MACHINE LEARNING 30412 – Project 2025 DATA

In this project, you will explore and work with a dataset containing information on over 15.000 football players. The dataset includes a wide range of features such as player positions, attributes, technical stats, and club affiliations.

Your task is to develop a machine learning models that predicts a player's market value in Euros (`value_eur`) based on their available characteristics. This problem will require you to go through the full machine learning pipeline, including data preprocessing, feature selection and engineering, model training, and performance evaluation.

Files

You can download from Blackboard the following csv files:

- *train.csv*: training data that you can use to train your algorithms;
- *test.csv*: test data for which you have to make predictions;
- *submission.csv*: an example of the structure of the final csv file that you'll have to upload as part of your final submission. The file currently contains random placeholder values for value_eur, which you will need to replace with your own predictions. The format of the csv file which you upload has to be exactly the same of this csv.
- *column_descriptions*.csv: a brief explanation of each available feature is provided. Attributes that describe players' playing style or on-field performance are labeled as "technical attributes", their exact definitions are not essential for the scope of this project.

Metric

For each test input you will have to predict the market value of the player (`value_eur`).

The metric with which we will evaluate your predictions is the Root Mean Squared Error (RMSE).