Forecasting the All-NBA Teams

**1.0 Introduction**

At the end of every NBA season teams, players and coaches are recognized for the performance through the NBA season. Outside of the 12 annual awards given at the end of year, players' are anointed to All pro teams (All NBA, All Rookie and All Defensive Teams) on the basis of their individual performance. It’s an honor that is bestowed to the best players in the league at their respective positions: guard, forwards and centers.

Each year three teams composed of two guards, two forwards, and a center voted on by sportswriters and broadcasters across North America. Players receive five points for the first team, three points for the second team, and one point for the third. The five players with the highest total scores make the first team, then next five make the second team, and so forth.

To my knowledge, I don’t believe there is some methodology to the way voters vote for players to receive honors. The award is supposed to reflect the best players in their respective positions but is this always true. Sometimes voters, vote on players by merit, not on individual performance or how one player’s performance helped their team. Given the availability of historical basketball data, can we predict which players will be voted into each team by the end of the 2014-15 season?

**2.0 Data Mining**

Data mining is the process of automatically discovering information from large data. In general, data mining techniques predict the bad, identify the good, automate an existing process or identify patterns in data. These techniques are useful finding non-trivial patterns that might have otherwise remained unknown.

**2.1 Data Mining Tasks**

Data mining tasks can be divided into two major categories: Predictive and Descriptive

Predictive Tasks: the goal of this is to predict the value of a particular attribute (dependent variable) based on the values of other attributes (independent variables). This type of modeling is also known as supervised learning. If the dependent variable being assessed is continuous, the predictive modeling technique is regression. For example, using a set a variables to predict price range or blood pressure. Values under this modeling technique are infinite.

On the hand, if the response variable is categorical, classification techniques are used to predict the function of the explanatory variables. For example, predicting a team will win or lose, or determining cancer class of tissue sample.

Descriptive Tasks: The goal under these tasks are to understand patterns that summarize the relationships in the data. There are no dependent variables, just a vector of predictors. The objective is to find groups features that have similar charactertistics or find a combination of features that explain the variation in the data. These tasks are useful explanatory tasks that can be used as a preproccesing step for supervised learning procedures. For example: Correlations, trends, clusters, anomalies.

**2.2 Data Mining meets basketball**

**Defining the problem**

Using data mining techniques, can we predict what players will make the 2015 All-NBA team? This question can be assessed using classification techniques. Using basketball statistical attributes used to characterize an individual players performance as the indepdent variables can we predict whether a player is voted one of the all-nba teams(categorical depedennt varibles).

Problem statement and hypothesis

Description of your data set and how it was obtained

Description of any pre-processing steps you took

What you learned from exploring the data, including visualizations

How you chose which features to use in your analysis

Details of your modeling process, including how you selected your models and validated them

Your challenges and successes

Possible extensions or business applications of your project

Conclusions and key learnings