DEPARTMENT OF COMMERCE

MSCBA 711: Minor Project – II

"Sport Analysis of Pro-Kabaddi League using Random Forest Classification Technique "

Report Submitted by

POOJASHREE RAJAGOPALAN (212626011)

VISHAL RAWAT (212626014)

RUTVIK JOSHI (212626025)

In the partial fulfilment for the award of the degree of

MASTER OF SCIENCE (BUSINESS ANALYTICS)

Under the Guidance of

DR. R. VENKATAMUNI REDDY

Professor

Department of Commerce

Manipal Academy of Higher Education

Manipal

DECLARATION

We hereby declare that this report titled "Sport Analysis of Pro-Kabaddi League using Random Forest Classification Technique**"** is being submitted for the partial fulfilment of the requirements for the award of M.Sc. Business Analytics program under the supervision and guidance of Prof. Dr. R. Venkatamuni Reddy of the Department of Commerce, Manipal Academy of Higher Education, Manipal, Karnataka and has no resemblance with any other person's work.

POOJASHREE RAJAGOPALAN (212626011)

VISHAL RAWAT (212626014)

RUTVIK JOSHI (212626025)

Department of Commerce

Manipal Academy of Higher Education (MAHE)

Manipal, Karnataka – 576 104

CERTIFICATE

This is to certify that project report titled "Sport Analysis of Pro-Kabaddi League using Random Forest Classification Technique" is submitted to Department of Commerce, Manipal Academy of Higher Education (MAHE), Manipal, towards partial fulfillment of the requirements for the award of M.Sc. Business Analytics degree.

Poojashree Rajagopalan (212626011), Vishal Rawat (212626014) and Rutvik Joshi (212626025) has worked under my supervision and guidance. No part of this report has been earlier submitted for the award of any Degree, Diploma, Fellowship or any other similar title or prizes. The work has not been published in any journal or magazine.

Dr. R. Venkatamuni Reddy

Professor

Department of Commerce

Manipal Academy of Higher Education (MAHE)

Manipal, Karnataka – 576 104

Date: 15/11/2022

ACKNOWLEDGEMENT

We would like to express our deepest gratitude to our guide Dr R. Venkatamuni Reddy, Professor, Department of Commerce, Manipal Academy of Higher Education for providing his valuable guidance and suggestions throughout this project.

With Warm Regards,

POOJASHREE RAJAGOPALAN (212626011)

VISHAL RAWAT (212626014)

RUTVIK JOSHI (212626025)

M.Sc. Business Analytics

Department of Commerce

Manipal Academy of Higher Education (MAHE)

Manipal, Karnataka – 576 104

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

In recent times, Pro Kabaddi League is one of the most talked-about sport in India. Spanning over 9 seasons, this League brings a Midas touch to the traditional Kabaddi game with minor changes in its rules. Kabaddi Data can be easily interpreted with the help of Data Visualisation aiding to discover patterns and trends. Data analysis involves looking at data, then cleaning, converting, and modelling it to highlight important information that can aid in decision-making. Hence this study seeks to analyse the performances of the teams over the seasons and understand the players' Retention by building a model using Random Forest Classification. The study analyses the key metrices like Raid Points, Tackle Points, Super Tackles, Super Raids etc. The Model achieves about 80% on the training set and 67% on the test set with a cross validation score of 60.81%. The results thus achieved infer that the model performed fairly well and it can be used for further study on the areas of improvement of Kabaddi players.

Keywords: **Pro Kabaddi League, Kabaddi, Guidelines, Data Visualisation, Patterns, Trends, Performance, Retention, Random Forest Classification**

INTRODUCTION

In India, the diversity of Sports is expanding. Professional sports are becoming more popular in India, thanks to the accomplishments of various leagues. Decision- making, Determination, Emotional Maturity, Discipline and Focus are all crucial to Sports, taking all things into consideration. Kabaddi is a team sport originated from and very popular in India. Kabaddi's uniqueness is that a whole group shields against a sole player from the opposite team. The International Kabaddi Federation, which is composed of more than 30 national associations, is the organisation in charge of overseeing the game's regulations all over the world.

It is a contact sport that is played all over the world, among two teams of 7 players each. The general gameplay is such that there are 2 - 20 minutes of play with a break of 5 minutes each during which the teams will exchange their sides. A single player (the Raider) enters the arena of the opposition to touch as many players (Defenders) as possible and return to their arena. This should be done without being tackled by the players of the opposition team.

The points are calculated such that if a defender is tagged, then they score a point. An additional bonus point is rewarded if the Raider ventures past the Bonus line of the rival teams' side. But if the Raider is stopped and tackled effectively, then the rival point gains a point. The tagged players are all considered to be out of the game but can be revived and restored if the team tags or tackles the opposite team players. A Super Raid is set to happen if the raider gains three or more points during the play.

The Pro Kabaddi League commenced in the year 2014 on air and it began entirely based on the Indian Premier League of the sport Cricket. Mashal Sports launched the Pro Kabaddi League to advance the sport of Kabaddi by bringing together athletes from around the world to compete in league matches. It included 8 franchises representing Indian towns. The second season had a bi-annual change. The guidelines of the game were slightly modified according to the taste and preferences of the national audience. It gained national recognition with a viewership of 435 million viewers across the country. The league featured a present-day competitive Kabaddi match with slight changes in the principles to viewers all over the globe.

OBJECTIVES OF THE STUDY

- To analyse the performance of the top 3 teams of the Pro-Kabaddi League.
- To build predictive modelling using Random Forest Classification Technique for the Pro-Kabaddi Leagues' players of the top 3 teams.

LIMITATIONS OF THE STUDY

1. The study's primary flaw is that it only included the top 3 teams which had consistent success over the course of 8 seasons indicating that the sample size of the study is quite low.

2. The Pro Kabaddi League's official website was used to create the dataset from scratch. The lack of data for season 5 is another drawback of the research.

SCOPE OF THE STUDY

The scope of this study is limited to only 3 teams of Pro Kabaddi League i.e., Bengaluru Bulls, Patna Pirates and U Mumba. Seasons 1-8 except Season 5 were taken into consideration for the analysis. This extensive study will further help researchers in figuring out the performance of other teams over the seasons of the Pro Kabaddi League. Further research can be done in analysing the performance of the players individually and also to understand the various metrics of areas of improvement.

RESEARCH METHODOLOGY

- **Research Design:** Descriptive Research Design was adopted to conduct this study. This approach was used to describe how the players and the top 3 teams performed over the seasons of the Pro Kabaddi League.
- **Source of Data:** This research is based on Secondary Data sources.
- **Method of Data Collection:** Dataset was built entirely from scratch using the data found on the official Pro Kabaddi League website.
- **Tools:** The tools used to conduct this study were Python and Power BI.
- **Method of Testing:** Random Forest Classifier

    - **Random Forest Classifier –** Random Forest is a classification technique which consists of many decision trees on different subsets of the dataset and averages the results to increase the dataset's predicted accuracy. It is an ensemble learning technique that builds many decision trees during training to do classification, regression, and other such operations meaning, it can handle binary, categorical, and numerical features. It uses bagging technique to prevent overfitting of our model. Unlike linear regression models, Random Forest is not affected by multicollinearity because it makes use of bootstrap sampling and feature sampling, i.e., row and feature sampling. Also, there is less effect of outliers on the prediction of Random Forest.

REVIEW OF LITERATURE

According to **Parmar (2018)**, the study titled “KABADDI: From an intuitive to a quantitative approach for analysis, predictions and strategy”: The study directed at developing a model that could forecast the game's outcome while confirming the theories and tactics that have already been proven. Each team ascertained on an average of 1 all-out every game, 2 Super Tackles per 3 games resulting in all variables positively skewed and platykurtic. It was noticed that among the teams which took part in the 2016 Kabaddi World Cup, Kenya had the greatest defence, while India had the best offence. The best team, according to the average Total points, was found to be team India, followed by Bangladesh and South Korea. Since the winning team and the overall total number of points scored were seen to be unaffected by the outcome of the coin toss, it showed no bearing on how the game turned out. Hypothesis testing was conducted by the researcher to comprehend team-wise performance applying general thumb rules of the game. It was revealed that Argentina, Bangladesh, India, Iran, Korea, Poland, and Thailand followed the thumb rule “attack is better than the defence”. It was interestingly noted that Australia did not follow any thumb guidelines, while India did. Countries like Australia, England, Japan, Kenya, and the USA did not adhere to the thumb rule. The study's key conclusions were around predictive models as well as hypotheses and visualisation concluding that the victorious team had:

- More Raid points than Defence points.
- More Defence points than the bonus points.
- More Defence points than the all-out points.
- More all-out points than bonus points.

According to **Gurule, Muley (2019)**, their study titled “Analysis of Success Ratio of Attacking and Defensive Skills in Pro Kabaddi League”: Each team in the Pro Kabaddi League (2016 season) had 25 players, so 200 players from the eight competing teams that played in 60 matches during the season served as the study's sample. It was discovered that the raiders attempted 2432 raids in the entire season out of which, 49% were successful. Players attempted to tag a player from the rival team of which 59% were successful. Thus, the study concluded that teams used their attacking and defensive skills deliberately and successfully to win the games. However, teams with stronger defenses had a better chance of winning the game.

According to **Nimbalkar, Mundhe and Dr. VijayKumar (2018)**, their study titled “Data Analysis and Visualization on Pro-Kabaddi League”: The study focused on analyzing and visualizing the Pro Kabaddi League data for the first 4 seasons with supervised learning and build a deterministic model using K-Nearest Neighbors algorithm and Logistic Regression classification technique. When the techniques were applied, the technique Logistic Regression showed less effectiveness than K-Nearest Neighbors and K-Means algorithms. K-Means was found to be the most accurate and effective algorithm overall, with an accuracy rate of 81.81%. According to the visualizations, the Player with the most Raid points was Rahul Chaudhari with 524 points and Patna Pirates players had consistently outperformed the rest of the teams. Super Tackles was realized to be the Key metrics of the game as most of the points were gained in this metric.

According to **Bagchi, Raizada, Mhatre and Augustine (2018)**, the study titled "Forecasting the winner of pro kabaddi league matches" was undertaken to develop a model to make predictions of the outcomes of the Pro Kabaddi League matches. They found that logistic regression model managed to predict better outcomes than other classification techniques. Out of four predictor variables they included three predictor variables in the model i.e. All Out Points, Tackle Points and Raid Points of first half data. The predictor variables according to the statistical significance were numerically weighted for better predictions of match outcomes. The study suggests that the model managed to predict 68.8% of the match outcomes correctly.

According to **Thabtah, Zhang and Abdelhamid (2019)**, the study titled "NBA Game Result Prediction Using Feature Analysis and Machine Learning" was undertaken for development of a model to predict the NBA match outcomes by developing an intelligent framework based on machine learning techniques and feature selection. The study suggests that the defensive rebounds, three-point percentage, free throws made, field goal percentage and total rebounds were found to be influential factors affecting the match outcomes of the NBA. Though after feature selection the feature vectors were reduced through feature selection method which had a 2-4% increase in the accuracy score of predictions by the model. However, the accuracy of the predictions made by the model was not as expected but the feature provides some significant insights and may help the coaches to improve the team's capabilities.

DATA ANALYSIS AND INTERPRETATION

A. TOP 3 TEAMS' PERFORMANCE OVER THE SEASONS

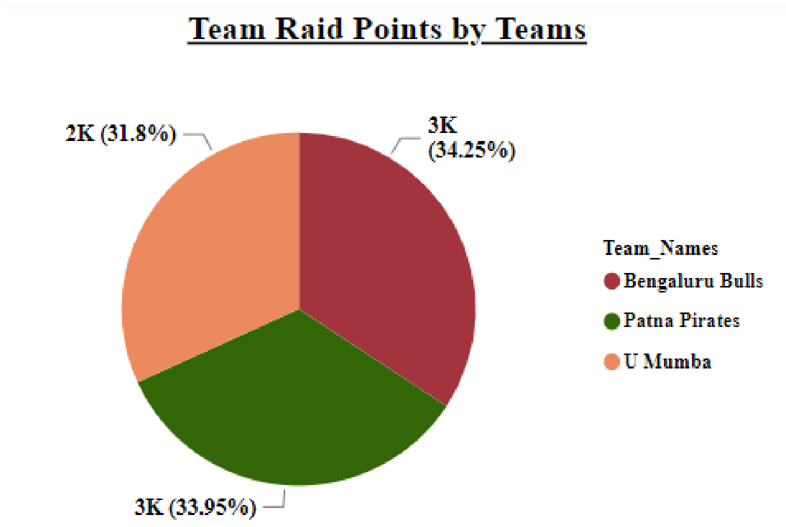

Figure 1: Team Raid Points by Teams

The above Pie Chart illustrates the Team Raid Points scored by the three teams for all the seasons. Bengaluru Bulls are found to have an edge of 0.30 % over the second highest team Patna Pirates. U Mumba is found to have scored only 31.8% of the Total Team Raid Points scored by the top 3 teams.

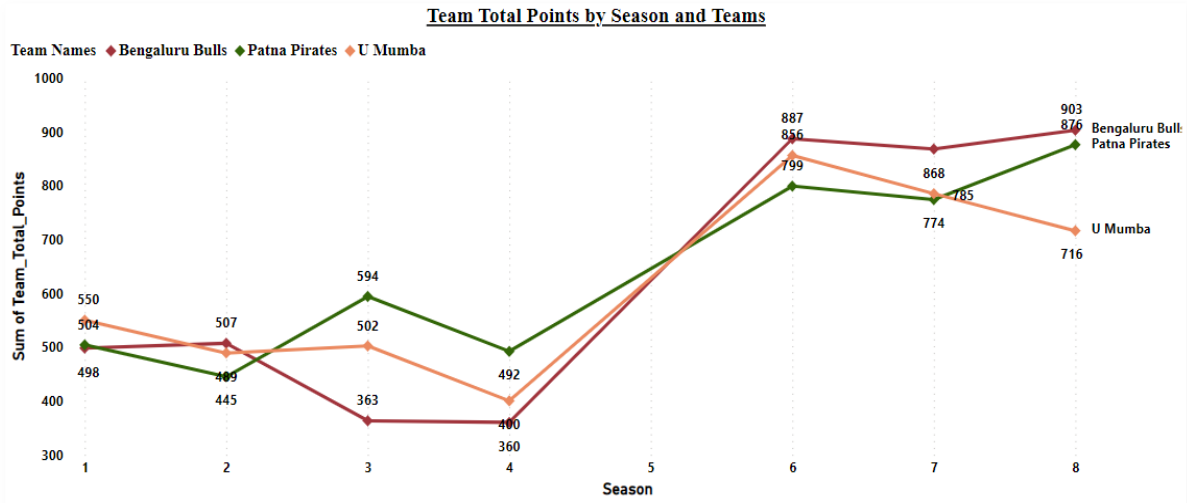

Figure 2: Team Total Points by Season and Teams

The above Line Chart is of the Team Total Points over the seasons. The Total Points for all the 3 teams have increased over the seasons. From the visualization, we find that in Season 1, the Team with highest Team Total Points is U Mumba with 550 points and the other two teams Bengaluru Bulls and Patna pirates had 498 and 504 respectively. During Season 4, it is shown to have a steep decrease in all the teams' performance and the performances increase during Season 6 and then on. In Season 8, Bengaluru Bulls was found to be the best performing team with 903 Team Total Points as compared to the other two teams which had 876 and 716 Team Total Points respectively.

SEASON	BENGALURU BULLS	PATNA PIRATES	U MUMBA	TOTAL DOD POINTS
1	39	42	39	**120**
2	81	63	79	**223**
3	68	71	59	**198**
4	61	72	42	**175**
6	73	54	89	**216**
7	54	78	79	**211**
8	38	90	53	**181**
TOTAL	**414**	**470**	**440**	**1324**

Table 1: Total Do-Or-Die Points of the top 3 teams over the seasons

A Do-Or-Die Point refers to the point scored during a third raid when two consecutive raids are not scored by the team. If this DOD point is not scored, then the team will be deemed out and the rival team will receive 1 point. The above table summarizes the Total Do-Or-Die Points scored by the top 3 teams season by season. We can determine that Patna Pirates have scored the most DOD points over the seasons and the most DOD points were scored in the Season 2 by the top 3 teams. It is also found that the most DOD points were scored by Patna Pirates in Season 8 (90 Points) followed by U Mumba in Season 7 (89 Points) and Bengaluru Bulls in Season 2 (81 Points).

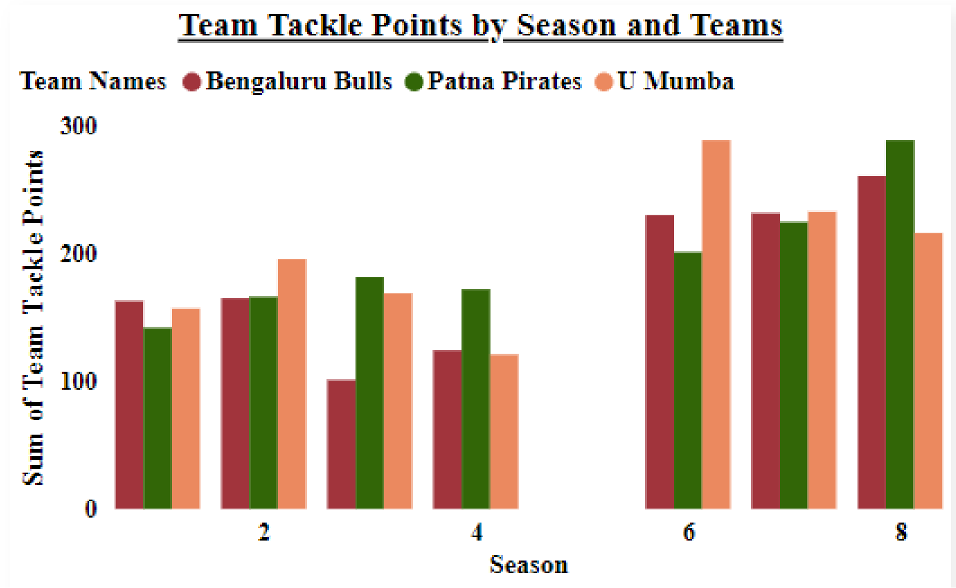

Figure 3: Tackle Points of the top 3 teams over the seasons

The above bar visualization represents relationships between the sum of Team Tackle Points by Season and Teams. It is evident that over the seasons the Team Tackle Points have increased for all the three teams. Bengaluru bulls and U Mumba had almost the same tackle points in Season 2 and Season 4. In Season 6 U Mumba had much more tackle points than the other two teams. In Season 3 Bengaluru Bulls had much lower tackle points than the other two teams. In Season 8, Patna Pirates are found to have more tackle points than the rest.

B. PREDICTIVE MODELING OF PLAYERS

	SUCCESSFUL RAIDS	RAID POINTS	SUCCESSFUL TACKLES	SUPER TACKLES	TOTAL POINTS	TACKLE POINTS	SUPER RAIDS	DOD RAID POINTS
COUNT	296	296	296	296	296	296	296	296
MEAN	20.628	26.381	13.057	1.104	38.922	13.780	0.706	4.439
STD	39.713	50.428	16.313	1.784	51.410	17.559	1.853	7.829
MIN	1	1	1	0	1	1	0	0
25%	2	2	2	0	5	2	0	0
50%	3	5	5	0	21.5	5	0	1
75%	17.25	24	19	2	52	20	0	6
MAX	262	346	86	9	360	89	14	41

Table 2: Descriptive Statistics

The above table displays 8 variables with 296 records of each collection over seasons. The Mean Successful Raids scored by the players over the seasons is 21 and the Median is 3 while the Mean Raid Points after Successful Raids are 26.38 and median is 5. The data is skewed to the right i.e., Positively Skewed, which explains why the Mean is greater than Median.

The Mean Successful Tackles over the seasons for all players is 13 and Median is 5 while the Mean Super Tackles is 1 and Median is 0. The Mean Tackle Points over the seasons are 13.78 and Median is 5.

The Standard Deviation is high for all relative to the Mean i.e., the values are spread out over a wider range. There are various categories of players i.e., Raiders, Defenders and All-rounders in a team and this might be one of the reasons for a higher standard deviation.

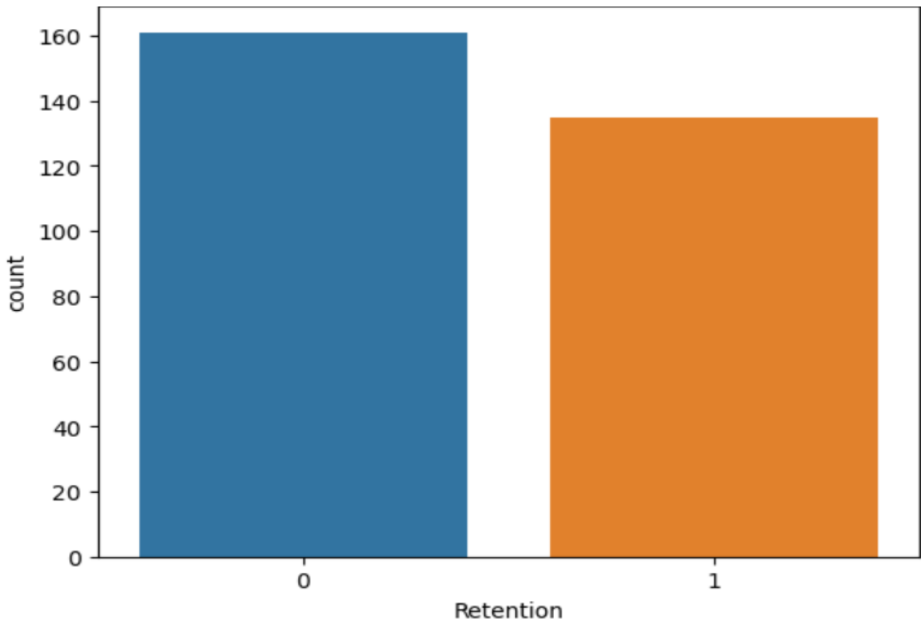

Figure 4: Retained vs Non-Retained Players

The above figure shows a count plot which is used to count records or counts for each value. Here, 1 refers to the Retained Players and 0 refers to the Non- Retained players in the Retention variable, the count being 135 and 161 respectively. This can also mean that 45.6 ~ 46% of the players were retained by the top 3 teams whereas 54.3 ~ 54% were not retained by the teams and they were either chosen by other teams of the league or they didn’t play in the season at all.

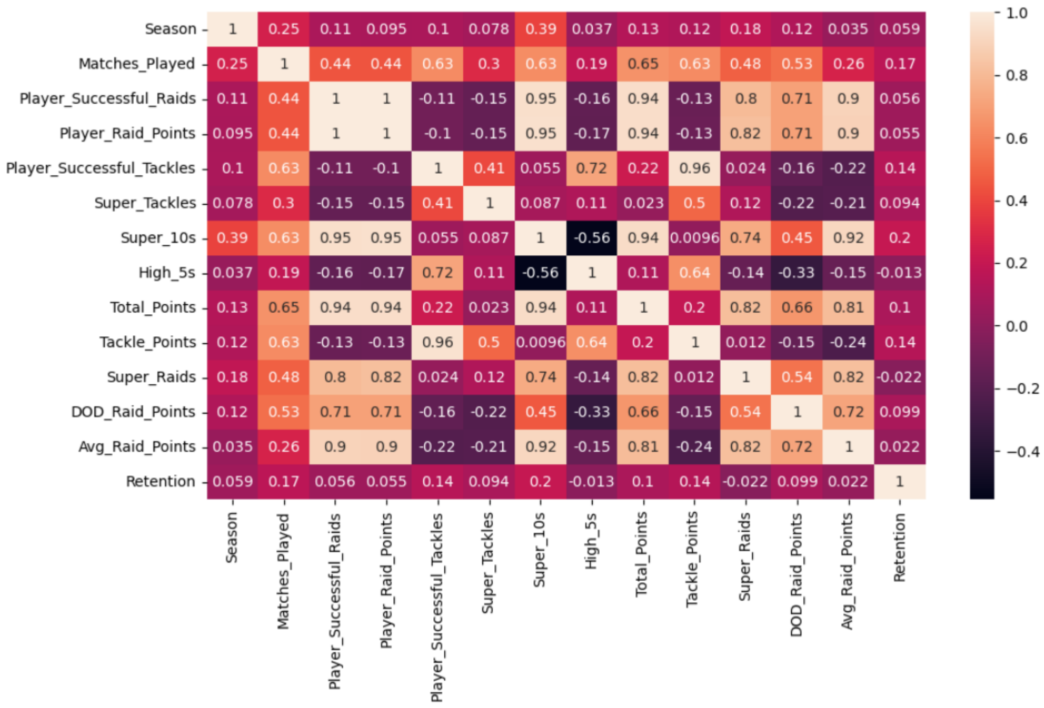

Figure 5: Correlation between all the variables

Heatmap shows the Correlation between the variables of the dataset. The correlation lies between -1 and +1. The value 0 means there is no correlation between the variables or there is no linear trend and values less than 0.30 mean that the variables have no possibility of correlation. The values closer to +1 tells us that the variables are positively correlated to each other. It means that if one increases, so does the other. The values closer to -1 tells us that the variables are negatively correlated to each other. It means that if one decreases, the other increases. Values from 0.75 to 0.90 tells us that there is sufficient degree of correlation between the variables, both negative and positive. Variables having values from 0.60 to 0.75 mean that they are moderately correlated, whereas variables from 0.30 to 0.60 tell that there is a slight possibility of correlation between the variables both positive and negative.

As observed from the above heatmap, Player Successful Raids has a strong positive correlation of 0.95 with Super 10s meaning that a Super 10 is achieved when a Raider scores 10 or more raid points in a single game. Therefore, the more the Successful Raids, the more the Super 10s. The High 5s is also found to have a positive correlation of 0.72 with Player Successful Tackles meaning that a High 5 is achieved when a player scores 5 or more tackle points in a game. Therefore, the more the Successful Tackles, the more the High 5s.

In case of Retention Player Successful Raids, Player Raid Points, Super Tackles, High 5s, Super Raids, DOD Raid Points have negligible correlation while Matches Played, Player Successful Tackles, Super 10s, Total Points, Tackle Points have weak positive correlation i.e., these variables increase the chances of Retention of a player by the team.

```
# Accuracy on Training data
print("Training Accuracy is: ", rfc.score(X_train, y_train))

# Accuracy on Test data
print("Testing Accuracy is: ", rfc.score(X_test, y_test))
```

```
Training Accuracy is:  0.8019323671497585
Testing Accuracy is:  0.6741573033707865
```

Figure 6: Training and Testing Accuracy Score

Classifier Accuracy is done to evaluate how well and how accurately the Classification Model is performing. It is the ratio of the Correct Predictions to the Total Predictions of the model. From the above figure, we observe that the Training and Testing Accuracy Score is determined to be 0.801 and 0.674 respectively. A score greater than 0.6 is ascertained to be idealistic and realistic. Thus, we can say that our model has performed well as per real-time standards.

ACTUAL VALUES	PREDICTED VALUES
1	1
0	0
0	0
1	1
0	1
0	0
0	1
0	0
1	1
1	1
0	0
0	1
0	0
0	0
0	1
1	0
1	0
1	1
0	0
0	1

Table 3: Actual Vs Model Predicted Values

```
from sklearn.model_selection import cross_val_score

print(np.mean(cross_val_score(rfc, scaled_X, y, cv=3)))
```

```
0.6081220366934653
```

Figure 7: Cross Validation Score of the Model

Cross validation score is a method where instead of using just one train/test split, the cross-validation method provides a more comprehensive knowledge of model performance throughout the entire dataset. For this model, the Cross Validation Score is 0.6081 which is slightly less than but in case asked to report the accuracy of the model even though it is lower we'll prefer the cross-validation metric over non-cross-validation metric because when compared cross-validation score express how the system performed on different train and test splits (80% of the data). The average cross validation accuracy is more consistent than normal accuracy score.

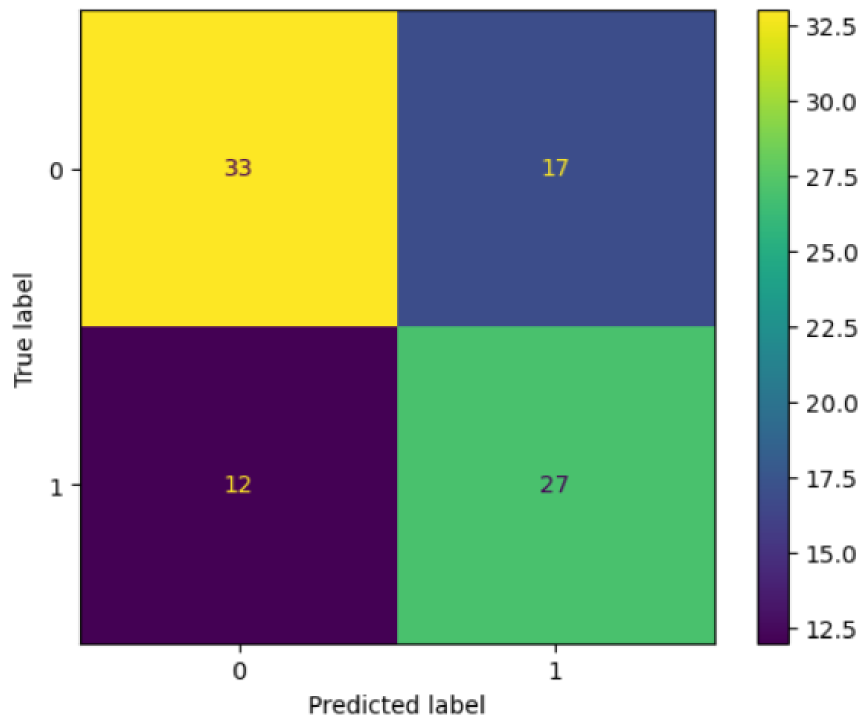

Figure 8: Confusion Matrix of True vs Predicted Values

The above figure demonstrates the Confusion Matrix of the True and Predicted Values of the model. It validates the accuracy of the classifiers' output. The left side of the quadrants are Negatives and the right side of the quadrants are Positives. The order of the matrix from the top left are True Negative, False Positive and the order of the matrix from the bottom left are False Negative and True Positive.

The following predictions made are:

- True positive: 27 records of the players who were retained were predicted correctly by the model.
- False-positive: 17 records predicted wrongly that a player is retained whereas in actual, he was not retained.
- False-negative: 12 records predicted wrongly that a player is not retained whereas in actual, he was retained.

- True Negative: 33 records of the players who were not retained were predicted correctly by the model.

```
Classifier metrics on the test set:
Accuracy: 67.42%
Precision: 61.36%
Recall: 69.23%
F1 Score: 65.06%
```

Figure 9: Classification Metrics of the Test Set

The Precision Score answers: Of all players that labelled as retained, how many actually got retained? Higher precision associate to the low False Positive rate. We have got a precision score of 0.6136 which is reasonably good as it is greater than 0.5.

Recall Score answers: For how many out of all the true positives in the data the model correctly predicted. We got a recall score of 0.6923 which is pretty good as it is greater than 0.5.

The F1 score considers both precision and recall. This is the weighted average of precision and recall meaning it will take both False Negatives and False Positives into account. We have an F1 score of 0.6506 which is a good and significant score as it is also more than 0.5.

FINDINGS AND SUGGESTIONS

From the above tables and figures, we infer that:

1. Bengaluru Bulls' improvement in Total Points may be attributable to their strong performance as Raiders leading to high Raid Point totals. Patna Pirates, who have made significant progress as Raiders, are immediately behind them. U Mumba could be associated to their successful tackling strategy resulting in a rise in Tackle Point totals.

2. Do Or Die, an important factor to win a match, is mastered by Patna Pirates.

3. Retention rate is low (46%) indicating that players from other teams were opted during the auctions over the seasons (54%).

4. The Random Forest Classification Model fared better than average when used to classify players based on retention, with an accuracy rate of 60.81% when comparing retained players to non-retained players.

Based on the above findings, we may recommend that additional study be done in the areas where the top players need to develop so as to improve the Retention Rate. As a result, this may help the teams perform better in the upcoming seasons.

CONCLUSION

The study's goal was to evaluate the top three Pro Kabaddi League teams and their players over the course of 8 seasons. The top 3 teams' performance was examined using important metrics such raid points, tackle points, DOD points, super tackles, and super raids. It was found that the top 3 teams have shown improvements in these areas over the due course of the game. It also helped us to classify and categorise the players into retained and non-retained players using Random Forest Classification Model.

REFERENCES

Parmar, M. (2018). KABADDI: From an intuitive to an quantitative approach for analysis, predictions and strategy, https://doi.org/10.31236/osf.io/b78xj

Gurule, J. Muley, N. (2019). Analysis of Success Ratio of Attacking and Defensive Skills in Pro Kabaddi League; Think India (Quarterly Journal) ISSN: 0971-1260 Vol-22, Special Issue-13

Nimbalkar, V. Mundhe. D, VijayKumar. (2018). Data Analysis and Visualization on Pro-Kabaddi League IJCRT | Volume 6, Issue 2 April 2018 | ISSN: 2320-2882

Bagchi, A. Raizada, S. Mhatre, A. & Augustine, A. (2018). Forecasting the winner of pro kabaddi league matches. Indian Journal of Physical Education, Sports Medicine & Exercise Science, 18(2), 177-177.

Thabtah, F. Zhang, L. & Abdelhamid, N. (2019). NBA game result prediction using feature analysis and machine learning. Annals of Data Science, 6(1), 103-116.

https://en.wikipedia.org/wiki/Kabaddi

https://www.prokabaddi.com/

Kaan Koseler & Matthew Stephan (2017) Machine Learning Applications in Baseball: A Systematic Literature Review, Applied Artificial Intelligence, 31:9-10, 745-763, DOI: 10.1080/08839514.2018.1442991

Sanjit Sardar and Ajay Kumar Pandey. 2016. An estimation of kabaddi performance on the basis of selected physical fitness components. Indian Journal of Physical Education, Sports and Applied Sciences 6, 4 (Oct 2016), 27–35. http://sportscientistsviews.com/wp-content/uploads/2016/12/5.pdf

https://blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures/

https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62

https://www.jmp.com/en_in/statistics-knowledge-portal/what-is-correlation/correlation-vs-causation.html#:~:text=It's%20possible%20to%20find%20a,fact%2C%20such%20correlations%20are%20common

Thomas Davenport. 2014. Analytics in sports: The new science of winning. International Institute for Analytics 2 (2014), 1–28.

S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas. 2006. Machine learning: a review of classification and combining techniques. Artificial Intelligence Review 26, 3 (01 Nov 2006), 159–190. https://doi.org/10.1007/s10462-007-9052-3

Peter Dizikes. 2013. Sports analytics: a real game-changer. (Mar 2013). http://news.mit.edu/2013/sloan-sports-analytics-conference-2013-0304

Isabelle Guyon and André Elisseeff. 2003. An introduction to variable and feature selection. Journal of machine learning research 3 (Mar 2003), 1157–1182.

Ms Harpreet Kaur and Lalit Sharma. 2015. Athletic Intelligence and Different Nature of Sports: an Analysis. International Journal of Movement Education and Sports Sciences 3, 1 (Jan 2015).

ANNEXURE

212626011-014-025-MSCBA_SemIII_MinorProjectII

ORIGINALITY REPORT

4%	4%	1%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	link.springer.com Internet Source	1%
2	Submitted to University of Arizona Student Paper	1%
3	medium.com Internet Source	<1%
4	www.ijraset.com Internet Source	<1%
5	indianexpress.com Internet Source	<1%
6	www.osmania.ac.in Internet Source	<1%
7	www.jagranjosh.com Internet Source	<1%
8	ijern.com Internet Source	<1%
9	towardsdatascience.com Internet Source	<1%
10	Submitted to Liverpool John Moores University Student Paper	<1%

Exclude quotes	On	Exclude matches	< 7 words
Exclude bibliography	On		