**COURSE SYLLABUS**

***DATA MINING***

1. **Program identification details**

|  |  |
| --- | --- |
| 1.1 Higher education institution | „Ovidius” University of Constanta |
| 1.2 Faculty | Faculty Mathematics and Computer Science |
| 1.3 Department | Mathematics and Computer Science |
| 1.4 Field of studies | **Computer Science** |
| 1.5 Cycle of studies (degree) | Master |
| 1.6 Degree program/qualification | **Cyber Security and Machine Learning** |
| 1.7 Academic year | 2022-2023 |

1. **Course identification details**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 2.1 Course title | | | **Data Mining** | | | | |
| 2.2 Course code | | | **FMI.CSML.I.2.06** | | | | |
| 2.3 Instructor | | | Assoc. Prof. Pelican Elena | | | | |
| 2.4 Teaching assistant | | | Assoc. Prof. Pelican Elena | | | | |
| 2.5 Year | I | 2.6 Semester | 2 | 2.7. Evaluation type | E | 2.8 Course type \*/\*\* | DCA/DI |

*\* DF – fundamental course, DD – field course, DS – specialty course, DC – complementary course, DAP – advanced study course, DSI – synthesis course, DCA – advanced knowledge course.*

*\*\* DI – mandatory course; DO – optional course.*

1. **Estimated workload (hours per semester)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 3.1 Number of teaching hours/week | | 3 | of which:  3.2 course | 2 | 3.3 applications*\*\*\** | 2 |
| 3.4 Total of teaching hours within the program/semester | | 42 | of which:  3.5 lecture | 14 | 3.6 seminar | 28 |
| **3.7 Student workload for individual study** | | | | | | 108 |
| ***Distribution of workload*** | | | | | | [hours] |
| Individual study of texbooks, handbooks/reader, bibliography and notes | | | | | | 30 |
| Additional research (library, electronic resources, fieldwork) | | | | | | 20 |
| Homework (preparing seminar presentations, portfolios, critical essays, research papers, etc.) | | | | | | 50 |
| Individual consultations (optional) | | | | | | 0 |
| Evaluations / exams | | | | | | 8 |
| Other activities | | | | | | 0 |
| **3.8 Total hours per semester** | *150* | |  |  |  |  |
| **3.9 Number of credits** | 6 | |  |  |  |  |

*\*\*\* S - seminar; L - laboratory; P – project*

1. **Prerequisites (if any)**

|  |  |
| --- | --- |
| 4.1 Curriculum-related | Machine Learning, Linear Algebra, Statistics, Optimizations, Fundamental Algorithms, Data Structures |
| 4.2 Skills-related | Python Programming |

1. **Requirements (if any)**

|  |  |
| --- | --- |
| 5.1. For running the course | Lecture Room/Lecture Hall |
| 5.2. For running the seminar / laboratory /project  *\*The type is to be chosen according to the discipline* | Computer Laboratory |

**6 . Acquired specific skills**

|  |  |
| --- | --- |
| Professional skills | Students must be able to deeply understand the methods and features of important classes of data mining algorithms, as well as the advantages and drawbacks of the main solutions in the field, but also their impact on the performance of the system/ application.  Responsible completion of professional tasks under limited authonomy and qualified assistance. |
| Cross-cutting skills | Autonomous completion of complex tasks, and also detection and solving related problems during development and deployment of the application.  Efficient development of activities organized in an interdisciplinary group and development of empathic capacities for interpersonal communication, relationships and collaboration with various groups.  Rigorous and efficient work rules, norms and values ​​of professional ethics within their own work strategy, for the optimal and creative capitalization their own potential. |

1. **Course goal and objectives**

|  |  |
| --- | --- |
| 7.1 The general objective of the course | Introduction of the principles and mechanisms of data mining. |
| 7.2 Specific objectives | In-depth understanding of the principles of data mining algorithms and of the differences between those and machine learning ones.  The ability to correctly interpret the results obtained by applying studied models, according to the performance scores corresponding to the analysied problem.  Ability to implement specific algorithms, to use specific software tools and to modify them as required.  Development of software applications based on algorithms in this field. |

1. **Contents**

|  |  |  |
| --- | --- | --- |
| **8.1 Lecture** | **Teaching methods** | **Number of hours** |
| Overview.  Datasets. Introduction to Weka and other similar products.  Data pre-processing (data cleaning, data transformation, dimension reduction, features selection)  Data Classification (classifiers based on instances, decision trees and rules etc)  Classifiers based on the majority class (ZeroR), Siple rules of classification (OneR)  Creating decision trees  Support vector classifiers  Classifiers based in neural networks  Clustering  Partition algorithms (kMeans, Fuzzy cMeans) Hierarchical algorithms (agglomerative and divisive ones)  Density based algorithms (DBSCAN)  Probabilistic algorithms (Expectation Maximization)  Association Rules  Regression models and time series analysis  Ensemble-like methods  Recommendation systems  Text Mining. Web Mining  Ethics in data mining | Interactive methods of teaching-learning  Problematization  The active and interactive methods  Methods which contribute to the development of critical thinking  Interaction, problematization, Dialogue  Information synthesizing/ essentialising  Problematization  Independent and cooperative learning  Generalization  Conversation  Argumentation | 1 hour  1 hour  2 hours  2 hours    1 hour  2 hours  1 hour  2 hours  1 hour  1 hour |
| **Bibliography:**   1. C.C. Aggarwal, Data Mining – The Text Book, Springer, 2015 2. T. Hastie, R. Tibshirani, J. Friedman - The Elements of Statistical Learning. Data Mining, Inference, and Prediction, Second Edition, Springer 2009 3. M. H. Dunham. Data Mining. Introductory and Advanced Topics, Pearson Education 2003 4. F. Gorunescu, Data Mining. Concepts, Models and Techniques, Springer, 2011 5. C. D. Manning, P. Raghavan and H. Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008. 6. I.H. Witte, E. Frank, M.A. Hall. Data Mining: Practical Machine Learning Tools and Techniques, 3rd Ed, Morgan Kaufmann Publishers, 2011 7. Lecture and Lab Notes uploaded in the corresponding channel in MS Teams | | |

|  |  |  |
| --- | --- | --- |
| **8.2 Applications\* (Laboratory / project)**  *\*The type is to be chosen according to the discipline* | **Teaching methods** | **Number of hours** |
| Overview.  Datasets. Introduction to Weka and other similar products.  Data pre-processing (data cleaning, data transformation, dimension reduction, features selection)  Data Classification (classifiers based on instances, decision trees and rules etc)  Classifiers based on the majority class (ZeroR), Siple rules of classification (OneR)  Creating decision trees  Support vector classifiers  Classifiers based in neural networks  Clustering  Partition algorithms (kMeans, Fuzzy cMeans) Hierarchical algorithms (agglomerative and divisive ones)  Density based algorithms (DBSCAN)  Probabilistic algorithms (Expectation Maximization)  Association Rules  Regression models and time series analysis  Ensemble-like methods  Recommendation systems  Text Mining. Web Mining | Interactive methods of teaching-learning  Problematization  The active and interactive methods  Methods which contribute to the development of critical thinking  Interaction, problematization, Dialogue  Information synthesizing/ essentialising  Problematization  Independent and cooperative learning  Generalization  Conversation  Argumentation | 1 hour  1 hour  4 hours  4 hours    4 hours  4 hours  2 hours  4 hours  4 hours |
| **Bibliography:**   1. C.C. Aggarwal, Data Mining – The Text Book, Springer, 2015 2. T. Hastie, R. Tibshirani, J. Friedman - The Elements of Statistical Learning. Data Mining, Inference, and Prediction, Second Edition, Springer 2009 3. M. H. Dunham. Data Mining. Introductory and Advanced Topics, Pearson Education 2003 4. F. Gorunescu, Data Mining. Concepts, Models and Techniques, Springer, 2011 5. C. D. Manning, P. Raghavan and H. Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008. 6. I.H. Witte, E. Frank, M.A. Hall. Data Mining: Practical Machine Learning Tools and Techniques, 3rd Ed, Morgan Kaufmann Publishers, 2011 7. Lecture and Lab Notes uploaded in the corresponding channel in MS Teams | | |

**9. Correlation between the content of the course and the needs/expectations of the epistemic community, professional associations and/or significant employers relevant for the program**

|  |
| --- |
| Preparing students to practice in the field of computer science and to meet the requirements of IT companies, especially for a profession in the field of cybersecurity and machine learning. |

**10. Evaluation**

|  |  |  |  |
| --- | --- | --- | --- |
| Type of activity | 10.1 Evaluation criteria | 10.2 Evaluation methods | 10.3 Percentage of final grade |
| 10.4 Course |  |  |  |
| 10.5 Applications\*  (Seminar/Laboratory / Project)  *\*The type is to be chosen according to the discipline* | Active participation, quizzes |  | 10% |
| Homeworks | 30% |
|  |  | Examination (2 Projects) | 60% |
| 10.6 Minimum standard of achievement for the acquisition of the ECTS credits | | | |
| Minimum knowledge of one item per studied chapter and presentation of one scientific paper. | | | |

Date of completion Course Instructor, Teaching Assistant,

September 20, 2022 Assoc. Prof. Pelican Elena Assoc. Prof. Pelican Elena

Date of approval in the Department Head of Department

Assoc. Prof. Puchianu Crenguta

September 27, 2022

Dean,

Assoc. Prof. Nicola Aurelian