xgb_df

2024-12-07

Can we predict the total number of goals scored in a game?

Let's try with xgboost

```r
set.seed(42)
df <- read.csv("match_data_clean.csv")
#Remove vars that would make such a prediction too easy
#Also remove, data, time, team name, etc.
df_reg <- df[,-which(names(df) %in% c("team1", "team2",
                                      "number.of.goals.team1",
                                "number.of.goals.team2","date",
                                "number.of.goals.team1",
                                  "hour", "category",
                                  "conceded.team1","conceded.team2",
                                  "goal.inside.the.penalty.area.team1",
                                "goal.inside.the.penalty.area.team2",
                                  "goal.outside.the.penalty.area.team1",
                                "goal.outside.the.penalty.area.team2",
                                  "own.goals.team1", "own.goals.team2",
                                "assists.team1", "assists.team2",
                                  "penalties.scored.team1",
                                "penalties.scored.team2", "outcome",
                                "elimination"))]
write.csv(df_reg, "xgb_df.csv", row.names = FALSE)
```