Journal of Web Development and Web Designing
Volume 1 Issue 1

Predicting the NBA Playoff Using SVM

Amruta Jadhav, Sandhya Das, Savita Khatode, Rohini Degaonkar
Department of Computer Science and Engineering, J.N.E.C, Aurangabad, India
E-mail: sandhyad930@gmail.com

Abstract

In this paper, we examine the use of Support Vector machine as a tool for predicting the success of basketball teams in the National Basketball Association (NBA). Statistics for 2012–2015 NBA games were collected and used to examine and predict the outcome of NBA playoffs. The best results were able to correctly predict the winning team is 88 percent. Further, we investigate which subset of features input to the SVM model are the most salient features for prediction.

Keywords*: SVM algorithm, playoffs, prediction, NBA*

INTRODUCTION

The National Basketball Association (NBA) is that the comfort station sport league in North America and is wide thought-about to be the most effective comfort station sport league within the world. It is thirty groups within which twenty nine groups are from U.S. a nd one team from Canada. The league adopted the name National Basketball Association on August 3, 1949. The current league organization divides thirty teams into Eastern Conference and Western Conference. Eastern Conference has three divisions Atlantic, Central, Southeast. Western Conference has three divisions Northwest, Pacific, Southwest. Atlantic has five teams Boston Celtics, Brooklyn Nets, New York Knicks, Philadelphia 76ers Toronto raptors. Central has five teams Chicago Bulls, Cleveland Cavaliers, Detroit Pistons, Indiana pacers, Milwaukee Bucks. Southeast has five teams Atlanta Hawks, Charlotte Hornets, Miami Heat, Orlando Magic, Washington Wizards. Northwest has five teams Denver Nuggests, Minnesota Timberwolves, Oklahoma City Thunder, Portland Trail Blazers, Utah Jazz. Pacific has five teams Golden state Warriors, Los Angeles Clippers, Los Angeles Lakers, Phoenix Suns, Sacramento Kings. Southwest has five teams Dallas Mavericks, Houston Rockets, Memphis Grizzlies, New Orleans Pelicans, San Antonio Spurs [1, 2].

In February, the regular season pauses to celebrate the annual NBA All-Star Game. Fans vote throughout the United States, Canada and on the Internet and the top vote-getters at each position in each conference are given a starting spot on their conference's All-Star team. Around the middle of April, the regular season ends. The Rookie of the Year Award is awarded to the most outstanding first-year player. The Most Improved Player Award is awarded to the player who is deemed to have shown the most improvement from the previous season. The Defensive Player of the Year Award is awarded to the league's best defender. The Coach of the Year Award is awarded to the coach that has made the most positive difference to a team. The Most Valuable Player Award is given to the player deemed the most valuable for (his team) that season.

RELATED WORKS

Several researches have been done to predict the results of the National Basketball Association. Students from Cernegie Mellon University (Jackie B. Yang, 2012) used support vector machines that made use of the kernel functions. Their model did not perform well on the attributes they used. The students started with 14 attributes per team but they did not investigate the possibility that using all of these attributes might lead to over fitting. This might be the case in this research looking at the performance of 55%. In 55% of 30 actual played games, the winner was correctly predicted [3, 4].

Wei (Wei, 2012) has examined the utilization of naive Bayes. Wei used home and away winning rates. Whereas Jones (Jones, 2007) showed the existence of the house advantage; mistreat ment this home and away winning rates did not influence Wei's predictions. The predictions performed even worse whereas accounting for the seasons home/away records. This would possibly indicate the playoffs home advantage being very totally different from the season's home advantage. While the best performing team gets the home advantage it is interesting to find out whether the best teams wins because of the home advantage or because they are simply the better team.

DATASETS AND FEATURES

This project collects data from some non-profit websites, which publish NBA game statistics data available to public. The following section will introduce sources of data used in this project.

Data Sources

In this project, data used for model training and testing, Basketball-reference.com is our major data source. Basketball-reference.com is one of the most famous one for those sports data miners, and has been used in many experiments. It was created in 2003 and data requests are comprehensive, relatively well organized, straightforward and easy to navigate and utilize.

For this project about 3 years of NBA game statistics data has been collected, which came from 2012–2013 season to 2014–15 season. These 3 years of data would have better consistency and applicability to new season predictions. Attributes use in our dataset are explained in Table 1.

Building the Model

The next step in the process of predicting is building a model on the information that is found. With the information that is available, different models will be run. Given previous research, the Naive Bayes may perform well but other algorithms have to be tried. The models will be trained on 8 of the 9 years and tested on the remaining one. The model that performs best will be used to predict the 2012 playoffs that have just finished. Simulating all of the playoff games in one time will lead to difficulties measuring performance. For example: when Miami plays Boston and the model predicts Miami to survive, while in fact Boston is going to win, the next round will be different in the model compared to reality. In this case it is impossible to measure whether the model was right on this next round because this round is never played in this particular setting. This is why the playoffs will be predicted round by round.

Table 1: *NBA Statistics Explanation.*

Attributes	Explanation
MP	Minutes Played
FG	Field Goals
FGA	Field Goal Attempts
3P	3-Point Field Goals
3PA	3-Point Field Goal Attempts
3P%	3-Point Field Goals Percentage
FT	Free Throws
FTA	Free Throw Attempts
FT%	Free Throw Percentage
ORB	Offensive Rebounds

DRB	Defensive Rebounds
TRB	Total Rebounds
AST	Assists
W/L%	Win Loss Percentage
PER	Player Efficiency Rating
FG	Field Goals
FGA	Field Goal Attempts
FG%	Field Goals Percentage
2P	2-Point Field Goals
2PA	2-Point Field Goal Attempts
2P%	2-Point Field Goals Percentage
STL	Steals
BLK	Blocks
TOV	Turnovers
PF	Personal Fouls
PTS	Points
PTS/G	
G	Season Game

Playoffs

As mentioned earlier, the playoffs are a best-of-seven knockout tournament. Eight teams from the west will play each other, as well as eight teams in the east. The western and eastern champions finally play each other in the NBA final. The NBA playoffs consist of four rounds. First round, conference semi-finals, conference finals and NBA finals. The teams will be seeded given their strength during the regular season. The seeds light are very important because the lower seed means this team, that performed better, will have the home advantage. Given these facts, the teams seeding number, playoff round and the number of won games during the best-of-seven, until then, are added as attributes.

SVM

We also modelled our problem with a support vector machine to predict whether or not a streak will occur. To date, few supervised learning algorithms have outperformed SVMs; perhaps the most alluring property of SVMs is that we can utilize their symbiosis with kernels. This lets us create high or infinite dimensional feature vectors, allowing us to capture the subtle interactions between features, which is particularly important for our dataset because it contains so many intertwined and interdependent features. Thus, the SVM presents an ideal classification model that does not limit our ability to utilize a dense, high dimensional feature vector to determine the state of the game.

One disadvantage of SVMs, however, is that they are easily susceptible to over fitting because they work in such a high dimensional feature space. Since it is likely that the extracted data will become linearly separable, we run the danger of the model learning hidden attributes of our data rather than the more general trends we are looking for. SVMs constructed in this manner work to maximize the geometric margin for every point by constructing a separating hyper plane in high-dimensional space (as most data sets are not linearly separable) that classifies as many points correctly with the largest margin possible.

RESULTS AND DISCUSSION

Our model produces 88% accuracy on testing dataset. Figure 1 shows SVM plotting points on attribute Win/Loss and PER. Separate hyper plane can be drawn easily on this attribute on labelled class result. Figure 2 shows SVM plotting points on attribute Win/Loss and PTS. Separate hyper plane can be drawn easily on this attribute on labelled class result. Figure 3 shows SVM plotting points on attribute ORB and DRB. Separate hyper plane cannot be drawn easily on this attribute on labelled class result.

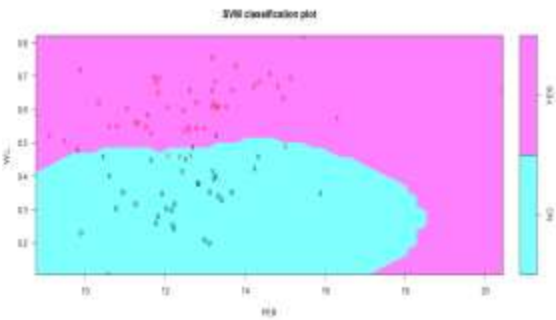

Fig. 1: *A Win/Loss and PER.*

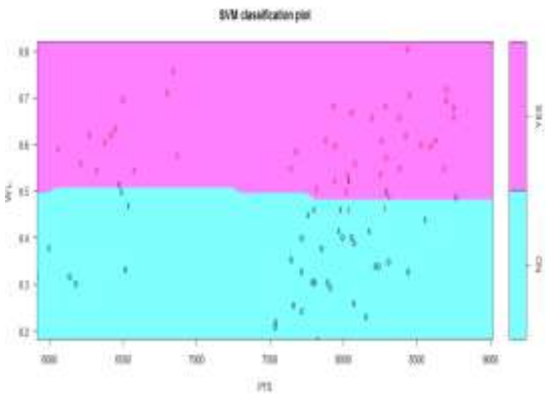

Fig. 2: *Win/Loss and PTS.*

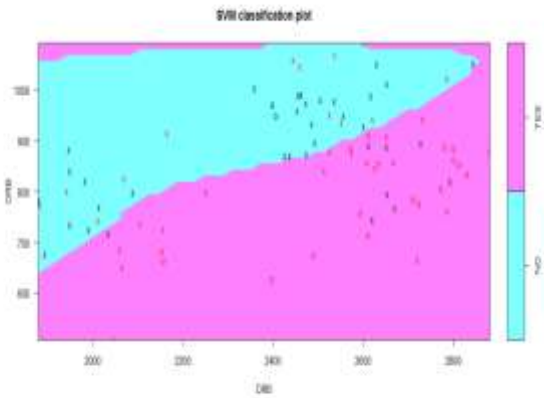

Fig. 3: *ORB and DRB.*

CONCLUSION

In this report, we have shown that machine learning techniques can be successfully applied to NBA games to predict the winner of any given game with around 88% accuracy. This research aims to predict the result of the basketball game using data mining technique with comprehensive statistical data and game related data and finally generates a persuasive model. By taking advantages of machine learning approach, ideally the predictive model would output reasonable prediction accuracy and this model can be used as a reference to make game strategy before games or can be used for sports betting.

ACKNOWLEDGMENT

We consider it as a great privilege to express my heartfelt gratitude to many respected personalities who guided, inspired and helped me in successful completion of this project. We also express my gratitude to our guide Rohini Narwade, of Computer Science and Engineering and Kishori Pawar, project coordinator for providing us with adequate facilities, ways and means by which we were able to complete this project.

REFERENCES

1. E. Boser, I. Guyon, V. Vapnik. A training algorithm for optimal margin classifiers. *In Proceedings of the Fifth Annual Workshop on Computational Learning Theory*. 1992; 144p.
2. Cortes, V. Vapnik. Support-vector network. *Machine Learning*. 1995; 20: 273–297p.
3. T. Lin, C.-J. Lin. A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods. Technical report. Department of Computer Science, National Taiwan University, 2003. URL http://www.csie.ntu.edu.tw/~cjlin/papers/tanh.pdf.
4. Vapnik. The Nature of Statistical Learning Theory. Springer-Verlag, New York, NY; 1995.