DAR F23 Project Status Notebook

Hockey Analytics

Amy Enyenihi

2023-10-02

Contents

Weekly Work Summary

NOTE: Follow an outline format; use bullets to express individual points.

- RCS ID: enyena
- Project Name: Hockey Analytics
- Summary of work since last week
  - Describe the important aspects of what you worked on and accomplished I have spent time looking through how the existing features were created and figuring out how to go about creating a new one.
- NEW: Summary of github issues added and worked
  - Issues that you've submitted: N/A
- Summary of github commits
  - branch name(s): dar-enyena
  - no new commits since assignment02
- List of presentations, papers, or other outputs
  - We presented our classification and feature selection presentation.
- List of references (if necessary)
- Indicate any use of group shared code base
- Indicate which parts of your described work were done by you or as part of joint efforts
- **Required:** Provide illustrating figures and/or tables

Personal Contribution

- Clearly defined, unique contribution(s) done by you: code, ideas, writing...
- Include github issues you've addressed

Analysis: Question 1 - Data Frame Construction

Question being asked

Provide in natural language a statement of what question you're trying to answer

How is the shots_stats.df data frame created?

Data Preparation

No data preparation was needed. We are analyzing code written by Mohamed. I met with him on the weekend to go over it, but he also spent time in class showing the whole group the process.

I included the region of the code we were looking at but it is all commented out to allow the notebook to be knitted

```
## Get stats for the shot in every sequence
#shots_stats.df <- seq_along(sequences) %>%
#  map_dfr(goalShotStats) %>%
  # Some models can't use logical data
#  mutate_if(is.logical, as.numeric)
```

Analysis: Methods and results

There was no statistical analysis.

```
#N/A
```

Discussion of results

The lines of code we were looking at were actually simple to understand once explained. The seq_along function returns a sequence of the same length of the item passed in the function. The function map_dfr iterates through sequences and performs the function goalShotStats on each item. The resulting stats are added to the growing data frame until each item is sequences has been passed through.

goalShotStats is a helper function that is found in a separate R file. This is where the actual stats are calculated and cleaned up for the data frame.

Analysis: Question 2 - Factoring the Goal Outcomes

Question being asked

Provide in natural language a statement of what question you're trying to answer

How can we edit shots_stats_goal.df data frame to factor all possible outcomes instead of categorizing everything as a goal or no goal?

Data Preparation

Provide in natural language a description of the data you are using for this analysis

No data preparation was needed since there was existing code that did any necessary prep.

```
# N/A
```

Analysis: Methods and Results

Describe in natural language a statement of the analysis you're trying to do

Provide clearly commented analysis code; include code for tables and figures!

The current existing data frame factors the result of a shot one of two ways, 0 or 1. There are really four different outcomes: miss, goal, defender block, or save. Since there was talk in the group about further analyzing circumstances surrounding the defender block outcome, I changed the code to account for all four outcomes.

Again, I have commented out the code to allow the notebook to be knitted. I intend on figuring out what to do to avoid this going forward with future status notebooks

```
#*Adding a feature that factors each outcome as a number 1 through 4
# outcomes.goal <- (info$outcome) %>% as.numeric %>% as.factor

#*Append to shots_stats.df
# shots_stats_goal.df <- cbind(shots_stats.df, outcomes.goal)

#*Save this dataframe on the file system in case we want to simply load it later (to save time)
# saveRDS(shots_stats_goal.df, "shots_stats_goal.df.Rds")
```

Discussion of results

Provide in natural language a clear discussion of your observations.

Because there was no analysis completed, there is no result to take away from this change yet. The data frame has simply been modified however, this feature modification can hopefully be useful for future analysis.

Summary and next steps

Provide in natural language a clear summary and your proposed next steps.

All my previous work has been about understanding Mohamed's code and getting up to speed with his notebooks. After that, I was able to understand what it takes to add a feature. I started with a small change. My next steps will either be adding a more complex feature or potentially completing some type of analysis using the new outcome feature.