Basketball Players Performance Analytic as Experiential Learning Approach in Teaching Undergraduate Data Science Course

Nurfadhlina Mohd Sharef Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Serdang, 43400 Selangor, Malaysia nurfadhlina@upm.edu.my Aida Mustapha
Faculty of Computer Science and
Information Technology,
Universiti Tun Hussein Onn Malaysia,
Parit Raja 86400, Batu Pahat,
Johor, Malaysia
aidam@uthm.edu.my

Muhammad bin Nor Azmi Kelab Bola Keranjang Presint 11 (Putrajaya) Precinct 11, 62300, Putrajaya pelontarp11@gmail.com

Rosdiadee Nordin
Faculty of Engineering and Built
Environment,
Universiti Kebangsaan Malaysia,
43600 Bangi, Selangor, Malaysia
adee@ukm.edu.my

Abstract—Sports analytic is an informative and beneficial tool especially for a coaching team who closely follow the development of their team. For a sports coach, deciding the best team composition is crucial to ensure winning. This requires understanding the dynamics of the players and the game parameters. However, small sports communities do not usually have the chance to benefit from this advanced technology. In a Data Mining course at Universiti Putra Malaysia, a data analytic project of basketball players' performance has been conducted through experiential learning approach. This paper presents the experience and output of shooting performance analytic, attack profiling analysis, team performance analysis, player rating prediction and game outcome prediction. The achievements of the students on the project are high with low marks deviation, plus the students said that the pedagogy encourages them to be highly committed and independent. Engaging lessons have also been happening as students compassionately share about their findings.

Keywords— Basketball performance analytic, Undergraduate Data Science Course, Sports Analytic, Experiential Learning,

I. INTRODUCTION

One of the challenges in coaching a winning basketball team is to decide the team composition for a planned game. This leads to the importance of basketball players' performance and game analytic as insights for the coach to improve the upcoming game strategies or to support the training system to hone player's skills. However, game parameters and player performance data are multi-faceted, whereby the performance of an individual player is measured from various angles, such as points, assist, or rebound, making the raw data too complicated and time consuming for the coach to analyze, especially in a crucial situation, such as deciding on the next move in a close game situation [1][2]. It is imperative for the coach to identify and predict which player will perform well in certain score metrics or techniques, in combination with other players.

At present, sports analytic has experienced rapid growth and it begins with those sports fans who are searching for predictive statistics, tools and related techniques to better measure both player and team performance [3],[4],[5],[6]. The

benefits of data analytics in basketball games can be seen as threefold [1][2][3]. First and foremost, the obvious use is to predict the game results or outcomes. Next, it helps the coach to evaluate players' performance with regards to their pace or movement throughout the game, as well as during the upcoming matches. Thirdly, it provides a solution for supporting the basketball coaches in making tactical or technical decisions; whether pre-game, in-game or post-game.

Basketball players' performance analytic is a good example of experiential or challenge-based learning activity in Data Mining course, one of the Data Science course series, where students come out with solutions based on real-world problems. The application of experiential learning [7] approach enables students to learn by doing, and collaboratively with their peers, through an iterative cycle of experiencing, reflection, conceptualization experimentation. The objective of this paper is to present the data analytic solution for basketball players performance analysis produced by students within two weeks as graded assessments in the Data Mining course at the Faculty of Computer Science and Information Technology, Universiti Putra Malaysia. The problem dataset is based on games records by a community basketball club called Pelontar XI (PXI) from Putrajaya, Malaysia.

There are five steps in data analytic, namely (i) data retrieval, (ii) data preprocessing, (iii) exploratory data analysis, (iv) machine learning models development, evaluation and testing, and (v) visualization and communication. Exploratory data analysis is aided by visualization and allows identification of the descriptive or basic summary of data, as well as diagnosis of happening in the data. Data analytic typically can be addressed using descriptive, diagnostic, predictive and prescriptive analytic. Descriptive analytic revolves around the utilization of statistics technique to summarize the trends in the data while diagnostic analytic allows identification correlations, clustering and association. Predictive analytic support interventive mechanisms, while prescriptive analytic provides ultimate strategy and optimization. The choice of machine learning models depends on the goal and characteristics of analytic. For example, the Apriori and k-means algorithms are used for descriptive analytic, including association rules and cluster identification. Decision Tree, Random Forest and Regression algorithms are used for predictive analytic.

This paper presents the utilization of data analytic methods for shooting performance analytic, attack profiling analysis, team performance analysis, player rating prediction and game outcome prediction. The work is similar to [1] but applied in a local context and referring to [2],[3] as a guide on basketball game parameters mining. A dashboard is delivered to visualize the analysis for strategical game and team. The remaining of this paper is organized as follows. Section 2 reviews all works related to sports analytics in basketball as well as on team performance. Section 3 presents the methodology used to perform sports analytics along with the dataset and different sports analysis. Section 4 presents the experimental results and finally, Section 5 concludes with some direction for future work.

II. EXPERIENTIAL LEARNING IMPLEMENTATION

The Data Mining course (SSK4604) is offered at the Faculty of Computer Science and Information Technology, Universiti Putra Malaysia as an elective course under the specialization of Bachelor of Computer Science (Computer System). There are 33 students in the course, and most of them are in fourth semester, out of eight. Many are also taking Business Analytic and Python programming courses, which are complementary skills in the Data Mining course. The Data Mining course (SSK4604) is offered at the Faculty of Computer Science and Information Technology, Universiti Putra Malaysia as an elective course under the specialization of Bachelor of Computer Science (Computer System). Various approaches have been taken to ensure that the content of the course is up-to-date and sync with the industry needs including implementing an experiential learning approach for the project-based assessment. Besides, students are exposed to popular tools used by the industry for data science, such as PowerBI, Tableau, Weka, Python and R. Datasets from competitions, research and supplied by the industry are also used for assignments and projects. Students work in groups to develop a blog to showcase their data analytic modeling experience and output, as their junior data scientist's online portfolio.

The learning outcomes of this course are that upon finishing the course, the students are able to:

- 1. Generate rules to represent knowledge and make decisions using data mining tools.
- 2. Demonstrate the ability to make decisions using data mining techniques.
- 3. Analyze problems, propose alternative solutions and make presentations.
- 4. Pro-active collaboration in groups.

Lessons have been conducted through various modes comprising of face to face, blended and online, mainly through Universiti Putra Malaysia's formal learning management system called PutraBLAST. Flipped classroom and digital tools are used to make the lessons engaging, such as FlipGrid, Padlet, Whatsapp, PutraBLAST and Quizizz. Emphasis has been given on the students' innovation development skills to complement their theoretical understanding. Many peer-based learning activities were

conducted, including celebrating diversity where teams of various demographic backgrounds are formed and tasks that require learning of cross-culture and background took place. One of the tasks given was to discuss the train services in each of their countries. This exercise allows students to identify the strengths and weaknesses of the service, which is to practice their analytical skills.

Sports analytics enable a coach to identify a specific player that will contribute the most in a specific game during a certain situation in the game. This way, coaches can tailor the training program for an individual player based on each strength and weakness. The decision to change player during a losing period can only be made if the coach has already analyzed the strengths of each player and predicted who will deliver the best 3-point shot during the remaining game time. For example, if the game is finishing and the team is losing, the coach can decide to change the player to an offensive player or player with the highest record of FG-2PT or FG-3PT. Similarly, if the game is finishing and the team are leading by a few points only, the coach can maintain or change the player to player with defensive skills.

To be able to deliver the project for the analytic challenge, the students had familiarized themselves with the basketball game analytic. The direct engagement with PXI through briefing, shared materials and presence during the students' project presentation boost the students' confidence that they solve real-world problem. The expected outcome of the project is to prepare students for:

- Understanding data mining tasks based on a focused domain scenario.
- 2. Developing skills for exploring, analyzing, visualizing data by applying suitable tools.
- 3. Planning and managing data mining projects by utilizing elevator pitch technique, and project reporting through an online blog.
- 4. Developing teamwork skills through management, development, implementation, and reporting of the project.

Students often start the task with loads of worry, but through time, they are motivated to develop the best solution. Many self-learning activities are conducted, where they typically refer to example codes online. The instructor coaches the progress, while the students undergo the learning experience cycle. The solution development (criteria of evaluation is shown in Table 1) carries 15% of the total marks while the presentation carries 5%. On average students scored 18% with 3% standard deviation on the Project.

Table 1. Criteria for Project Assessment

Criteria	Marks
Summary	3
Dataset description	5
Objectives	5
Data preprocessing	10
Data exploration	10
Model development and experiment analysis	10
Look and feel	2
Coding/Model + progress +demonstration	30
Total	75

III. BASKETBALL PLAYERS PERFORMANCE ANALYTIC

A. Dataset

The dataset used for the analysis comprised of game records of National Community Basketball League (NCBL) season 2018/2019 for PXI. This dataset has a total of 28 sheets of player performance records for 13 matches and 4 friendly matches. It provides the game statistics, competition scores, team performance summary as well as player statistics such as points, rebound, assist, and attempt. The statistics also include the total number of games played by the player and how many games they won.

B. Requirements by PXI

PXI is open on analytic of team or player; and they want to understand which analytic is valuable during training or game. The complete list of attributes is shown in Table 2. Five expectations have been set by PXI, which are:

- Set of players that is either offensive of defensiveoriented
- Prediction/probability of winning/losing based on a different set of players
- Association rules on winning or losing
- Clustering players based on their strengths or weaknesses

Table 2. List of Attributes for Pelontar XI Dataset

No.	Attribute	Description		
1	No	Number of the player		
2	Player	Name of the player		
3	PTS	Points made in a game		
4	REB	Rebound made in a game		
5	AST	Assist made in a game		
6	STL	Steal made in a game		
7	BLK	Block made in a game		
8	T/O	Turnover made in a game		
9	RTG	Rating		
10	FG(TOT)	Total field goal (both 3- and 2-points)		
11	FG(TOT)	Total field goal made in a game		
	MADE			
12	FG(TOT)	Total field goal missed in a game		
	MISS			
13	FG(TOT)	Total field goal attempted in a game		
	ATT			
14	FG%	Percentage of the total field goal		
15	FG(2-PT)	Total 2-points field goal		
16	FG(2-PT)	Total 2-points made in a game		
	MADE			
17	FG(2-PT)	Total 2-points missed in a game		
	MISS			
18	FG(2-PT)	Total 2-points attempted in a game		
	ATT			
19	2-PT%	Percentage of the 2-points field goal		
20	FG(3-PT)	Total 3-points field goal		
21	FG(3-PT)	Total 3-points made in a game		
	MADE			
22	FG(3-PT)	Total 3-points missed in a game		
	MISS			
23	FG(3-PT)	Total 3-points attempted in agame		
	ATT			

24	3-PT%	Percentage of the 3-point field goal	
25	FT	Total free throw	
26	FT	Total free throw made in a game	
	MADE		
27	FT MISS	Total free throw missed in a game	
28	FT ATT	Total free throw attempted in a game	
29	FT%	Percentage of the free throw made	
30	FOULS	Number of fouls	
31	MIN	Game length in minutes	
32	W/L	Game result of win or lose	
33	OPPT	Opponent	
34	ATT	Total number of scoring attempt	

C. Exploratory Basketball Games Data Analytic

The exploratory data analysis focuses on identifying patterns in the data and often performed by visualizing the summary of statistics, for example, top players for each game points (Fig. 1). Players' performance in each game according to game points is also explored (Fig. 2). Players rating is identified as well where the more games a player plays, the higher the RTG. For those who play more games but have a lower RTG may indicate that the player is a substitute. Although he participated in many games, his playing time and scoring rate are not high (Fig. 3). Performance in each game is summarized by the average of players' scores in each game point.

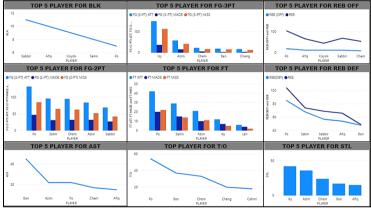


Fig. 1 Top Players according to Total Score of each Game Points

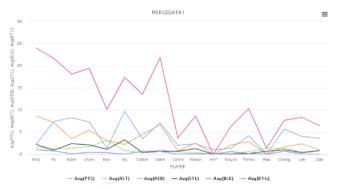


Fig. 2 Players' Score in each Game according to Game Points

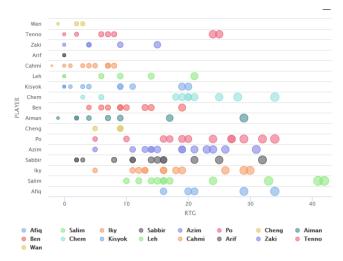


Fig. 3 Statistics of Player's Rating in Played Games

D. Machine Learning Models Development

Two data mining methodology have been chosen as the primary analysis, which are clustering and classification. Data clustering and classification are both a method of data analysis. Clustering aims to divide a data set into different homogeneous packets, in the sense that the data of each subset share common characteristics, which most often correspond to proximity criteria (computer similarity) that are defined by introducing distance measurements and classes between objects. Meanwhile, classification aims to accurately predict or assign each new testing sample or case into pre-determined categories or classes. The clustering methodology is used for (1) 3- and 2-point shooting analysis, (2) attack profiling analysis, and (3) team performance analysis. Meanwhile, the classification methodology is used for (1) game match outcome and (2) player rating or performance prediction. A total of nine data mining algorithms have been developed, which are Naïve Bayes, Decision Trees, Random Forest, k-Nearest Neighbor, Linear Regression, Logistic Regression, Deep Learning, Neural Networks, and k-Means.

E. Shooting Performance Analysis

One way to predict a win in a basketball game is by analyzing shooting performance for 2- and 3-point field-goals. High shooting performance will lead to higher winning chances as a game outcome. In analyzing shooting performance, this experiment focused on clustering the PXI performances based on attempts (ATT) made against the opponents they faced during the season. The values used are the total of the team's member game points statistics. The objective of clustering is to identify player's performance profiles, which would be useful to the coach as they would be able to change game strategy based on the outcome of their players' performance against each team.

Fig. 4 and Fig. 5 show the clustering results for 2-point and 3-point shooting, respectively. The experiment produced four clusters, which are Yellow (Y) that represents extremely competitive opponents, Blue (B) and Red (R) that represent competitive opponents, and Green (G) represents least competitive opponents. From the results, 2-point shooting attempts are considered easy to attempt and succeed in, but

the analysis showed that one team has totally shut the 2-point zone from the PXI attack.

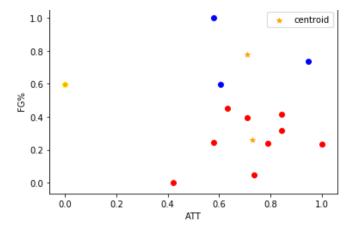


Fig. 4 Clustering Results for 2-point Shooting based on Attempts

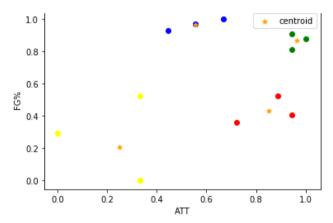


Fig. 5 Clustering Results for 3-point Shooting based on Attempts

Analysis on shooting performance for 2- and 3-point field-goals is a team based-analysis, which is different from a player based-analysis in the sense that it allows the coaching staff to create new training and new strategies based on teamwork rather than specified or roles training for each player. The analysis projects an overall point of view of how the team performs with 2- and 3-points scoring against a specific opponent to strategize for higher winning chances. Typically, this type of data analysis permits a way better structure of the attacking plan facing a specific opponent, especially for the extremes where the efficiency plays a lot more involving a change of plan or eventually abuse strategy that works very well.

F. Attack Profiling Analysis

The objective of attack profiling analysis is to understand which players are active attackers and which one are more passive attackers (usually defensive profiles). In general, offensive players will have more rebounds (REB) and assist (AST) points while defensive players mainly focus on steals (STL) and blocks (BLK). For this analysis, *k*-Means algorithm is used to perform the analysis to cluster the player profiles in terms of offensive vs. defensive profiles. The resulting clusters are showed in Fig. 6.

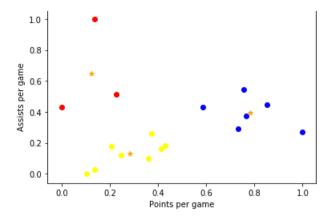


Fig. 6 Clustering Results for Player Profile

From the figure, the Yellow (Y) cluster indicates defensive/passive players, Blue (B) cluster indicates great scorer, and Red (R) cluster indicates great passer. This analysis is useful to the coach because each team formation should have a balanced strength in scoring, passing, and defending. The balance is important because the scorer makes the win, the passer enables the scorers and the defensive players are those who ensure a good defense but also offer an opportunity for offensive transition.

In a basketball game, there are attacking screens, offensive rebounds and screens for rebounds, which all are important in tactical aspects and need players that are great at more than just scoring and assisting. The impact of a player is not determined only by assisting and scoring, hence the reason why the clustering results suggest a balance between elements of the three clusters on the court to optimize the attacking and partially the defensive part.

G. Team Performance Analysis

Analysis of team performance is conducted via a clustering experiment using the *k*-Means algorithm from statistics of individual player performance. The attributes under study are AST, BLK, PTS, REB, RTG, STL, and T/O for each player across all 17 games. The clustering experiment produced five clusters, which can be described as follows:

- Class 0 belongs to the weaker team of the team. They
 have played a lot of times, but the data is very poor, it
 may be the recent state is poor, or their own ability.
- Class 1 belongs to the team's top players; their appearances and statistics are very bright.
- Class 2 belongs to the team's modest personnel. They have many people, but the statistics are mediocre, and the number of appearances is not significant.
- Class 3 belongs to the second echelon of the team.
 Although their abilities are not as good as those of the first class, they are still outstanding in the entire team.
- Class 4 belongs to the team's bench players, with a small number of appearances and average data. They are players who play temporarily when the team is in an emergency.

Clustering individual performance will allow the coach will get to know each player's performance and likewise, the

players will know their potential to play in the future, hence increasing the probability of winning.

H. Player Rating Prediction

In basketball, building a balanced team is the key to winning. Based on the PXI data, there is no specific information that state which player will make an excellent performance when put on the field. Hence, predicting which player who will give the most outcome when being in action would benefit the coach in assembling a winning team. In player rating prediction, this research employed six prediction algorithms, which are Decision Trees, *k*-Nearest Neighbor, Support Vector Machine, Random Forest, Deep Learning, and Neural Networks. Table 3 shows that the Neural Networks algorithm performed the best with the lowest root mean square error among these three models to predict player performance.

Table 3. Prediction Results for Player Rating

Algorithm	Root Mean Squared Error	
	(RMSE)	
Decision Trees	8.015 +/- 0.000	
k-nearest Neighbor ($k = 5$)	6.750 +/- 0.000	
Support Vector Machine	6.736 +/- 0.000	
Random Forest	6.714 +/- 0.000	
Deep Learning	3.000 +/- 0.000	
Neural Networks	0.512 +/- 0.000	

In predicting the player rating, important factors contributing to achieve a high rating is calculated. Fig. 7 shows that the rating of players can be predicted and the factor of contribute to achieving high rating mostly comes from Rebound (REB). This suggests that a coach can pick the top five players in the game to plan a suitable training depending on the weakness of each player.

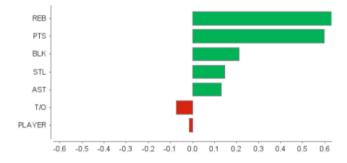


Fig. 7 Impact of Attributes in Predicting Player Rating

In this experiment, important factors for each player is calculated. As an example, Fig. 8 shows that Block (BLK) and Assist (AST) statistics contributed mostly for winning a game.

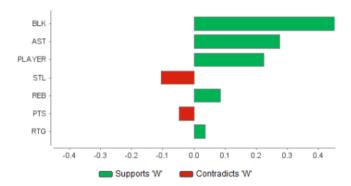


Fig. 8 Impact of Attributes in Predicting Team Factor

I. Game Outcome Prediction

Classification is a data mining function that assigns items in a collection to target categories or classes. The goal of classification is to accurately predict the target class for each case in the data. The purpose of this experiment is to predict the game outcome, whether win or lose using seven classification algorithms, which are Naïve Bayes, Decision Trees, and *k*-Nearest Neighbors, Random Forest, Linear Regression, Logistic Regression, and Deep Learning. The experiments were carried out using the RapidMiner with 70:30 holdout validation method for training and testing, respectively. The attributes selected were AST, BLK, FG (2-PT) MADE, FG (3-PT) MADE, FT MADE, PTS, REB, STL, T/O, with Win/Lose as the class. The results are shown in Table 4.

Table 4. Game Outcome Prediction Results for Class Win

Algorithm	Accuracy	Precision	Recall
	(%)	(%)	(%)
Naïve Bayes	72.00	77.42	77.42
Decision Trees	62.00	62.00	100.00
5-Nearest Neighbor	58.00	64.71	70.97
9-Nearest Neighbor	66.00	66.67	90.32
Random Forest	64.29	66.67	80.00
Linear Regression	63.35	67.62	75.53
Logistic Regression	64.60	68.47	80.85
Deep Learning	60.00	64.00	76.19

The results showed that Naïve Bayes achieved the highest accuracy performance in predicting a win for PXI games, followed by k-Nearest Neighbor with k=9 and Decision Trees. The accuracy of the k-Nearest Neighbor algorithm increased to 66.00% when the number of cluster k is changed to 9 from 5. Next, the impact of attributes for each prediction model is shown in Fig. 9, while an example on simulation of a new game and winning chance is in Fig. 10.

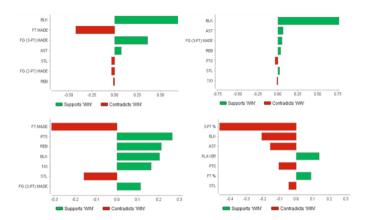


Fig. 9 Impact of Attributes in Predicting Win for All Algorithms (a) Naïve Bayes (b) Decision Trees (c) k-Nearest Neighbor (d) Random Forest



Fig. 10 Prediction of Winning based on Simulated Data

From the figure, it is shown that the main attribute to support winning prediction is Block (BLK) for both Naïve Bayes and Decision Trees. Meanwhile, the Free Throw (FT) Made attribute has the least predictive impact for both Naïve Bayes and *k*-Nearest Neighbor algorithm. In Random Forest, 3-PT% is an important factor that influences a game. This means the coach can plan a suitable offensive strategy to make the player gain the score other than shooting 3-point, but it only a temporary solution because 3-point shooting is one of the biggest 'swords' in the game. The best solution is making a training plan for those who are lower 3-point shooting percentage.

IV. CONCLUSIONS AND FUTURE WORK

This paper presents the experience on implementing data science for sports analytics, for coaches to make informed decisions not only in strategizing for the upcoming game but even while the game is still on. Shooting performance analytic, attack profiling analysis, team performance analysis, player rating prediction and game outcome prediction solution have been presented in the paper to facilitate insightful team composition.

Through data mining, various game parameters, one of the main findings from the developed solution is that the PXI Basketball team more likely focused on rebound, unlike NBA players that focused on scoring. This work is constrained to the availability of data on training and previous basketball matches, which limits the use of the dataset to analyze player-related trends such as their performance level.

The experiential learning approach has enabled students to be exposed to real-world challenges and meeting the stakeholders demand. Interestingly, there is no single group that provides a similar solution. This indicates high creativity and critical thinking by the students. Students indeed undergo high order thinking skills challenge as they explore the most effective machine learning model and proposing the best solution. Besides, working in a team allows them to build strong collaborative skills. The presentation session and the blog development showcased their solution and sharpened their written and online citizenship character.

Acknowledgement. We would like to express our appreciation to Pelontar 11 (PXI) from Putrajaya for inspiring and supporting this work. We hope that basketball, and other sports analytic will grow even further in Malaysia, even for small community benefits. The experience has been meaningful in especially in introducing students to real world challenges. We would also like to acknowledge the following undergraduate students at Universiti Putra Malaysia for contributing to the analysis of this case study: Malik Dehli, Marie-Aude Morel, Dallas Rapidash, Siti Nurrainy Miswan Hanis, Raja Nur Amanina Raja Azizan Suhaimi, Mow Fei Kim, Nurazizah Osman, Chua Yong Chi, Ain Izyani Zulkifli, Wan Mardhiah Wan Ayub, Tan Eng Tian, Florian Picard, Pierre Legentil, Eliot Sadrin, Paul Thong-Vanh, Chen Yaping, Ni Xiaoyan, Li Jiayu, Prissha Shamugam, Thasshwinthiran Gopal, Kirtana Nair Gopalan, Zheng Tianyi, Yu Jingjing, Muhd Alif Asraf Mustaffa Kamal, Mohd Danial Mohd Dariff, Mohamad Imran Badaruzzaman, Muhammad Azhan Syakirin Azmani, Mohd Syatir Shalehuddin, Nurfatin Ainina Ishak, and Muhammad Hafiz Mustavudeen.

REFERENCES

- [1] Sarlis, V., and Tjortjis, C., (2020), Sports analytics Evaluation of basketball players and team performance, *Information Systems*, 93(101562).
- [2] Asserman, E. B., Herzog, M.M., Collins, C. L., Morris, S. N., Marshall, S. W., (2018), Fundamentals of Sports Analytics, *Clinics in Sports Medicine*, 37(3), pp. 387-400
- [3] Cummins, C., White, R., Jones, B., (2019), Application of sports technology and analytics to performance and injury, *Journal of Science and Medicine in Sport*, 22(2).
- [4] Miljković, Dragan & Gajić, Ljubisa & Kovacevic, Aleksandar & Konjovic, Zora. (2010). The use of data mining for basketball matches outcomes prediction. 8th International Symposium on Intelligent Systems and Informatics (SISY), pp. 309-312.
- [5] Manovska, Lj & Stamatoski, A. & Lacmanovic, Izabela & Pecev, Predrag & Srecković, M.. (2012). Potentials of using data mining in basketball coaching. *International Conference on Applied Internet and Information Technologies (ICAIIT 2012)*, At Zrenjanin, Republic of Serbia, Volume: pp. 353-356.
- [6] Kapadia, K., Abdel-Jaber, H., Fadi Thabtah, F., Hadi, W., (2019), Sport analytics for cricket game results using machine learning: An experimental study, *Applied Computing and Informatics*
- [7] Kolb, D. A. (1984). Experiential learning: Experience as the source of learning and development (Vol. 1). Englewood Cliffs, NJ: Prentice-Hall.