

Comparative Review of Statistical Parameters for Men's and Women's Basketball Leagues in Serbia

Laszlo Ratgeber*, **Branko Markoski****, **Predrag Pecev****, **Dejan Lacmanović****, **Zdravko Ivanković****

*PTE-ETK University of Health Sciences Pécs - Doctor School, Pécs, Hungary, ratgeber@ratgeber.hu

**University of Novi Sad, Technical Faculty "Mihajlo Pupin", Zrenjanin, Serbia, markoni@uns.ac.rs, pecev@tfzr.uns.ac.rs, dlacman@tfzr.uns.ac.rs, zdravko@tfzr.uns.ac.rs

Abstract: Basketball is one of the most popular sports. During a basketball game, statisticians note a large amount of information, helping coaches and players to improve their game and to analyse opponents in order to prepare for the following games. Due to amount and complexity of the information, basketball is the ideal discipline for the application of data mining techniques, especially neural networks that enable one to extract conclusions and knowledge from these data. In this paper, we compared the First senior leagues for men and for women in Serbia during the 2011/2012 season, and from these statistics, we calculated influence of certain parameters on a game's outcome. We concluded that the most influential parameter, in both leagues, is the offensive rebound. In men basketball then follows the three-point shot, the two-point shot and the one-point shot. In women's basketball, after the defensive rebound, the two-point shot follows, and then turnovers.

Keywords: data mining; neural networks; statistics; basketball

1 Introduction

Basketball, as a team sport, puts specific demands on players at particular positions, regarding their anthropological status. Therefore, the players' selection according to certain criteria is one of main functions of the coach. Playing experience, arising from basketball practice, points that every position in the game demands a certain level of development for certain dimensions of the players' anthropological status, which has an influence on the efficiency of playing basketball. At the beginning of a season, coaches are mostly interested in using different statistical reports for analysing and evaluating individual players. Once they have insight into the strengths and weaknesses of their players, their interest

then moves towards the team as a whole. They want to know how good the team is. The team statistics therefore become most important. After all, a basketball is still a team sport. Finally, bearing in mind that different statistic reports may be used also to analyse the opponent's play, during the season the coaches' interest moves towards their opponents. More often than not, analysing the opponents' play well means the difference between winning and losing. Aspiration to win causes insight: that every detail of the game enables way to the coveted triumph. Information hunger slowly rises regarding the play and players of opponent teams, in order to better prepare for the following game, and this enables their team to impose their style and tempo in order to win. By gathering information regarding the opponent's players and their game, we begin to build a game philosophy, system and technology of scouting the opponent. By gathering information, the analysis and the systematization, we aspire toward a working model that will include, apart from gathering information, their presentation to the team, practicing individual and team tactics during micro-cycles between games and the control of its success and application during the game itself. Coaches are not the only ones to use statistics. On the contrary, whole population of sport fans and audience is able, using mass production of technology and media as TV and Internet, to follow the efficiency of teams and individual athletes. To many people, such as reporters and commentators, statistics is an important tool in doing their job, and for some, such as sport managers, it is a vital part of their profession. After all, over the last few decades, sport has become more than a game; it is a large business, with considerable amounts of money invested [3]. Professional sports organizations are multi-million-dollar companies and certain decisions are worth large amounts of money. With this kind of capital, a single wrong decision may potentially set them back years. Due to high risk and need to make correct decisions, the sport industry is just the right environment to apply data mining technologies. The final result is not the only thing important anymore.

2 Data Analysis in Sports

Huge amounts of data are present in every sport. It is extremely important to determine which information to store and to find a way for its best usage [1]. By finding the appropriate methodology to extract sense from information and to turn this information into practical knowledge, sports organizations provide themselves an advantage in comparison to other teams [2]. Such approaches to knowledge seeking may be applied into a whole organization – from players who may improve their performance using techniques of video analysis, to scouts who use statistical analysis and projection techniques in order to identify which talent will develop the most and become a good player. Most sports organizations use a third or fourth type of connection between data and their use, while only a handful uses data mining techniques. Although introduction of data mining in sports is relatively recent, the influence of those teams who applied these techniques is

extraordinary [4]. Evaluations are being done based on strong analyses and scientific investigations. Since more and more sports organizations have embraced the digital era, it is possible that sports will soon become a struggle for better algorithms or better metrics for performances measurement, and analysts will become equally as important as players. The essence is in finding knowledge included in given data [5]. Statistics may also sway decisions in the wrong direction, if there is no knowledge regarding the basis of a problem, as a consequence of imprecise measurements of performances or due to an over-enhancement of certain qualities by the sports community [6]. For example, a certain player may have exceptional individual statistics, but he still may have little influence on the team as a whole. Sports statistics suffers from imprecision, since statistical metrics may not measure completely the influence of all players. For instance, a defensive rebound is a measure how many times a certain player in defence catches a ball after an unsuccessful shot by the opponent players. In order to have a defensive rebound, another player from his team must block the opponent players and therefore they are equally important in this action. Having in mind the way of noting rebounds, only the player who caught the ball is noted in statistics and rewarded a defensive rebound. The practice of data mining application in organized sports was not developed overnight. There were certain events over the past several decades, slowly bringing these changes. During 1980s, Dean Oliver began to ask questions regarding basketball metrics. His idea was to create statistics for a whole team, and not just for individual players. Oliver published his thoughts and created performances measurements for the rest of the basketball community [7]. In his work, he was especially interested in the performance of players within a team and the performance of the team as a whole, and also in how certain players function together. His work was recognized by professional sports organizations and he was hired as a consultant by the NBA team, the Seattle SuperSonics. Sport analysis is usually connected to the moves of a player on the court. Following objects and trajectory analysis usually start from the analysis of the video from a game, where reference points are recognized (the size of a player is significant in comparison to that of the ball so they are easily distinguished) [8] [9]. Exposing hoaxes in sports is also possible using data mining. The NCAA organization revealed that 1% of all games played are fixed [10]. Exposing hoaxes is a relatively complex operation with data [11]. When basketball game data are collected, the first step is to find knowledge in this information. The creating of predictions has been the goal for numerous individuals and organizations for a long time. Making predictions is comprised of a number of techniques, but simulation and machine learning are the core techniques [12]. Neural networks are the most dominant system in machine learning used in sports [3]. Data sets collected from games are analysed using neural networks in order to find patterns and tendencies due to competition and financial gain. Other techniques are decision trees, genetic algorithms, ID3, and a regressive variant of the Support Vector Machine (SVM) called the Support Vector Regression (SVR).

3 Data Mining

Large quantities of data exist in all areas of sport. These data may show the individual quality of a player, the events that have taken place in the game and/or how the team functions as a unit. It is crucial to determine which data should be stored and how their usage could be maximized. By finding the correct way to extract the true meaning of the data in order to turn them into practical knowledge, sports organizations ensure themselves an advantage over other teams. Most of the analysed data are acquired from local database storage using data mining techniques such as C5.0 decision tree and its algorithm, while some data are acquired using Web Data Mining Methods [13] [14].

Modelling is done by using neuron nets. Neuron nets have been inspired by the recognition of the complex system of learning within the human brain, which is made of closely connected units of neurons. The incoming parameters are the following: *p1_procenat*, *p2_procenat*, *p3_procenat*, *def_rebound*, *of_rebound*, *assist*, *steal*, *lost* and *block*. The outgoing parameter is the result. Due to this fact, the net has nine incoming and one outgoing knots. Apart from this, there is a hidden layer in the net. The used net is a feed-forward neuron net. Each layer of the net is connected with all the knots in the previous level as well as with all the knots in the following level of the net. The net training is done with the method of the back propagation of errors, which is based on the generalised delta rule. For every syllable brought to the net during the training, the information goes through the net in advance, so that it can anticipate the outgoing layer. This anticipation is compared with the real outgoing value of the given datum, and the difference between the real and the anticipated value is then sent back through the net in order to adjust the heavy factors and to improve the anticipation of the syllables which follow. During the net training, the incoming number of data is divided in the range 75:25, into the data which will be used throughout the net training and the data which will be used throughout testing. The point is to prevent the net from saving the input as well as to prevent the wrong result as the other data are modelled. After the net training is done with the input data, we can see how much the final outcome of the game is determined by certain parameters.

The basket consists of the hoop, which is 45.7 cm (18 inch) in diameter, set at the height of 305 cm (10 feet) and mounted on the board. A team scores points by putting the ball through the hoop during a game. Shots may be scored in several ways, and the hardest to achieve are the long-distance shots, so they bring the most points. On the floor is a line at 625 centimetres from a basket. Points (in some leagues this boundary is moved even further from the basket), and shots from outside this line bring three points. Within this line, every score brings two points.

A foul happens when a player during a game is irregularly disrupted by the opponent players. If the team committing the foul has already exceeded the limit (four fouls committed during a period or quarter) or if the foul is done during an

attempt to score, then a player has a chance to score from the free-throw line. Every shot scored from this line brings one point. Players have the opportunity to try two or three free throws, depending on whether the foul was committed while the player was trying to score two or three points, respectively. When throwing free throws or when shooting for two points or three points, a player may make the shot or miss the shot, i.e. score or miss. In basketball statistics, shooting percentage is the relation between shots and scores, and there are separate percentages for one-point shots, two-point shots and three-point shots. In addition to shots, which are one of major factors in basketball, there are other important factors that decide which team the team will win. Rebounds have relatively high influence on the outcome of the game. Defensive rebounds are highly important, since they prevent the opponent from having another attack after missing the basket or from scoring easy points if they obtain the ball directly under the basket. The same goes for offensive rebounds. The assist is passing the ball to teammate who is in position immediately to score. Good assists can have effect on spectators and the team itself, enabling the team to increase a game standard. Steals and turnovers are also important, since they enable a team who intercepts a ball to score easy points from counterattack.

3.1 Data Preparation Phase

The following table presents the various basketball parameters that are very important for the outcome of a match. Column headers of Table 1 have the following meanings: field *p1_percent* represents shots for one point, field *p2_percent* represents shots for 2 points, field *p3_percent* represents 3-point shots, field *def_rebound* represents jumps on the defence, field *of_rebound* represents a rebound, field *assist* represents assists, field *steal* represents steals, field *lost* represents turnovers, field *block* represents blocks, and the field marked *result* states whether a team won or lost.

Table 1
Data obtained by Data Mining

p1_percent	p2_percent	p3_percent	def_rebound	of_rebound	assist	steal	lost	block	result
0.5652	0.5789	0.3668	22	8	15	12	11	3	win
0.7273	0.3556	0.4211	20	14	9	19	12	4	loss
0.6500	0.5517	0.2083	24	15	13	5	19	1	win
0.7368	0.4722	0.3333	20	6	14	6	25	6	loss
0.8421	0.6061	0.5333	13	4	11	8	13	2	loss
0.6538	0.4773	0.2857	14	8	12	9	12	2	win
0.8077	0.4958	0.3847	24	12	8	13	18	1	loss
...

This table shows that the values for offensive and defensive rebounds, assists, turnovers, steals and blocks are whole number values in a certain interval. The last column shows the result, i.e. the final score of the game, telling whether a team won or lost the game.

3.1.1 Effect of Certain Parameters on Outcome of the Game

The following graphs were obtained by analysis of statistical parameters in the Serbian First basketball league for men and the Serbian First "A" Basketball league for women. The First basketball league for men is a professional senior league, with 120 games played in season 2011/2012, including play-off and play-out. The First "A" Basketball league for women is a professional senior league for women, with 110 games played in season 2011/2012.

On the acquired statistical data, the analysis was done, showing influence of particular parameters on game outcome. The same parameters were observed in both leagues, with aim of doing comparative analysis. In all graphs, the colour blue represents the number of losses and red is the number of wins gained by teams in the period observed.

Effect of the One-Point Shot

Figure 1 below shows the effect of the one-point shot on the final outcome of the game. The upper graph is for results acquired for the Serbian First basketball league for men, and one below is for First "A" basketball league for women. The upper graph shows that the number of wins rises sharply when one-point shot percent exceeds 65% limit. If one-point shot percent is below 65%, teams in most cases lose that game. From the dependencies obtained, we may say that one-point shot has a significant effect on final outcome of the game.

This graph also shows that during a season, sometimes relations between shot percent and a game outcome deviates from the general rule. It happened once that a team had one-point shot percent over 85%, and they still lost the game. There was also a game when a team had one-point shot percent of only 38%, and still they won. Statistical minimum for one-point shot percent is 25%, and statistical maximum is 96.2%. The average value for this shot, regarding all games in the league, was 71.4%. The standard deviation is 0.11.

In the Serbian First "A" basketball league for women, the one-point shot percent does not influence final score of the game so much. This is visible from the lower graph, since an increase in shot percentage with the number of wins rises slower. In addition, with a decrease in the percent, the number of losses gradually increases.

The statistical minimum for one-point percent is 0%, which means that a team did not score even one free throw, and the statistical maximum is 95.2%. The average value for this shot, regarding all games in the League, is 66.8%. The standard deviation is 0.125.

For every coach, the one-point shot is very important. At every break during games, players are prompted to practice free throws, in order to improve this segment of their game.

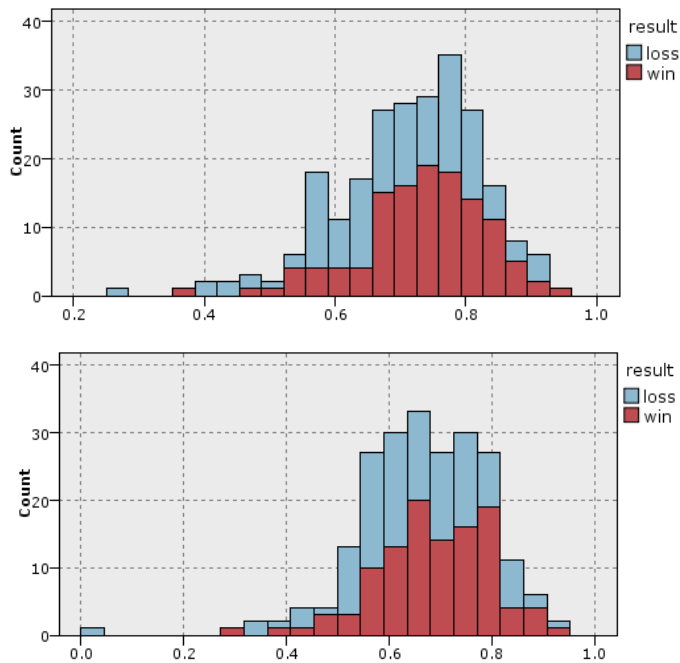


Figure 1

Effect of one-point shot percent on the final outcome of the game in Serbian First basketball league for men (above) and the Serbian First “A” basketball league for women (below)

Effect of the Two-Point Shot Percent

Figure 2 shows effect of the two-point shot on the final outcome of the game in the Serbian First basketball league for men (above) and the Serbian First “A” basketball league for women (below). It is visible that number of wins is higher when the two-point shot percent is over 58%. Below this level, the number of losses is higher. Therefore, the two-point shot percent has an effect on the game outcome. Here also are exceptions from general rule. It happened that teams had a two-point shot percent over 70%, and they still lose the game. In addition, it happened that a team with only 40% of two-point shots won the game.

The statistical minimum for the two-point shot percent 39.5% and statistical maximum is 83%. The average value for this shot, regarding all games in the league, is 58.2%. The standard deviation is 0.083. The two-point shot percent has a higher minimal value and a lower maximal value than the one-point shot.

Regarding women’s league graph, it is visible that number of wins is higher when the two-point shot percent is over 50%. When the two-point shot percent is below

40%, the number of losses is considerably higher. Therefore, this percent has a significant effect on the game outcome. By comparing these graphs, we may conclude that the two-point shot has more influence in women's basketball than in men's.

These graphs also show particular values that do not comply with the general rule. It happened that a team had a two-point shot percent over 83% and still lost the game. It has also happened that a team with about 30% still won the game.

The statistical minimum for the two-point shot is 24.3%, and the statistical maximum is 85.2%. The average value for this shot, regarding all games in the league, is 48.3%. The standard deviation is 0.105.

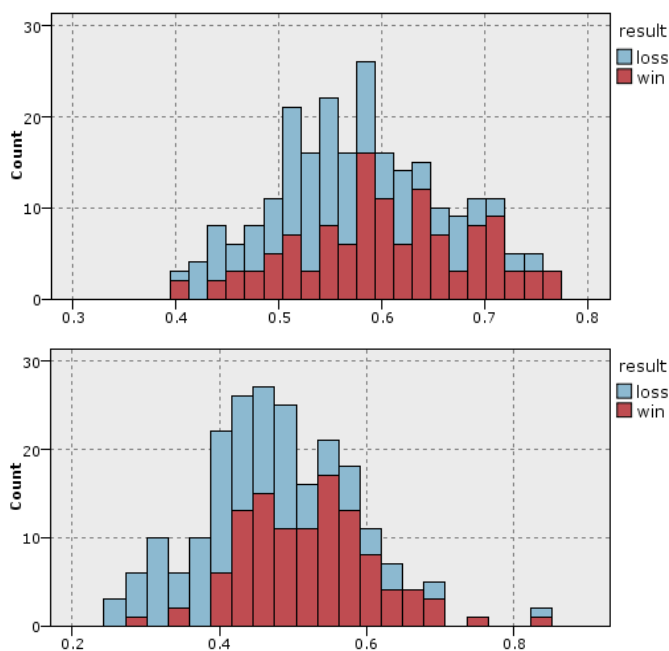


Figure 2

Effect of two-point shot percent on the final outcome of the game in the Serbian First basketball league for men (above) and the Serbian First "A" basketball league for women (below)

Effect of the Three-Point Shot

Figure 3 shows the effect of the three-point shot on the final outcome of the game in both leagues observed. From the upper graph (the league for men), it is visible that the number of wins is higher when the three-point shot percent is over 38%. When three-points shot percent is under 30%, the team loses the game more often. From this we may conclude that three-point percent has a significant influence on the outcome of the game. The statistical minimum for the three-point shot percent is 0%, meaning that a team did not realize a single three-point shot, and the

statistical maximum is 83%. The average value for this shot, regarding all games in the league, is 34.9%. The standard deviation is 0.349. The three-point shot has more extreme values than the two-point shot because of the lower numbers of shots during a game.

From the lower graph, showing the three-point shot percentages for the Serbian First “A” basketball league for women, it is visible that the number of wins is higher when the shot percent is over 43%. If the percent is over 60%, the team wins all games. When the shot percent is under 43%, there is less influence on the game outcome, since the number of wins and losses in these situations is approximately the same. From this we may conclude that the three-point shot percent has an effect on the outcome of the game, but this effect is significantly lower than in the Serbian First basketball league for men.

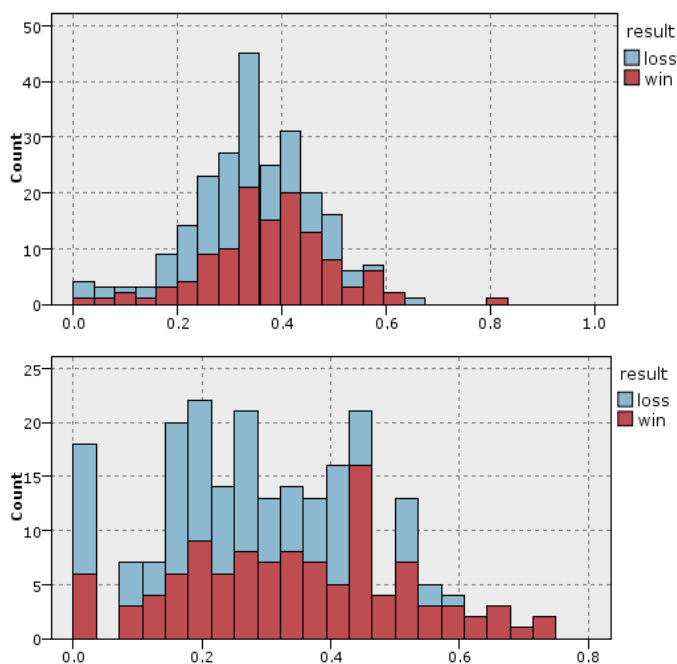


Figure 3

Effect of the three-point shot percent on the final outcome of the game in the Serbian First basketball league for men (above) and the Serbian First “A” basketball league for women (below)

In the sample observed, there was the extreme case when a team had a three-point shot percent of 0%, without a single successful shot, and still they won. This happened in 6 out of 110 observed games.

The statistical minimum for the three-point shot percent is 0%, and statistical maximum is 75%. The average value for this shot, regarding all games in the league, is 30%. The standard deviation is 0.166.

Effect of Defence Rebounds

There is a dilemma in some situations: determining whether it is a rebound or a steal. There are obviously offensive and defensive rebounds, but what about a situation when no one gets the ball in rebound? Many coaches think that if a ball is “rolling” on the court after being deflected from the hoop, a player who takes it must have a steal and not a rebound. Some think that steal should be given to the player who deflects a ball to his teammate, if he is not able to catch it after shot. If this is steal, then is the so-called tap in an offensive rebound or a steal?

Figure 4 shows the effect of defensive rebounds on the win or loss of the team. The graph for the Serbian First basketball league for men shows that the number of wins is higher when the number of defensive rebounds is over 22 for the team.

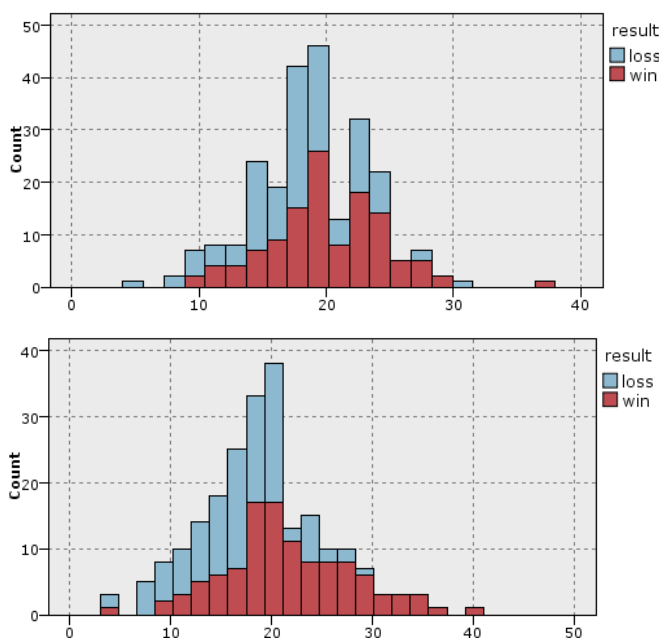


Figure 4

Effects of defensive rebounds on the final outcome of the game in the Serbian First basketball league for men (above) and the Serbian First “A” basketball league for women (below)

If the number of defensive rebounds is under 10, the team will lose the game. Regarding this distribution of wins and losses depending on defensive rebounds, we may conclude that they have a significant effect on the game outcome. With an increase in rebounds, the number of wins also increases, and vice versa.

The graph shows also a case when the team had 31 defensive rebounds, and they still lost the game. The statistical minimum for defensive rebounds is 4, and the maximum is 38. The average value, regarding all games in the league, is 18.98. The standard deviation is 4.637.

From the graph for the Serbian First “A” basketball league for women it is visible that the number of wins increases when the number of defensive rebounds is over 22 for the team. If the number of defensive rebounds is under 16, the team will most often lose the game. Therefore, we may conclude that defensive rebounds have a significant effect on the outcome of the game, similar as in the Serbian First basketball league for men. The statistical minimum for defensive rebounds is 3, and the maximum is 41. The average value regarding all games in the league is 19.23. The standard deviation is 6.213.

Effect of Offensive Rebounds

Figure 5 shows the effect of offensive rebounds on the final outcome of the game. From the upper graph (the Serbian First basketball league for men), it is visible that the number of wins is higher when the number of offensive rebounds is over 15. Because of the almost identical number of wins and losses, offensive rebounds do not have much effect on the final outcome of the game. There is also the case when the team had 24 offensive rebounds and they still lost the game.

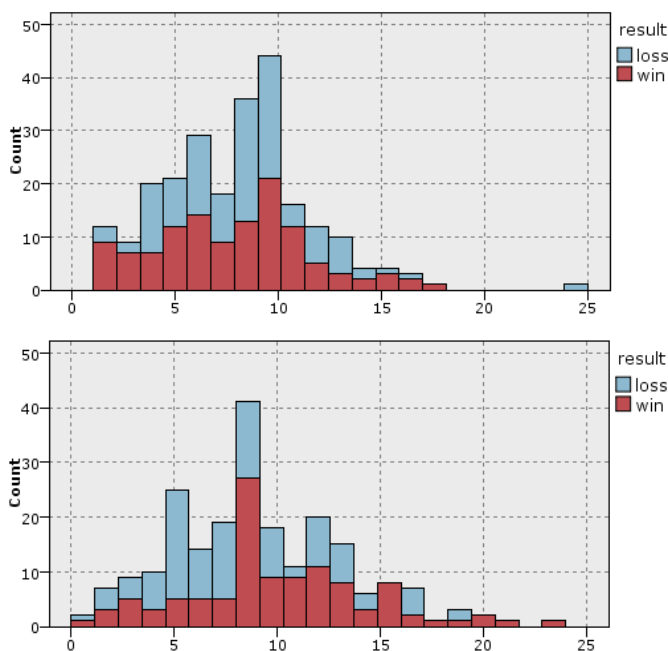


Figure 5

Effects of offensive rebounds on the final outcome of the game in the Serbian First basketball league for men (above) and the Serbian First “A” basketball league for women (below)

The statistical minimum for offensive rebounds is 1, and the maximum is 24. The average value, regarding all games in the league, is 7.93. The standard deviation is 3.529.

From the graph for women's basketball (below) it is visible that a team will certainly win if they have over 20 rebounds. Offensive rebounds have no major effect on winning or losing in this league either. There was a case when a team had 18 offensive rebounds, and still they lost the game. There was a game when one team had no offensive rebounds, and yet they won the game. The statistical minimum for offensive rebounds is 0, and the maximum is 24. The average value, regarding all games in the league, is 9.04. The standard deviation is 4.264.

Effect of Assists

Figure 6 shows the effect of assists on winning or losing by a team in the Serbian First basketball league for men (above) and the Serbian First "A" basketball league for women (below). In the men's league, when there are over 15 assists in a game, the team is wins in most cases. When a team has less than 8 assists, they will mostly lose.

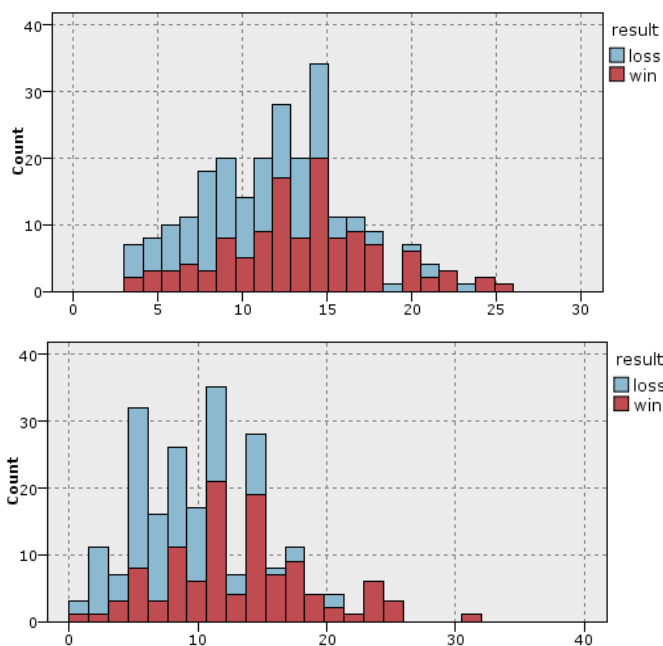


Figure 6

Effects of assists on the final outcome of the game in the Serbian First basketball league for men (above) and the Serbian First "A" basketball league for women (below)

The graph also shows cases when a team had 19, 20, 21 or 23 assists, and yet they lost the game. We may conclude that assists have no major effect on the outcome of the game. The statistical minimum for assists is 3, and the maximum is 26. The average value, regarding all games in the league, is 12.08. The standard deviation is 4.472.

In the women's league, when the number of assists is over 11, the team will mostly win. If a team has fewer than 8 assists, they will lose the game in most cases.

The graph also shows that a team who has more than 22 assists certainly wins the game. There was the case when a team had no assists, and they still won the game. The number of assists has more effect in the women's league than in the men's.

The statistical minimum for assists in the women's league is 0, and the maximum 32. The average value, regarding all games in the league, is 10.87. The standard deviation is 5.537.

Effect of Steals

Figure 7 shows the effect of steals on wins and losses in games in the two leagues observed. When the number of steals is over 15 in the men's league, this team will mostly win. If the number of steals is under 6, this team will mostly lose the game. From this graph, we may conclude that steals have an effect on outcome of the game, but less than defensive rebounds or shot percentages.

The statistical minimum for steals is 0, when a team has no steals during a game, and the maximum is 21. The average value regarding all games in the league is 9.18. The standard deviation is 4.192.

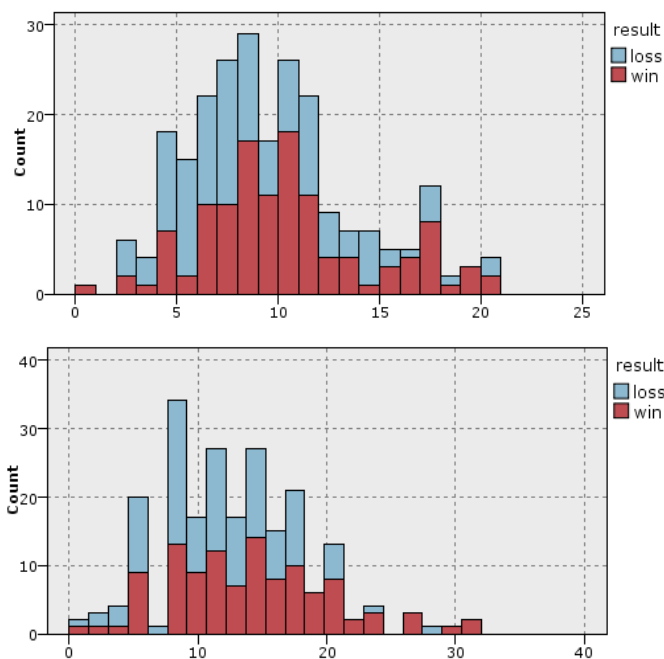


Figure 7

Effect of steals on the final outcome of the game in the Serbian First basketball league for men (above) and the Serbian First "A" basketball league for women (below) steals

In women's competition, when the number of steals is over 18, the team will mostly win. With a decrease in the number of steals, the number of losses increases.

The statistical minimum for steals is 0, and the maximum is 32. The average value regarding all games in the league is 12.99. The standard deviation is 5.644.

Effect of Turnovers

Figure 8 shows the effect of number of turnovers in the competitions observed. In the men's league, if the number of turnovers is over 15 (graph above), the team mostly losses the game. If a team has less than 10 turnovers, they mostly will win the game. This graph shows that turnovers have a significant effect on the outcome of the game.

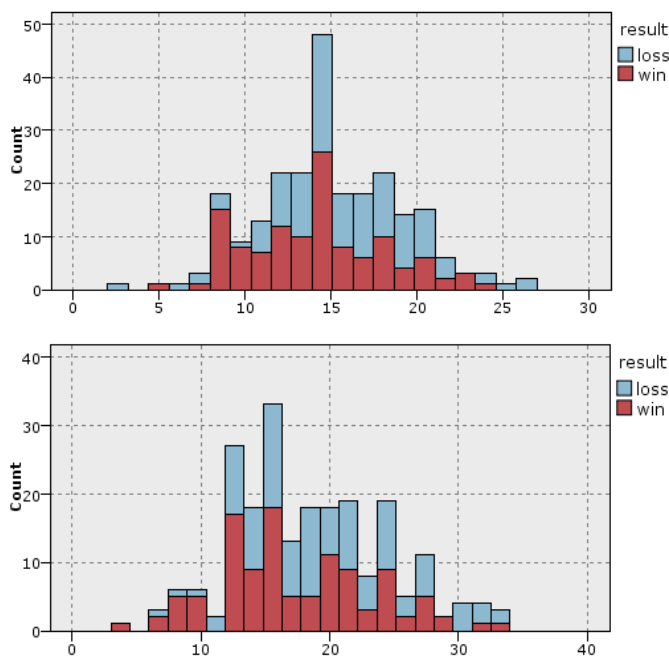


Figure 8

Effects of turnovers on the final outcome of the game in the Serbian First basketball league for men (above) and the Serbian First "A" basketball league for women (below)

The statistical minimum for turnovers is 2, and the maximum is 27. The average value, regarding all games in the league, is 14.92. The standard deviation is 4.144. In the women's league, if the number of turnovers is over 16, the team will mostly lose. If the team has less than 10 turnovers, they will mostly win the game. This graph shows that turnovers have a significant effect on the outcome of a game. The statistical minimum for turnovers is 3, and the maximum is 34. The average value, regarding all games in the league, is 18.49. The standard deviation is 6.058.

Effect of Blocks

Figure 9 shows the effect of blocks on the final outcome of the game. From the graph above, it is visible that if there is more than 4 blocks, the team will mostly win the game. This statistical parameter has no significant effect on the outcome of the game if its value is fewer than 4. The statistical minimum for blocks is 0, and the maximum is 8. The average value, regarding all games in the league, is 2.22. The standard deviation is 1.835. The graph for women's league is almost identical to the one for the men's league, so we may conclude that in women's basketball blocks have no major effects on the final outcome of a game.

The statistical minimum for blocks is 0, and the maximum is 8. The average value, regarding all games in the league, is 2.02. The standard deviation is 1.728.

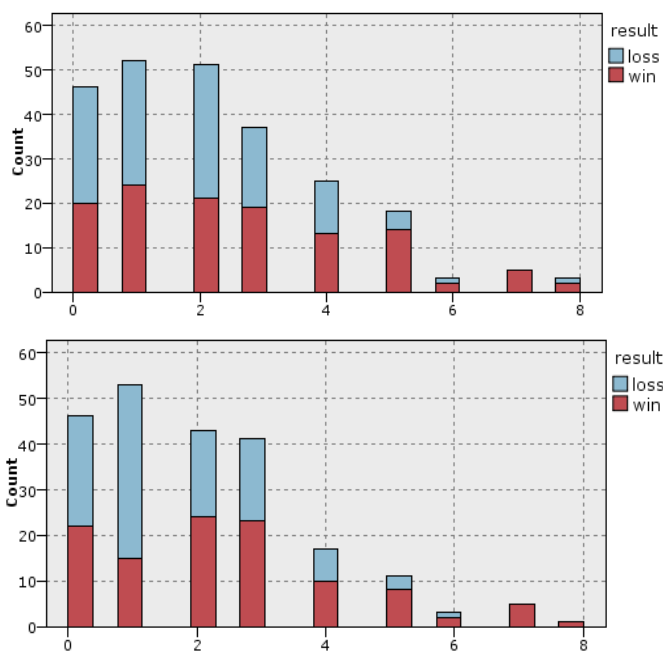


Figure 9

Effects of blocks on the final outcome of the game in the Serbian First basketball league for men (above) and the Serbian First "A" basketball league for women (below)

3.1.2 Comparative Analysis of the Basic Basketball Parameters

Table 2 shows the comparative list of one-, two- and three-point shots for the Serbian First basketball league for men and the Serbian First "A" basketball league for women. It can be seen that in the women's league, the statistical minimum for one-point shot percent is 0%, meaning that a team did not score any free throws, and consequently they lost the game. In men's league, this percent

varies from 25% to a maximum 96.2%. Regarding the average values of shots from all positions, we can conclude that men are more precise. Precision difference varies from 4.6% for the one-point shot, to 4.9% for the three-point shot to 9.9% for the two-points shot. Considering the three-points shot, Jamie Angeli [15] says that these shots are hardest to score, since a player must be alone in order to score successfully from a distance. Many players stay after training sessions in order to perfect this shot [16] [17].

Table 2
Basic statistical parameters for one-, two- and three-point shots

	1P				2P				3P			
	Min%	Max%	MEAN	S.DEV	Min%	Max%	MEAN	S.DEV	Min%	Max%	MEAN	S.DEV
MEN	25.0	96.2	71.4	0.110	39.50	77.40	58.2	0.083	0.00	83.30	34.9	0.122
	1P				2P				3P			
	Min%	Max%	MEAN	S.DEV	Min%	Max%	MEAN	S.DEV	Min%	Max%	MEAN	S.DEV
WOMEN	0.0	95.2	66.8	0.125	24.30	85.2	48.3	0.105	0.00	75.00	30.0	0.166

Table 3 shows three very important parameters: defensive rebounds, offensive rebounds and assists. It can be seen that the average number of rebounds per game is higher for women. Women have on average 0.244 more defensive and 1.112 more offensive rebounds. This is to be expected, since they have lower shot percentages from all positions. Data regarding assists are especially interesting. According to FIBA statistician manual [18], an assist is a pass leading directly to another player's score, and only if a shooter reacts by immediate shot movement. A pass to another player who is in a good shooting position but who considers other options before shooting, is not an assist. Shooting distance and the manner of shot are not factors deciding whether the pass is also an assist. Passing to a player who is alone on the middle-court and who dribbles to basket is assist, but if this player must dribble past an opponent player, then it is not an assist. Therefore, the deciding factor for determining an assist must be the sum of his action and of other player's imminent intention to shoot and score. In USA, this is quite different, whether in NBA, NSA or WNBA. There the assist is noted if a ball was passed to another player when he/she is alone in a shooting position, no and it does not matter whether he/she was fouled after that. In Table 3, it is visible that men have on average 1.21 more assists per game.

Table 3
Basic statistical parameters for defensive rebounds, offensive rebounds and assists

	DEF. REBOUND				OFF. REBOUND				ASSIST			
	Min	Max	MEAN	S.DEV	Min	Max	MEAN	S.DEV	Min	Max	MEAN	S.DEV
MEN	40	38	18.988	4.637	1.00	25	7.929	3.529	3	26	12.083	4.472
	DEF. REBOUND				OFF. REBOUND				ASSIST			
	Min	Max	MEAN	S.DEV	Min	Max	MEAN	S.DEV	Min	Max	MEAN	S.DEV
WOMEN	3	41	19.232	6.213	0	23	9.041	4.264	0	32	10.873	5.537

The next set of parameters is a basis for good defence and a counter-attack. Table 4 shows that women had more steals than men, on average 3.816 more per game. Regarding turnovers, women are more successful too: on average, 3.57 per game. Therefore, we can conclude that in the Serbian First “A” basketball league for women, better defence is played. The number of blocks is almost identical in both leagues.

Table 4
Basic statistical parameters for steals, turnovers and blocks

MEN	STEAL				TURNOVER				BLOCK			
	Min	Max	MEAN	S.DEV	Min	Max	MEAN	S.DEV	Min	Max	MEAN	S.DEV
	0	21	9.179	4.192	2	27	14.925	4.144	0	8	2.217	1.835
WOMEN	STEAL				TURNOVER				BLOCK			
	Min	Max	MEAN	S.DEV	Min	Max	MEAN	S.DEV	Min	Max	MEAN	S.DEV
	0	32	12.995	5.664	3	34	18.495	6.058	0	8.	2.027	1.728

3.2 Model Evaluation Phase

In the Serbian First basketball league for men, the network correctly predicted 195 outcomes from 240 possible (120 games where every team must either win or lose), which is 81.25% of total number of input data. Therefore, the model failed 45 times to predict the outcome correctly, which is 18.75% of all cases. From 120 wins documented, the algorithm correctly predicted 99, while for 21 wins it predicted losses. Regarding losses, the algorithm correctly predicted 96 from 120 losses, while for 24 losses it predicted wins. The confidence matrix is shown in Table 8. In the Serbian First “A” basketball league for women, the network correctly predicted 304 from 362 outcomes or 83.98% of the total number of input data. Therefore, the model did not correctly predict 58 outcomes, which is 16.02% of all outcomes. From 181 wins documented, the algorithm correctly predicted 144, while for 37 wins it predicted losses. Regarding losses, the algorithm correctly predicted 160 losses, while for 21 losses it predicted wins. The confidence matrix is shown in Table 9.

The model including the most relevant basketball parameters has a relatively high prediction precision for game outcomes based on the input parameters. More than eighty percent of the input data would correctly predict the outcome of the game.

Table 6
Model precision for the Serbian First basketball league for men

True	195	81.25%
False	45	18.75%
Total	240	

Table 7

Model precision for the First "A" basketball league for women

True	181	82.27%
False	39	17.73%
Total	220	

Table 8

Confidence matrix for the Serbian First basketball league for men

	Loss	win
loss	96	24
win	21	99

Table 9

Confidence matrix for the Serbian First "A" basketball league for women

	loss	win
loss	82	28
win	11	99

The results obtained by the neural network were confirmed by the C5.0 decision tree.

A prediction correctness of over 80% confirms that the model used was correct. Greater correctness was prevented by the fact that, in keeping basketball statistics, a large number of events are not noted. Programs like BSV are used in real time, during the game, so there must be a selection as to which events will be documented and which not. In order to obtain more complete knowledge regarding a game and find some new patterns, we need a richer data set, or the application of software solutions that would note all relevant events during later viewing of the game.

Conclusion

The game of basketball is progressing rapidly. The number of quality players and teams is growing quickly. At high levels of competition, there are no teams that can count on a safe win for every game. Good preparation for the game may mean the difference between average and good results. Scouting opponents is an important and indispensable element in these preparations. Special attention must be paid not only to the manner of collecting and recording data during a game or by viewing video clips, but also to a way of processing data and presenting them to the team. Increasing professionalism and competition are prompting clubs to develop systematic approaches to all activities. With data mining techniques, it is possible to analyse both certain players and teams as a whole. Using modelling data from the Serbian First basketball league for men and the Serbian First "A" basketball league for women, we obtained knowledge regarding the way of playing and the decisive elements that influence the final outcome of the game. In

both leagues, defensive rebounds have the most influence on winning the game. Men's basketball is based on a faster game and more shots, so precision and good training is very important. Women's basketball pays more attention to defence, so other important elements are steals and turnovers. In addition to these, it is important to maintain a high level of two-point shots, not to miss "safe shots". In team sports, there are many methods and ways of preparing athletes for competition. The results were obtained from the application of a neuron net to the collected data, whereas the check was done through applying C5.0 decision trees, which confirmed the results.

These methods include physical, technical, tactical, psychological and integral types of preparation. Each of them has fundamental importance in the formation of athletes and teams, leading to a successful performance at the competition and good results. Today, good scouting cannot be imagined without the use of modern information technologies. The coaches at the game often ask how many points were received from the zone defence and how the particular player shoots from different positions. Coaches, assistant coaches, players and scouts are also interested in how much each player scored, information about everyone's points given different types of defence, and how many points were received from another attack or counterattack, or whether the team will take on a zone defence in time. These are additional parameters that influence the final outcome of the game and which are not a subject of the modelling in this paper. The reason for this is that such information is not recorded in the kept statistics on one basketball game, but may be recorded in a deeper analysis of the game.

Acknowledgement

This work was partially supported by the Serbian Ministry of Education and Sciences (Grant No: 171039).

References

- [1] Lyons, K. "Data Mining and Knowledge Discovery", Australian Sports Commission Journals 2, 2005
- [2] O'Reilly, N., P. Knight "Knowledge Management Best Practices in National Sport Organizations". International Journal of Sport Management and Marketing 2(3), pp. 264-280, 2007
- [3] Schumaker R., Solieman O., Chen H. "Sports Data Mining", Springer 1st edition, 2010
- [4] Stefani, R. "A Taxonomy of Sports Rating Systems". IEEE Transactions on Systems, Man, and Cybernetics - Part A 29(1): 116-120. 1999
- [5] Choo, C. W. "The Knowing Organization: How Organizations Use Information to Construct Meaning, Create Knowledge, and Make Decisions". International Journal of Information Management 16(5): 329-340, 1996

-
- [6] Fielitz, L. & D. Scott. "Prediction of Physical Performance Using Data Mining". *Research Quarterly for Exercise and Sport* 74(1): 1-25, 2003
- [7] Dean Oliver, "Basketball on paper – Rules and tools for performance analysis", Brassey's, Washington DC, 2005
- [8] Chang, C. W., S. Y. Lee. "A Video Information System for Sport Motion Analysis". *Journal of Visual Languages and Computing* 8(3): 265-287
- [9] Marakas, G. "Modern Data Warehousing, Mining, and Visualization: Core Concepts". Prentice Hall, Upper Saddle River, NJ, 2003
- [10] Wolfers, J. "Point Shaving: Corruption in NCAA Basketball". *AEA Papers and Proceedings* 96(2): 279-283, 2006
- [11] Dobra, J., T. Cargill, B. Goff, R. Tollison. "Efficient Markets for Wagers: The Case of Professional Basketball Wagering. In *Sportometrics*", Texas A&M University Press, College Station, TX, 215-249
- [12] Witten, I. H. & Frank, E. "Data Mining: Practical machine learning tools and techniques, 2nd Edition". Morgan Kaufmann, San Francisco, 2005
- [13] Shih-Yang Yang, Po-Zung Chen, Chu-Hao Sun, "Using Petri Net to Enhance Web Usage Mining", *Acta Polytechnia Hungarica*, Vol. 4, No. 3, 2007, pp. 113-125
- [14] Lajos Izsó, Péter Tóth, "Applying Web-Mining Methods for Analysis of Student Behaviour in VLE Courses", *Acta Polytechnia Hungarica*, Vol. 5, No. 4, 2008, pp. 79-92
- [15] Jamie Angeli, *Scouting America's Top Basketball Programs*, Volume 1, 2003
- [16] Ratgeber, L. *Play from a Game: (Head Coach). Mizo Pecs 2010, 2007/2008, Mizo Pecs 2010 vs. Euroleasing Sopron*
- [17] *Serbian Basketball Coaches: Zeljko Obradovic, Aleksandar Dordevic, Branislav Prelevic, Bozidar Maljkovic, Dusko Vujosevic, Svetislav Pesic*, Book group 2008
- [18] *FIBA - Basketball Statisticians' Manual*, 2010