Universidade do Minho
Departamento de Informática

Dados e Aprendizagem Automática
1° Ano, 1° Semestre
Ano letivo 2025/2026

Group Practical Work
October 2025

Theme

Design and optimization of Machine Learning models

Learning Objectives

The aim of this assignment is to sensitize and motivate students to design and develop a Machine Learning project using, among others, the learning models covered throughout the semester.

Work Group

Work groups can be formed by students from different PL class sections but cannot exceed 4 members. Students from MIA and MECD can only form groups with elements of their course. . Students should use this link to register their group: https://docs.google.com/spreadsheets/d/1JfKpad2RJGuDT8u4UUwsr5DedZG63jf8BQc_zwFjJJE/edit?usp=sharing

Introduction

The modeling of road traffic flow is a well-known problem with stochastic and non-linear characteristics. However, several models have emerged in the literature showing remarkable potential for this type of prediction. With this in mind, a dataset was built containing data on vehicle traffic in a Portuguese city over a period of more than one year. The objective of this work is, among others, to develop Machine Learning models capable of predicting road traffic flow at a given hour in the mentioned city.

Assignment

The assignment includes 2 TASKS.

ANALYSIS AND VALIDATION TASK :

- Explore, analyse and prepare the dataset provided, seeking to extract relevant knowledge in the context of the problem in question.

- Obtain and critically analyze some modelling results.

COMPETITION TASK :

- The groups will have to work on the dataset available at **https://www.kaggle.com/competitions/DAA-TG**

    o The link above redirects to the Kaggle platform where a competition has been created. The dataset to be used in the competition, as well as all the details and how it works, are available at that link;

    o The first step is to access the Kaggle platform using the following link to register for the competition:

https://www.kaggle.com/t/48a8e6021e9b49dcb3047f27848e85ae

Students should then form teams with the other members of the working group. The team name should follow the format **GRUPO_<CURSO(s)>_<X>** where **<CURSO(s)>** corresponds to the master's programme(s) (MEI, MECD, MES, MIA, MMC or MCS) and **<X>** to the group number. Submissions cannot be made on the Kaggle platform as long as the group is incomplete;

- Design and optimisation of Machine Learning models for the competition dataset:

  o Students must submit the results obtained on the Kaggle platform in order to obtain the model's Accuracy score.

  o There is a **daily limit of 3 valid submissions**, so you should endeavour to start submitting as soon as possible. The competition closes at the end of **2 January 2026**;

- Obtain and critically analyze the results;

- Interpret the results acquired and define their usefulness in the context of the problem underlying the dataset worked on. Determine and explain the most relevant results.

Delivery and Assessment

The results obtained should be the subject of a report, limited to 20 pages, which presents, among other things:

- The areas to be tackled, the objectives and how it is proposed to achieve them;

- The methodology used and how it was applied;

- Detailed description and exploration of the dataset and any processing carried out;

- Description of the models developed, what their characteristics are, how and on what hyperparameters the model was tuned, training characteristics, among other details that should be provided;

- Summary of the results obtained and their critical analysis;

- Presentation of suggestions and recommendations after analysing the results obtained and the models developed.

The whole process must be accompanied by examples and indications that make it possible to reproduce all the steps taken and the results obtained.

During the class period on **2 and 3 of december 2025**, a checkpoint will be made on the work carried out by the working groups, with each group using the means it considers most appropriate to demonstrate the results obtained.

On **5,6 and 7 of january 2026** there will be presentation sessions for the work carried out in both TASKS. The working groups will have to choose the slot they want for their presentation, which will be made available in the coming weeks. Each group will have 10 minutes to make their presentation, using whatever means they deem most appropriate.

The report, as well as all other elements produced, must be compressed into a single zip file which must be submitted by a member of the group by **3 january 2026** on the University of Minho's e-learning platform (in "*Conteúdo/Instrumentos de Avaliação em Grupo/Submissão TPG*").

Peer Assessment

Each group should carry out a collective analysis of the contribution and effort that each member has made to the progress of the work. From this analysis they should be able to identify the members who worked above, at and below the average. For this assessment component, 1 value is provided for each student, reflecting their individual contribution to the development of this assessment tool.

So, one member of the group should send an email, with the other members of the group in CC, to valves@di.uminho.pt, filipa.ferraz@di.uminho.pt, dad@di.uminho.pt, fm@di.uminho.pt and gjfbarbosa@di.uminho.pt. The subject should be "**AP DAA - GRUPO_<CURSO(s)>_<X> - Avaliação Por Pares**".

In the text of the email, each member of the group should indicate their delta (the amount to be added to the mark for this component). Remember that deltas can be negative, zero or positive and that, in each group, the sum of the deltas must always be equal to 0.00 and, individually, can never exceed one.

<u>Example 1</u> (corresponds to an equal effort by all):

PG1234 João DELTA = 0
PG5678 António DELTA = 0
PG9123 Maria DELTA = 0
PG4567 Rita DELTA = 0

<u>Example 2</u> (António receives 1 additional value, Rita keeps her classification, João and Maria are deducted 0.5 values each):

PG1234 João DELTA = -0.5
PG5678 António DELTA = 1
PG9123 Maria DELTA = -0.5
PG4567 Rita DELTA = 0

Code of Conduct

The participants in this academic work declare that they have acted with integrity and confirm that they have not resorted to the practice of plagiarism or any form of misuse or falsification of information or results in any of the stages leading to its preparation. They also declare that they are aware of and have respected the University of Minho's Code of Ethical Conduct.

References

In addition to the material provided in class, it is advisable to consult sources such as:

- Machine Learning. T. Michell, McGraw Hill, ISBN ISBN: 978-1259096952, 2017.

- Introduction to Machine Learning. Alpaydin, E. ISBN: 978-0-262-02818-9. Published by The MIT Press, 2014.

- Computational Intelligence: An Introduction. Engelbrecht A., Wiley & Sons. 2nd Edition, ISBN: 978-0470035610, 2007.

- The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Hastie, T., R. Tibshirani, J. Friedman, 12nd Edition, Springer, ISBN: 978-0387848570, 2016.

- Machine Learning: A Probabilistic Perspective. K.P. Murphy, 4th Edition, The MIT Press, ISBN: 978-0262018029, 2012.