





Data-driven Basketball Web Application for Support in Making Decisions

Tomislav Horvat¹^a, Ladislav Havaš¹^b, Dunja Srpak¹^c and Vladimir Medved²^d

¹Department of Electrical Engineering, University North, 104 Brigade 3, Varaždin, Croatia

²Department of General and Applied Kinesiology, Faculty of Kinesiology, University of Zagreb, Zagreb, Croatia

Keywords: Basketball, Information System, Making Decisions, Statistics Analysis, Web Application.

Abstract: Statistical analysis combined with data mining and machine learning is increasingly used in sports. This paper presents an overview of existing commercial information systems used in game analysis and describes the new and improved version of originally developed data-driven Web application / information system called Basketball Coach Assistant (later BCA) for sports statistics and analysis. The aim of BCA is to provide the essential information for decision making in training process and coaching basketball teams. Special emphasis, along with statistical analysis, is given to the player's progress indicators and statistical analysis based on data mining methods used to define played game point's difference classes. The results obtained by using BCA information system, presented in tables, proved to be useful in programming training process and making strategic, tactical and operational decisions. Finally, guidelines for the further information system development are given primarily for the use of data mining and machine learning methods.


1 INTRODUCTION


Nowadays, sports statistics and analysis, more particular the information and communication technologies are omnipresent in sport and have become a very important factor in making decisions in sport. The term decision making in sport refers to decision making during, before or after the games, decision making related to changes in the training process, changes related to the preparation of specific sports tactics or decision making related to the new player engagement, the finding of sports talents etc. The precondition of good sport analysis is sufficient amount of relevant data. Nowadays, at the time of the existence of the global Internet network, access to information, more particularly information related to sports events is publicly available.


This paper presents the new and improved version of originally developed data-driven Web application / information system called Basketball Coach Assistant (later BCA) for sports statistics and analysis, with the aim to provide the essential


information for decision making in training process and coaching basketball teams. The first version of the BCA information system, called *AssistantCoach*, was presented at the International Congress on Sport Sciences Research and Technology Support in Lisbon (Horvat et al., 2015). The second version, more precisely the new added information system functionalities, were described in paper Lacković et al., 2018. The application was later used for the purpose of outcome predicting in Euroleague, the most elite basketball competition in Europe (Horvat et al., 2018). The general definition of the term data-driven refers that progress in an activity is completed by the data, rather than by intuition or by personal experience.

Prior to getting into the core of papers' topic, most popular presently available information systems that are used in basketball analytics will be presented. In 1997, the Advanced Scout application was introduced with the aim of revealing interesting patterns using data mining methods in NBA game data (Bhandari et al., 1997). The application was

^a  <https://orcid.org/0000-0002-8358-3218>

^b  <https://orcid.org/0000-0002-5051-4486>

^c  <https://orcid.org/0000-0003-1497-9080>

^d  <https://orcid.org/0000-0002-8298-5602>

of the BCA information system are, usually by user experience or experts opinion, regularly upgraded and modified according to user experience. BCA information system offers flexibility, which means that user can define a "focus/observed" team. User defines focus team that represents the team whose analysis and notes are displayed. In addition, the information system BCA offers to the user the possibility to define the time period of analysis. This paper presents its current form and embedded features. As stated earlier, BCA information system views and subviews were introduced in paper Lacković et al., 2018. Figure 1 shows BCA information system architecture and a schematic overview of the Web application views. The aim of this paper is to introduce the statistical and analytical capabilities of the information system that can help coaches in making decisions. Due to its infrastructure and design, information system BCA can be easily adapted to many common team sports such as soccer, football, handball, baseball etc.

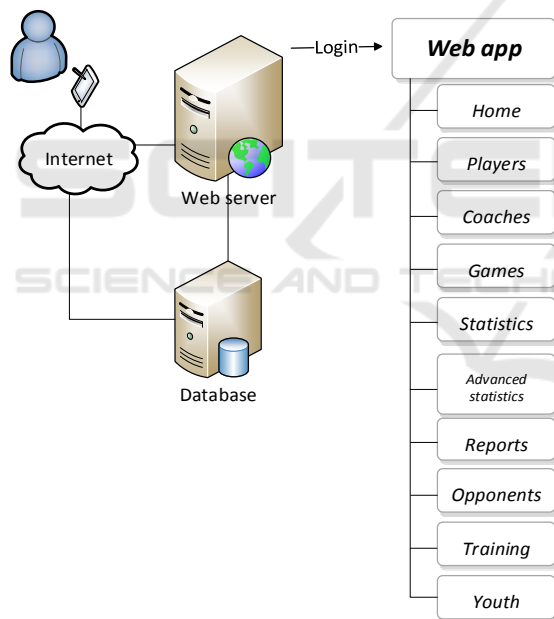


Figure 1: A schematic overview of the Basketball Coach Assistant information system.

The BCA information system allows the monitoring of the senior team and all the club younger age categories. Most of the views shown in Figure 1 refer to the senior team, while the younger categories are analysed in view called "Youth". The available statistical analysis can be used for all age categories of the analysed club and will be presented in the next chapter.

2.1 Data Acquisition

The precondition of good sport analysis is sufficient amount of relevant data. Nowadays, there is no barrier in collecting data as game statistics are publicly available through the global Internet network. The BCA information system allows users to manually enter game statistics while advanced users can use embedded scraping scripts to automatically scrape whole Web domain and to input the statistical data directly into the database. Users are automatically enabled to collect game statistics data from the basketball-reference.com (NBA) and www.euroleague.net (Euroleague) Web site by using Web scraping embedded scripts. Web scraping is a process of data scraping used for extracting data from websites. Web scraping scripts may access World Wide Web directly through a Web browser. The page content must be extracted, transformed and in this case loaded (ETL process) into the MySQL database. Figure 2 shows Web scraping process.

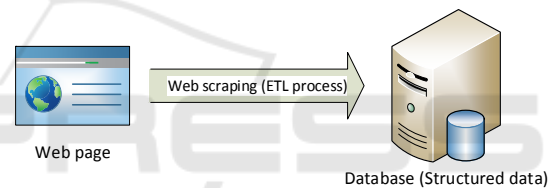


Figure 2: Web scraping process.

Very important information related to the BCA information system is that BCA enables the input of statistical data from the analysed and opponent team that gives the user even greater opportunities and thus more advanced analysis. In addition to raw statistics, as noted above, the users can also input their own notes related to players or teams. Defining of a unified Web scraping script was not possible due to the specificity of the game statistics view. In addition, some basketball leagues official websites do not record certain basketball game statistical parameters, which results in a lack of data that can't be predicted. In that case the missing statistical parameters are excluded from the statistical analysis.

3 STATISTICS AND ANALYSES

This chapter presents the statistical analysis capabilities associated with the analysed team. As already mentioned above, the precondition of good

statistical analysis is sufficient amount of relevant data.

Information system allows their users to record and to grade every training (group or individual) and therefore also player's performance rated by 1-5. Based on obtained results information system user can make useful conclusions that will help in further decision making. Figure 3 shows the performance analysis of a player in training during user defined time period.

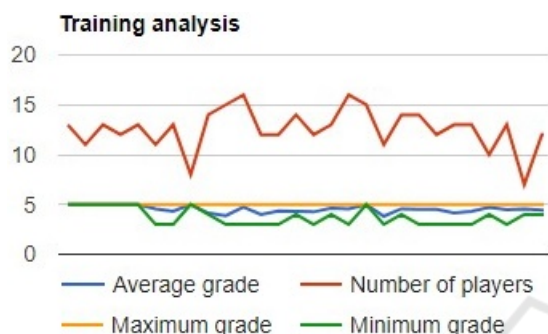


Figure 3: Training performance evaluation.

As can be seen on Figure 3, the information system provides the user an overview of average training grade, number of players on the training and maximum/minimum grade of the player's training performance. The main quality training indicator is the trend line visible for each analysed statistical parameter.

The most well-known, and therefore the most used statistical information are scored points per game. In addition to the basic information regarding the average points per game, the BCA information system offers to the user an indication of the minimum and maximum player's game points performance and the game number played by player. The most interesting information regarding the points per game is certainly the share of player's points in each game played. The player's points share is a very important information, especially for the opposing team, when preparing the upcoming mutual game(s). The statistics shown above are available for all parameters of basketball games. Most of the basketball leagues record 13 basic basketball elements to the user usually shown in a form of so-called boxscore tables. Figure 4 shows statistical information about the scored points. This analysis proved to be very useful when preparing upcoming games, more precisely in the analysis of the opponent's teams. Figure 4 also shows the progress of the player. Players with progress in defined time period are marked with green arrows. By using least

squares method the line of best fit for a set of data, providing a visual demonstration of the relationship between the data points, was obtained and thus enabled defining player's progress or decline in game performance. In order to be able to track player's progress, information system BCA needs at least two played games.

Points (Basketball club)						
#	PLAYER		AVG (% of all game points)	MAX	MIN	GAME NUMBER
1.	Player 1	↓	18.75 (18.45%)	34	10	8
2.	Player 2	↑	11.75 (11.75%)	32	1	8
3.	Player 3	↑	9.00 (11.28%)	28	2	7
4.	Player 4	↑	6.89 (4.15%)	18	0	9
5.	Player 5	↑	5.88 (7.54%)	18	0	8
6.	Player 6	↓	4.75 (7.18%)	16	0	8
7.	Player 7	↑	3.50 (7.6%)	5	2	2
8.	Player 8	↓	3.29 (12.09%)	8	0	7
9.	Player 9	↓	2.43 (5.94%)	7	0	7
10.	Player 10	↑	2.14 (10.76%)	5	0	7
11.	Player 11		2.00 (1.68%)	2	2	1

Figure 4: Analysed time period player performance (average points per game).

Figure 5 shows the basic player's training attendance. The information system allows the user to choose the time period or number of training to be displayed and thus analyse a defined time period and make useful decisions.

	28.11. (19:30h)	27.11. (19:30h)	26.11. (18:30h)	16.11. (17:30h)	14.11. (19:30h)	11.11. (18:30h)	09.11. (17:30h)	%
Player 1	✓	✓	✓	✗	✗	✓	✓	71%
Player 2	✗	✓	✗	✓	✓	✓	✗	57%
Player 3	✓	✓	✓	✓	✓	✓	✓	100%
Player 4	✓	✓	✓	✓	✓	✗	✓	86%
Player 5	✗	✓	✓	✓	✓	✓	✓	86%
Player 6	✓	✓	✗	✓	✓	✓	✓	86%
Player 7	✓	✓	✓	✓	✓	✓	✓	100%
Player 8	✓	✗	✓	✗	✓	✓	✓	71%
Player 9	✓	✓	✓	✗	✓	✓	✓	86%
Player 10	✓	✓	✓	✗	✓	✓	✓	86%
Player 11	✓	✓	✓	✗	✗	✓	✗	57%

Figure 5: Training attendance in defined time period.

Figure 6 and Figure 7 show the relationship

between the training number, the training attendance and the average training grades compared to the played games outcome. The obtained statistic allows coach to recognize good or bad trends related to the relationship between the training performance and the played games outcome. Training analysis also allows detection of player overtraining and detection of game number in a particular minicycle.

Week	Training number	Attendance	Average grade	Game number	Games
03.09. - 09.09.	4	81.16%	4.54	1	Team 2 - Team 1 51:54 ✓
10.09. - 16.09.	4	81.25%	4.37	2	Team 1 - Team 3 71:69 ✓ Team 1 - Team 3 16:45 ✗
17.09. - 23.09.	3	62.5%	4.57	2	Team 1 - Team 4 68:61 ✓ Team 1 - Team 4 43:26 ✓
24.09. - 30.09.	3	72.92%	4.31	0	-
01.10. - 07.10.	4	82.81%	4.28	0	-
08.10. - 14.10.	3	77.08%	4.35	0	-
Season average	3.5	77.03%	4.39	-	-

Figure 6: Analysed time period week by week training performance and game outcome success.

Month	Training number	Attendance	Average grade	Game number	Games
9 (2018.)	14	75.55%	4.45	5	Team 2 - Team 1 51:54 ✓ Team 1 - Team 3 71:69 ✓ Team 1 - Team 3 16:45 ✗ Team 1 - Team 4 68:61 ✓ Team 1 - Team 4 43:26 ✓
10 (2018.)	17	78.68%	4.45	4	Team 5 - Team 1 26:38 ✓ Team 6 - Team 1 23:28 ✓ Team 1 - Team 7 26:44 ✗ Team 1 - Team 4 70:78 ✗
Season average	15.5	77.03%	4.39	-	-

Figure 7: Analysed time period month by month training performance and game outcome success.

Another proved to be very important statistical analysis is player performance in games with different point's difference. The information system based on the game's outcome during defined time period defines the point difference classes and calculates average player performance. Player effort is different from game to game, especially if there is a game between unbalanced opponents. In this case, the more important team players play less and thus their performance falls. Figure 8 shows statistical information about the scored points during defined time period and based on information system defined points difference. The most interesting information regarding Figure 8 is certainly the share of player's points in each game played during time period in information system defined point's difference classes. The statistics tables shown on Figure 8 are

available for all recorded parameters of basketball game. The average fan is most interested in the three most widely used basketball game statistics parameter such as scored points, rebounds and assists. Advanced users such as coaches, managers, scouts and the rest of the coaching staff use all available basketball statistics parameters for the purpose of advanced player or team performance analysis.

Games with points difference [-8, +7]			
#	PLAYER	POINTS	GAME NUMBER
1.	Player 1	21.8 (26.55%)	5
2.	Player 2	10.8 (13.07%)	5
3.	Player 3	5.8 (6.45%)	5
4.	Player 4	5.75 (7.86%)	4
5.	Player 5	5 (7.79%)	4
6.	Player 6	4.25 (5.99%)	4
7.	Player 7	2.75 (2.2%)	4
8.	Player 8	2.5 (2.9%)	4
9.	Player 9	2 (2.71%)	5
10.	Player 10	1 (0.71%)	4
11.	Player 11	1 (0.83%)	4

Figure 8: Average player performance based on BCA information system defined point's difference classes during defined time period.

Figure 8 beside player's name shows average points per game with the percentage of the player's points share in relation to team scored points and number of played games in the defined time period. The player's points share is taken only for games played by the analysed player.

4 CONCLUSIONS

Various statistics and analysis have proved to be very useful and have begun to create additional values that contribute to the sports success of players and teams when talking about team sports and individuals in individual sports. More and more professional teams employ statisticians and experts in data mining and the use of machine learning. This paper presents the new and improved version of originally developed Web application / information system called Basketball Coach Assistant for sports statistics and analysis, with the aim to provide the essential information for decision making in training process and coaching basketball teams. In practice, the

player's share statistics as well as player's progress indicator proved to be very useful. It was especially useful to define the point's difference and the share of the player's performance over the whole team which enabled coach to easier program the training and selecting the tactics for upcoming games.

5 FUTURE WORK

The developed information system is a suitable starting point for further development of statistical analysis. A very important matter will be to statistically link the training effect and played games' outcomes through various analyses, but also to involve artificial intelligence (data mining, machine learning) that will suggest to coaches certain changes related to training and game. Certain steps have been made and coaches certainly get useful information and facts.

REFERENCES

- Bhandari I., Colet E., Parker J., Pines Z., Pratap R., Ramanujam K., 1997. Advanced Scout: Data Mining and Knowledge Discovery in NBA Data, *Data Mining and Knowledge Discovery*, 1, 121-125.
- Blaikie A.D., David J.A., Abud G.J., Pasteur R.D., 2011. *NFL & NCAA Football Prediction using Artificial Neural network*, <https://www.semanticscholar.org/paper/NFL-%26-NCAA-Football-Prediction-using-Artificial-Blaikie-Abud/5207ead80e566abd29bf2c171143fa6473c28b6a>.
- Cheng G., Zhang Z., Kyebambe M.N., Kimburgwe N., 2016. Predicting the Outcome of NBA Playoffs Based on the Maximum Entropy Principle, *Entropy*, 18(12).
- Delen D., Cogdell D., Kasap N., 2012. A comparative analysis of data mining methods in predicting NCAA bowl outcomes, *International Journal of Forecasting*, 28(2), pp. 543 – 552.
- Elfrink T., 2018. Predicting the outcomes of MLB games with a machine learning approach, *Business Analytics Research Paper*.
- FIBA Organizer. <http://www.fibaorganizer.com>. Accessed June 2019.
- Horvat, T., Havaš, L., Medved, V., 2015. Web Application for Support in Basketball Game Analysis. *icSports 2015*, Lisboa, 225-231.
- Horvat, T., Job, J., Medved, V., 2018. Prediction of Euroleague games based on supervised classification algorithm k-nearest neighbours. *Proceedings of the 6th International Congress on Sport Sciences Research and Technology Support: K-BioS*, pp 203-207, Seville, Spain (20. - 21.9.2018).
- Igiri C.P., Nwachukwu E.O., 2014. An Improved Prediction System for Football a Match Result, *IOSR Journal of Engineering*, 4(12), pp. 12-20.
- Krossover. <https://www.krossover.com>. Accessed June 2019.
- Lacković K., Horvat T., Havaš L., 2018. Information System for Performance Evaluation in Team Sports. *The International Journal of Business Management and Technology*, 2 (1).
- Lam M.W.Y., 2018. One-Match-Ahead Forecasting in Two-Team Sports with Stacked Bayesian Regressions, *Journal of Artificial Intelligence and Soft Computing Research*, 8(3), pp. 159-171.
- McCabe A., Travathan J., 2008. Artificial Intelligence in Sports Prediction, *Fifth International Conference on Information Technology: New Generations*.
- Ping-Feng P., Lan-Hung C., Kuo-Ping L., 2017. Analyzing basketball games by a support vector machines with decision tree model, *Neural Computing & Applications*, 28(12), pp. 4159-4167.
- Prasetio D., Harlili D., 2016. Predicting football match results with logistic regression, *International Conference On Advanced Informatics: Concepts, Theory And Application (ICAICTA)*, Penang, Malaysia.
- Soto Valero C., 2016. Predicting Win-Loss outcomes in MLB regular season games – A comparative study using data mining methods, *International Journal of Computer Science in Sport*, 15(2), pp. 91 – 112.
- Synergy. <https://corp.synergysportstech.com>. Accessed June 2019.
- Zaveri N., Shah U., Tiwari S., Shinde P., Kumar T.L., 2018. Prediction of Football Match Score and Decision Making Process, *International Journal on Recent and Innovation Trends in Computing and Communication*, 6(2), pp. 162-165.