Machine learning for networking

01DSMUV, 01DSMBG, 01DSMUW

A.A. 2024/25

Course Language

Inglese

Degree programme(s)

Master of science-level of the Bologna process in Cybersecurity - Torino
Master of science-level of the Bologna process in Communications Engineering - Torino
Master of science-level of the Bologna process in Cybersecurity - Torino

Course structure

Teaching	Hours
Lezioni	40
Esercitazioni in laboratorio	40

Lecturers

Teacher	Status	SSD	h.Les	h.Ex	h.Lab	h.Tut	Years teaching
Vassio Luca	Ricercatore a tempo det. L.240/10 art.24-B	IINF-05/A	35	0	15	0	2

Co-lectures

Espandi

Context

SSD	CFU	Activities	Area context
ING-INF/05	8	D - A scelta dello studente	A scelta dello studente

Anno accademico di inizio validità

2024/25

Course description

This course explores how Machine Learning can help engineers solve problems in the world of networking. The course introduces the data science process and then provides theoretical and practical knowledge about the machine learning approach and algorithms commonly used to analyze large and heterogeneous data. The students will also acquire Python programming competencies and learn how to use its main libraries related to ML. Many practical examples will be focused on how to address inference problems in the field of communication networks and cybersecurity.
A significant part of the courses will be devoted to laboratory activities allowing the students to practice the theoretical notions on real problems, from traffic classification to anomaly detection. Many laboratory sessions, based on a learning-by-doing approach, allow experimental activities on all the phases of a machine learning pipeline (e.g., data preparation and cleaning, data visualization and characterization, ML algorithm selection, tuning, and result evaluation).

Expected Learning Outcomes

Knowledge and abilities:
• Knowledge of Python programming language and the main Python libraries for machine learning;
• Knowledge of the main phases characterizing a data science and ML process;
• Knowledge of the different data exploration, visualization and pre-processing techniques;
• Knowledge of the basic theoretical principles of machine learning;
• Knowledge of the principal models for supervised and unsupervised learning;
• Knowledge of the main theoretical properties, domains of application, and limitations of different machine learning approaches;
• Knowledge of networking problems that can be approached with ML;
• Ability to design, implement and evaluate analytics scripts in the Python language.
• Ability to manage large datasets of data, from pre-processing to visualization.
• Ability to employ the Python machine learning libraries to devise complete solutions for inference problems;
• Ability to design, implement and evaluate a machine learning pipeline;
• Ability to apply different methods to real (networking and cybersecurity) tasks, to critically evaluate their effectiveness and to analyze which strategies are better suited to different applications;

Pre-requirements

The students should have basic knowledge of:
• Programming skills (whatever the language)
• Communication networks
• Probability theory and statistics
• Linear algebra
• Calculus
• Operational research

Course topics

Introduction to Machine Learning and its application to Networking (0.5 CFU)
• Definitions of pipeline and taxonomy of Machine Learning tasks
• Problems in networking: from traffic classification to anomaly detection

Python usage and libraries (2.0 CFU)
• The Python language
• Numerical libraries: Numpy, Pandas and Matplotlib
• ML libraries (Scikit-learn, PyTorch)

Data exploration and preprocessing (1.5 CFU)
• Data visualization
• Data transformation and feature extraction
• Dimensionality reduction techniques

Basics of ML (1 CFU)
• Empirical risk minimization
• Loss functions and performance metrics
• Gradient-based learning
• Model selection and validation

Supervised and unsupervised ML (3 CFU)
• Classification
• Regression
• Clustering
• Algorithms: from linear models to deep neural networks
• Regularization

Sustainable development goals

Additional information

Course structure

The course will include 40 hours of lectures and 40 hours of laboratory activities.
The lectures will focus both on theoretical and practical aspects of the course topics and will include open discussions aimed at developing suitable solutions for different problems. Some simple practical exercises will be solved in the classroom. The course includes laboratory sessions on data science processes and machine learning algorithms for engineering applications. The laboratories will allow the students to apply the methods presented during lectures to real data and tasks, with a particular focus on networking and cybersecurity applications.
Students will prepare a written report on a group project assigned during the course.

Reading materials

Copies of the slides used during the lectures, exercises, and manuals for the activities in the laboratory will be made available. All teaching material is downloadable from the course website or the teaching Portal.

Suggested books:
[1] A. Jung, Machine Learning: The Basics, Springer, 2022
[2] Jake VanderPlas, Python Data Science Handbook: Essential Tools for Working with Data, O'Reilly, 2016
[3] Christopher M. Bishop, Pattern Recognition and Machine Learning. Springer, 2006
[4] Kent D. Lee, Python Programming Fundamentals, Springer, 2015

Study materials

Lecture slides; Lab exercises; Video lectures (current year);

Assessment and grading criteria

Exam: Written test; Individual project; Group project;

The exam includes two mandatory parts. The two mandatory parts are (i) a written exam and (ii) the evaluation of a group project.
The final score is defined by considering both the evaluation of the group project and the written part. The teacher may request an integrative oral test to confirm the evaluations that were obtained.

The written examination lasts 90 minutes and will consist of open and closed questions and exercises covering the topics presented during the lectures. A single-sided page of notes is allowed. Textbooks and electronic devices of any kind are not allowed. The written examination will assess the following:
• The theoretical understanding of the basic principles of the presented machine learning approaches
• The knowledge and understanding of the different approaches that have been presented during the lectures
• The ability of the students to apply ML techniques to a simple numerical case study.
• The ability of the students to design, implement and evaluate code in the Python language and its ML libraries

The projects will address machine learning tasks. For each project, a dataset will be provided, and the students will have to develop a pipeline using suitable models for the specific tasks based on the topics and tools presented during lectures and laboratories. Each group will have to provide a technical report detailing the methodology employed and critically analyzing the results. The report will assess:
• The degree of understanding of the theoretical principles of different machine-learning approaches
• The ability of the students to analyze a specific problem, assessing which approaches, among those that have been presented, are more suited to solve the task
• The working knowledge of the Python language and the major data mining and machine learning libraries
• The ability of the students to apply the studied methods to devise suitable solutions for the specific case study
• The ability of the students to critically evaluate the effectiveness of the proposed approaches
After submitting their report, the students will have the possibility to peer-review other reports from the other group to obtain bonus points.

The final grade will be given by the weighted average of the written exam (40%) and the project report grade (60%). Each part will have a grade between 0 and 30 cum laude. Both parts must be sufficient to pass the exam.
• Individual written exam (40%, at least 18/30)
• Project (60%, at least 18/30)

In addition to the message sent by the online system, students with disabilities or Specific Learning Disorders (SLD) are invited to directly inform the professor in charge of the course about the special arrangements for the exam that have been agreed with the Special Needs Unit. The professor has to be informed at least one week before the beginning of the examination session in order to provide students with the most suitable arrangements for each specific type of exam.