

1) Describe, generally, from start to finish how you approached, executed, and completed the project. Include all relevant materials (e.g. code).

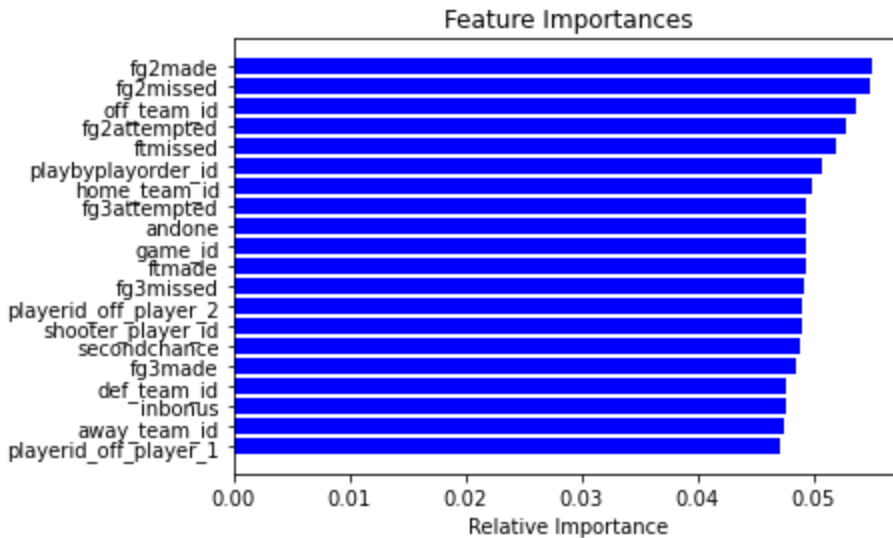
Offensive rebounds are very important in the sport of basketball and are often underrated. Getting an offensive rebound can have a difference between winning or losing a game. In this project I will create a machine learning algorithm that can predict offensive rebounds. This model will help coaches formulate more effective game plans, by knowing which plays are likely to result in offensive rebounds. Offensive rebounds are much less common than defensive rebounds. They are also harder to come by because of offensive/defensive player formations. As a result of this it was no surprise that the dataset had a severe class imbalance issue. In order to correct this, I decided to upsample the minority class, which in this case was a successful offensive rebound.

Missing data in the target column was another important issue that had to be addressed. After exploring the data, I learned that the missing data was successful shots. Shots that didn't result in an offensive/defensive rebound. All the missing data was then filled with 'No' because they didn't result in an offensive rebound. I also decided to make use of the features in the player location dataset. I believe that knowing player locations before and after a shot will be a big factor in whether a play results in an offensive rebound. This is why player formations are so important when it comes to rebounds. Defensive formations are created on the basis of stopping the opponent from scoring and recouping the basketball. As a result of a larger feature space, I decided to use mutual information as a method for feature selection. The top 20 most important features were then passed through a Random Forest Classifier. A Random Forest Classifier was used because it's a robust model that's able to effectively capture patterns within data. With this model I was able to gain further insights into the important features. According to its feature importance method, some of the top features appear to be 'ftmissed', 'ftmade', 'inbonus', 'fg2attempted', and 'fg2missed'.

Link to code:

<https://drive.google.com/drive/folders/1uilhodxjZwsQRESvUB2wKrY99pERA9sA?usp=sharing>

2) Include and describe a visualization from the project. The visualization should highlight a feature or insight from your model. Explain the decisions you made in constructing the visualization.



In order to gain a better insight into my models predictions, I decided to plot the features that were considered important. This was achieved by using the Random Forest Classifier models attribute called 'feature_importance_'. With this plot we can see which features we can keep and which don't provide as much value. According to the chart above we can see that features that include whether a shot was made or missed were valuable. This makes sense because there can't be an offensive/defensive rebound with a made shot. Another important feature was offensive player location during a shot. The location of a player on the court can make a huge difference on obtaining an offensive rebound. This is why defensive formations are optimized to stop shots and get defensive rebounds.

3) A general manager of a team wants you to discuss the findings of your rebounding project with the team's head coach. Please write a short email to the head coach introducing your project and summarizing one or two key findings from your research.

Dear Head Coach,

Recently, our team here in the analytics department was tasked with creating a machine learning model that can predict offensive rebounds. In order to achieve this we have been collecting data from your previous games. The data collected consisted of commonly tracked basketball statistics such as offensive/defensive rebounds, free throws, 2/3pt field goals, player locations on the court, etc. With this information we were able to uncover a couple of important factors when it comes to obtaining an offensive rebound. It turns out that if you'd like to obtain more offensive rebounds one of the best times to do it is when you're in the bonus. This is because the opposing team is going to be playing less aggressively in order to avoid giving away any free throws. Therefore, in this scenario we can afford to be the aggressors in the paint.

Another important factor in obtaining offensive rebounds is our team's offensive positioning during and after a shot. If we can manage to have a select few players get inside defenders and have first looks at rebounds it would improve our chances. Again, this would vary from different in-game situations, because of the delicate balance between playing offense and defense. Whether your goal is to increase the amount of offensive rebounds our team gets or create player formations to stop opponents offensive rebounds, this project will help you gain the upper hand next season!

Thanks,

Christopher Feliz

Data Scientist

OKC Thunder Data Science & Solutions