Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 1 (January) (2025)

A Cluster Analysis of the Performance of Allrounders in T-20 International Cricket

Qamruz Zaman (Corresponding Author)
Department of Statistics, University of Peshawar, Pakistan.
Email: cricsportsresearchgroup@gmail.com

Muhammad Ibrahim Khattak
Department of Statistics, University of Peshawar, Pakistan.
Email: Ibrahim.ktk444@gmail.com

Sidra Nawaz
Department of Statistics University of Peshawar, Pakistan.
Email: sidrabintenawaz@gmail.com

Mohib Ullah Khan
Department of Statistics University of Peshawar, Pakistan.
Email: mohibullahkhan578@gmail.com

Ghazala Sahib
Department of Statistics, Shaheed Benazir Bhutto Women University Peshawar (SBBWUP), Pakistan. Email: ghazalasahib@yahoo.com

Abstract
In T-20 cricket, all-rounder comparisons across nations reveal diverse trends. Robust cricketing nations like Australia, England, and New Zealand feature 5-6 all-rounders, emphasizing versatility. Contrarily, Afghanistan and India showcase fewer (around 4) specialized all-rounders. Smaller cricketing nations, Kenya, Nepal, and UAE, exhibit 1-2 all-rounders due to evolving cricket landscapes. Zimbabwe stands out with 7, while Ireland, Pakistan, and Scotland have 4-5. The chi-square test highlights height-weight correlation among T-20I all-rounders. Continent-wise, strong performances emerge from Africa (Zimbabwe, South Africa), Americas (West Indies), Asia (India, Pakistan, Sri Lanka), and Australia-New Zealand (East Asia Pacific). Hierarchical clustering delves into player attributes, yielding insights into categorization and impacts. T-20 cricket is dynamic, with varying all-rounder presence, enriched by statistical analysis and clustering techniques.
Keywords; T-20I, All-rounder, Height, Weight, Cluster analysis

Introduction

Statistics is a branch of mathematics that deals with the collection, analysis, interpretation, presentation, and organization of data. It involves gathering information or observations about a particular phenomenon, summarizing the data, and drawing meaningful conclusions or making predictions based on the information available. In simpler terms, statistics helps us make sense of numerical information and provides tools and techniques to understand and describe patterns, relationships, and trends in data. In order to understand and use numerical data more effectively, statistics is a method of analysis. It is vital in a variety of fields, including economics, where it helps in our understanding of concepts like income and unemployment rates. It is also important for comprehending issues like population growth, housing, education, healthcare, and sports [1]. Scientists from disciplines such as physiology and medicine have aided in enhancing athletic performance. Now, professionals in the fields of law, sociology, economics, and statistics are participating in the sports industry. This agenda centers on the application of statistical analysis to sports research. Sports and statistics are closely related since there is a wealth of data in sports. Most sports keep track of various individual and team statistics, which are used to rank teams and players [2]. The gathering, analysis, interpretation, and presentation of data are all topics in the mathematical discipline of statistics. It offers a framework for generating data-driven decisions and predictions and is applied in a variety of sectors, including business, medicine, social sciences, and more. Descriptive statistics, inferential statistics, probability, regression analysis, and Bayesian statistics are important ideas in statistics. Large datasets can be mined for useful insights using statistical techniques, and these insights can then be used to guide data-driven decision-making [3].

Cricket, a widely beloved sport globally, has maintained its popularity for an extensive period. Spanning over 250 years, the sport has been guided by a series of Codes of Law. Similar to its cousin, baseball, cricket places a significant emphasis on strategic elements, which greatly contributes to its allure among fans. The game's inherent strategic nature has led to a deep appreciation among enthusiasts. For statisticians, cricket proves to be a fascinating field due to its finite outcomes and ball count, providing an ideal foundation for modeling. The sport has also amassed an extensive dataset of match-related information, further adding to its appeal for analytical purposes [4]. The three main game types are test cricket, one-day international cricket (ODI), and Twenty20. The longest format of the sport, test cricket, is regarded by experts as the toughest test of skill. ODI matches are played over 300 legal deliveries (balls) per side, whereas T-20 matches are played over 120 legal deliveries (balls) per side. A standard cricket team has eleven players, and the team that wins the coin toss determines who gets to bat first. Batting, bowling, and fielding are the three essential aspects of cricket. Selecting one member from each of the three previously stated groups will help balance an 11-man squad [5].

The International Cricket Council (ICC) functions as the regulatory authority overseeing cricket matches on the global stage. Its role encompasses the coordination of significant tournaments worldwide. Presently, the ICC boasts 10 full-member nations, namely Australia, Bangladesh, England, India, New Zealand, Pakistan, South Africa, Sri Lanka, West Indies, and Zimbabwe [4]. A player who is good at throwing the cricket ball is referred to as a bowler in the cricket community. On the other hand, a player who is skilled at hitting the ball is referred to as a batsman. A player who does well as a bowler and a batsman is an all-rounder [5].

The field of cricket as a subject of research has transformed in parallel with the evolution of the game itself. Recent advancements within the sport have been significantly influenced by commercial considerations. The conventional cricket match involves a two-innings per team setup and spans several days. However, this form of the game has experienced declining spectator interest due to its often-inconclusive outcomes and perceived dullness. In response to this, a

ISSN Online: 3007-3154
ISSN Print: 3007-3146

Vol. 3 No. 1 (January) (2025)

condensed variant emerged – the one-day game – featuring a single innings per team and a limited number of overs. This modification proved to be a commercial triumph, as it captured the enthusiasm of viewers and led to higher attendance at matches. A more contemporary iteration, known as Twenty20, further shortened the game, restricting each team to a maximum of 20 overs for batting and bowling combined. The commercial prosperity of this format prompted the inception of a global tournament, with its inaugural edition taking place in South Africa in 2007, followed by the second held in England in 2009 [6].

The International Cricket Council (ICC), the organization that oversees cricket, seeks to increase the appeal of the sports. They intend to introduce T-20, a more rapid version of the game, as one approach to doing this. The objective is to increase the excitement and attractiveness of cricket in order to draw in new followers [7].

The England and Wales Cricket Board (ECB) introduced Twenty20 cricket, a modified version of the sport, in the United Kingdom in 2003. This introduction was primarily for professional inter-county tournaments. The first-ever men's international Twenty20 match occurred on February 17, 2005, at Eden Park in Auckland. During this match, Australia achieved victory against New Zealand. In this format, two teams engage, each getting a single innings, with a maximum of 20 overs to bat. Unlike the longer formats, a Twenty20 game concludes in approximately two and a half hours, featuring innings of around 75 minutes each. This adjustment aligns the game's duration more closely with other popular team sports. A comprehensive understanding of Twenty20 cricket is crucial to grasp the conventional and proposed performance metrics. In cricket, each team, comprising eleven players, is referred to as a "side." Notably, in Twenty20 matches, each side only bats once, sequentially. One team completes its batting "innings," followed by the opposing team. When in the bowling phase, both teams are restricted to a maximum of 20 overs each. Additionally, individual bowlers can only bowl a maximum of four overs. A pre-game coin toss determines which team chooses to bat or bowl first. Victory is achieved by the team that accumulates the highest run total [8].

Cricket initially involved using a ball crafted from wool or cork, encased in leather, while the bat resembled a hockey stick. The wickets, made of wooden stumps, completed the setup. Matches unfolded on circular or oval fields, where one team aimed to accumulate runs as the opposing side attempted to dismiss them. Tracing its roots to 16th century England, cricket emerged among rural shepherds and laborers. Documentation from 1598 references a match among schoolboys in Guildford, Surrey, marking one of its earliest mentions. The sport's evolution gained momentum in the 17th century, with the initial recorded game occurring in 1646. Predominantly embraced by the upper class, matches often carried substantial stakes. The game's regulations were still in flux, resulting in varying rules across different locations [9].

The game is conducted on a spacious oval field, bordered by a rope that outlines its perimeter. Positioned at the center is a rectangular area known as the "pitch," featuring a wooden wicket at either end. These wickets are located just outside the pitch but still relatively close. The pitch spans approximately 66 feet in length and 10 feet in width. Each wicket consists of three vertical components called "stumps" and two horizontal elements known as "bails." Striking the wicket with the ball can dislodge one or both of the crosspieces, a common method of dismissing a batsman. At the commencement of an innings, the batting team sends its first two players onto the field according to their batting order. This initial duo forms the first "partnership" and remains at the crease until one of them is dismissed. Upon such an occurrence, the following player in the batting order replaces the outgoing batsman. Subsequently, the second partnership begins and carries on until one of the two batsmen is dismissed. This pattern continues, with the fourth player in the order entering the game and so forth [4].

Let's delve into how two players, named B1 and B2, take turns batting. B1 positions on the left end of the pitch while B2 stands on the right. The role of the 'striker', the active batter, alternates based on certain factors. Assuming B1 is currently the striker, among the eleven fielding players, one is the bowler (not throughout the innings) and another is the wicketkeeper. The bowler stands at the opposite end of the pitch from the striker, so in this scenario, the bowler delivers from the pitch's right end.

Thus, B1 stands just outside the pitch's boundary on the left end, with the wicket placed between B1 and the wicketkeeper. The wicketkeeper's role includes catching the ball if the striker (B1) doesn't hit it. When the striker hits the ball, both batsmen, B1 and B2, exchange positions while the fielding team retrieves the ball to target the wickets. It's important to note that B1 and B2 can change places multiple times, each exchange scoring a run for their team and the striker.

However, switching places must be done cautiously to avoid being 'out'. If the ball hits the wickets and dislodges the horizontal pieces while a batsman is not safely inside the pitch boundary, they are declared 'out'. A batsman is 'in' if they and their bat are within the pitch. Since batsmen carry their bats while running, the implication is that they needn't physically cross the outer pitch line to be 'safe'. The bat is an extension of the batsman's body, so merely touching the bat's tip to the safe area keeps them secure [8].

There are multiple methods by which the fielding team can dismiss batsmen. Similar to baseball, if a hit ball is caught by a fielder before touching the ground, the batsman is considered out. Another way is if the bowler delivers the ball in a manner that knocks over the wicket and the batsman fails to prevent it. If the ball strikes the batsman's leg and the umpire determines that it would have hit the wicket if not for the leg, the batsman is out through "lbw" (leg before wicket). This rule exists because a batsman cannot use their body to protect the wicket. While there are additional methods for batsmen to be dismissed, the aforementioned are the most common ones.

An important term to understand is that when a batsman is dismissed, it is phrased as "a wicket has been taken/lost." Each time a player is out, the batting team loses a wicket, which significantly impacts the course of an innings. In essence, there are two primary scenarios for an innings to conclude: either when 10 wickets are lost or when 20 overs have been bowled (overs are explained in the following paragraph).

To provide more accuracy, it's worth noting a third situation that can lead to an innings ending prematurely: this occurs when the team batting second surpasses the run total of the team that batted first. The rationale behind concluding an innings after 10 wickets are lost is that the batting side is left with just one player who hasn't been dismissed. Consequently, no additional batsmen are available to form new partnerships [10].

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 1 (January) (2025)

Let's delve into the concept of "overs". An over encompasses the act of a fielding team's bowler delivering six balls. It's important to note that during an inning, the maximum limit is set at 50 overs, equivalent to 300 balls. However, there are situations where the total number of balls delivered can exceed 300. This usually occurs due to occurrences known as "no balls" and "wides". The former arises when the bowler executes an unlawful action while delivering the ball, while the latter is declared by an umpire when the ball is deemed unreachable by the batsman. In both scenarios, the batting team is awarded one run, and the delivered ball is not counted as a legitimate part of the over. This clarifies the possibility of surpassing the standard 300-ball limit in an inning [11].

A victory in a cricket game can be achieved either by runs or by wickets. According to cricket terminology, if the team that bats first wins the match, the extent of their victory is measured in terms of runs. Conversely, if the chasing team wins, the measurement is in wickets. For instance, if the initial batting team scores 250 runs and their opponents manage 220 runs in response, we say that team 1 has triumphed over team 2 by 30 runs. Conversely, if team 2 outdoes team 1's run total with seven wickets remaining, the accurate expression would be 'team 2 has won by three wickets.' This phrasing signifies that team 2 had three more wickets in hand for batting, although they didn't require them.

An all-rounder is a player who has the ability to be selected for the squad based on either their batsman-ship or bowling abilities. Fielding is significant, but batting and bowling are valued more highly. A true all-rounder who is good at both bowling and batting is unique. The majority of all-rounders, often known as batting all-rounders or bowling all-rounders, perform well in just one area. Finding all-rounders is helpful for cricket coaches, selectors, and players and essential for a team's success [12].

An all-rounder in cricket is a versatile player who excels in both batting and bowling aspects of the game. They are considered to be a valuable asset to their team as they contribute significantly with both their batting and bowling skills. An all-rounder possesses the ability to score runs consistently with the bat, often being capable of playing aggressive or defensive strokes depending on the match situation. They are also adept at bowling, capable of taking wickets and maintaining a good economy rate. All-rounders are known for their versatility and adaptability, providing balance and depth to their team by contributing in multiple areas of the game. Their presence on the field adds an extra dimension to the team's strategy and increases their chances of success in various match scenarios.

In order to delve into cricket, statistics are necessary. It makes it easier to evaluate a player's effectiveness, monitor team chemistry, and forecast game results. The most frequently used cricket statistics are the batting average, strike rate, bowlers' economy rate, and fielding stats like catches and run-outs. These data give selectors a thorough insight into a player's skill set and aid their decision-making. By highlighting the strengths and weaknesses of the opposition's players and teams, statistical analysis can also assist teams in developing their game plans for upcoming contests. In general, statistics offer cricket fans, analysts, coaches, and players a useful tool for better understanding the game [13].

Statistics plays a crucial role in cricket by providing valuable insights and analysis of player and team performance. It helps in evaluating individual players' batting averages, bowling averages, strike rates, and other key metrics, which provide a quantitative measure of their skill and consistency. Statistics also aid in understanding team strategies, such as run rates, wicket-taking rates, and fielding efficiency, helping teams make informed decisions during matches. Additionally, statistics help identify trends, patterns, and strengths/weaknesses of players and teams, enabling better game planning and strategizing. Overall, statistics in cricket provide a quantitative foundation for assessing and comparing players and teams, contributing to a deeper understanding of the game.

Literature Review

All-rounders were divided into four categories by Saikia and Bhattacharjee (2011) using information from the Indian Premier League: performance, batting all-rounder, bowling all-rounder, and under-performer. Based on their findings, they discovered that the Nave Bayes method has a classification accuracy of 66.7% [14]. Tan and Ramachandran (2010) used a methodology based on both batting and bowling data information from the Indian Premier League: performance, batting all-rounder, bowling all-rounder, and under-performer. Based on their findings, they discovered that the Nave Bayes method has a classification accuracy of 66.7% [15]. A Bayesian parametric model was applied in another study by Stevenson and Brewer (2019) to forecast how cricket players' skills develop over a game [16]. Christie conducted a study on the physical demands of fast bowlers in 2012 and stressed the significance of judging a player's performance based on physical prowess [17]. Saikia et al. (2016) developed an all-round performance measurement that integrated batting and bowling statistics [18]. In order to anticipate a batsman's performance, Wickramasinghe developed a method in 2014. He used a multi-level, hierarchical linear model. To develop predictions about the player's performance, our model considered both player-level and team-level variables. It is difficult to choose a team to play against a certain opponent without taking both teams' strengths and limitations into account [19]. Bandulasiri et al. (2016) discovered that decision-making, batting, and bowling make up a normal one-day international cricket match. A team benefits from having a talented all-around player since it allows for compositional flexibility [20]. According to Van Staden (2008), an all-rounder makes the captain's job easier because they can score runs with the bat and bowl a ball more accurately when necessary [21]. Using machine learning techniques, Wickramasinghe, I. (2020) created a system to identify all-rounders in the One Day International format as genuine all-rounders, batting all-rounders, bowling all-rounders, or average all-rounders [5].

The market for cricket betting holds a valuation in the billions of dollars, creating a compelling drive for predictive models capable of surpassing bookmakers' odds. This research aimed to assess the feasibility of predicting cricket match outcomes, specifically focusing on the English twenty over county cricket cup. Utilizing a blend of original and engineered features, a comprehensive set of over 500 team and player statistics was established. The modeling process initially concentrated solely on team features and later incorporated both team and player attributes. Evaluating model performance across individual seasons spanning from 2009 to 2014, each model was trained on data from the preceding season. The most effective model combined a straightforward prediction approach with intricate hierarchical features, notably outperforming a benchmark set by the gambling industry [22].

In the realm of cricket, all-rounders assume a pivotal role, being players who contribute significantly with both bat and ball as required. Despite their versatile skills, these players are often categorized as batting all-rounders or bowling all-rounders, based largely on subjective assessments. This research delves into the realm of machine learning techniques to effectively classify all-rounders into these categories by analyzing their performance statistics. The study specifically explores a range of classification methods, including logistic regression, linear discriminant function, quadratic discriminant function, naive Bayes, support vector machine, and random forest. The performance of these methods is evaluated using accuracy and the area under the ROC curve as metrics. Among the six techniques, logistic regression, linear discriminant function, quadratic discriminant function, and support vector machine demonstrate exceptional performance, suggesting their potential in developing an automated classification system for cricket all-rounders. With cricket's surging popularity and the sport's increasing financial significance, the application of such a predictive tool could offer substantial advantages to decision-makers within the cricketing world [23].

The captivating allure of the latest iteration of cricket, coupled with the potential to be part of the esteemed Olympic extravaganza, positions T-20 cricket as the foremost cricket format in the future. This study's outcomes affirm that test cricket, spanning five days, tends to lack excitement for followers. One-day matches could be pursued on weekends, yet T-20 cricket contests, typically held in the evening under floodlights after work and school hours, hold greater appeal for a wider audience. Furthermore, it's evident that sponsor interest aligns with public interest, further enhancing the prominence of the T-20 format in the times ahead [24].

"With the continuous advancements in the field of Data Science, businesses are swiftly embracing the latest technologies to foster their growth. The market is brimming with competition, driving companies to excel in aspects such as management, quality evaluations, and service delivery. The optimal approach to achieve these benchmarks revolves around conducting precise and comprehensive data analysis. At the forefront of this evolution is machine learning, a burgeoning discipline that utilizes historical data to forecast future outcomes, enabling informed decision-making. Cricket, a globally cherished sport played and followed in 104 countries, captivates a massive fan base. Enthusiasts yearn for their respective teams to excel and emerge victorious. To facilitate a team's triumph, it's imperative to assess strengths and past performances. Predicting the victor of a cricket match hinges on numerous factors, including players' batting prowess, team dynamics, venue considerations, and weather conditions. This study delves into the analysis of various features aimed at predicting the winner of an IPL cricket match prior to its commencement. The prediction is achieved by training machine learning models using carefully chosen attributes. Several algorithms, namely Random Forest, SVM, Naive Bayes, Logistic Regression, and Decision Tree, are employed to build predictive models on distinct training and testing datasets. The outcomes of this predictive model offer valuable insights for cricket boards, aiding in team assessment and performance analysis. Beyond this, the model holds potential for applications in gambling and match reporting media, serving as a valuable asset [25].

Presently, in One Day International (ODI) cricket matches, the prediction of the first innings score relies solely on the Current Run Rate, calculated by dividing the total runs scored by the number of overs bowled. This approach overlooks critical factors like the number of wickets fallen and the match venue. Additionally, the second innings lacks a method to foresee the match's outcome. This study introduces a novel model encompassing two methodologies. The first approach forecasts the first innings score by not only considering the Current Run Rate, but also factoring in variables such as the number of wickets fallen, the match venue, and the batting team. The second methodology anticipates the match result during the second innings, incorporating attributes akin to the first method, along with the target set for the batting team. To execute these methodologies, Linear Regression Classifier and Naïve Bayes Classifier are employed for the first and second innings, respectively. Both methodologies divide the 50-over match into five-over intervals. During each interval, the mentioned attributes are recorded for all uninterrupted matches played between 2002 and 2014, independently for each team. The outcomes highlight the superiority of the Linear Regression classifier over the Current Run Rate method, showcasing reduced prediction errors in estimating the final score. Furthermore, the Naïve Bayes classifier displays an escalating accuracy in predicting match outcomes, starting at 68% accuracy from the initial 0-5 overs and progressively increasing to 91% accuracy by the end of the 45th over [26].

The objective of this research paper is to create a deterministic framework aimed at establishing a target score in T-20 Cricket when the team is batting first. The development of the model is rooted in mathematical methodologies, employing recursive functions and drawing from secondary statistical data derived from the Indian Premier League (IPL), a prominent T-20 cricket tournament. This data encompasses runs scored and wickets fallen at various stages, as well as pitch characteristics. Verification of the model took place across 120 matches held during IPL 2016 and 2017. The efficacy of this newly devised model was established through comparative analysis with pre-existing models. Its significance lies in its potential utility for key stakeholders, including team coaches and captains, aiding in strategic decision-making throughout the match. Furthermore, this model holds promise for future exploration, serving as a foundation for potential regulations set forth by national and international cricket boards. This could involve utilizing the model to calculate target scores in scenarios involving interruptions during gameplay [27].

Anticipating match outcomes has become a significant focal point, especially in T-20 cricket matches, given their current surge in popularity. This foresight can be approached in two distinct manners: 1) prior to the match commencement, and 2) while the match is in progress. Both of these prediction approaches rely on a range of dynamic ground-related historical variables and team performance metrics. This article presents a novel predictive model that takes into account present team and player statistics, as well as historical ground characteristics. Our predictive model is constructed using a multi-layer perceptron, allowing for adjustable weighting of factors. To assess its effectiveness, we evaluated this model using a dataset comprising historical ball-by-ball match information accessible online. Intriguingly, our proposed model demonstrated strong

ISSN Online: 3007-3154
ISSN Print: 3007-3146

Vol. 3 No. 1 (January) (2025)

performance, achieving an 85% accuracy in pre-match predictions and an 89% accuracy in in-match predictions [28].

This study presents a Twenty20 cricket simulator designed for matches among International Cricket Council teams. The simulator relies on input data such as the likelihoods of various batting outcomes, contingent on factors like the batsman, bowler, overs played, and wickets lost. The development of these batting probabilities combines classical estimation methods with a hierarchical empirical Bayes approach. This approach involves drawing information from analogous scenarios to enhance the accuracy of the probabilities. The initial focus is on determining batting probabilities for the first innings. However, in the second innings, the batting approach is influenced by the target score achieved in the first innings. To account for this, the batting probabilities are adjusted using the target score in the second innings simulation. This finding suggests that teams might not be optimally adapting their second innings batting strategies. The effectiveness of the simulator is evaluated using various diagnostic measures to assess its goodness of fit to real-world data [29].

This paper undertakes an examination of Cricket Players' performance utilizing the Factor Analysis technique. The study involves two datasets: one comprising 85 batsmen and 85 bowlers from IPL9, 2016 (20 overs), and the other consisting of 95 batsmen and 95 bowlers from ICC World Cup, 2015 (50 overs). The results of the analysis indicate that within this context, batting prowess holds more significance than bowling ability. This finding aligns with a prior study conducted on a similar genre of the sport [30].

The intersection of Data Mining and Machine Learning with Sports Analytics has emerged as an innovative realm within Computer Science, laden with intriguing challenges. The primary objective of this research pertains to the creation of a real-time result prediction system for T-20 cricket matches. This entails the utilization of diverse statistical and machine learning techniques to uncover the most optimal outcomes. Among the array of methods explored, the renowned Decision Tree algorithm takes center stage, accompanied by the versatile Multiple Linear Regression, facilitating a comprehensive comparative analysis of their findings. The distinctive nature of T-20 cricket, characterized by its dynamic momentum shifts, presents a formidable task in forecasting match outcomes. The novelty of this study lies in its pioneering approach to T-20 cricket prediction, a domain largely unexplored. Given the contemporary popularity of T-20 cricket matches, the motivation to undertake this challenge becomes even more compelling. Leveraging the Decision Tree algorithm, the research endeavors to craft a predictive system grounded in historical match data between competing teams. This predictive system bears the potential to revolutionize in-game decision-making for teams, enabling strategic choices such as optimal player placement and bowler selection during the crucial middle overs. In essence, this research significantly broadens the horizons of sports analytics, transcending its prior confines and extending its reach into the realm of T-20 cricket—a domain that was previously underrepresented in analytical pursuits [31].

The realm of sports analytics has greatly profited from the expansion and popularity of Machine Learning (ML) algorithms. The progress in Machine Learning and Data Mining has empowered sports analysts in effectively assessing the performance of players. A thorough examination of existing literature on

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 1 (January) (2025)

methods for evaluating player performance reveals a distinct requirement for formulating a novel evaluation metric tailored to Twenty20 (T-20) Cricket. This study introduces the concept of a Deep Player Performance Index (DPPI), designed to appraise a T-20 Cricket player's ability in batting and bowling. DPPI takes into account a player's current form and their designated role within the team structure. The DPPI concept has a two-fold purpose. Firstly, it facilitates a comparison among players fulfilling similar roles in different teams, both for sports enthusiasts and researchers. Secondly, when aggregated, the DPPI values of players occupying various positions within a team provide an estimate of the overall team strength. The development of DPPI involves a modification of the existing player performance evaluation guidelines from the Fédération International de Football Association (FIFA), suitably adapted to the context of T-20 Cricket. The proposed DPPI metric is constructed using a combination of K-Means clustering and the Random Forest algorithm. The outcomes are then juxtaposed against the prevailing player performance evaluation indices for the 2019 season of the Indian Premier League (IPL). Empirical findings convincingly demonstrate that DPPI excels in capturing a player's proficiency in batting and bowling compared to other existing indices. As a result, DPPI emerges as a valuable tool catering to fantasy Cricket users, Cricket enthusiasts, coaches, and team managers, imparting enhanced insights into the performance of players [32].

The batting average has traditionally served as the primary measure of a batsman's skill in cricket, especially at the first-class level. However, the limited-overs format, like one-day games, has introduced a significant shift in how we evaluate performance due to the restricted number of balls available. In one-day matches, simply having a high batting average isn't sufficient if the batsman maintains a low strike rate. In this format, scoring runs at a slow pace, even if no wickets are lost, tends to lead to defeat rather than victory. As a result, assessing batting performance in one-day cricket requires a more complex approach, considering both strike rate and the probability of getting out – much like the risk-return concept used in portfolio analysis. We've developed a novel graphical method that places strike rate on one axis and the likelihood of getting out on the other. This approach provides direct insights into batting prowess, particularly in the context of one-day matches, where time constraints play a crucial role. Within this two-dimensional framework, we've formulated a selection criterion for batsmen that takes into account both their average and strike rate. To illustrate its application, we've analyzed the batting performances of the 2003 World Cup using this criterion. Our findings highlight the consistent strong showings of Australian and Indian batsmen. Furthermore, we've ranked the top 20 run-scorers in the tournament based on their batting proficiency [33].

The emergence of T-20 cricket has sparked a transformative shift in the sport. The Indian Premier League (IPL), an annual tournament organized by the Board of Cricket Control of India, has gained immense popularity, amassing a massive fan base. The tournament's unique structure involves franchises bidding for players to represent their teams, with substantial financial investments made during the auction process. Within the context of the IPL, player rankings hold significant importance as they guide franchises and team managers in making well-informed decisions when selecting their squads. This study delves into a machine learning-centered approach to introduce a novel metric termed the

Deep Performance Index (DPI). This index aims to offer a comprehensive evaluation of both batsmen and bowlers, considering the intricate demands of T-20 cricket. The methodology employs the Recursive Feature Elimination algorithm, a machine learning technique, to identify and prioritize meaningful performance factors pivotal in shaping the DPI. Notably, the research illustrates that the DPI outperforms established T-20 cricket ranking systems in effectively encapsulating performance-centric data for both batsmen and bowlers [34].

Significance of the Study

Nowadays, T-20 cricket is more popular than ODI and Test cricket due to its shorter format. Thus far, a lot of work has been done in the field of T-20 cricket, but nothing is understood about the player's physical nature. This study aims to investigate how an all-around player's physical attributes affect their performance. The all-round cricketer's height and weight are these physical characteristics. The research is limited to the bowler batting solely due to the nature of the study. Due to the lack of studies based on height and weight to examine the performance of all-rounders in T-20 cricket. The results of this study will assist the team selection committee in selecting players who are all-round performers.

Aim and Objectives

This research seeks to analyze the influence of height and weight on the performance of all-rounders, with a particular focus on T-20 cricket. The study aims to achieve the following objectives:

- To check the significance of the bowling all-rounders in T-20 international cricket
- To evaluate the performance of batting all-rounders in T-20 international cricket
- To check the importance of wicketkeeper all-rounders in T-20 international cricket

To perform the clustering of T-20 all-rounders and to determine the association of different factors related to all-rounders, e.g., height and weight, etc.

Methodology

Source of Data

The information for this study was gathered from the MRFTyers (ICC Ranking), Cricbuzz, and ESPN Cricinfo official websites [35-37]. The International Cricket Council (ICC) list of the top 100 T-20 international all-rounders is used as the basis for analysis. The majority of the all-rounders' height and weight will make up the data, along with additional metrics like average difference, economy rate, strike rate, and bowling average.

Clustering analysis is used to evaluate the performance of all-around players. To predict the average differences of all-rounders in T-20 cricket, clustering analysis, and the K-mean algorithm is employed. A batting all-rounder is an all-rounder who consistently scores runs as a hitter with a higher strike rate than average and consistently allows runs as a bowler with an economy rate higher than average.

A bowling all-rounder is also one who scores runs as an all-rounder who bowls at a rate that is lower than average, concedes runs as a bowler at an economy rate

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 1 (January) (2025)

that is lower than average, and bats with a strike rate that is lower than normal. Additionally, an all-rounder is referred to as a bowling all-rounder if their strike rate as a batter is below average and their economy rate as a bowler is below average, and they are referred to as underperformers if their strike rate as a batter is below average and their economy rate as a bowler is higher than average.

Clustering Analysis

Clustering is a methodology that involves partitioning a population or dataset into multiple groups, with the intention of having data points in the same group exhibit greater similarity to each other compared to those in different groups. The primary goal is to uncover clusters or groups characterized by shared characteristics. This analytical approach, often utilized for exploratory data analysis, aims to unveil the inherent data structure. The challenge lies in identifying subsets within the data where data points within a given subset (cluster) are highly alike while being significantly distinct from data points in other clusters. Essentially, the objective is to discern cohesive subgroups within the data, aiming to achieve maximum comparability among data points in each cluster. This comparability is often evaluated using metrics like Euclidean distance or correlation-based measures [38].

The clustering phase within the data mining process holds significant importance. Data classification stands out as a pivotal step in the journey of knowledge discovery, being one of its core elements. During exploratory data analysis, an unsupervised learning process takes place, aimed at unveiling intricate patterns that are difficult to classify, from the data. The principal objective of cluster analysis is to group objects within clusters in a way that maintains closer proximity among objects in the same cluster compared to those in different clusters. This results in the formation of data clusters, where items possessing shared characteristics are brought together. Various clustering models are at one's disposal, adaptable to different data types and the desired traits of clusters. Recently, an array of innovative algorithms has emerged with the purpose of harmonizing diverse clustering methods and integrating contrasting clustering approaches. This development has been driven by the growing demand to analyze large, sequential datasets characterized by intricate connections, spanning various applications across a wide spectrum. To address the challenge of handling scattered data points and achieving efficient clusters with minimal outlier impact, a multitude of clustering techniques has been developed.

Clustering has been used to accomplish a variety of objectives, including comprehending how data is distributed, developing hypotheses, detecting features and discovering abnormalities, constructing natural classifications, and even summarizing data. The term 'clustering algorithms' encompasses various existing techniques for clustering. When employing partitioning clustering methods such as k-means, analysts need to pre-determine the suitable number of clusters to be generated within a dataset. The common k-means method will be used to extract the clusters from the datasets. For this objective, several different metrics have been proposed in the literature. The optimum clusters in the datasets will be found using the well-known silhouette, gap, and elbow statistics [39-40].

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 1 (January) (2025)

K-means clustering algorithm

Data mining frequently uses the clustering method K-Means to perform cluster analysis. With K-Means clustering, n observations are split up into k clusters, each of which is comprised of the centroid of the closest cluster.

One of the most fundamental non-supervised learning algorithms, the k-means method, was applied to solve the well-known cluster problem. It is a partitioning clustering algorithm, and its objective is to divide the given data items into k distinct clusters using an iterative procedure that eventually converges to a local minimum. The resulting clusters are compact and self-contained as a result. There are two sections to the algorithm. The first stage involves choosing k centers at random from a specified set of k centers.

The following procedure is used to transfer each piece of data to the close-by data center (Euclidean distance): The number of clusters that the method identified in the data is indicated by the letter "k" in the word "k-means". With this method, the sum of the square distances between each data point and the cluster centroid is kept to a minimum.

$$Within\ sum\ of\ square = \sum_{i=1}^{N}(C_i - X_i)^2 \qquad (1)$$

Where Ci is the observation and Xi is the cluster center. It is important to remember that less varied clusters have more comparable data points inside each cluster.

The **k-means** approach yields locally optimal results in terms of clustering error. A lot of clustering software uses this speedy iterative approach. The cluster centers are moved at each level of the point-based clustering algorithm's progression to minimize clustering error the method's main drawback is that it depends on where the cluster centers are first placed. To obtain almost optimal solutions using the k-means method, several runs with various cluster center beginning placements must be scheduled [41].

K-means clustering, an extensively used unsupervised learning technique involving 'k' clusters, facilitates the segmentation of a provided dataset into k distinct groups predetermined by the analyst. In this method, each cluster is characterized by its centroid, computed as the average of points assigned to the cluster, which acts as the central reference point. To identify clusters with the least intra-cluster variance (sometimes referred to as total within-cluster variation), which is the fundamental tenet of k-means clustering. In order to use clustering, the first step is to determine how many clusters (k) will eventually form in the solution. The process starts by randomly choosing k objects from the data set to serve as the initial centers of the clusters. Centroids or clusters are alternate names for the chosen elements [42].

Optimal number of Cluster

A significant priority in the fields of pattern recognition and machine learning is cluster analysis. With samples in one cluster being as similar as possible and samples in other clusters being as dissimilar as possible, clustering is the process of grouping samples into several groups based on some similarity criterion. When using partitioning and clustering techniques, it is crucial to select the appropriate number of clusters in a data set. The analyst must choose the right number of clusters to generate using this family of clustering algorithms. Unfortunately, this scenario cannot be measured using a typical metric. It is

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 1 (January) (2025)

rather arbitrary and depends on the techniques used to calculate item similarity.

Analysis and Discussion
To provide a brief comparison among the T-20 allrounders from the countries have listed, we can look at the number of players and their potential impact based on their performances. In the realm of T-20 cricket, a comparison of all-rounders' player count reveals interesting trends. Countries with robust cricket traditions, such as Australia, England, and New Zealand, present a higher concentration of all-rounders, each having around 5 to 6 players. These nations have historically emphasized the development of versatile players who can contribute both with bat and ball. On the other hand, countries like Afghanistan and India feature fewer T-20 all-rounders, around 4 each. This could be attributed to their stronger focus on specialized roles, with an emphasis on either batting or bowling prowess. Nevertheless, their all-rounders are known for their potential game-changing abilities. Smaller cricketing nations like Kenya, Nepal, and the United Arab Emirates exhibit a limited number of all-rounders, around 1 to 2 players. This could be due to the evolving nature of cricket in these regions and the corresponding emphasis on building core strengths. Zimbabwe stands out with 7 all-rounders, indicating a commitment to nurturing players who can excel in multiple facets of the game. Ireland, Pakistan, and Scotland have a balanced representation of 4 to 5 players, reflecting their strategies to maintain a mix of versatile players. Overall, while the count of T-20 all-rounders varies among countries, the presence of such players enhances team flexibility and performance, contributing to the dynamic nature of T-20 cricket on the global stage. Table-1 explain the country wise distribution of all-rounders.

Table 1: Country wise distribution of all-rounder

Countries	**No. of Players**	**Countries**	**No. of Players**	**Countries**	**No. of Players**	**Countries**	**No. of Players**
Afghanistan	4	Scotland	6	Kenya	2	Uganda	2
Australia	5	South Africa	4	Nepal	2	Bahrain	1
Bangladesh	6	Sri lanka	4	Netherlands	4	Jersey	1
England	6	West Indies	6	Oman	5	Qatar	1
India	4	Zimbabwe	7	P.N.G	4	Romania	1
Namibia	5	Ireland	5	U.A.E.	2	Guernsey	1
New Zealand	5	Pakistan	5	Bermuda	2		

The given pie chart (Figure-1) presents the breakdown of All-rounders into different categories. Among a total of 100 all-rounders, 47 are classified as batting all-rounders, 52 are categorized as bowling all-rounders, and 1 falls under the wicket keeper all-rounders group.

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 1 (January) (2025)

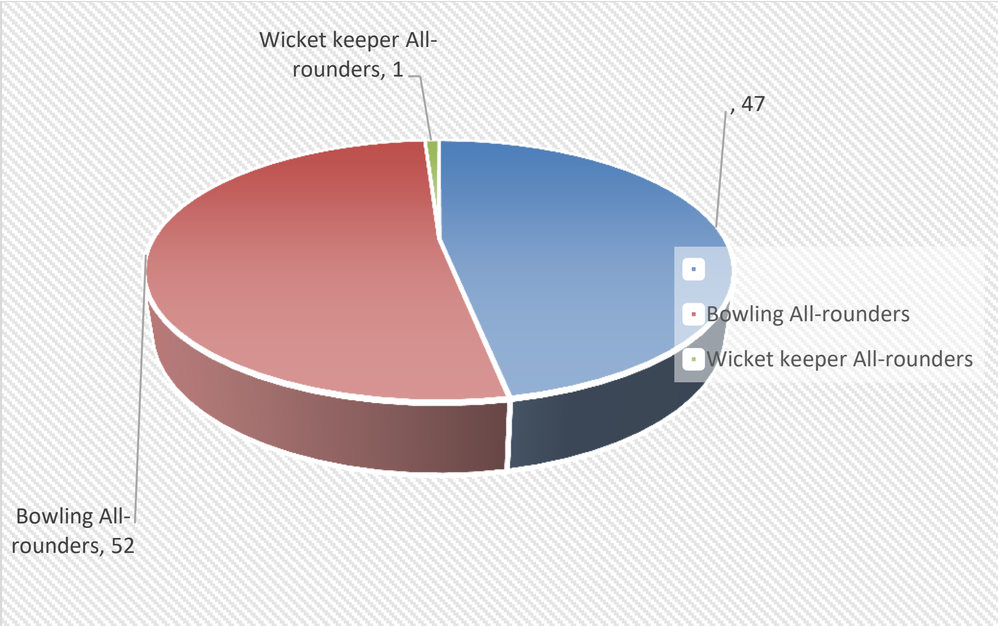

Figure 1: All-rounders role

The world map consists of six major continents: Africa, the Americas, Asia, Europe, the Pacific, and the South Pacific Ocean. These continents collaborate within the framework of the ICC. Among these, the African continent hosts five associated countries. Notably, Zimbabwe and South Africa have demonstrated consistent strong performance. Their dedicated support has led to substantial growth and has benefited numerous ICC members. The Americas region currently boasts two affiliated members who are actively developing growth and promotional strategies for the area. Over the years, the West Indies have consistently upheld their strong membership in this region. Within the Asian region, there are ten associated members of the ICC. Notably, India, Pakistan, and Sri Lanka have made significant advancements among all the associate members, showcasing impressive global performances. Afghanistan and Bangladesh have also displayed strong showings on the international stage. The East Asia Pacific territory, while the smallest in terms of member count with eleven, has witnessed remarkable achievements at the ICC primarily due to the contributions of Australia and New Zealand over the past decade. Papua New Guinea stands as the singular representative from the South Pacific Ocean region within the ICC. Among all regions, Europe boasts the highest membership count, totaling 33. The growth trajectories among European members, excluding the Full Member England, differ significantly. Countries like Scotland and Netherlands are ambitiously working towards ascending to the top echelon of Associate members on the global stage. In the current timeframe, we encompass 7 European countries in our consideration due to their presence (Table-2).

Table 2: Country wise distribution of all-rounders

Africa	**Americas**	**Asia**	**East Asia Pasific**	**South Pacific Ocean**	**Europe**
Namibia	**West Indies**	**Afghanistan**	**Australia**	**P.N.G**	**England**
Johannes	Kieron	Mohammad Nabi	Glenn Maxwell	Charles	Moeen Ali

Jonathan Smi	Adrian Pollard			Jordan Alewa Amini	
David Wiese	Jason Omar Holder	Karim Janat	Mitchell Ross Marsh	Norman Vanua	Liam Livingstone
Merwe Gerhard Erasmus	Fabian Anthony Allen	Gulbadin Naib	Marcus Peter Stoinis	Asadollah Vala	Christopher James Jordan
Craig George Williams	Andre Dwayne Russell	Rashid Khan Arman	Ashton Charles Agar	Sese Bau	Benjamin Andrew Stokes
Jan Nicolaas Frylinck	Akeal Jerome Hosein	**Bangladesh**	Josh Reginald Hazlewood		Adil Usman Rashid
South africa	Christopher Henry Gayle	Shakib Al Hasan	**New Zealand**		Samuel Matthew Curran
Aiden Markram	**Bermuda**	Mohammad Mahmudullah	Mitchell Josef Santner		**Ireland**
Dwaine Pretorius	Delray Millard Wendell Rawlins	Mahedi Hasan	James Douglas Sheahan Neesham		Gareth James Delany
George Fredrik Linde	Kamau Sadiki Leverock	Afif Hossain Dhrubo	Timothy Grant Southee		Paul Robert Stirling
Kagiso Rabada		Soumya Sarkar	Kane Stuart Williamson		Mark Richard Adair
Zimbabwe		Mohammad Saifuddin	Colin de Grandhomme		Curtis Campher
Sean Colin Williams		**India**			Andrew Robert McBrine
Ryan Ponsonby Burl		Hardik Himanshu Pandya			Kevin Joseph O'Brien
Wessley Nyasha Madhevere		Virat Kohli			Simi Singh
Sikandar Raza Butt		Ravindrasinh Anirudhsinh Jadeja			**Guernsey**
Luke Mafuwa Jongwe		Rohit Gurunath Sharma			Luke Thomas Bichard
Uganda		**Pakistan**			**Romania**
Dinesh Magan Nakrani		Shadab Khan			Taranjeet Singh

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 1 (January) (2025)

Riazat Ali Shah		Syed Imad Wasim			**Jersey**
Kenya		Faheem Ashraf			Benjamin Ward
Collins Omondi Obuya		Hasan Ali			**Netherlands**
Rakep Rajendra Patel		Mohammad Nawaz			Pieter Marinus Seelaar
		Srilanka			Roelof Erasmus van der Merwe
		Pinnaduwage Wanindu Hasaranga de Silva			Colin Niel Ackermann
		Madagamagamage Dasun Shanaka			Maxwell Patrick O'Dowd
		Chamika Karunaratne			**Scotland**
		Mashtayage Danushka Gunathilaka			Richard Douglas Berrington
		Bahrain			Michael Alexander Leask
		Sarfaraz Ali			Calum Scott MacLeod
		Qatar			Christopher Nicholas Greaves
		Muhammad Tanveer			Mark Robert James Watt
		Oman			Safyaan Mohammed Sharif
		Zeeshan Maqsood			
		Khawar Ali			
		Syed Aamir Kaleem			
		Aqib Ilyas Sulehri			
		Mohammad Nadeem Nepal			
		Dipendra Airee			
		Karan KC			
		U.A.E.			

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 1 (January) (2025)

		Rohan Mustafa			
		Kashif Daud			

The study utilized the chi-square test to investigate the potential association between the height and weight of T-20 All-rounders. The computed chi-square (χ^2) statistic was determined to be 39.600. With a single degree of freedom, this resulted in a P-value of 0.000. Since this derived P-value is smaller than the conventional significance level of 0.05, it can be inferred that a statistically significant relationship exists between the height and weight of T-20 All-rounders.
Additionally, referring to Table-3, it becomes evident that within the group of 72 players, 22 of them exhibit both below-average height and weight. Interestingly, there are no players who surpass the average weight while simultaneously having a height below the average range. Moreover, 10 players weigh less than the average, but possess a height greater than the average. Notably, 40 players showcase both above-average height and weight.

Table 3: Cross tabulation of the Height and Weight of All-rounders

Height record	Weight record		Total
	Less than Average	Greater than Average	
Less than Average	22	0	22
Greater than Average	10	40	50
Total	32	40	72
Chi-square Value	39.600	p-Value	0.0000

The bar graph (Figure-2) illustrates the distribution of heights and weights among the top-ranked ICC players. Among this group, 22 players possess heights that are below the average, whereas 50 players exhibit heights exceeding the average benchmark. Similarly, 32 players weigh less than the established average, while 40 players showcase weights that surpass this average threshold.

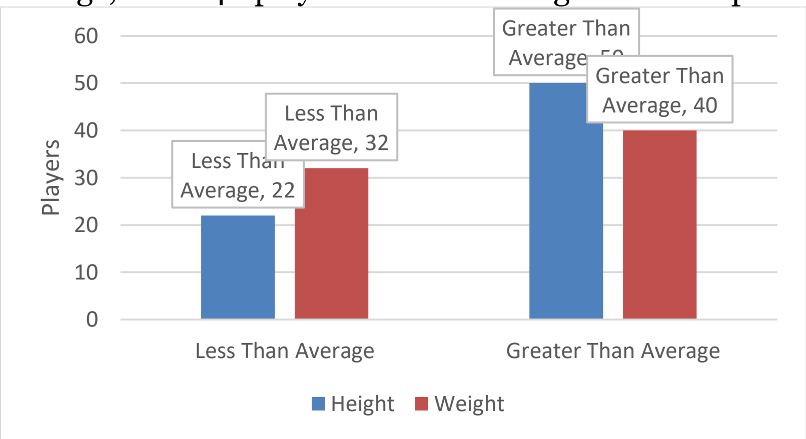

Figure: 2: Bar graph of the height and weight of All-rounders
The utilization of the chi-square test aimed to investigate any potential association between the height and weight of T-20 batting all-rounders. Upon computation, the resulting chi-square (χ^2) value was determined to be 2.106 (Table-4). Considering the presence of a sole degree of freedom, this

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 1 (January) (2025)

computation yielded a corresponding P-value of 0.147. Given that this derived P-value of 0.147 surpasses the generally acknowledged significance threshold of 0.05, we can infer that there is no statistically meaningful link between the height and weight of T-20 batting all-rounders.
The table presents height and weight records categorized as "Less than Average" and "Greater than Average." In "Less than Average" height, 7 individuals and 2 in "Greater than Average" weight intersect, totaling 9 with both below averages. In "Greater than Average" height, 13 coincide with "Greater than Average" weight, summing to 26 surpassing averages. The table offers insights into players' height and weight distributions relative to averages among 35 individuals.

Table 4: Cross tabulation of the Height and Weight of Batting All-rounders

Height record	Weight record		Total
	Less than Average	Greater than Average	
Less than Average	7	2	9
Greater than Average	13	13	26
Total	20	15	35
Chi-square value	2.106	P-Value	0.147

The bar chart (Figure-3) provides a visual representation of how heights and weights are distributed among the ICC's top-ranked players. Within this set, 9 players have heights that fall below the average height, while 26 players have heights that are above this average reference point. Likewise, 20 players weigh less than the established average weight, while 15 players have weights that go beyond this average standard.

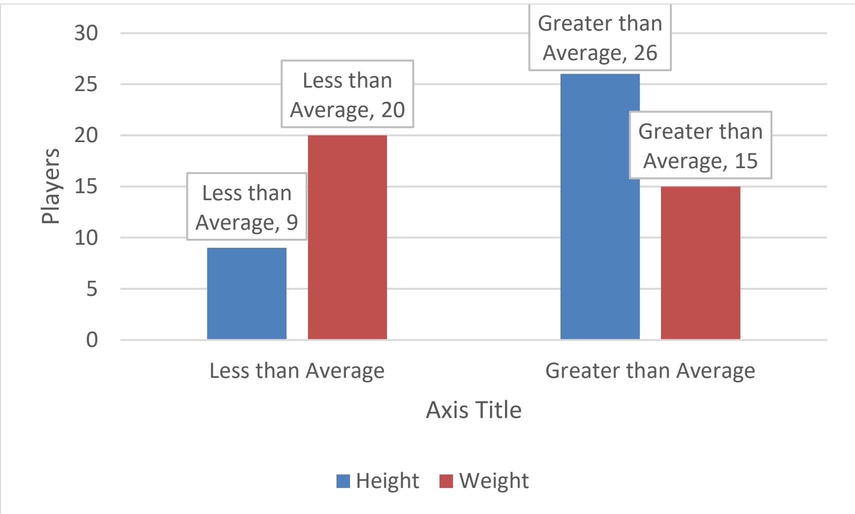

Figure 3: Bar chart of the Height and Weight of Batting All-rounders

The presented table compares individuals' height and weight records as "Less than Average" or "Greater than Average." A Chi-square test examines whether an association exists between these categorical variables. Among 36 individuals, patterns emerge: 22 with both "Less than Average" height and weight, 9 with

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 1 (January) (2025)

higher weight and normal height, 3 with tallness and lower weight, and 11 with both "Greater than Average" height and weight. The Chi-square analysis evaluates whether these observed patterns significantly differ from what would be expected if height and weight were independent. A significant result implies a meaningful relationship between these traits.

The relationship between height and weight among T-20 bowling all-rounders was investigated using the chi-square test. The computed chi-square (χ^2) statistic was found to be 4.915, resulting in a P-value of 0.027, with one degree of freedom taken into account (Table-5). Given that the obtained P-value (0.027) is less than the conventional threshold of 0.05, we can infer a meaningful statistical link between height and weight among T-20 bowling all-rounders.

Table-5: Cross tabulation of the Height and Weight of Bowling All-rounders

Height record	Weight record		Total
	Less than Average	Greater than Average	
Less than Average	13	9	22
Greater than Average	3	11	14
Total	16	20	36
Chi-square Value	4.915	P-Value	0.027

The bar chart displays the breakdown of heights and weights within the highest-ranked ICC players (Figure-4). Among them, 13 players have heights below the average, and 3 players exceed it. Likewise, 9 players weigh less than the average, while 11 players surpass this typical weight.

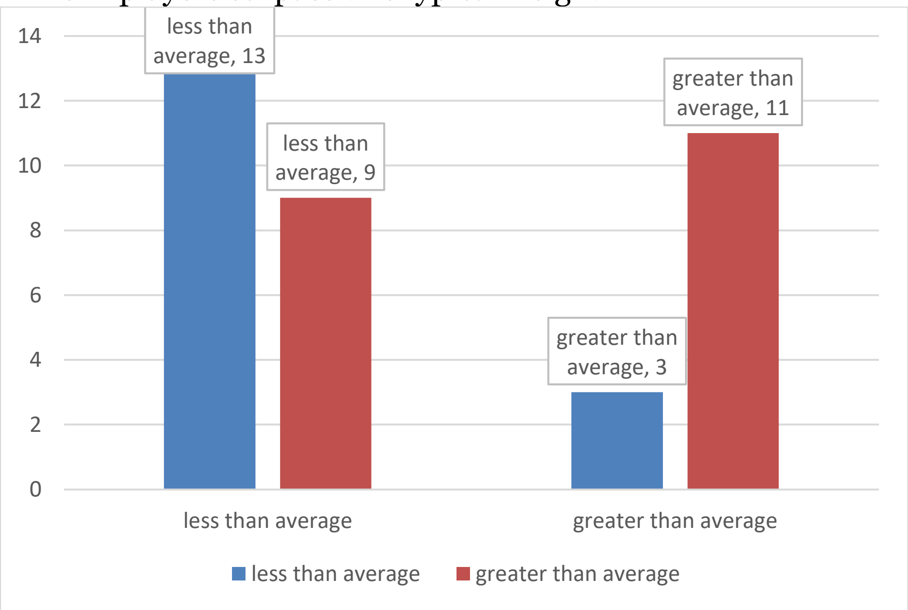

Figure -4: Bar chart of the Height and Weight of Bowling All-rounders

The dataset pertains to T-20 all-rounders and encompasses diverse variables, including Height, Weight, Shirt Number, Highest Score, Batting Average, Runs, Economy, Bowling Strike Rate, Bowling Average, Average Difference, and Latest

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 1 (January) (2025)

Ranking as all-rounders. Employing an agglomerative hierarchical clustering approach, the data was grouped into two clusters based on variable coefficients. The clusters, named "Cluster 1" and "Cluster 2," represent distinct stages of data assignment. The "Cluster Combined" column reflects aggregated coefficients within each cluster (Table-6). "Stage Cluster First Appears" indicates a data point's initial cluster affiliation, while the "Next Stage" column indicates its subsequent transition. This technique seeks to unveil patterns by associating similar all-rounders based on attributes. Clustering aids in profiling players and deciphering performance disparities. Insights into variable impacts on overall rankings emerge. Further analysis can yield valuable insights into player categorization and trends, enriching our understanding of T-20 all-rounder dynamics.

Table 6: Cluster Analysis

Agglomeration Schedule						
	Cluster Combined			Stage Cluster First Appears		
Stage	Cluster 1	Cluster 2	Coefficients	Cluster 1	Cluster 2	Next Stage
1	460	567	400.933	0	0	3
2	846	1049	551.726	0	0	6
3	460	747	744.523	1	0	4
4	460	581	930.101	3	0	8
5	105	501	1280.489	0	0	8
6	738	846	1299.849	0	2	10
7	289	339	1354.301	0	0	18
8	105	460	1405.700	5	4	9
9	105	759	1414.983	8	0	11
10	390	738	1566.416	0	6	11
11	105	390	1666.319	9	10	12
12	52	105	1742.785	0	11	17
13	984	1060	1929.362	0	0	15
14	22	219	1960.560	0	0	18
15	984	1262	2070.458	13	0	16
16	984	1186	2115.570	15	0	17
17	52	984	2128.313	12	16	19
18	22	289	2154.574	14	7	21
19	52	597	2540.759	17	0	20
20	52	327	3217.100	19	0	22
21	1	22	3320.680	0	18	23
22	52	229	3923.624	20	0	23
23	1	52	8780.026	21	22	24
24	1	79	9330.408	23	0	25
25	1	1290	45984.108	24	0	0

The dataset encompasses attributes of T-20 all-rounders, subjected to hierarchical clustering for attribute-based grouping (Figure-5). The dendrogram's y-axis signifies cluster dissimilarity. The visualization depicts

ISSN Online: 3007-3154
ISSN Print: 3007-3146

Vol. 3 No. 1 (January) (2025)

clusters progressively emerging from individual data points at the dendrogram's base. As merging proceeds upwards, similarity among data points leads to larger clusters. Merging height indicates cluster resemblance; lower heights imply higher similarity. Clustering choice hinges on vertical cutting points, influencing cluster count. The forming clusters symbolize groups of all-rounders with analogous attributes, like height, weight, shirt number, highest score, batting average, runs, economy, bowling strike rate, bowling average, average difference, and latest ranking. The "Next Stage" column traces clusters' evolution, aiding insight into the merging process. Overall, this hierarchical clustering offers valuable insights into the inherent patterns and relationships within T-20 all-rounders' attributes.

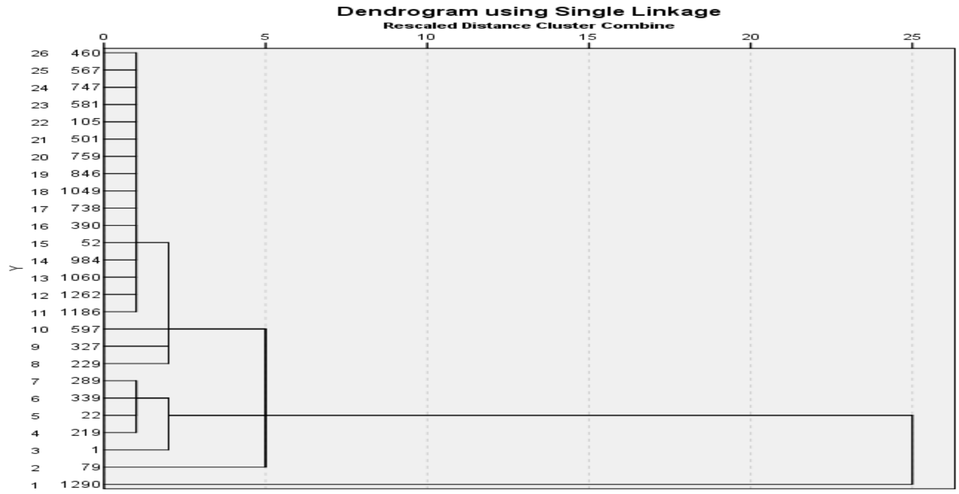

Figure-5: Dendrogram using single linkage

Summary, Conclusion and Recommendation

In the domain of T-20 cricket, a comparison of all-rounders across various countries reveals intriguing trends in terms of player count and potential impact. Robust cricketing nations like Australia, England, and New Zealand boast a higher concentration of all-rounders, typically around 5 to 6 players. Their historical emphasis on versatile players capable of contributing with both bat and ball is evident. In contrast, countries like Afghanistan and India feature fewer T-20 all-rounders, around 4 each. This could be attributed to their focus on specialized roles, either batting or bowling. Nonetheless, their all-rounders are known for game-changing abilities. Smaller cricketing nations such as Kenya, Nepal, and the UAE have fewer all-rounders, usually 1 to 2 players, due to evolving cricketing landscapes and core strength development. Zimbabwe stands out with 7 all-rounders, reflecting their commitment to nurturing players excelling in multiple aspects. Ireland, Pakistan, and Scotland exhibit balanced representations of 4 to 5 players, indicating strategies to maintain a mix of versatility. The diverse distribution of T-20 all-rounders enhances team flexibility and performance, contributing to the dynamic nature of T-20 cricket on the global stage. The statistical analysis using the chi-square test demonstrates a significant correlation between the height and weight of T-20 all-rounders, implying that these traits are linked in this context. Additionally, the analysis of height and weight distributions among top-ranked ICC players shows variations in relation to averages, shedding light on player profiles. Considering the continental context, various cricketing regions collaborate under the ICC framework. Notably, Zimbabwe and South Africa demonstrate strong performances from the African continent, while the Americas see contributions from the West Indies. Among Asian nations, India, Pakistan, and Sri Lanka have made significant advancements. Afghanistan and Bangladesh also showcase strong international performances. Australia and New Zealand drive achievements in the East Asia Pacific territory. Europe, with the highest ICC membership count, has varying growth trajectories among its members. Smaller cricketing nations like Scotland and Netherlands strive for higher ranks within the associate members. To analyze player attributes further, a hierarchical clustering approach categorizes T-20 all-rounders based on variables such as height, weight, shirt number, highest score, batting average, runs, economy, bowling strike rate, bowling average, average difference, and latest ranking. The resulting clusters provide insights into player categorization, performance disparities, and attribute impacts on overall rankings. In conclusion, the world of T-20 cricket presents a fascinating landscape of all-rounders across different nations. The player count and impact vary, reflecting cricketing traditions and strategies. Statistical analyses highlight correlations between player attributes, while clustering techniques unveil patterns within the data. This comprehensive exploration provides valuable insights for scholars and cricket enthusiasts alike, enriching the understanding of T-20 all-rounder dynamics and contributing to the broader discourse on cricketing strategies and player development.

Reference

[1] Shinh, V(2005). A Role of Statistics in Computer Applications and Marketing; 65-77

[2] Albert, J., & Koning, R. H. (Eds.) (2007). *Statistical thinking in sports*. CRC Press.

[3] Wild, C. J., & Pfannkuch, M (1999). Statistical thinking in empirical enquiry. *International statistical review*. *67*(3), 223-248.

[4] (Beaudoin, D. (2003). *The best batsmen and bowlers in one-day cricket*. Simon Fraser University.)

[5] Wickramasinghe, I (2020). Classification of all-rounders in the game of ODI cricket: Machine learning approach. *Athens Journal of Sports*. *7*(1), 21-34.

[6] Sharp, G. D., Brettenny, W. J., Gonsalves, J. W., Lourens, M., & Stretch, R. A. (2011). Integer optimisation for the selection of a Twenty20 cricket team. *Journal of the Operational Research Society*, *62*(9), 1688-1694.

[7] McHale, I. G., & Asif, M (2013). A modified Duckworth–Lewis method for adjusting targets in interrupted limited overs cricket. *European Journal of Operational Research*. *225*(2), 353-362.

[8] (Suleman, M. T., & Saeed, M. A. (2009). Option on human performance: A case study of indian premier league. *A & C Black Publishers Ltd*.)

[9] Sports cheetah. (2023). the Evolution of Cricket: A Journey through Its Origins and Development. https://sportscheetah.com/cricket/the-evolution-of-cricket//

[10] (Pawley, A. (1991). How to talk cricket: On linguistic competence in a subject matter. In *Currents in Pacific linguistics: Papers on Austronesian languages and ethnolinguistics in honour of George W. Grace*. Pacific Linguistics.)

[11] (Swartz, T. B. (2017). Research directions in cricket. In *Handbook of*

statistical methods and analyses in sports (pp. 461-476). Chapman and Hall/CRC.)

[12] Bailey, T (1989). *The Greatest Since My Time: The Best Cricketers of the Last Twenty Years and the Major Changes in the Game*. Hodder & Stoughton.

[13] Lewis, A. J (2005). Towards fairer measures of player performance in one-day cricket. *Journal of the Operational Research Society*. *56*, 804-815.

[14] Saikia, H., & Bhattacharjee, D (2011). On classification of all-rounders of the Indian premier league (IPL): a Bayesian approach. *Vikalpa*, *36*(4), 51-66

[15] Tan, A., & Ramachandran, R (2010). Ranking the greatest all-rounder's in test cricket. *Available at www. cricket society. com/ranking_the_greatest_all-ro. pdf, Accessed on October*.

[16] Stevenson, O. G., & Brewer, B. J. (2019). Modelling career trajectories of cricket players using Gaussian processes. In *Bayesian Statistics and New Generations: BAYSM 2018, Warwick, UK, July 2-3 Selected Contributions* (pp. 165-173). Springer International Publishing.

[17] Christie, C. J (2012). *The physical demands of batting and fast bowling in cricket* . (pp. 978-953). ISBN.

[18] Saikia, H., Bhattacharjee, D., & Radhakrishnan, U. K (2016). A new model for player selection in cricket. *International Journal of Performance Analysis in Sport*. *16*(1), 373-388.

[19] Wickramasinghe, I. P (2014). Predicting the performance of batsmen in test cricket. *Journal of Human Sport and Exercise*. *9*(4), 744-751.

[20] Bandulasiri, A., Brown, T., & Wickramasinghe, I (2016). Factors affecting the result of matches in the one-day format of cricket. *Operations Research and Decisions*. *26*(4), 21-32.

[21] Van Staden PJ (2008). Comparison of bowlers, batsmen and all-rounders in cricket using graphical display. Technical Report 08/01. Department of Statistics, University of Pretoria, South Africa.

[22] Kampakis, S., & Thomas, W. (2015). Using machine learning to predict the outcome of english county twenty over cricket matches. *arXiv preprint arXiv:1511.05837*.

[23] Manage, A. B., Kafle, R. C., & Wijekularathna, D. K. (2020). Classification of all-rounders in limited over cricket-a machine learning approach. *Journal of Sports Analytics*, *6*(4), 295-306.

[24] Subhani, M. I., Hasan, S. A., & Osman, M. A. (2012). Will T-20 clean sweep other formats of Cricket in future?

[25] Vistro, D. M., Rasheed, F., & David, L. G. (2019). The cricket winner prediction with application of machine learning and data analytics. *International journal of scientific & technology Research*, *8*(09).

[26] Singh, T., Singla, V., & Bhatia, P. (2015, October). Score and winning prediction in cricket through data mining. In *2015 international conference on soft computing techniques and implementations (ICSCTI)* (pp. 60-66). IEEE.

[27] Modekurti, D. P. V. (2020). Setting final target score in T-20 cricket match by the team batting first. *Journal of Sports Analytics*, *6*(3), 205-213.

[28] Yasir, M., Chen, L. I., Shah, S. A., Akbar, K., & Sarwar, M. U. (2017). Ongoing match prediction in T-20 International. *International Journal of Computer Science and Network Security*, *17*(11), 176-181.

[29] Davis, J., Perera, H., & Swartz, T. B. (2015). A simulator for Twenty20 cricket. *Australian & New Zealand Journal of Statistics*, *57*(1), 55-71.
[30] Shah, S., Hazarika, P. J., & Hazarika, J. (2017). A study on performance of cricket players using factor analysis approach. *International Journal of Advanced Research in Computer Science*, *8*(3), 656-660.
[31] Munir, F., Hasan, M. K., Ahmed, S., & Md Quraish, S. (2015). *Predicting a T-20 cricket match result while the match is in progress* (Doctoral dissertation, Brac University).
[32] Prakash, C. D., & Verma, S. (2022). A new in-form and role-based Deep Player Performance Index for player evaluation in T-20 cricket. *Decision Analytics Journal*, *2*, 100025.
[33] Barr, G. D. I., & Kantor, B. S. (2004). A criterion for comparing and selecting batsmen in limited overs cricket. *Journal of the Operational Research Society*, *55*(12), 1266-1274.
[34] Prakash, C. D., Patvardhan, C., & Singh, S. (2016). A new machine learning based deep performance index for ranking IPL T-20 cricketers. *International Journal of Computer Applications*, *137*(10), 42-49.
[35] ICC Men's T-20 World Cup [Online] 2021/22. Available from ESPN cricinfo. https://www.espncricinfo.com/series/icc-men-s-T-20-world-cup-2021-22-1267897 (accessed 13/05/2023)
[36] Men's T-20 player rankings [Online]. Available from ICC CRICKET. https://www.icc-cricket.com/homepage (accessed 13/05/2023)
[37] ICC player ranking: Men's ranking: Top 100 batsmen [Online]. Available from Cricbuzz. https://www.cricbuzz.com/cricket-stats/icc-rankings/men/batting (accessed 13/05/2023)
[38] Kodinariya, T. M., Makwana, P. R (2013). Review on determining number of Cluster in K-Means Clustering. International Journal.1(6), 90-95.
[39] Rousseeuw, P. J. Silhouettes (1987). a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics. 20, 53-65.
[40] Tibshirani, R., Walther, G., Hastie, T Estimating the number of clusters in a data set via the gap statistic. Journal of the Royal Statistical Society: Series B (Statistical Methodology), **(2001)**. 63(2), 411-423.
[41] Likas, A., Vlassis, N., Verbeek, J. J. The global k-means clustering algorithm. Pattern recognition, **(2003)**. 36(2), 451-461.
[42] Napoleon, D., Pavalakodi, S (2011). A new method for dimensionality reduction using k-means clustering algorithm for high dimensional data set. International Journal of Computer Applications. 13(7), 41-46.