COURSE DETAILS

" MACHINE LEARNING AND BIG DATA "

SSD ING-INF/05

DEGREE PROGRAMME: AUTONOMOUS VEHICLE ENGINEERING (MOVE)

ACADEMIC YEAR 2022-2023

GENERAL INFORMATION – TEACHER REFERENCES

TEACHER: GIANCARLO SPERLÌ
PHONE:
EMAIL: GIANCARLO.SPERLI@UNINA.IT

GENERAL INFORMATION ABOUT THE COURSE

INTEGRATED COURSE (IF APPLICABLE): NO
MODULE (IF APPLICABLE): -
CHANNEL (IF APPLICABLE): -
YEAR OF THE DEGREE PROGRAMME: I
SEMESTER: II
CFU: 9

REQUIRED PRELIMINARY COURSES

None.

PREREQUISITES

None.

LEARNING GOALS

The course aims to present the main machine learning techniques, covering all aspects from data preparation to performance evaluation, through practical exercises with commercial and/or open-source tools. An introduction to Big Data and Data Analytics lifecycle is also provided, concerning the design of large and complex databases and the process of modelling, acquiring, sharing, analysing, and visualizing the information embedded into Big Data.

EXPECTED LEARNING OUTCOMES (DUBLIN DESCRIPTORS)

Knowledge and understanding

The student must know the main Machine Learning (ML) algorithms and must demonstrate the ability to choose the most suitable ML algorithm to solve a specific problem, based on the requirements of the problem itself. The student must also know the techniques to be used for properly evaluating the performance of ML algorithms.
The students must know the main Big Data frameworks in order to acquire, model, share, analyze and visualize large amount of information. The student must also demonstrate that he/she is able to choose the most suitable framework to deal with different tasks.

Applying knowledge and understanding

The student must demonstrate to be able to solve real problems by using Machine Learning techniques. The student must also demonstrate that he/she can properly evaluate the performance of a machine-learning based system.
The student must demonstrate to be able to manage, model and analyze large amount of data through different Big Data frameworks for dealing with different tasks, also evaluating the performance of the designed architecture.

COURSE CONTENT/SYLLABUS

Data Mining and Machine Learning. Knowledge representation: Trees, Rules, Clusters. (0.5 CFU)
Basic Machine Learning methods: Statistical Modelling, Linear Models, Instance-based learning, Clustering. (0.75 CFU)
Performance Evaluation: Cross-Validation, Cost-sensitive classification, ROC curves. (0.5 CFU)
Advanced Machine Learning: Decision Trees, Support Vector Machines, MLP (0.75 CFU)
Data transformation: attribute selection, PCA (0.25 CFU).
Deep Learning: training and performance evaluation of Deep Networks, Convolutional Neural Networks. (0.75 CFU)
Introduction to database systems. Data model for Big Data. NoSQL database: Key-value - Column-family, Graph database systems. (1.5 CFU)
Introduction to Big Data systems (BD): definition of a BD system. The Hadoop ecosystem. Yarn. Pig. Hive. Giraph. Spark. (2.5 CFU)
Introduction to Big Data Analytics (BDA): BDA Lifecycle: knowledge discovery in the database, data preparation, model planning, model building, data visualization. (1 CFU)
Examples of commercial and open-source Tools: Oracle, IBM Business Analytics, Microsoft Power BI, Microsoft Azure. AWS. SAP Hana (1 CFU).

READINGS/BIBLIOGRAPHY

Data mining: practical machine learning tools and techniques. — 4th ed. / Ian H. Witten, Frank Eibe, Mark A. Hall, Christopher J. Pal —The Morgan Kaufmann, 2017.
Mining of Massive Datasets", J. Leskovec, A. Rajaraman, J.D.Ullman, 2014 (online book).

TEACHING METHODS

Lectures and laboratory activities.

EXAMINATION/EVALUATION CRITERIA

Exam type	
written and oral	
only written	
only oral	X
project discussion	
other	

a) Evaluation pattern:

N.A.