SYLLABUS

1. Data about the program of study

1.1 Institution	The Technical University of Cluj-Napoca
1.2 Faculty	Faculty of Automation and Computer Science
1.3 Department	Computer Science
1.4 Field of study	Computer Science and Information Technology
1.5 Cycle of study	Master
1.6 Program of study / Qualification	Cybersecurity Engineering / Master
1.7 Form of education	Full time

2. Data about the subject

2.1 Subject name	***Big Data and Machine Learning for Cybersecurity***			Subject code	**9.10**
2.2 Course responsible / lecturer	Prof. dr. eng. Camelia LEMNARU - Camelia.Lemnaru@cs.utcluj.ro				
2.3 Teachers in charge of seminars / Laboratory / project	Assoc.prof.dr.eng.. Ciprian OPRIȘA - Ciprian.OPRISA@cs.utcluj.ro				
2.4 Year of study	I	2.5 Semester	2	2.6 Type of assessment (E - exam, C - colloquium, V – verification)	E
2.7 Subject category	Formative category: DA – advanced, DS – speciality, DC – complementary				DS
	Optionality: DI – imposed, DO – optional (alternative), DF – optional (free choice)				DO

3. Estimated total time

3.1 Number of hours per week	4	of which:	Course	2	Seminars	0	Laboratory	2	Project	0
3.2 Number of hours per semester	56	of which:	Course	28	Seminars	0	Laboratory	28	Project	0
3.3 Individual study:										
(a) Manual, lecture material and notes, bibliography										32
(b) Supplementary study in the library, online and in the field										18
(c) Preparation for seminars/laboratory works, homework, reports, portfolios, essays										43
(d) Tutoring										0
(e) Exams and tests										2
(f) Other activities:										0

3.4 Total hours of individual study (suma (3.3(a)…3.3(f)))	94
3.5 Total hours per semester (3.2+3.4)	150
3.6 Number of credit points	6

4. Pre-requisites (where appropriate)

4.1 Curriculum	Data bases
4.2 Competence	Statistics and probabilistic calculus

5. Requirements (where appropriate)

5.1. For the course	blackboard, beamer, computers
5.2. For the applications	blackboard, beamer, computers, specific software

6. Specific competence

6.1 Professional competences	develop information security strategy perform ICT security testing manage system security identify ICT security risks define security policies educate on data confidentiality provide ICT consulting advice perform data analysis implement ICT security policies ensure compliance with legal requirements ensure information privacy monitor developments in field of expertise keep up with the latest information systems solutions
6.2 Cross competences	The graduate • develop an analytical approach • taking a proactive approach • developing strategies to solve problems • being open minded • coordinate engineering teams

7. Expected Learning Outcomes

Knowledge	• computer programming • cyber attack counter-measures • digital systems • security engineering • software anomalies • cloud technologies • database development tools • network standards • operating systems • cloud monitoring and reporting • cyber security • information confidentiality • building systems monitoring technology • business intelligence • project management • cyber attack counter-measures • open source model • decision support systems

Skills	The student is able to: • analyse ICT systems • create flowchart diagrams • define technical requirements • develop software prototypes • execute software tests • interpret technical texts • keep up with the latest information systems solutions • monitor system performance • use software design patterns • use software libraries • debug software • develop creative ideas • perform project management • ensure proper document management • maintain database security • address problems critically • assess ICT knowledge • implement anti-virus software • manage cloud data and storage • manage databases • store digital data and systems • train employees • use scripting languages for programming • collect cyber defence data • create project specifications • give live presentations • provide information • define technology strategy • implement cloud security and compliance • implement spam protection
Responsibilities and autonomy	The student has the ability to work independently in order to: • develop an analytical approach • take a proactive approach • develop strategies to solve problems • be open-minded • coordinate engineering teams

8. Discipline objective (as results from the *key competences gained*)

8.1 General objective	Acquiring the ability to analyse large datasets. Considering the increasing number of malicious programs in the wild, the goal is to learn how to handle large collections of data, design, implement and evaluate malware detection and classification models.
8.2 Specific objectives	1. Acquire the ability to use scripting languages and databases to handle large datasets. 2. Design and implement distributed systems, understand and use the Map-Reduce paradigm. 3. Understand and learn algorithms and techniques for searching in large collections of data. 4. Understand and learn Machine Learning algorithms suitable for malware classification and detection.

9. Contents

9.1 Lectures	Hours	Teaching methods	Notes
Introduction to Big Data	2	Blackboard	

Big Data Storage	2	illustrations and explanations, beamer presentations, discussions, short challenges	
Map-Reduce	2		
Extracting Features from File Collections	2		
Finding Similar Items	2		
Clustering	2		
Advanced Clustering Techniques	2		
PageRank	2		
Linear and Logistic Regression	2		
SVMs and Decision Trees	2		
Bagging	2		
ML pipeline. Feature Selection	2		
Dimensionality reduction	2		
Recap	2		

Bibliography:

- J. Leskovec, A. Rajaraman and J. D. Ullman. Mining of massive data sets. Cambridge University Press, 2020
- Pattern Recognition and Machine Learning (Bishop, Christopher – 2007 – Springer)
- Data Science for Business: What you need to know about data mining and data-analytic thinking (Provost, Foster – 2013 – O'Reilly)
- G. Fourny. The Big Data Textbook: From clay tablets to data lakehouses. ETH Zurich, 2024
- T. Erl, W. Khattak and P. Buhler. Big Data Fundamentals: Concepts, Drivers & Techniques. Prentice Hall Press, 2016

9.2 Applications - Seminars/Laboratory/Project	Hours	Teaching methods	Notes
Sentiments Analysis from X Data	2	Brief reviews, blackboard illustrations and explanations, tutorials, roadmaps, short live demos and guidance of code development, discussions, homework	
From SQL to Search Engines	2		
Extracting OpCode Sequences from Binary Programs	4		
n-grams Storage and Filtering	2		
Advanced Search	4		
Classic Clustering	2		
Advanced Clustering	2		
PageRank	2		
Classifiers	2		
Deep Learning	4		
Recap	2		

Bibliography

- J. Leskovec, A. Rajaraman and J. D. Ullman. Mining of massive data sets. Cambridge University Press, 2020
- MongoDB: The Definitive Guide (Chodorow, Kristina – 2013 – O'Reilly) (2nd ed)
- Learning Python (Lutz, Mark – 2013 – O'Reilly) (5th ed)

**Se vor preciza, după caz: tematica seminariilor, lucrările de laborator, tematica și etapele proiectului.*

9. Bridging course contents with the expectations of the representatives of the community, professional associations and employers in the field

This aspect will be achieved by recurrent discussions with the relevant industry employers (cybersecurity domain). Big Data courses are delivered within other master programs, but very few focus on computer and information security. Both malware and spam detection and classification require, from a practical standpoint, working with large collections of data, which requires big data analysis and machine learning. For example, there are several master programs which teach big data and business analytics, teaching methods which can be successfully applied to the data/computer security domain:

- CS246, Mining Massive Data Sets, Stanford https://web.stanford.edu/class/cs246/
- Big Data Analytics and Information Technology Missouri University of Science andTechnology, USA
- Big Data, Masters in Computer and Information Security, University of Liverpool, UK

10. Evaluation

Activity type	Assessment criteria	Assessment methods	Weight in the final grade
Course	Ability to define and explain concepts and methods specific to course's field. Attendance frequency, interest, and interactivity during lecture classes.	Written exam, including online quiz tests (e.g. on Moodle platform) and presentation(s) of different subjects / paper in the course's field during semester time.	60%
Seminar	-	-	-
Laboratory	Capability and ability to give correct and functional solutions to problems specific to course's field. Attendance frequency, interest, and interactivity during lab classes.	Evaluate lab activity. Evaluate lab assignments (homework). Evaluate solutions of problems given in a final lab exam.	40%
Project	-	-	-

Minimum standard of performance:
Lecture. **Attending minimum 50%** of lecture classes, to be allowed to take the final examination. Minimum final grade must be 5 for the exam to be considered passed.
Lab. Attending all lab classes (one lab could be recovered during the semester, and one more during re-examination sessions). Minimum lab grade must be 5 to be allowed at final exam. By the end of the course, the students should be able to work with big datasets, both structured and unstructured, using sequential and distributed algorithms (e.g. Map-Reduce). The main operations students should have assimilated are: search in large collections of data, classification and clustering, building and evaluating prediction models.

Date of filling in: 01.09.2025	**Responsible**	**Title First name Last name**	**Signature**
	Course	Prof.dr.eng. Camelia LEMNARU	
	Applications	Assoc.prof.dr.eng. Ciprian OPRIȘA	

Date of approval in the department 17.09.2025	Head of department, Prof.dr.eng. Rodica Potolea
Date of approval in the Faculty Council 19.09.2025	Dean, Prof.dr.eng. Vlad Mureșan