Research Task: understand what effect different factors have on a future match being conceded

Data Features:

Response variable: DNF(whether or not match being conceded)(Categorical)

Explanatory variables: Gender(Categorical), Tournament(Categorical), Surface(Categorical), Round(Categorical), Best.of(Categorical), WRank(Quantitative), LRank(Quantitative).

There are one response variable and seven explanatory variables.

Analysis Strengths: Some explanatory variables have obvious correlation with the response variable. The data has a large sample.

Analysis Weaknesses: The model dose not explain causation. Some explanatory variables are correlated with each other.

Alternative Example:

Understand what effect different factors have on if a player would enter the basketball Hall of Fame.

The response variable is whether or not is in the Hall of Fame.

The explanatory variables are player’s career stats.

Hall of Famers usually have high achievements in basketball. They often have good stats and have won championships. There are many stats which reflect their performance.

This website provides each player’s career stats.

<https://stats.nba.com/players/list/>

This website has the list of players who are in the Hall of Fame

http://www.hoophall.com/hall-of-famers/browse/?players=true&yr=&country=&loc=&team=&college=&position=&height=