Predicting Overtime in NBA Games

Overview

6.1% of NBA games go into overtime. Based on how well-matched two teams, the probability would intuitively go up for very evenly-matched teams, and down for teams with unevenly matched teams. Using

The Problem

TV advertising slots are valued based on the number of viewers at the time the ad is screened. Overtime extends the attention of the audience, increasing the value of the ad slots that otherwise would be in the less-valuable "post-game" time slot. A model that provides more articulate predictions can be used for the valuation

Data Collection

Statistics for every NBA game since 1990 were scraped from basketball-reference.com. From the data scraped, the following (pre-game) team standings data points were either gathered or derived:

Wins	Losses	Win-Loss %	Games Behind in Conference	Avg. Points Scored per Game	Avg. Points Allowed per Game

In addition, the score at the end of the first quarter, the half, and the third quarter scores are gathered to create models to use as the game progresses

Analytics and Model Building

Logistic regression and *Random Forest* machine learning method were used to create predictive models based on the given inputs. For each method, for models were constructed, corresponding to (1) Pre-game (2) 1st Quarter, (3) Halftime, and (4) 3rd Quarter. Each model progressively adds the scores as they become available. The models are then tested with training and testing data subsets, and their accuracy and AUC figures are computed in addition to their predictive abilities.

Results

Predictive ability time moves forward. Logistic regression proves to be more accurate, but less articulate. At each point in time, the standard deviation of predictions increases (up to 3.7% by the end of the 3rd quarter.) Random forest model prove more articulate (greater standard deviation of predictions, but less accurate in training vs testing measurements. In either case, the improvement over the flat 6.1% prediction of overtime gives a meaningful input into the models that determine ad slot pricing; thereby allowing TV networks to capture market value that otherwise would be left on the table.