STM JOURNALS

Recent Trends in
Sports

Volume 2, Issue 2, 2025
July–December
DOI (Journal): 10.37591/RTS

https://journals.stmjournals.com/rts

Research RTS

Evaluating the Performance of Test Cricket Players Using Principal Component Analysis and the Weighted Average Method

Arjun Prakash Mane[1,]*, Jaydip Yuvaraj Patil[2]

Abstract

This study evaluates cricket player performance using Principal Component Analysis (PCA) and a weighted average approach. In order to achieve this, we analyzed detailed batting and bowling datasets from the International Cricket Council (ICC) to calculate player performance based on various performance indicators. The datasets included comprehensive statistics from multiple matches and tournaments, allowing for an in-depth evaluation of players' skills and contributions. PCA ranked players according to their participation level, with the first three principal components accounting for nearly 90% of the data's variability, which demonstrates the robustness of the approach in capturing the majority of information inherent in the data. Our analysis reveals Sachin Tendulkar and Muthaiya Muralidharan as top-ranked batsmen and bowlers, respectively, based on the combined assessment of their performance metrics. This study provides a comprehensive evaluation of cricket player performance, highlighting the effectiveness of both PCA and weighted average approaches in producing accurate and meaningful rankings. The findings clearly demonstrate the applicability of this methodology in selecting the best cricket squad, while also offering detailed insights into the strengths and weaknesses of top players. Overall, this research makes a significant contribution to the field of sports analytics by providing a robust and reliable framework for evaluating cricket player performance. The study further informs decision-making processes in team selection and strategy development, supporting coaches, analysts, and managers in making data-driven choices for optimizing team performance and achieving competitive success.

Keywords: PCA, Weighted average method, heat map, scree plot, ranking

INTRODUCTION

Test cricket is the most prestigious and longest format of the game, played by ICC Full Members. Each team plays two innings with an infinite number of overs, making it a comprehensive assessment of a team's stamina and ability. The increasing number of nations playing Test matches is driven by revenue growth, with cricket boards seeking to maximize income. As the most popular sport worldwide, cricket requires each player to perform well for the team to succeed.

Assessing individual player performance is crucial but challenging due to the interdependence of evaluation factors. Test cricket is regarded as the pinnacle of the game, with matches lasting up to five days or longer. The physical and psychological toll of these prolonged competitions is significant, and teams must demonstrate exceptional skill and endurance to succeed. Previous studies have employed various methods to evaluate player performance, including Principal

***Author for Correspondence**
Arjun Prakash Mane
E-mail: manearjun184@gmail.com

[1]Biostatistician, Department of Applied Sciences or Allied Health Sciences, Manav Rachna International Institute of Research and Studies, Aravali Hills, Sector 43, Faridabad, Haryana, India
[2]Assistant Professor, SNBP College of Arts, Commerce, Science and Management of Studies, Morwadi, Pimpri, Pune, Maharashtra, India

Received Date: June 11, 2025
Accepted Date: September 10, 2025
Published Date: October 25, 2025

Citation: Arjun Prakash Mane, Jaydip Yuvaraj Patil. Evaluating the Performance of Test Cricket Players Using Principal Component Analysis and the Weighted Average Method. Recent Trends in Sports. 2025; 2(2): 20–30p.

Component Analysis (PCA). Researchers have used PCA to assess batting performance, construct integrated models, and measure captaincy impact on match outcomes. PCA has also been applied to Big Data Regression models to address multicollinearity issues [1-5].

This study aims to evaluate cricket player performance using PCA and a weighted average approach. By analyzing key performance indicators, we can gain insights into the strengths and weaknesses of top cricket players and inform decision-making in team selection and strategy development. The findings of this study can contribute to the field of sports analytics and provide a robust framework for evaluating cricket player performance.

The complexity of cricket performance analysis necessitates the use of advanced statistical techniques like PCA. By applying PCA to cricket data, we can identify the most important factors contributing to player performance and develop a comprehensive evaluation framework. This study will contribute to the growing body of research on cricket analytics and provide valuable insights for coaches, players, and team managers [8-9].

METHODOLOGY

PCA

Python 3.11.4 was used for all statistical analyses, with a significance level of 5%. PCA is a multivariate analytic method that is undoubtedly the most well-established and ancient. It was first presented by Pearson in 1901, and Hoteling developed it independently in 1933 [9, 10]. According to research, the main goal of principal component analysis is to reduce the dimensionality of a data set with several associated variables while preserving as much of the data set's variance as feasible [10, 11]. The principal components, a fresh set of uncorrelated variables arranged so that the first few retain the majority of the variance present in all of the original variables, are used to achieve this reduction.

The process outlined below is used to perform the PCA:
Step 1: Standardization
Step 2: Correlation Matrix Calculation
Step 4: Feature Selection Step 3: Eigen values and eigen vector computation
Step 5: Prioritization

Weighted Average Method

A weighted average is one in which a weight is provided to each quantity that needs to be averaged. The average relative relevance of each quantity is determined by these weightings [12, 13].

This can be written as;

Weighted Average = $\frac{\sum_{i=1}^{n} WiXi}{\sum_{i=1}^{n} Wi}$

Where;
n = Number of players
w_i = Weights applied to x values
x_i = Model values of corresponding ranks

Standardization

In order to achieve an even distribution around a mean value of zero and a standard deviation of one, our features are translated and scaled throughout the standardization process. We want to standardize our data so that comparing the covariance for each set of variables will be easy. Wider number range features will have higher covariance if we don't. While not strictly necessary, this part can be quite helpful and is required for some models [6, 14].

$Z = \frac{\text{Value} - \text{Mean}}{\text{S.D}}$

Once the standardization is done, all the variables will be transformed to the same scale.

Data Collection

We use the statistics of players who have scored at least 2000 runs from stats.espncricinfo.com for our batting analysis. As we need to use the whole dataset for bowling analysis, we have used the data of bowlers who have at least 200 wickets. The information spans December 1877 to 2019.

Batting Attributes

In this analysis, we consider the variables such as Total runs, Total innings, Average, etc. A batsman's performance in ICC Test matches is evaluated through several key metrics. These include their name, career duration, total runs scored, match appearances, top score, batting average (runs per innings, adjusted for not outs), innings played, not out instances, centuries, fifties, and ducks (dismissals without scoring). Together, these statistics offer a thorough assessment of a player's abilities and achievements in the longest format of the game.

Bowling Attributes

In this analysis, we consider the variables such as Total Wickets, Economy, and Average, etc. The effectiveness of a bowler in ICC Test matches is evaluated through various performance indicators. These metrics encompass the bowler's identity, duration of their career, total wickets claimed, number of matches participated in, bowling average, economy rate, strike rate, total deliveries bowled, frequency of claiming five or more wickets, instances of securing 10 or more wickets, number of innings bowled in, and total runs surrendered. Collectively, these factors offer a thorough insight into a bowler's proficiency and accomplishments in the game [15].

RESULTS

Results of the batter's performance

Correlation Matrix

Actually, the initial step in PCA is creating the correlation matrix. In this phase, we must create a 9 x 9 matrix with each row denoting a feature related to batting performance and each column denoting a similar feature. For comprehension, we must use the visualization tool Heat Map. A heat map illustrates the relationship between each pair of features by using color. As a result, Das et al. stated that dark color blocks show strong positive relationship between the attribute [16].

Figure 1 shows the interrelationship between each pair of features is depicted by a heat map. Heat mapping demonstrates the strong positive correlation between each pair of features. Additionally, the variables (Runs and Fifties), (Runs and Hundreds), (Average and Hundreds), (Inn and Fifties), and (NO and Matches) indicate a strong positive relationship, while the variable (Average and number Ducks) shows a negative relationship between them.

Role of Eigenvalues in PCA

The total amount of variation that can be accounted for by each principal component is represented by the eigenvalues. In 2018, Shah et al illustrated that the eigenvalues account for the data's variance along the new features [4, 5, 17].

Table 1. Variation explained by each PCA

Eigen Values	Variation (%)	Cumulative Variation
5.645180379	62.72	62.72
1.862320351	20.69	83.42
0.635644413	7.06	90.47
0.374074072	4.16	94.64
0.271413252	3.02	97.65
0.147292032	1.64	99.29
0.051310337	0.57	99.86
0.00935369	0.1	99.96
0.003411474	0.04	100

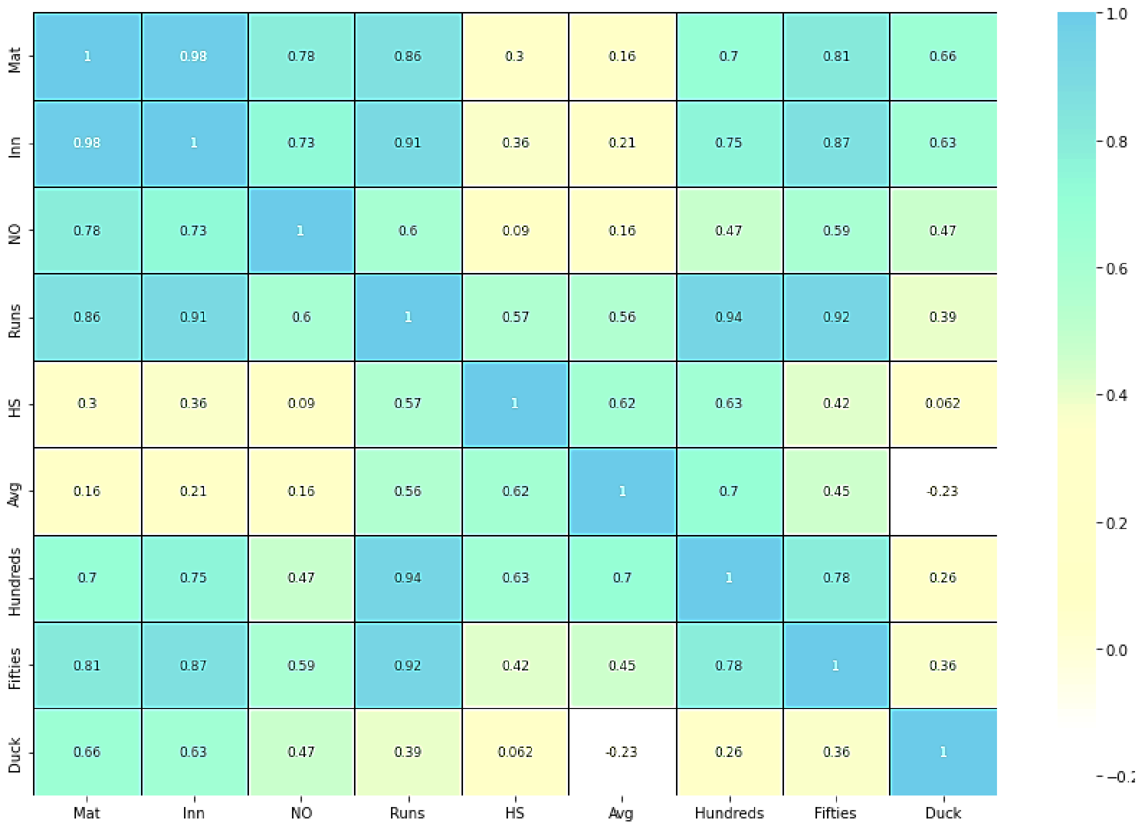

Figure 1. Heat map represents the relationship within an each pair of feature.

Table 1 demonstrates that the first principal component is capable of explaining around 62.72% of the variation in the dataset (PC). Further evidence that PCA is beneficial comes from the fact that the top three principal components of the dataset's variation account for around 90.47% of the variation.

Scree Plot

In a nutshell, a Scree plot is a line plot of the Eigen values vs the number of primary components. In a PCA, the Scree plot is used to decide how many principal components to retain [17, 18].

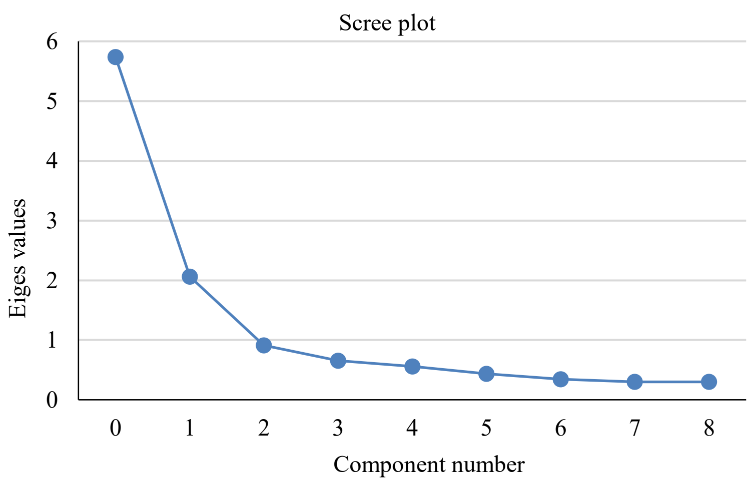

Figure 2. Scree plot represents the variation explained by each PCA

From Figure 2, we can see that, elbow point is at abscissa three, hence we consider the first three principal components in the analysis. Here first three PCs explain near about 90.47% of the variation in the dataset. Then we can build the principal component model on these three components and apply it for the ranking.

Principal Components and Their Models

The first three principal components in analysis can be taken as.

Table 2. First three PCA

Feature	pca1	pca2	pca3
Mat	-0.388	-0.25	0.0195
Inn	-0.399	-0.197	-0.024
NO	-0.302	-0.275	0.4866
Runs	-0.411	0.0975	0.0235
HS	-0.227	0.4402	-0.605
Avg	-0.205	0.5752	0.2173
Hundreds	-0.376	0.2508	-0.003
Fifties	-0.384	0.0263	0.1656
Duck	-0.219	-0.475	-0.567

Table 2 shows that the first three PCs, also known as the first three eigen vectors, are present. In this instance, pca1 stands for the first Eigen vector that corresponds to the first eigen value, and pca1 explains the 62.72% variation within the dataset. When creating the principle components model, all of the elements in the pca1 are taken into account as coefficients of the associated features [7, 19].

L1 = -Mat*(0.388)-Inn*(0.399)-NO*(0.302)-Runs*(0.411)-HS*(0.227)-Avg*(0.205)-Hundreds * (0.376)

-Fifties*(0.384)-Duck*(0.219)

Similar to this, pca2 represents the second eigen vector and explains the 20.69% variation in the dataset. Here, all of the pca2's elements are taken into account as the respective features' coefficients for creating the principle component model [7];

L2 = -Mat*(0.25)-Inn*(0.197)+NO*(0.275)+Runs*(0.0975)+HS*(0.4402)+Avg*(0.5752)+ Hundreds*(0.2508)

+Fifties*(0.0263)-Duck*(0.475)

Similar to this, pca3 represents the second eigen vector that corresponds to the third eigen value and pca3 explains the 7.06% variation within the dataset [7].

L3 = Mat*(0.0195)-Inn*(0.024)+NO*(0.4866)+Runs*(0.0235) -HS*(0.605) +Avg*(0.2173)-Hundreds*(0.003)+Fifties*(0.1656)-Duck*(0.567)

Weighted Average Method

From Table 3 and L1, we can see that the first principal component model gave less accuracy or less variation about the dataset. There is no sufficient variation about dataset to say that our model is efficient for ranking. This drawback can be overcome by taking the required principal components and then creating a combined principal component model by using the weighted average [13]. Here L1 is the first PC model, which explains 62.72% of the variation. This variation can be taken as the weight of Model L1. Also, we say that 20.69% is the weight of model L2 and 7.06% is the weight of model L3.

Principal Weighted Average (PWA) Model = $\frac{L1*62.72 + L2*20.69 + L3*7.06}{90.47}$

This principal weighted average model explains 90.47% variation in the dataset, and it is a very efficient model for the analysis of ranking [7]. Here the elements in principal components are negative, hence the lower value of the PWA Model indicates the stronger performance of batsmen and the higher values of PWA Model indicates weak performance which is shown as below [7, 13].

Table 3. Ranking of the test cricket batsmen's using PWA model

Ranks	PWA	Player Name
1	-6.821325394	SR Tendulkar(INDIA)
2	-5.829400881	JH Kallis(SA)
3	-5.688437415	RT Ponting(AUS)
4	-5.627771514	S Chanderpaul(WI)
5	-5.514141612	SR Waugh(AUS)
6	-5.075007829	R Dravid(INDIA)
7	-4.851128353	AR Border(AUS)
8	-4.435630798	AN Cook(ENG)
9	-4.364988606	DPMD Jayawardene(SL)
10	-3.997302329	KC Sangakkara(SL)

To Rank of Batsmen's by using Principal Weighted Average Model

According to Tables 3 Tendulkar was the top-ranked cricketer among batsmen, and all of his attributes that affected his batting performance were mostly positive. Tendulkar batting average is lower than Kallis, but his other aspects of batting performance are strong. It is evident from this that our PWA model functions after taking into account all the factors that affect batting performance [17].

RESULTS OF THE BOWLER'S PERFORMANCE

Correlation Matrix

Building the correlation matrix is the actual first step of PCA. In this step, we have to create 10X10 matrixes where each row represents features about bowling performance and each column also represents features about bowling performance. As a result, Das et al. stated that dark color blocks show strong positive relationship between the attribute [16].

Figure 4 Heat Map represents the inter relationship between each pair of features. Here the variables (Runs and Innings), (Wickets and Balls) and (Runs and Wickets) shows the strong positive relationship while the (SR and Econ), (Econ and Balls) and (Average and Wickets) shows the negative relationship between them.

Role of Eigen values in PCA

Eigen values represent the total amount of variation can be explained by each principal component. In other words, the Eigen values explain the variance of the data along the new features [4, 5, 17].

Table 4. Variation explained by each PCA

Eigen Values	Variation (%)	Cumulative Variation (%)
5.611157634	56.11	56.11
1.85696321	18.57	74.68
1.508166442	15.08	89.76
0.838455107	8.39	98.15
0.109757342	1.1	99.25
0.05226266	0.52	99.77
0.009828297	0.1	99.87
0.002034635	0.02	99.89

0.00629494	0.06	99.95
0.005079733	0.05	100

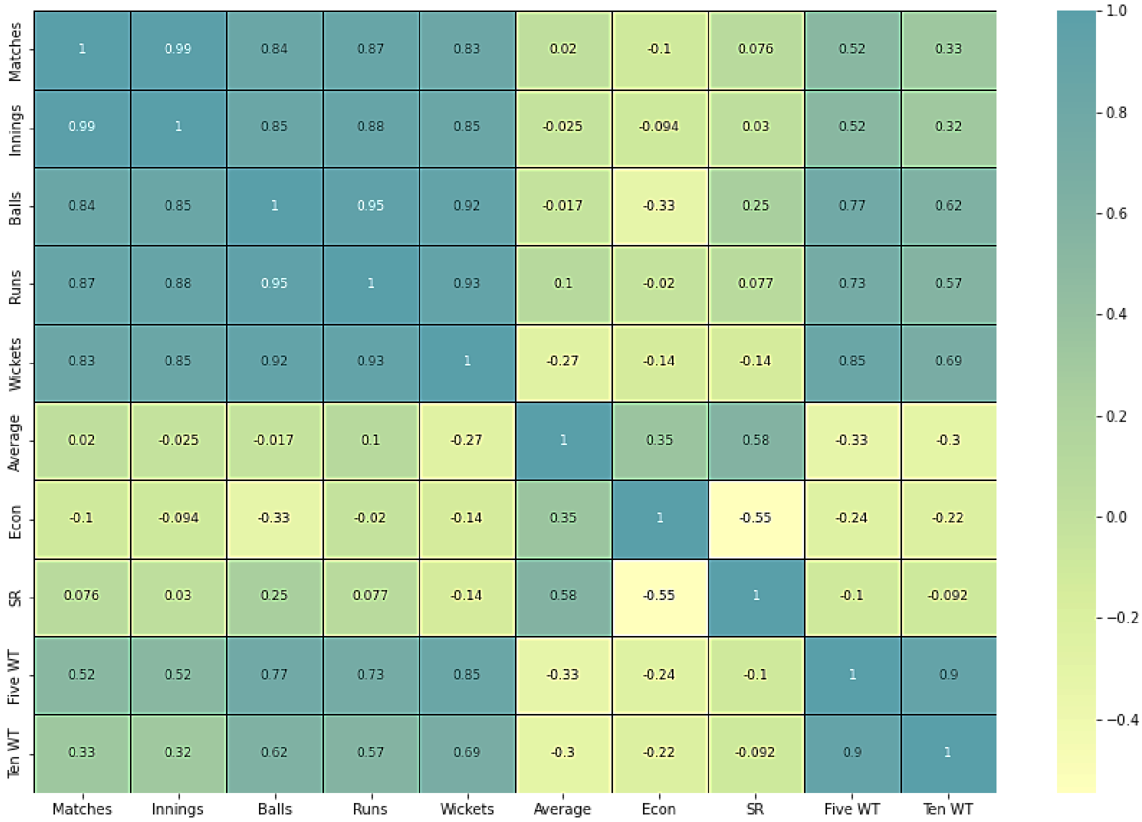

Figure 3. Heat map represents the relationship within an each pair of feature.

Table 4 demonstrates that the first main component may account for roughly 56.11 percent of the dataset's volatility (PC). Additionally, we can see that the top three principal components of the dataset's fluctuation explain around 89.76% of the variation, which makes PCA effective.

Scree Plot

A scree plot is essentially a line plot of the eigenvalues versus the number of PCs. To choose how many principal components to keep in an analysis, use the Scree plot [17, 18].

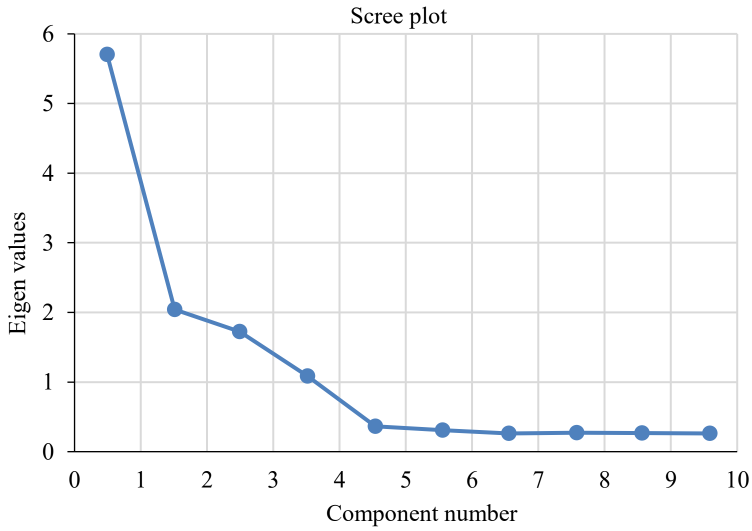

Figure 4. Scree plot represents the variation explained by each PCA.

From Figure 4 we can see that, elbow point is at abscissa three, hence we consider first three principal components in analysis. Here first three PCs explain near about 89.76% variation about the dataset. Then we can build the principal component model on these three components and apply for the ranking.

Principal Components and its Models

These are first three principal components in analysis can be taken as;

Table 5. First three PCA

Feature	pca1	pca2	pca3
Matches	-0.36747182	-0.180218348	0.193259754
Innings	-0.370074693	-0.150962698	0.206589614
Balls	-0.407974709	-0.146423294	-0.074612462
Runs	-0.399145803	-0.132286816	0.175635861
Wickets	-0.412894928	0.104382252	0.07095014
Average	0.067998726	-0.614326562	0.257906869
Econ	0.099595704	0.081026311	0.748392749
SR	-0.013921561	-0.616197791	-0.427655884
Five WT	-0.358592534	0.236430156	-0.153148681
Ten WT	-0.297389778	0.27447231	-0.21347311

The first three PCs, also known as the first three eigenvectors, are shown in **Table 5**. Here, PCA1 explains the dataset's variation of 56.11%. Here, the PCA (pca1) takes into account all of the elements as coefficients of the relevant attributes to produce the principal component model [7, 19].

L1 = -Mat*(0.36747182)-Inn*(0.370074693)-Balls*(0.407974709)-Runs*(0.399145803)-Wickets*(0.412894928)
+Avg*(0.067998726)+Econ*(0.099595704)-SR*(0.013921561)-Five WT*(0.358592534)-Ten WT(0.297389778)

Similar to this, pca2 represents the second eigen vector and pca2 explains the 18.57% variation within the dataset. When creating the main component model, all of the PCA2's constituents are taken into account as the feature's coefficients.

L2 = -Mat*(0.180218348)-Inn*(0.150962698)- Balls*(0.146423294)-Runs*(0.132286816) +Wickets*(0.104382252)
-Avg*(0.614326562)+Econ*(0.081026311)-SR*(0.616197791)+Five WT*(0.236430156)+Ten WT(0.27447231)

Similarly to this, PCA3 explains the 15.08% variation in the dataset [7];
L3 = Mat*(0.193259754)+Inn*(0.206589614)-Balls*(0.074612462)+Runs*(0.175635861)+ Wickets*(0.07095014)
+Avg*(0.257906869)+Econ*(0.748392749)-SR*(0.427655884)-Five WT*(0.153148681)-Ten WT(0.21347311)

Weighted Average Method

From Table 6 and L1, we can see that first principal component model gave less accuracy or less variation in the dataset. There is not sufficient variation in the dataset to say that our model is efficient for ranking. This drawback of the first principal component model we can remove first, taking the required principal components. Secondly, we can create a combined principal components model by using the weighted average method [13]. Here, L1 is the first PC model, which explains 56.11% variation. This variation we can take as a weight of Model L1. Also, we say that 18.57% is the weight of model L2 and 15.08% is the weight of model L3 [7].

Principal Weighted Average (PWA) Model = $\frac{L1*56.11 + L2*18.57 + L3*15.08}{89.76}$

This principal weighted average model explains 89.76% variation in the dataset, and it is a very efficient model for analysis of ranking [7]. Here the elements in the principal components are negative hence the lower value of PWA Model indicates the stronger bowling performance and the higher values of PWA Model indicates weak performance which is shown as below [7, 13].

To Rank the Bowlers by using Principal Weighted Average Model

According to Table 6 the Muralidaran was the top-ranked cricket bowler, and all of his characteristics that affected his bowling performance were generally the best. Additionally, Anil Kumble and Shane Warne are placed second and third, respectively. Additionally, JM Anderson played more matches than Muralidaran, Warne, and Kumble did, yet his performance in other circumstances lagged behind those three players. Here, it is evident that our PWA model is effective since it takes into account all factors that affect bowling performance [17-21].

Table 6. Ranking of the test cricket bowlers using PWA model

Ranks	PWA	Player Name
1	-5.613811573	M Muralidaran
2	-4.314340033	S Warne
3	-3.982412967	A Kumble
4	-3.738597566	JM Anderson
5	-2.536179362	SCJ Broad
6	-2.244386988	CA Walsh
7	-2.198597949	GD McGrath
8	-2.003016424	DL Vettori
9	-1.950615204	N Kapil Dev
10	-1.907724333	Harbhajan Singh

DISCUSSION

This study focuses on Test cricket, a four-inning format with two innings each for batting and bowling. The ICC ranks players using a scale of 0-1000, considering factors like ducks, not outs, and maximum score. We employed Principal Component Analysis (PCA) to analyze player performance data. Initially, the first principal component accounted for around 60% of the variation, which was insufficient. To address this, we used a weighted average approach, considering the first three principal components, which accounted for over 90% of the variation. This allowed us to create a combined principal component model for ranking players. Our analysis utilized the first three principal components, which effectively captured the data's variation. The weighted average technique was then applied to obtain the ranking for the PWA model. This approach enabled us to develop a robust ranking system for cricket players.

CONCLUSIONS

Measuring athletic performance is an exciting task in any sport. The role of the batsmen’s and bowlers in cricket it is directly impacted on the result of the match and eventually the profit of the countries. Unfortunately, there is no measure to quantify the performance of each individual player. In this analysis of the batting and bowling performance of the ICC test cricket, we can conclude that Tendulkar and Muralidaran are the players mainly in the first position as a batsmen and bowler respectively. Also, we can see that Kallis, Ponting and Shane Warne, Anil Kumble are ranked at 2nd and 3rd positions as a batsmen and bowler respectively.

STATEMENTS AND DECLARATIONS

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Author Contributions

Each author had a part in the study's conceptualization. Mr. Jaydip Y. Patil and Mr. Arjun P. Mane handled the material preparation, data collecting, and analysis. All contributors provided feedback on earlier drafts of the manuscript, with Mr. Arjun P. Mane penning the original draft. The completed work was read and approved by all authors.

REFERENCES

1. Yin W, Ye Z, Shah WUH. Indices development for player’s performance evaluation through the Super-SBM approach in each department for all three formats of cricket. Sustainability. 2023;15(4):3201. doi:10.3390/su15043201
2. Gupta K. An integrated batting performance analytics model for women’s cricket using principal component analysis and Gini scores. Decision Analytics J. 2022;4:100109. doi:10.1016/j.dajour.2022.100109
3. Mahmoudi MR, Heydari MH, Qasem SN, Mosavi A, Band SS. Principal component analysis to study the relations between the spread rates of COVID-19 in high risk countries. Alexandria Eng J. 2021;60(1):457–464. doi:10.1016/j.aej.2020.09.013
4. Shah P, Patel M. Ranking the cricket captains using principal component analysis. 2018. doi:10.13140/RG.2.2.33455.38564
5. Manage A, Scariano S. An introductory application of principal components to cricket data. J Stat Educ. 2013;21. doi:10.1080/10691898.2013.11889689
6. Zhang Z, Castelló A. Principal components analysis in clinical studies. Ann Transl Med. 2017;5(17):351. doi:10.21037/atm.2017.07.12
7. Das N, Mukherjee I, Paul G, Priya R. A multiple criteria decision making approach for ranking cricket captains. 2023. doi:10.1109/GCON58516.2023.10183420
8. Jadwani Y, Denholm-Price J, Hunter G. A machine learning-based approach to analyse player performance in T20 cricket internationals. 2023.

9. Yenuga BR. Selection of best players in ODI cricket using ranking based indexing method [Doctoral dissertation]. Dublin: National College of Ireland; 2022.
10. Gløersen Ø, Myklebust H, Hallén J, Federolf P. Technique analysis in elite athletes using principal component analysis. J Sports Sci. 2018;36(2):229–237. doi:10.1080/02640414.2017.1298826
11. León-Mantero C, Casas-Rosal JC, Pedrosa-Jesús C, Maz-Machado A. Measuring attitude towards mathematics using Likert scale surveys: The weighted average. PLoS ONE. 2020;15(10):e0239626. doi:10.1371/journal.pone.0239626
12. Pessanha Santos N. The expansion of data science: Dataset standardization. Standards. 2023;3(4):400–410. doi:10.3390/standards3040028
13. Mangtani H. Futuristic accession of test cricket: A statistical tour. Int J Sci Res Publ. 2021;11(2). doi:10.29322/IJSRP.11.02.2021.p11004
14. Holubova H. A comparative analysis of the principal component method and parallel analysis in working with official statistical data. Stat Transit New Ser. 2023;24:199–212. doi:10.59170/stattrans-2023-011
15. Manage A, Kafle R, Wijekularathna D. Classification of all-rounders in limited over cricket - a machine learning approach. J Sports Anal. 2020;6:1–12. doi:10.3233/JSA-200467
16. Khan S, Ishtiaq M, Bibi K, Amin R, Ahmed A, Chaudhary I. Statistical analysis of cricket leagues using principal component analysis. J Eng Appl Sci. 2021;2(1):30–40. doi:10.33897/fujeas.v2i1.451
17. Hussain A, Yan Q, Ullah I, Ullah N, Qudoos A. Region-based teams ranking in the game of cricket using PageRank algorithm. Int J Comput Appl. 2019;177:10–15. doi:10.5120/ijca2019919458
18. Patil R, Duraphe A, Motarwar P, Suganya G, Premalatha M. Cluster-centric based hybrid approach for cricket sports analytics using machine learning. In: Kottursamy K, Bashir AK, Kose U, Uthra A, editors. Deep sciences for computing and communications. Cham: Springer; 2023. p. 233–246. doi:10.1007/978-3-031-27622-4_21
19. Souza AS, Bezerra MA, Cerqueira UMFM, et al. An introductory review on the application of principal component analysis in the data exploration of the chemical analysis of food samples. Food Sci Biotechnol. 2024;33:1323–1336. doi:10.1007/s10068-023-01509-5
20. Akhtar S, Scarf P, Rasool Z. Rating players in test match cricket. J Oper Res Soc. 2015;66:684–695. doi:10.1057/jors.2014.30
21. Bharadwaj F, Saxena A, Kumar R, Kumar R, Kumar S, Stević Ž. Player performance predictive analysis in cricket using machine learning. Revue d'Intell Artif. 2024;38(2):449–457. doi:10.18280/ria.380208

