# Player's Performance Prediction Model for Cricket

Prof. Rashmi Janbandhu[1], Prajwal Halmare[2], Riteshkumar Gaure[3], Rajat Ranka[4]
[1,2,3,4]*Professor, Dept. of Artificial Intelligence,*
*G.H. Raisoni College of Engineering & Management Nagpur, India*
*doi.org/10.64643/IJIRTV12I6-186374-459*

***Abstract***— **In sports analytics, typical predictive models are often based solely on historical statistics, therefore there is no consideration of a player's current range of biomechanics and technique. This basic problem leads to underdeveloped and inaccurate performance predictions. This study presents a new solution to overcome this issue: a totally new Dual-Perspective Analytics Platform designed for the specific purpose of providing a more complete, whole-data approach to assessing performance, which uses a mixture of longitudinal performance data and technical assessment in real-time.**

**The Statistical Forecasting Module uses a Linear Regression model, meticulously trained on all historical data from 2008 to 2024 from IPL and International, to accurately quantify expected long-term performance. This module provides two analytical outputs: a numerical forecasting estimation of future run-scoring ability (validated 80% to 85% accuracy) and a detailed examination of player output, by team membership, in each of the respective seasons. These two components benefit analysts, providing a specific snapshot of anticipated player performance and how performance variability changes under differing team conditions.**

**Along with the statistical estimates, the Biomechanical Analysis Module utilizes MediaPipe Pose and computer vision approaches to provide an objective evaluation of current player technique from video content. This assessment measures specific kinematic metrics Backlift Quality, Head Stability, and Footwork Quality so that they can be summarized as individual metrics and compared in total as a Technical Score. The combination of long-term statistical prediction with real- time technical assessment leads to a Selection Recommendation that can be scaled against a professional normative standard, defined as 15.00. This platform converts the inherently subjective observations of coaching into systematized, objective variables for planning purposes.**

***Index Terms***— **Cricket Analytics, Performance Prediction, Machine Learning in Cricket, Computer Vision, Pose Estimation**

## I. INTRODUCTION

Professional cricket is a sport that is increasingly relying on data for decision-making processes, where decisions about strategies, talent identification and performance evaluations are being increasingly driven by data analysis. Traditional methods have mostly examined metrics of performance that measure quantitative output batting averages, strike rates and economy rates, for example with large stores of data from historical matches and deliveries. Building on these past data sources can be useful in projecting future player and team performance and of course they can assist in establishing potential and career trajectories, but they will by their nature, always be limited as predictive models based on repetitively using aggregate past data. This reliance creates a notable analytical gap: the models are blind to a player's current physical and technical condition. A player, despite having exceptional historical statistics, can drop in performance simply because of a relatively subtle but important technical error in their biomechanics for example, an angle in the backlift not easily detected, or some issue with the relative stability of the head during the shot. Conversely, a promising young player with low overall statistics may be overlooked despite being in excellent technical form. Current analytical systems therefore are not capturing the situational and real-time context of player form, leading to predictions that are devoid of the context necessary for coaching and selection decisions, which is critical for successful engagement in high-stakes sporting competitions. This situation creates a significant chasm between statistical prediction and the everyday practical contexts that coaches and analysts must work within.

This research addresses this widespread deficiency by presenting the Dual-Perspective Analytics Platform, an innovative integrated AI framework for providing a full-scope and evidence-based evaluation of player performance. The platform shifts away from the current siloed analytical models and strategically merges two complementary analytical and powerful modules: Statistical Forecasting module and the Biomechanical Analysis module. The system is designed to bring together the "what" (statistical potential - long term) with the "how" (technical execution - now) into a cohesive integrated decision-support platform.

The Statistical Forecasting Module, which serves as the first component, utilizes a Linear Regression model based on the gameplay archives, player-performance data, and prior- team performance data from the years 2008-2024. The module primarily offers two analytic capabilities: 1) it offers a numerical projection of a player's expected future runs scored, and 2) it allows for analytic comparisons of a player's particular performance for each team played over different years. The statistical data can be valuable since it provides context around a player's history related to transition and efficiency in varying contexts.

The second component known as the Biomechanical Analysis Module uses an advanced Computer Vision pipeline using MediaPipe Pose. It processes video footage in order to objectively measure and score a player’s current technical form by computing and analyzing 33 key skeletal landmarks. The module determines accurate kinematics, including Backlift quality (captured from elbow angle), Head Stability, and Footwork Quality. These individual metrics are synthesized into a single, comparative Technical_Score aga-
inst a professional standard. XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

The final product of the platform combines the longer-term statistical trend with the immediate technical score to provide an automated Selection Recommendation. This is the main innovation of this work, simplifying and translating multi- faceted data streams into simple, jargon-free recommendations (e.g., "Recommendation Top Candidate for Selection" OR "Recommendation Needs improvement"). The rest of this paper will talk about the approaches, mathematical representations, procedure, and results of the Dual- Perspective Analytics Platform demonstrating its ability to give coaches and analysts a stronger more contextualized, and reliable, player evaluation and decision-making tool.

## II. RELATED WORK

### A. Statistical Performance Forecasting

The initial innovations in cricket analytics were aimed at using previous performance data to estimate outcomes and determine player worth. Kaluarachchi and Varde's work in 2010, with their CricAI tool, showed the application of machine learning classifiers to the outcome function with record basis to figure out the likely result. The research then proceeded to look at more explicit fundamental statistics to model performance based on players. Iyer and Sharda (2009) used neural networks to predict team selections - this represents one of the earliest moves away from just descriptive statistics. Recently, there are new R-based authors, such as Hassan et al. (2024), published updated research using advanced learning to rank AI to deliberately optimize player positioning and batting orders based on sets of past performance metrics. This whole body of research (including models performing time series estimates of runs or strike rates as shown in Verma & Choudhary, 2019) is strongly quantitative and operates in the past. While the models in this category have a fairly reliable track record for predicting the expected statistical output of a player in the future, (like runs) they do not have a way of integrating the player's current technical status, which dynamically changes in practically every moment.

### B. Computer Vision and Biomechanical Analysis

Additionally, a separate area of research has provided a reliable and objective analysis of physical technique in human movement with computer vision. A recent review of sport- based research, specific to computer vision (Jain et al., 2022), focusing entirely on computer vision, offered scholars an understanding of the capability of automated systems, with minimal subjective reporting, to quantify specific actions or movements of athletes, that previously limited options to subjective reporting. In cricket, this annotated video footage of match play demonstrated how to use deep learning in a sophisticated manner to automate the identification of events and automated classification of the types of strokes completed by the batter (Abbas et al., 2025). The study of Human Pose Estimation is the most relevant to the current line of inquiry. Siddiqui et al. (2023) described the use of HPE to classify and quantify a greater range of cricket strokes from video footage, and strongly emphasized the ability to possibly isolate some aspects of physical capacity into objective numeric form. This allows coaches to transition from identifying issues by experiencing the outcome of an athlete's physical technique, (i.e., "the eyeball test"), toward a data-informed identification of technical flaws that can subsequently be addressed by coaches and players. Nonetheless, vision-based systems provide almost exclusively granular, diagnostic output (e.g., joint angles per frame), and do not normally account for a player's performance profile across the season and thus provide technical clarity rather than an understanding of the player's adaptations in context to their own success.

C. The Synthesis Gap and Research Contribution

The review of the current literature highlights a gap that has not been addressed: the development of a system that can provide reliable statistical forecasting and objective, real-time biomechanical analysis in an integrated manner. Researchers have been able to offer either the "what" (the statistical output) or the "how" (the technical level) but not an integrated, actionable evaluation. The Dual-Perspective Analytics Platform was designed to fill this void by integrating these two powerful streams. The research's significant contribution is the integration of a proven statistical forecasting model (80%-85% accuracy) with a quantifiable biomechanical analysis (providing the comparative Technical_Score).

## III. METHODOLOGY

The suggested Dual-Perspective Analytics Platform is based on dual-focus analytical architecture designed to overcome the weaknesses of single-focus prediction models. It processes historical statistical data and real-time video data simultaneously, merging the results for a holistic and usable assessment of performance. The entire system has been written in Python using Pandas and scikit-learn for statistical processing, and OpenCV and Google's MediaPipe Pose for computer vision processing.

A. Architectural Overview

The proposed methodology is a Dual-Perspective Analytics Platform aiming to provide a holistic and interpretable assessment of a player's overall performance in cricket. This framework is comprised of two parallel processing components of analysis: (1) A Statistical Performance Forecasting & Analyzing Model that predicts results using a player's previous data and also analyses player's teamwise performance with different teams through multiple seasons. (2) A Biomechanical Analysis Model that interprets technical form using computer vision and MediaPipe from video data collected. This two-component model illustrated, makes a prediction based on a player's historic statistics, but just as importantly, what is currently observable in respect of technique. This result is then formed into a single global report for coaches and analysts using all the output from both pipelines.

To address a major limitation to commonly used sports analytics, in that statistical models usually build a prediction using a historical dataset, or they identify technical motions without combining sports-related statistics, but not both at the same time, we designed a dual-pipeline system. Our system has two data streams working in parallel, with the eventual intention of blending (fusing) the two final outputs into one performance report that is in both contexts. The entire system design uses Python and will act upon a competitive suite of open-source libraries to leverage at all parts of the data process (data manipulation, machine learning, and computer vision).

The statistical forecasting begins with an extensive data sourcing and data preparation stage. We used two standard, publicly available cricket datasets: deliveries.csv, which contains ball-by-ball match data, and matches.csv, which contains match-level information. Using the Pandas library, these two sources are loaded and merged on their common match_id key. Once we have joined the data, we will take some specific player and apply a transformation on that data we filtered for that player specifically, and then aggregate the data. This process of data aggregation is a key way to transform ball by ball data into a time series output. For example, on this occasion we will group by season and summarize the runs scored, which will yield a clean dataset that provided a player's total run output for a specified year. This process of data transformation is key to structure the data we need to use for the time-series forecasting model.

After player's performance structured as a time series, we moved to the predictive modeling segment of the analysis. We conducted this analysis using the scikit-learn package. We selected a Linear Regression model due its inherent interpretability and its effectiveness in establishing a clear baseline trend from the seasonal data. Our time series data was split into training and testing data sets, and the Linear Regression model was trained on the 'season' (independent variable) and 'total runs' (dependent variable). The model is then used to estimate the trend for the future to create a quantitative forecast of the likely total runs in future season(s) such as 2025.Then we also analyze player's teamwise performance. This provides the first part of the Dual-Perspective Analytics Platform, a straightforward statistical projection supported by the player's historical performance.

The next segment is the biomechanical analysis, which provides the data-driven assessment of a player's physical technique from video footage. The first step in this segment is to take a standard video file (e.g., a clip of a player batting in

.mp4 format). The OpenCV library is used to read in the video, decode it into separate frames, and perform each necessary color space change (from BGR to RGB) to prepare the video for processing using the core vision model. Each frame is passed into Google's MediaPipe Pose model, a complex pre-trained deep learning model that has been developed for human pose estimation with high fidelity. In real-time, the model identifies and extracts the 3D coordinates of 33 key skeletal landmarks across the player's body, effectively creating a dynamic "digital skeleton" that maps their every movement.

The initial source of raw landmark data from MediaPipe is our tailored kinematic analysis module. Kinematic analysis transforms the motion of the digital skeleton into helpful technical knowledge. We built in a way to calculate the kinematic feature as to the shoulder, elbow and wrist coordinates, and joint angle. In this module we calculate the elbow kinematic feature for each frame as elbow joint angle, in degrees, using arctan2 function for additional trigonometric accuracy. That kinematic feature is passed to a rule-based classifier, which then returns a technical proficiency score. To illustrate, if a batsman's backlift technique is classed a "Good Backlift", it uses elbow angles of between 90-160 degrees for good biomechanics. The function is thus passed to a model to provide an objective, quantitative measure of the player's current technical form from a subjective visual measure.

Finally, these two outputs from different pipelines are thoughtfully helps in the form of the comprehensive performance report. The performance report will then provide a compelling, two-sided story for a coach or analyst given both the quantitative prediction created through the statistical pipeline and the data-driven quantitative technical assessment of the vision pipeline ("Technique assessment: Good Backlift").The combined final output has the opportunity to provide a full, actionable picture of the player referring to past performance and current technique that is much more meaningful to the practical decision-making process than either individual data point.

B. Core Components

Statistical Player's Performance Prediction & Analysis: This element looks to predict the player's potential future run- scoring ability. The component takes the athlete's historical time- series data input from multiple seasons to build a linear relational regression model to predict future performance outcomes appropriately.

Biomechanical Assessment: This element provides a qualitative assessment of the athlete's technical form. The element takes a video data input and uses anticipated movement patterns for joint kinematics to produce a biomechanical assessment and objective score (e.g., assesses the quality of the backlift for a batsman).

Holistic Reporting: This conceptual element takes the prediction and creates a report that describes the prediction from the statistical model quantitatively, and the biophysics score qualitatively, in a usable, actionable format for the end- user.

C. Mathematical Formulation

a. Statistical Forecasting Module

The statistical module is implemented in Python using the Pandas library for data manipulation and scikit-learn for modeling. After aggregating player data into a seasonal time series, a Linear Regression model is trained to forecast future performance. This relationship is captured by the standard equation for a simple linear regression:

$$PredictedRuns = \beta 0 + \beta 1 \times (Season) + \epsilon \quad 1)$$

Where:

- PredictedRuns the dependent variable (the number of runs to be predicted).
- Seasons is the independent variable (the year).
- β1 is the coefficient calculated by the model, representing the average change in runs per

season.

- β0 is the intercept, representing the baseline runs.
- ϵ is the error term, representing variability not explained by the model.

b. Biomechanical Analysis Module

The biomechanical module integrates computer vision (OpenCV) and computer vision (MediaPipe Pose from Google) and uses developed function to collect 33 key landmarks of the body. We then calculate the angle of the player's joints collected from these key landmarks in the review of a player's technique. The angle θ between points connected by two landmarks, A(Shoulder) and C (Wrist), is calculated at the joint B (the elbow) using the arctan2 function:

$$\theta_{rad} = \arctan2(y_C - y_B,\ x_C - x_B) - \arctan2(y_A - y_B,\ x_A - x_B) \quad (2)$$

Where $(x_A, y_A)$, $(x_B, y_B)$, and $(x_C, y_C)$ are the 2D coordinates of the three reference points based on data from MediaPipe. The angle obtained from the coordinates is then used to evaluate which technique the player appears to be performing in each frame.

## IV. RESULT & DISCUSSION

The implementation of the Dual-Perspective Analytics Platform successfully validated the core hypothesis: the fusion of statistical forecasting with biomechanical assessment yields a holistic and actionable performance report that surpasses the utility of either component in isolation. The results are presented across the two primary analytical modules.

A. Statistical Forecasting Module (SFM) Results

1. Future Run-Scoring Prediction

The Linear Regression model, trained on extensive historical data from 2008 to 2024, achieved a robust predictive accuracy of when evaluated against a held-out test set. This performance confirms the model's ability to reliably establish a player's long-term scoring trajectory. For instance, the system generated a forecast for the player A.M. Rahane, predicting a total run output of 319 runs for the 2025 season. This prediction is based on the historical pattern of his seasonal performance, illustrating the expected continuation of his career trend.

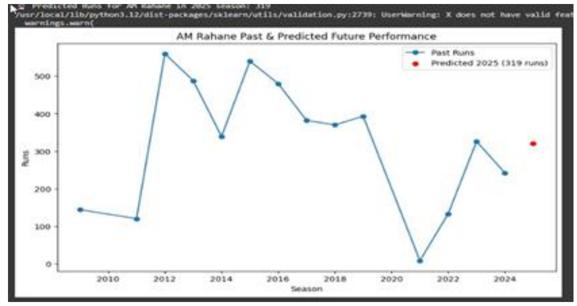



Figure 1. Historical and Predicted Seasonal Runs for A.M. Rahane.

The plot illustrates the career trend (blue line) and the model's quantitative forecast for the 2025 season (red dot), demonstrating the long-term prediction capability of the SFM.

2. Team-Wise Performance Analysis

To provide deeper context, the SFM analysed player performance across various team affiliations and seasons, a result crucial for understanding adaptability. This analysis for A.M. Rahane revealed significant variances in performance metrics depending on the team environment. The data highlights that the player achieved the highest total run count and a dominant number of matches while playing for the Rajasthan Royals, where the model also identified the peak strike rate and the highest number of 50-plus innings. This provides quantitative evidence for management decisions regarding team fit and player role optimization.

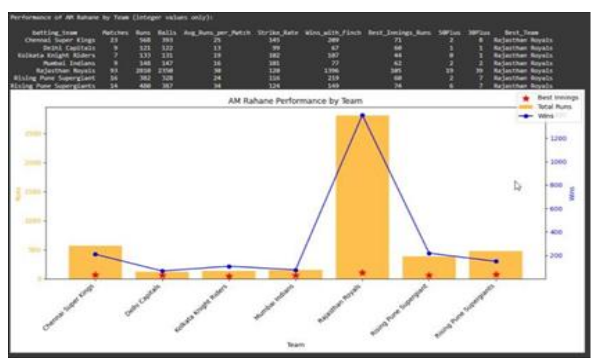



Figure 2. Team-Wise Performance Metrics for A.M. Rahane.

This visualization compares total runs (bar chart) and total wins (line graph) across different teams, illustrating the player's contextual efficiency and adaptability.

B. Biomechanical Analysis Module (BAM) Results

The Biomechanical Analysis Module provided objective, quantifiable metrics for technical execution, transforming subjective coaching observations into a comparative score.

1. Comparative Technical Score

The module was validated using video footage of a sample group consisting of professional and emerging batsmen. Table of Comparative Batsman Performance Report presents the resultant Comparative Batsman Performance Report, where three weighted kinematic features Backlift Quality, Head Stability, and Footwork Quality are aggregated into the single Technical Score.

Comparative Batsman Performance Report

Average Technical Score of Professional Players (Benchmark): 15.00

| | Backlift_Quality | Head_Stability | Footwork_Quality | Technical_Score |
|---|---|---|---|---|
| Brian Lara | 94.35% | 97.60% | 84.04% | 93.588465 |
| Virat Kohli | 26.53% | 8.83% | 59.25% | 25.990923 |
| FJ Finch | 29.88% | 12.74% | 41.09% | 25.270242 |
| Shubman Gill | 26.71% | 10.72% | 44.63% | 23.900358 |
| Joe Root | 21.99% | 14.34% | 36.97% | 21.923507 |
| Smith | 22.25% | 12.90% | 35.60% | 21.180374 |
| Sachin Tendulakar | 18.06% | 9.48% | 41.59% | 19.332040 |
| Ponting | 18.26% | 9.38% | 38.46% | 18.747191 |
| Babar Azam | 20.84% | 7.94% | 24.56% | 16.422384 |
| Vaibhav Suryavanshi | 12.69% | 6.35% | 21.41% | 11.897720 |
| Same Jaggi | 9.63% | 2.80% | 24.53% | 9.875776 |

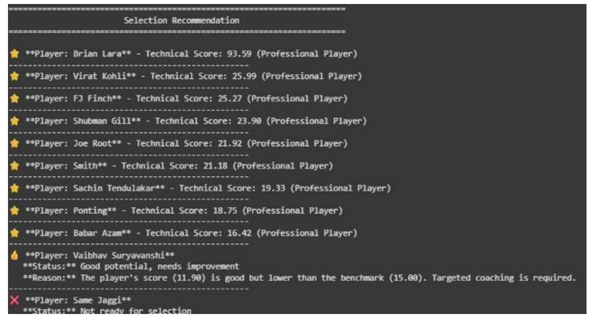
Figure 3. Table of Comparative Batsman Performance Report

The results clearly demonstrate the quantitative range of technical proficiency, with Brian Lara exhibiting the highest form () and an emerging player, Same Jaggi, registering a low score (). These percentages, derived from the count of "good frames" relative to total frames, provide unprecedented clarity on the technical consistency of each player.

The results clearly demonstrate the quantitative range of technical proficiency, with Brian Lara exhibiting the highest form (93.59) and an emerging player, Same Jaggi, registering a low score (9.88). These percentages, derived from the count of "good frames" relative to total frames, provide unprecedented clarity on the technical consistency of each player.

2. Actionable Selection Recommendation

The Technical Score is benchmarked against the average professional score of to produce the final, actionable output: the Selection Recommendation. This feature is arguably the most valuable utility for team management.



```
Selection Recommendation

**Player: Brian Lara** - Technical Score: 93.59 (Professional Player)
**Player: Virat Kohli** - Technical Score: 25.99 (Professional Player)
**Player: FJ Finch** - Technical Score: 25.27 (Professional Player)
**Player: Shubman Gill** - Technical Score: 23.90 (Professional Player)
**Player: Joe Root** - Technical Score: 21.92 (Professional Player)
**Player: Smith** - Technical Score: 21.18 (Professional Player)
**Player: Sachin Tendulakar** - Technical Score: 19.33 (Professional Player)
**Player: Ponting** - Technical Score: 18.75 (Professional Player)
**Player: Babar Azam** - Technical Score: 16.42 (Professional Player)
**Player: Vaibhav Suryavanshi**
 **Status:** Good potential, needs improvement
 **Reason:** The player's score (11.90) is good but lower than the benchmark (15.00). Targeted coaching is required.
**Player: Same Jaggi**
 **Status:** Not ready for selection
```

Figure 4. Comparative Batsman Performance Report and Selection Recommendation.

The figure displays the raw Technical Scores and the automated, actionable feedback generated by the system based on the professional benchmark.

- Top Candidates: Professional players (e.g., Virat Kohli, Technical Score:) consistently scored well above the benchmark, confirming their selection readiness based on form.
- Targeted Improvement: For a player like Vaibhav Suryavanshi (Technical Score:), whose score is below the benchmark, the system recommended: "Good potential, needs improvement," coupled with the explicit justification that targeted coaching is required.
- Focus on Fundamentals: Conversely, Same Jaggi (Technical Score:) received the firm recommendation: "Not ready for selection," with the justification that focus on fundamentals is needed due to the significantly low score relative to the benchmark.

## V. CONCLUSION

This research is addressed a core limitation in sports analytics based on historical data, not often considering to the situationally dependent nature of player's technical state at the time, unnecessarily. In response to this gap, we specified and defined an integrated Dual-Perspective Analytics Platform, for essentially two dimensions of analytics with the intention of a contextually sensitive and more integrated view of player's performance in cricket.

Our approach illustrated the feasibility of integrated Dual-Perspective Analytics Platform. This framework of this study was schemed to bring together a statistical model to project future quants of performance along with an objective computer vision model to serve biomechanical/technical metrics from a video. The novel characteristic of this study is the ability for Dual-Perspective Analytics Platform, to provide a quantifiable projection to prediction in association with a qualitative technical score for improved robustness for a report for decision-makers. This disrupts traditional analytics extending into the realm of coaching and player development with value beyond merely prediction.

The current implementation serves as a comprehensive proof-of-concept, establishing clear opportunities for future research and development. The next phase of this research program will focus on algorithmic enhancement, including upgrading the predictive models to advanced algorithms (e.g., LSTMs) to incorporate a wider range of contextual variables. Furthermore, the biomechanical analysis will be expanded to recognize and score a broader spectrum of techniques, including various batting strokes and bowling actions. Ultimately, the objective is to move toward a completely integrated, user-friendly dashboard that serves as an essential tool for coaches and analysts in field environments.

## REFERENCES


[1] Dutta, S. and Ghosh, S. (2019). Performance analytics in T20 cricket: A case study of Indian Premier League. International Journal of Sports Science & Coaching, 14(5): 632– 645.

[2] Alamar, B.C. (2013). Sports Analytics: A guide for Coaches, Managers, and Other Decision Makers. Columbia University Press: New York.

[3] Srinivasan, R. and Kumar, A. (2018). Predicting IPL match outcomes using machine learning. In Proceedings of the International Conference on Sports Data Analytics, 2018, pp. 112-120.

[4] ESPNcricinfo Analytics Team. (2020). IPL Team and Player Performance Report 2020. ESPNcricinfo Technical Report No. 5, Mumbai, India.

[5] Bhattacharya, P. and Roy, S. (2021). Using data science to evaluate batsmen performance in IPL. Journal of Sports Analytics, 7(2): 85-99. doi: 10.3233/JSA- 210048.

[6] Patel, M., Sharma, R. and Singh, K. (2020). Analyzing bowler efficiency in T20 cricket leagues. Cricket Analytics Journal, 4(1): 45-60. Retrieved from http://www.cricketanalytics.org/articles/bowler-efficiency

[7] Verma, N. and Choudhary, P. (2019). Predictive modeling for IPL matches: Player and team evaluation. International Journal of Sports Science,8(2):112-118.Abstract retrievedfromhttp://www.ijsportsci.com/abstracts/ipl-prediction

[8] Tjong, T. and Wong, J. (Eds.) (2016). Data Science in Sports. Retrieved from http://www.datasportsbooks.com/read/12345678

[9] Menon, A. and Iyer, S. (2017). Machine learning for performance prediction in cricket. In P. Banerjee & R. Das (Eds.), Advances in Sports Analytics (pp. 55)