DATA MINING PROJECT

Leonardo Martini, Na Zhu, Simone Caldaro

INPUT: Matches, Players, [weather, fans]

* Source web page:
* Number of *tables*:
* Mega/Gigabytes:

Tensor flow:

OUTPUT: Match result, its probability, connection with bet-odds

1. ***Pre-process***:

Given a match (team1 vs team2), we build a list of matches between these two.

*match\_list* [<d1, r1>,<d2, r2>,…,<d\_i, r\_i>] d\_i is the date of match i,

r\_i is the result of match i.

We must consider the order 🡪 team1 is the home team in each entry of *match\_list*.

1. ***TensorFlow***:
2. ***Process output data***:

# HOW REPRESENT MATCHES

**Team Forms**. The team forms are representative of a team’s recent performance. The team form values will be represented using four different representations.

* ***Home Team Form***: The total amount of points the home team got from the last *n* (the value of *n* will be determined experimentally) games.
* ***Away Team Form***: The equivalent information for the away team.

1. Representation 1: numeric values of the team forms, normalized to interval [0, 3].
2. Representation 2: discretized value of the team forms. We had reason to believe that the classifiers do not distinguish between values well enough while using r1, so we discretized r1 using the set of rules shown in Table 1
3. Representation 3: subtracted value between the home team form and away team form. This subtracted value is normalized to the interval [-3,3]; a negative value means away team superiority and a positive value means home team superiority while zero means an equal advantage.
4. Representation 4: discretized values of r3. This representation will be discretized by equal frequency into three bins.