

## ML 2016 Student Contest description

You are provided with a Learning Dataset describing the matches of a game. For every year, you are provided with the descriptions of (nearly all, but not all) matches, with the ID of the two players, the score of the winning and losing player, the location (one out of 3 possible values) and several other parameters you don't need to understand (your system will eventually learn if they are relevant or not).

You should use whatever learning algorithm, preprocessing, and algorithm modification you want, to learn a binary classifier to predict the winning player when ScoreP1 and ScoreP2 are unknown. You can also add features, delete features and summarize sets of features already in the dataset. Of course, predicting scores is much more complicated and we don't expect this.

We suggest that you learn and test on 90% of your learning set and take a 10% for final tuning.

ON JUNE 15 you will be given a test set TS (a random number of matches for which the scores will NOT be assigned), and you must predict the winning team for each match in TS, and return those prediction to me and Georgia.

You must also provide us with your software and a detailed description (at least 5 pages pdf) of your project (you have an extra week to do so, if you wish), your choices and thoughts. You are expected to work in teams of two.

Every project will be evaluated with a grade from 0 to 30, and the grade will be averaged with the score of your written exam. The winning team (the one with highest precision/recall in predictions) will be prized with a +2 on the final (averaged) grade, after we check all the details.

NOTE: Don't start too early! You should learn more algorithms first! And you are free, of course, to invent your own algorithm.