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.

## Team Home & Away Form

The home and away forms are representative of the teams’ recent home performance and recent away performance respectively. Their values will also be represented by the four representations used in the Team Form experiments.

* ***Home Team Form***: The total amount of points the home team got from the last n games.
* ***Home Team Home Form***: The amount of points gotten by the home team from the last n home games.
* ***Home Team Away Form***: The amount of points gotten by the home team from the last n away games.
* ***Away Team Form***: The equivalent information for the away team form; Away Team Home Form and Away Team Away Form.

1. Representation 1: This represents the numeric value of the  team home and away forms, normalized to interval [0,3]
2. Representation 2: This represents the discretized values of r1
3. Representation 3: This represents the subtracted value between the home team form and away team form; the home team home form and away team away form; the home team away form and away team home form. This subtracted value is normalized to interval [-3,3]. A negative value means away team superiority and a positive value means home team superiority.
4. Representation 4: This represents the discretized values of r3

## Match statistics

* ***Goals***: The subtracted difference in goals from home and away teams over 9 games.
* ***Shots***: The subtracted difference in shots from home and away teams over 9 games.

ShotOnHome – ShotTargetHome, ShotOnAway - ShotTargetAway

* ***Shots on target***: The subtracted difference in shots on target from home and away teams over 9 games.
* ***Goals Ratio***: The subtracted difference in goals ratio from home and away teams over 9 games.
* ***Shots Ratio***: The subtracted difference in shots ratio from home and away teams over 9 games.
* ***Form***: The subtracted difference in points from home and away teams over 9 games (represented as Form-9); 11 games (represented as Form-11) and all subsequent games (represented as Total Form).

Window:

* no window 🡪 consider all the matches happened previously (start from the season 2008/2009, first match. Even if no match before the previous stage are available).
* 1 year window 🡪 consider all the matches happened in the previous year.

Starting from the first match of season 2009/2010, consider the season 2008/2009.

(Window of 38 matches)

* 2 years window 🡪 consider all the matches happened in the previous year.

Starting from the first match of season 2009/2010, consider the season 2008/2009.

(Window of 76 matches)

* 3 years window 🡪 consider all the matches happened in the previous year.

Starting from the first match of season 2009/2010, consider the season 2008/2009.

(Window of 114 matches)

* …
* i years 🡪 38 x i matches
* 30 matches
* 25 matches
* 20 matches
* …
* 5 matches