Rocket League Analysis

Katie Lee, Hugo Leo, Linda Wang

March 4, 2022

Contents

1 Executive Summary

Rocket League is a multi-platform game where players control cars with a rocket booster and aim to score as many goals as possible within 5 minutes using an oversized ball. The ball never leaves the field and is a team game, but for our analysis, we focused on 2 versus 2 player mode. Players have the ability to control the direction, velocity, and rotation of their cars and can jump to hit the ball. This allows players the ability to take shots or control the car to dribble, pass, and block incoming shots.

Our dataset includes random frames of 2 versus 2 player Rocket League games that includes location data of all 4 players, car features, ball position, and whether or not a particular shot resulted in a goal.

For our analysis, we built an expected goals (xG) model using location data to predict goals for a given frozen game frame. xG is important in that it is an estimator of goals a team is expected to score in the long run. Using our xG model, we explored individual player ability and analyzed the question: are good players good because they use more attempts, or are good players good because they are truly better skilled?

2 Data Source

[to write]

3 Data Cleaning

- dropped car features, only used location data
- because we are interested in expected goals and evaluating shot ability, we filtered for shot = true only

```
data_condensed <- read.csv("../results/clean_shot_data.csv") %>%
  select(c(goal, ball_pos_x, ball_pos_y, ball_pos_z,
           distanceToGoal, shot_taker_pos_x))

data_condensed %>%
  kable(format = "latex", row.names = NA,
        booktabs = TRUE, digits = 2,
        caption = "Subset of Location Features") %>%
  kable_styling(position = "center", latex_options = "HOLD_position")
```

Table 1: Subset of Location Features

goal	ball_pos_x	ball_pos_y	ball_pos_z	distanceToGoal	shot_taker_pos_x
0	2099	616	120	4681	2220
0	1013	3655	348	1490	1139
0	-213	941	205	4343	-189
1	530	4797	283	501	532
0	-3907	-2882	1183	8568	-3998
1	350	4390	229	1455	278

4 Exploratory Data Analysis

As a first step, we explored the location data to get a sense of the placement of the goal. Based on Figure 1, it seems that the goal post is located at $(0, 5000, 0)$.
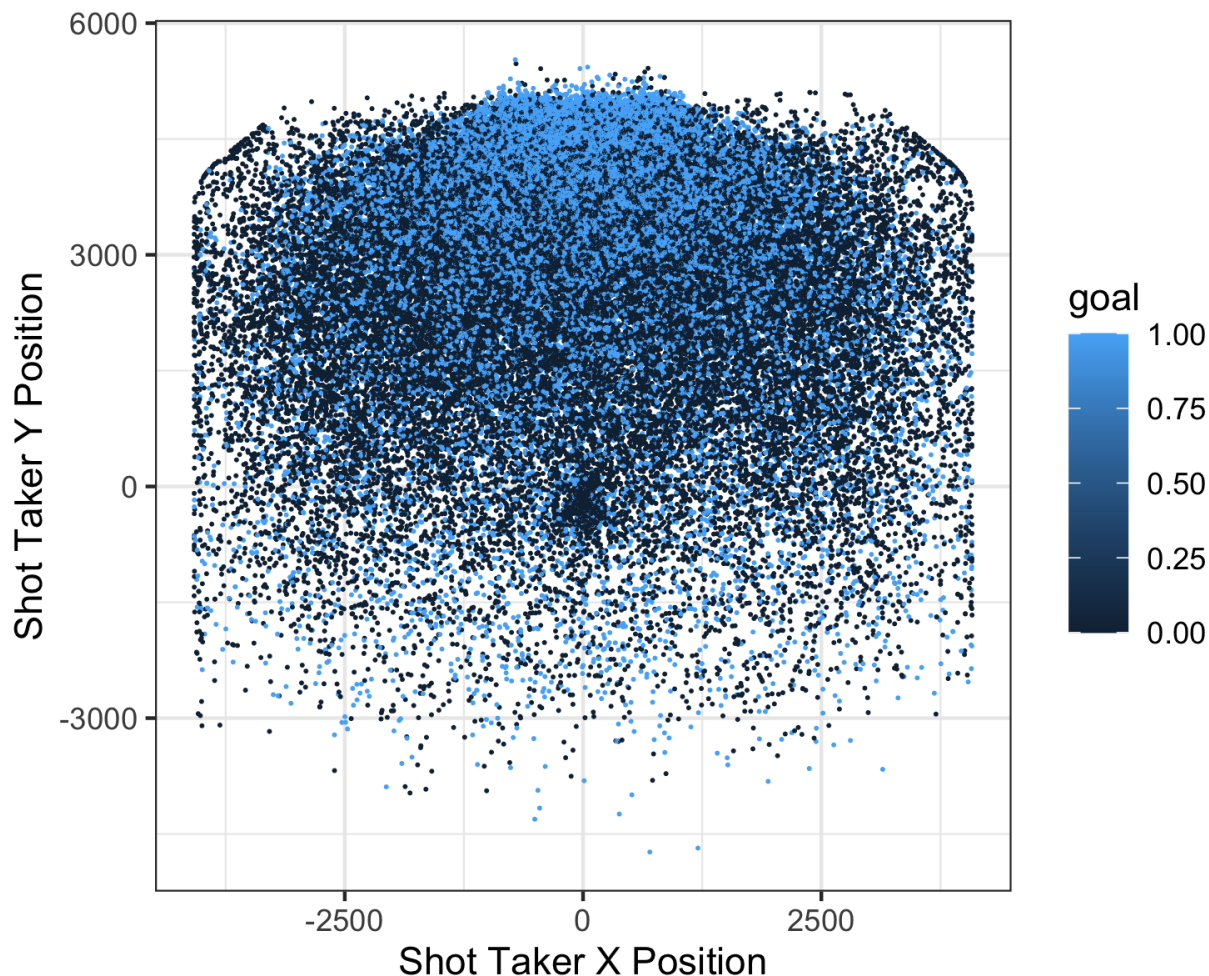
Figure 1: X-Y Field and Shot Outcome

5 Expected Goals Models

We built two expected goals models using logistic regression and boosted trees. Because of the nature of the dataset, we have data about the ball's trajectory, so we can easily compute whether a shot will result in a goal or not. However, for our model, we used position, velocity, angular velocity, and rotation of the shot taker, defender, and teammate to build our model. To train our model, we predicted the outcome of a shot (goal or no goal) for a particular frame. This implies that our xG estimates are for a set of positions of the shot taker, teammate, and defenders.

5.1 Analysis Goals

- how is xg used in soccer
- why is xg important

5.2 Feature Engineering

We used feature engineering to transform the raw data into features that can be used in supervised learning.

5.3 Logistic Regression

```
glm_fit_output <- read.csv("../results/glm_fit_output.csv") %>%
  select(X, coefficients.Estimate, coefficients.Pr...z..) %>%
  arrange(coefficients.Pr...z..) %>%
  head(10) %>%
  select(X, coefficients.Estimate) %>%
  rename("Coefficient Estimate" = coefficients.Estimate)

glm_fit_output %>%
  kable(format = "latex", row.names = NA,
        booktabs = TRUE, digits = 2,
        caption = "Subset of Statististically Significant Covariates") %>%
  kable_styling(position = "center", latex_options = "HOLD_position")
```

Table 2: Subset of Statististically Significant Covariates

X	Coefficient Estimate
ball_vel_y	0.00e+00
opp_1_pos_y	0.00e+00
opp_2_pos_y	0.00e+00
ball_ang_vel_x	0.00e+00
cos_theta_opp_2	3.40e+06
shot_taker_vel_z	0.00e+00
cos_theta_opp_1	2.85e+06
logDistanceToGoal	-5.50e-01
distanceToOpp1	0.00e+00
shot_taker_pos_z	0.00e+00