



Examining Substitution Decision Making in the Canadian Premier League Using Machine Learning

Jacob Klein
Department of Mathematical Sciences
Rochester Institute of Technology
Rochester, New York
jek8543@rit.edu



Abstract

One of the most influential changes that a manager can make in a game of soccer is the substitution. Using a game-by-game dataset from the Canadian Premier League (provided by StatsPerform), this work attempts to understand what statistics (visual statistics like turnovers and underlying statistics like expected goals) coaches may be making substitutions based off. We employ random forest modeling to see what variables are important to the decision-making process. This information and analysis can be replicated with similar datasets to assist owners and general managers in the coaching recruitment process.

References

- Decroos, Tom, et. all, "Actions Speak Louder Than Goals: Valuing Player Actions in Soccer", July 2019
- Muller, John, "Goals Added: Introducing a New Way to Measure Soccer", May 4 2020
- Myers, Bret, "A Proposed Decision Rule for the Timing of Soccer Substitutions", 2011
- Stats Perform, "Canadian Premier League Centre Circle Data", 2019

Introduction

Substitutions have played a vital role in determining soccer matches. Of the past three major international tournament finals, two were 1-0 victories where a substitute scored the winning goal. However, the priority with substitutions isn't always to bring an attacker on to try and win a game. It can be to change formational shape or to sure up defenses. The primary focus of this work, however, is looking at why players are brought off.

Understanding this can help to unlock what coaches are trying to accomplish with their substitutions.

Inspiration

The inspiration for this paper was Bret Myers' "A Proposed Decision Rule for the Timing of Soccer Substitutions". While the work done for that paper was focused on the timing of substitutions and on creating a strategy for managers, this paper focuses more on the statistics behind them and on just observing managerial behavior.

Dataset

The dataset featured in this work is from the 2019 Canadian Premier League season. The 2020 season is not used in this dataset because the teams were not limited to three substitutions. Each row within the dataset represents a player's statistics within a certain match.

Goalkeepers will be removed from the initial dataset, as the decision to take off a goalkeeper is, for the most part, entirely down to injury reasons (Within the dataset, only 1 goalkeeper was substituted off). The dataset will also remove players who are subbed on, as we can't build the decision model on players who can't be decided on. The dataset does not specify whether players were taken off because of injury or not. To account for this, we will remove any players who were taken off before halftime. The players must also be separated into their respective teams.

All statistics will be used as rate statistics (the statistic divided by the minutes played in a match). This is to make sure that player statistics are not lowered because they were subbed off.

Analysis

For the analysis, we are going to train models for every team based on four categories: General, Attackers, Midfielders, Defenders. We will use more general statistics and have all players within a team be used to train the model. For the latter three categories, we will only use players within the specified positions and focus on statistics that are relevant to the position.

General

All Positions

- * Successful Duels Per Minute -> SDuels
- * Failed Duels Per Minute -> FDuels
- * Successful Passes Per Minute -> SPasses
- * Failed Passes Per Minute -> FPasses
- * Total Times Dispossessed Per Minute -> Dispos

Attackers

- * Expected Goals Per Minute -> xG
- * Expected Assists Per Minute -> xA
- * Shots On Goal Per Minute -> OnGoal
- * Missed Shots Per Minute -> Missed
- * Touches in the Opponent's Box Per Minute -> TouchBox
- * Successful Passes in the Attacking Third Per Minute -> A3Passes
- * Times Offsides Per Minute -> Offsides
- * Successful Dribbles Per Minute -> Dribbles

Midfielders

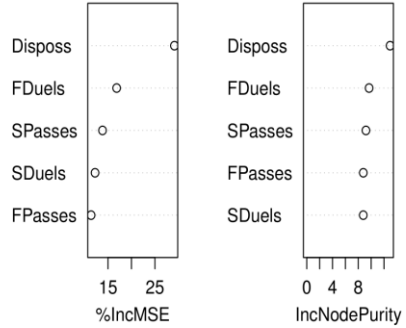
- * Successful Duels Per Minute -> SDuels
- * Successful Passes Per Minute -> SPasses
- * Successful Tackles Per Minute -> STackles
- * Interceptions Per Minute -> Intcpt
- * Recoveries Per Minute -> Recov
- * Passes Completed in the Middle Third Per Minute -> M3Passes
- * Touches in the Middle Third Per Minute -> M3Touch
- * Fouls in the Middle Third Per Minute -> M3Fouls

Defenders

- * Successful Duels Per Minute -> SDuels
- * Successful Tackles Per Minute -> STackles
- * Interceptions Per Minute -> Intcpt
- * Recoveries Per Minute -> Recov
- * Passes Completed in the Defensive Third Per Minute -> D3Passes
- * Touches in the Defensive Third Per Minute -> D3Touch
- * Fouls in the Defensive Third Per Minute -> D3Fouls
- * Total Clearances Per Minute -> Clear

General

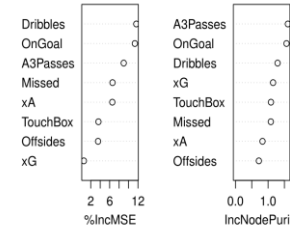
GeneralCavalryForest



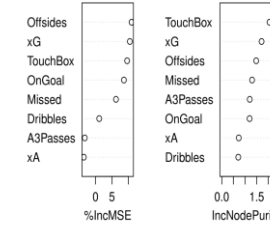
The total times that a player is dispossessed per minute seems to be the most influencing factor on the head coach of Cavalry FC, Tommy Wheeldon Jr., when making a substitution. However, we have only focused on five statistics that are not positional specific, so our analysis is limited.

Attackers

AttackingValourForest



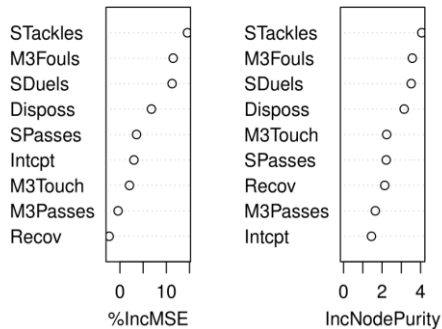
AttackingEdmontonForest



Rob Gale of Valour FC seems to be basing his attacking decisions based off of shot accuracy (high emphasis on shots on target) and keeping hold of the ball in dangerous positions (high emphasis on accurate passes in the final third and successful dribbles). Jeff Paulus of Edmonton FC looks to be basing his decisions based on off the ball-movement (high emphasis on offsides and touches in the opponent's box) and getting quality chances (high emphasis on xG).

Midfielders

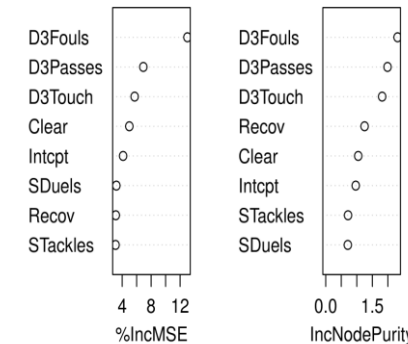
MidHFXForest



A lot of the metrics that appear on the top of both of the variable importance charts are centered around duels and tackles, rather than passes or interceptions. Steven Hart, coach of HFX Wanderers, may be looking for players who aren't winning their duels and are committing too many fouls in the middle of the field when he is deciding who to take off in the midfield.

Defenders

DefenderForgeForest



Above all else, there is a large focus on fouls in the defensive third. Bobby Smyrniotis of Forge FC seems to be making defensive substitutions on the basis of not conceding fouls in the final third.