Politecnico di Torino

Data Science & Machine Learning for engineering applications

AA 2023-2024

DataBase and Data Mining Group

Prof. Tania Cerquitelli

General Information

- SSD: ING-INF/05
- CFU: 6
- Professor: Tania Cerquitelli
  - Department of Computer and Control Engineering
  - Office: Cittadella Building, Corso Castelfidardo 39, IV Floor, DAUIN
  - Phone# (office): 011 090 7178
- Teaching Assistant: Salvatore Greco
- Email: name.surname@polito.it

Course Description

- This course explores **how data-driven algorithms** can help engineering **applications** become **smarter**.
- Emerging disciplines, such as **Data Science**, **Machine Learning**, **Deep Learning** could lead to new opportunities to develop cutting-edge and unconventional engineering applications.
- The course focuses on designing and implementing data-driven processes to extract knowledge from data and support decision-making.
  - It introduces the **data science process**, focusing on its main phases.
  - It provides a **theoretical and practical understanding** of data science and machine learning algorithms commonly used to analyze large and heterogeneous data generated in engineering scenarios.
  - It introduces the **Python language** and **state-of-the-art data science and machine learning libraries** to design and develop machine learning-based applications.

Teaching approach

- *Learning-by-doing* approach
- Theoretical lectures will be interleaved with laboratory (hands-on) sessions to allow understanding of algorithms through experimental activities on all the phases of a data science process
  - data preparation and cleaning,
  - data exploration and characterization,
  - selection and tuning of machine learning algorithm,
  - experiments (i.e., execution of instructions written in Python) and result assessment
- using the most widespread open-source tools and libraries.

Expected Learning Outcomes

- Students will acquire abilities by exploiting state-of-the-art machine learning algorithms tailored to engineering applications.
- **Theory-based knowledge**
  - Knowledge of the main phases characterizing a data science and machine learning process for a real-life engineering application.
  - Knowledge of the major data mining algorithms for classification, regression, clustering, and association rule mining.
  - Knowledge of the major machine learning and deep learning algorithms.
- **Hands-on skills** (abilities acquired through active engagement and practical learning)
  - Knowledge of the Python language.
  - Knowledge of the major data science and machine learning libraries.
  - Ability to design, implement, and evaluate a data science process.
  - Ability to design, implement, and evaluate a machine learning and deep learning algorithm.
  - Ability to use and tune data science and machine learning algorithms.

Course topics

- Data science process: main phases
  - Data collection, cleaning, transformation and enrichment, and feature engineering
- Data science algorithms: association rules, classification, and clustering
- Machine learning and deep learning algorithms
- Introduction to Python and data science and machine learning libraries (e.g., scikit-learn)
- Case study analysis
  - Design and development of data science process and machine learning and deep learning algorithms for engineering applications

Course structure

PoliTo DBMG

- The course includes lectures and practices on the course topics, particularly on data science process design, data preprocessing, and machine learning and deep learning algorithms.
- The course includes laboratory (hands-on) sessions on the data science processes and machine learning algorithms for engineering applications.
  - Laboratory sessions allow experimental activities (i.e., run) on the most widespread tools and libraries.

Exam structure & Grading criteria

- The exam includes
  - homework (4 points)
  - a group project (20 points)
  - a written part (10 points)
- The final grade is given by the sum of all three parts.
- The professor may request an integrative test to confirm the obtained evaluation.
- Constraints
  - Grade of the group project is greater than or equal to 12,
  - Grade of the written part is greater than or equal to 6,
  - Group project + written part must exceed or equal to 18.
  - Homework points will be considered only if the (Group project + written part) >= 18
- If the final score is strictly greater than 31 the registered score will be 30 with honor.

Assessment

- Group project and Homework will assess
  - the ability to design, implement and evaluate a complete data science process, including the evaluation and tuning of machine learning algorithms and result assessment for a specific engineering application
  - the working knowledge of the Python language and the major data science and machine learning libraries.
- Computer-based written test in class using POLITO platform will assess
  - the knowledge of the data preparation techniques and the major data mining algorithms for classification, regression, clustering, and association rule mining
  - the knowledge of machine learning and deep learning for engineering applications.

Homework

- During the course, we will assign homework
  - Participants can practice data science and machine learning algorithms in Python and the major data science and machine learning libraries and become proficient.
- 7 homework (hands-on activities to be delivered by the deadline)
  - The first 6 homework handed in by the deadline will give 0,5/30
    - They are very correlated to the hands-on activities shown during the lectures
  - The 7th homework handed in by the deadline will give 1/30
    - a specific use case that requires the use of advanced machine learning algorithms for an image dataset
  - Overall, 4 points on the final score
- The points for the homework are valid until the exam session in January 2025 (included)

Group project

- A team of 6-8 students
- To design and develop a data science process, based on machine learning algorithms, for solving a data analytics task related to a specific engineering application.
- The project is assigned after 7-8 weeks of lectures.
- A written report has to be delivered (further details will be provided) to present the work.
- The evaluation of the group project is based on the performance and accuracy of the proposed solution, in terms of standard quality measures (e.g., prediction accuracy), completeness (i.e., in depth analysis of each phase of the designed process and motivation for selecting given techniques and algorithms), robustness and sensitivity analysis.

Written part

- It covers the theoretical part of the course.
  - It includes multiple choice and box-to-fill questions related to the theoretical part of the course.
  - For multiple choice questions, wrong answers are penalized.
  - The written exam lasts 60 minutes.
  - Textbooks, notes, electronic devices of any kind are not allowed.