with students during class	
4	Principal component analysis: applying model on supplementary data sets, model generalization.	Oral presentation and interaction with students during class	
5	Factor analysis: concept definition, model assumptions, factorability.	Oral presentation and interaction with students during class	
6	Factor analysis: factor extraction, determining number of factors - Bartlett's test, factor rotation.	Oral presentation and interaction with students during class	

7	Canonical analysis.	Oral presentation and interaction with students during class	
8	Generalized canonical correlation analysis.	Oral presentation and interaction with students during class	
9	Correspondence analysis.	Oral presentation and interaction with students during class	
10	Multiple correspondence analysis.	Oral presentation and interaction with students during class	
11	Discriminant analysis: variability indicators, model significance.	Oral presentation and interaction with students during class	
12	Linear discriminant analysis.	Oral presentation and interaction with students during class	
13	Bayesian discriminant analysis. Knn classifier.	Oral presentation and interaction with students during class	
14	Hierarchical cluster analysis.	Oral presentation and interaction with students during class	

Bibliography

- Berry M. J. A., Linoff G. , Data mining techniques : for marketing, sales, and customer relationship management, 2nd ed., Wiley Publishing, Inc., 2004
- Harman, H.H., Modern Factor Analysis, University of Chicago Press, Chicago, Ill., 1967, Statele Unite ale Americii
- Ruxanda Gh. , Analiza datelor, Editura ASE, Bucuresti, 2001, România
- Saporta G. , Probabilités, analyse des données et statistique, Ed. Technip, Paris, 1990, Franța
- Cherkassky V., Mulier F. , Learning from Data: Concepts, Theory and Methods, John Wiley & Sons, Inc., New York, 1998, Statele Unite ale Americii
- Neagoe V., Stanasila O., Recunoasterea formelor si retele neuronale, Editura MATRIXROM, Bucuresti, 1999, România
- Ruxanda Gh. , Analiza factoriala – tehnica de investigare multidimensionala, Studii și Cercetări de Calcul Economic și Cibernetică Economică, nr. 3/2000, Anul XXXIII, Academia de Studii Economice, Bucuresti, 2000, România
- Benzecri J. P. , L'analyse des données, Dunod, Paris, 1979, Franța
- Harman, H.H., Modern Factor Analysis, University of Chicago Press, Chicago, Ill., 1967, Statele Unite ale Americii
- Cherkassky V., Mulier F. , Learning from Data: Concepts, Theory and Methods, John Wiley & Sons, Inc., New York, 1998, Statele Unite ale Americii
- Murtagh, F., Heck, A., Multivariate data analysis, Dordrecht, 1987, Olanda
- Gheorghe Ruxanda, Analiza datelor, ASE, Bucuresti, 2001, România

8.2. S(S)		Teaching/Work methods	Recommendations for students
1	Python Integrated Development Environment: PyCharm. Case study: principal component analysis.	Data set presentation, discussions upon the necessary processing algorithm, and software solution implementation for data analysis, in Python programming language.	
2	Principal components analysis. Case study in Python - PyCharm.	Data set presentation, discussions upon the necessary processing algorithm, and software solution implementation for data analysis, in Python programming language.	
3	Factor analysis. Case study in Python - PyCharm.	Data set presentation, discussions upon the necessary processing algorithm, and software solution implementation for data analysis, in Python programming language.	
4	Canonical analysis. Case study in Python - PyCharm.	Data set presentation, discussions upon the necessary processing algorithm, and software solution implementation for data analysis, in Python programming language.	
5	Correspondence analysis. Case study in Python - PyCharm.	Data set presentation, discussions upon the necessary processing algorithm, and software solution implementation for data analysis, in Python programming language.	
6	Discriminant analysis. Case study in Python - PyCharm.	Data set presentation, discussions upon the necessary processing algorithm, and software solution implementation for data analysis, in Python programming language.	
7	Cluster analysis. Case study in Python - PyCharm.	Data set presentation, discussions upon the necessary processing algorithm, and software solution implementation for data analysis, in Python programming language.	
<p>Bibliography</p> <ul style="list-style-type: none"> - Wes McKinney, Python for Data Analysis - 2nd Edition, O'Reilly Media, New York, 2018, Statele Unite ale Americii - Wes McKinney & PyData Development Team, pandas: powerful Python data analysis toolkit, 2017, https://pandas.pydata.org/pandas-docs/stable/pandas.pdf - The Python Standard Library, https://docs.python.org/3.7/library/index.html 			

- The Python Language Reference, <https://docs.python.org/3.7/reference/index.html>
- matplotlib Library Reference, <https://matplotlib.org/index.html>
- seaborn Library Reference, <https://seaborn.pydata.org/>

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 analysis, modeling, forecast activities in the economic field are based on data capturing the structure and the evolution of the economic phenomenon. But the data is affected by noise and adequate techniques are needed in order to extract the essence.
- The data mining methods and techniques allow researchers to go "beyond" the apparent data, to find latent elements of great significance and importance for knowledge: associations, hierarchy and so on.
- The data mining methods, techniques and procedures covered in this course are directly applicable to different fields from real economy (business intelligence, resource utilization optimization, capital markets, marketing, risk assessment, performance evaluation).
- The interest for data mining methods and techniques is increasing and there is a high demand for specialists in this field on the labor market.

10. Assessment

Type of activity	Assessment criteria	Assessment methods	Percentage in the final grade
10.1. S(S)	Individual project for multivariate analysis of a data set of choice, on which two of the studied data analysis methods are applied. The results obtained from the application of the two methods of analysis must be corroborated, interpreted and drawn conclusions from the analysis. Data processing is implemented in Python programming language.	Project presentation.	30.00
10.2. Final assessment	Final exam	Evaluation by an written exam	70.00
10.3. Modality of grading	Whole notes 1-10		
10.4. Minimum standard of performance	<p>The final grade is computed as the weighted arithmetic average of the grade obtained at the final exam (70%) and the grade obtained at the project (30%).</p> <p>It is also mandatory for entering the final exam to present the Python individual project at the seminar. Instead of the project, students can opt for a computer test. The test will consist in the implementation in the Python programming language of multivariate analysis methods applied on a given set of data. If the option is for the computer test, then the test result will have a weight of 50% in the final grade. The difference of 50% will be represented by the mark obtained in the written exam.</p> <p>Entrance to the final exam is conditional upon attendance of at least 4 seminars out of 7.</p>		

Date of listing,

Signature of the discipline leaders,

10/22/2019

Date of approval in the
department

Signature of the Department Director,

Signature of the Dean,