COURSE PROGRAMME

1. Information about the programme

1.1 University	**University "Alexandru Ioan Cuza" of Iasi**
1.2 Faculty	**Faculty of Mathematics**
1.3 Department	**Department of Mathematics**
1.4 Domain	**Mathematics**
1.5 Cycle	**Masters**
1.6 Programme / Qualification	**Applied Mathematics**

2. Information about the course

2.1 Course Name			Applied statistics and data minning				
2.2 Course taught by			Assoc. Prof. PhD. IULIAN STOLERIU				
2.3 Seminary / laboratory taught by			Assoc. Prof. PhD. IULIAN STOLERIU				
2.4 Year	I	**2.5** Semester	I	**2.6** Type of evaluation*	E	**2.7** Course type**	Ob

E – Exam / C – Colloquium / V – Verification
***OB – Obligatory / OP – Optionally / F – Facultative*

3. Total hours (estimated per semester and activities)

3.1 Number of hours per week	4	**3.2** course	2	**3.3** seminary/ laboratory	2
3.4 Total number of hours	56	**3.5** course	28	**3.6** seminary/ laboratory	28

Distribution	hours
Individual study using textbooks, course notes, bibliography items, etc.	32
Supplimentary study (library, on-line platforms, etc.)	36
Individual study for seminary/laboratory, homeworks, projects, etc.	30
Tutoring	19
Examination	2
Other activities	0
3.7 Total hours of individual activity*	119
3.8 Total hours per semester	175
3.9 Credit points	7

4. Pre-requisites - Curriculum (if necessary)

Scientific computing with MATLAB/Python

5. Conditions (if necessary)

5.1 Course	Beamer presentation
5.2 Seminary / Laboratory	MATLAB/Python application

6. Objectives

This course provides a comprehensive introduction to applied statistical methods and data mining techniques. Students will learn the theoretical underpinnings and practical applications of both fields, equipping them with the skills to analyze complex datasets, extract meaningful insights, and build predictive models. The first part of this course will present theoretical and practical statistical techniques for analysing data. The last lectures will introduces the principle methods of data mining, discusses how to evaluate generated models and present some practical examples of data mining applications.

7. Specific competencies/Learning outcomes

• conducts quantitative research
• carries out interdisciplinary research activities
• applies scientific methods
• performs data analysis
• applies statistical analysis techniques

8. Contents

8.1 Course	**Teaching methods**	**Remarks** (number oh hours, references)
Review of fundamental concepts from Statistics - Population, sample, parameters, statistics - Review of fundamental concepts - Types of data (nominal, ordinal, interval, ratio) - Descriptive statistics: measures of central tendency and dispersion, and others - Sampling distributions and the Central Limit Theorem.	Beamer presentation	2 hours
Parametric hypothesis testing for one or two populations - One-sample t-tests and z-tests (for means, variances and proportions) - Two-independent samples t-tests and z-tests (for means, variances and proportions) - Paired samples t-test (two dependent samples) - Parametric hypothesis testing using MATLAB	Beamer presentation	2 hours
Analysis of Variance (ANOVA) - Introduction to ANOVA: comparing means of multiple groups - One-way ANOVA: assumptions, calculations, and interpretation of results - Post-hoc tests (e.g., Tukey's HSD, Bonferroni correction) - Two-way ANOVA (Introduction to factorial designs and interaction effects	Beamer presentation	2 hours
Non-parametric methods - Introduction to non-parametric tests; when to use them - Non-parametric tests: Median test, Wilcoxon signed-rank test, Mann-Whitney test, Runs test, Wald-Wolfowitz test, Kruskal-Wallis test. - Randomization tests: Permutation tests, Bootstrapping, Jackknifing, Monte Carlo simulation	Beamer presentation	2 hours
Categorical data analysis - Pearson’s test of goodness-of-fit test,test of independence, test of homogeneity. Fisher’s exact test - Contingency tables and interpretation of results. - Measures of association for categorical variables (e.g., odds ratio, relative risk, Cramer’s contingency coefficient)	Beamer presentation	2 hours

8.1 Course	**Teaching methods**	**Remarks** (number oh hours, references)
Correlation and Regression - Correlation: Pearson, Spearman and Kendall, r, R2, significance. - Simple linear regression: model formulation, least squares estimation, interpretation of coefficients. - Multiple linear regression: Model formulation and interpretation of coefficients. - Model selection (brief overview of methods like adjusted R-squared, AIC, BIC) - Model evaluation and validation. - Other types of regression	Beamer presentation	4 hours
Logistic regression - Introduction and motivation. - Binary logistic regression: model formulation, odds ratios, interpretation of coefficients. - Model evaluation: goodness-of-fit, classification accuracy. - Assessing the overall fit of the model. - Evaluating classification performance. - Multinomial logistic regression (introduction, softmax function).	Beamer presentation	2 hours
Introduction to time series analysis - Definition and examples - Components of a time series - Basic forecasting methods: moving averages (calculating the average of a fixed number of past observations), exponential smoothing (assigning exponentially decreasing weights to past observations). - Introduction to ARIMA models. - The concept of autocorrelation and partial autocorrelation. - Mentioning the Box-Jenkins methodology. - Other advanced techniques (briefly): exponential smoothing with trend and seasonality (e.g., Holt-Winters), state space models.	Beamer presentation	2 hours
Introduction to Data mining - The Data Mining Process: CRISP-DM Model (business understanding, data understanding, data preparation, modeling, evaluation, deployment). - Types of data mining tasks: descriptive vs. predictive. - Major mining tasks: association rule mining, classification, prediction, clustering, anomaly detection. - Data preprocessing: cleaning, integration, transformation, reduction. - Applications of data mining in various domains. - Challenges and issues in data mining.	Beamer presentation	2 hours
Data preprocessing - Data quality: noise, outliers, missing values, inconsistent data. - Data cleaning techniques: handling missing values, smoothing noisy data, identifying and removing outliers. - Data integration: combining data from multiple sources. - Data transformation: normalization, standardization, discretization, concept hierarchy generation. - Data reduction: dimensionality reduction (feature selection, feature extraction - brief overview, PCA).	Beamer presentation	2 hours

8.1 Course	**Teaching methods**	**Remarks** (number oh hours, references)
Classification techniques - General approach to classification. - Decision trees: ID3, C4.5, CART algorithms. Overfitting and tree pruning. - Evaluating classification models: accuracy, precision, recall, F1-score, confusion matrix, ROC curves, AUC. - Bayesian classification: Naive Bayes classifier. - Support Vector Machines (SVM): basic concepts, linear and non-linear kernels (brief introduction). - K-Nearest Neighbors (KNN) algorithm. - Ensemble methods: Bagging, Boosting (e.g., AdaBoost, Random Forests - conceptual overview).	Beamer presentation	2 hours
Clustering techniques - Introduction to clustering: concepts and applications. - Types of clustering methods: partitioning, hierarchical, density-based (brief overview of each). - Partitioning methods: K-means algorithm, K-medoids. - Hierarchical clustering: agglomerative and divisive approaches. - Density-based clustering: DBSCAN. - Evaluating clustering results: silhouette coefficient, other internal and external measures. - Applications of clustering (e.g., customer segmentation).	Beamer presentation	2 hours
Association rule mining - Basic Concepts: itemsets, support, confidence, lift. - Market basket analysis and association rules. - Apriori algorithm: support, confidence, lift. Finding frequent Itemsets. - Generating association rules from frequent itemsets. - Mining multilevel and multidimensional association rules (brief overview). - Constraints-based association mining (brief overview). - Applications of association rule mining (e.g., market basket analysis). - Anomaly detection techniques: Statistical approaches, proximity-based approaches (brief overview).	Beamer presentation	2 hours

Bibliography

1. Devore, J. L., & Berk, K. N. (2021). Modern mathematical statistics with applications (3rd ed.). Springer.
2. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). Springer.
3. Bruce, P., Bruce, A., & Gedeck, P. (2020). Practical Statistics for Data Scientists (2nd ed.). O'Reilly Media.
4. Han, J., Pei, J., & Tong, H. (2022). Data Mining: Concepts and Techniques (4th ed.). Morgan Kaufmann.
5. Tan, P.-N., Steinbach, M., Karpatne, A., Kumar, V.: Introduction to Data Mining, 2nd edn. Pearson, UK (2018).
6. McKinney, W. (2018) Python for Data Analysis: Data Wrangling with Pandas, Num Py, and IPython. OReilly Media, Inc., Boston, MA.

8.2 Seminary / Laboratory	**Teaching methods**	**Remarks** (number oh hours, references)
- Introduction to statistical software (Random number generators in MATLAB)	Exercises solved on the blackboard and PC simulations	2 hours
- Applications in MATLAB for parametric hypothesis testing for one or two populations	Exercises solved on the blackboard and PC simulations	2 hours
- Applications in MATLAB for ANOVA	Exercises solved on the blackboard and PC simulations	2 hours

8.2 Seminary / Laboratory	Teaching methods	Remarks (number oh hours, references)
- Applications in MATLAB for nonparametric methods	Exercises solved on the blackboard and PC simulations	2 hours
- Applications in MATLAB for categorical data analysis	Exercises solved on the blackboard and PC simulations	2 hours
- Applications in MATLAB for correlation and regression	Exercises solved on the blackboard and PC simulations	4 hours
- Applications in MATLAB for logistic regression	Exercises solved on the blackboard and PC simulations	2 hours
- Applications with time series	Exercises solved on the blackboard and PC simulations	2 hours
- Data preprocessing: cleaning, integration, transformation, reduction.	Exercises solved on the blackboard and PC simulations	2 hours
- Data reduction: dimensionality reduction (feature selection, feature extraction - brief overview, PCA) using MATLAB or Python	Exercises solved on the blackboard and PC simulations	2 hours
- Various applications containing classification techniques in MATLAB or Python	Exercises solved on the blackboard and PC simulations	2 hours
- Various applications containing clustering techniques in MATLAB or Python	Exercises solved on the blackboard and PC simulations	2 hours
- Various applications on association rule mining in MATLAB or Python	Exercises solved on the blackboard and PC simulations	2 hours

Bibliography

1) I. Stoleriu, Statistica prin MATLAB, Editura MatrixRom, Bucuresti, 2010.
2) J. L. Devore, K.N. Berk, Modern Mathematical Statistics with Applications, Duxbury, 2007.
3) D. Wackerly, W.Mendenhall, R. L. Scheaffer, Mathematical Statistics With Applications, Duxbury Press, 7th edition, 2007.
4) M.R. Spiegel, L.J. Stephens, Schaum's Outline of Statistics, McGraw-Hill, 2007.
5) Gh. Mihoc, N. Micu: Teoria probabilităţilor si statistică matematică, Bucuresti, 1980

9. Coordination of the contents with the expectations of the community representatives, professional associations and relevant employers in the corresponding domain

Students could use the acquired information from this course in any area that employs statistical data

10. Assessment and examination

10.1 Continuous assessment		Percentage (min. 30%)		40
Course	Assessment type			
	Percentage			0
	Failure to pass the continuous assessment results in failure to pass the final assessment			
	Assessment methods	Details	Percentage	with reexamination
Seminary / Laboratory	Assessment type			Mixed assessment
	Percentage			100
	Failure to pass the continuous assessment results in failure to pass the final assessment			No
	Assessment methods	Details	Percentage	with reexamination
		Continuous oral assessment	50	No
		Homework	50	No

10.2 Final assessment	Percentage (max. 70%)	60

	Assessment type	Final written assessment

10.3 Special notes (special situations is assessment)	

10.4 Minimum performance standard	
Minimum 5(five) at final written exam	

Date, **Course coordinator,**
Assoc. Prof. PhD. IULIAN STOLERIU

Seminary coordinator,
Assoc. Prof. PhD. IULIAN STOLERIU

Aproval date in the department,

Head of the departament,
Prof. PhD. IONEL DUMITREL GHIBA