Cricket Analytics: The Past, Present, and Future Direction

Kailee Zeltner, Arja Kaushal, Claudia Reiter, Kaiyan Li, Noah Pettee, Andy Liu

Introduction

Where is Cricket Played?

Cricket is the second most followed sport in the world, second only to soccer (Ray). This may come as a surprise to many Americans, as cricket isn't particularly popular in the US. Cricket's popularity first exploded in England in the 18th century. At this time, England was a world leader and colonizer, and as the British Empire expanded to new colonies across the world, it brought its love for cricket (Beyer). As a result, cricket quickly picked up popularity in India, Pakistan, Sri Lanka, The West Indies, New Zealand, Australia, Zimbabwe and South Africa. While cricket grew to prominence in England, it is currently most popular in India, Pakistan, Sri Lanka, Bangladesh, and Australia (Wood). Cricket is not particularly prevalent or popular in the United States, largely because of the 19th-century-rise of baseball, which displaced cricket as the American pastime ("Cricket in the United States").

How is Cricket Played?

Cricket is a field game played with bat and ball by two teams of 11 players each. A cricket field is a large oval and in the center is the pitch, a rectangular area that's 66 feet long and 10 feet wide. At each of the two ends of the pitch there are wickets, which consist of three sticks with two horizontal pieces, called bails, resting on top.

During a game each team takes a turn batting and bowling, which are analogous to batting and pitching/fielding in baseball. Each turn is called an "inning" and the number of innings in a game is predetermined, but is typically two or four.

There are two batsmen up at one time at each wicket. Bowlers get to bowl six balls at one wicket which completes an over, then another player from the same team bowls six balls to the wicket at the other side of the pitch. When a team is bowling, the aim is to hit the wicket with the balls so that the bails fall off the wicket causing the batter to get out.

The batsman that is being bowled to is called the striker, and the striker's goal is to hit the ball away from the wicket so they can protect the bails from falling off. There are two types of hits - defensive and offensive. A defensive hit is when the ball is hit by the striker, but the two batsmen on the field cannot exchange places without getting out (and therefore a run cannot be scored). An offensive hit is when the two batsmen exchange places and keep doing so until they cannot reach the other wicket without getting out. Each time they exchange places, a run is scored.

This is not the only way to score runs. Around the perimeter of the field, there is a boundary. If a ball is hit beyond the boundary or flies through the air and hits the boundary, this is an automatic six points and the players do not have to run back and forth at the wickets. If the ball just reaches or goes past the boundary (not through the air), this is an automatic four points. The objective of Cricket is to score the most runs by the end of the game. Innings change when 10 out of 11 players on the batting team get out. If both teams cannot complete their number of innings before the time allotted runs out, the match is a draw. Some matches in Cricket can last a day or some can spread over five days, especially professional-level matches.

There are multiple ways to get a batter out. These include:

1. **Bowled:** When the bowler bowls, the ball hits the wicket and causes at least one of the bails to fall off (Common)
2. **Caught Out:** The batter hits a fly ball and a player of the opposing team catches it causing the player to be out (Common)
3. **Run Out:** The fielding team manages to hit the wickets and dislodge the bails if while the batsmen are running between the wickets, any part of the batsmen's bodies or bat are not grounded behind the crease line. (Common)
4. **Hit Wicket:** A player hits their stumps (the three sticks) with the bat or their body while attempting to hit the ball or run between wickets (Not Common)
5. **Stumped:** When a batsman leaves the crease to try and make a shot, but the batsman on the other side misses the ball and the wicketkeeper knocks the bails off causing the batsman to be out since he wasn't behind the crease. (Common)
6. **Leg Before Wicket:** A batsman attempts to block a bowl from hitting the wicket with their pads or other parts of their body (Common)
7. **Obstructing the Field:** Batsman uses their hand or leg to stop the ball from hitting the stumps, shout, or do some action to distract a fielder from making a catch. (Not Common)
8. **Handling the Ball:** Batsman handles the ball without permission from the fielding side (Not Common)
9. **Timed Out:** Batsman takes too long to appear on the field (specifically 3 minutes) (Not Common)
10. **Retired:** Batsman leaves the field before their innings are done, due to injury and the umpire does not let them come back to complete their innings. (Not Common)

Cricket Formats

There are three primary formats of cricket played around the world: Test cricket, One Day International, and T20.

The traditional form of cricket is called test cricket. In this version, games are played over the course of five days and each team bats twice and bowls twice. An inning ends when the batting team is completely out or if they believe they have an insurmountable lead and declare the inning to be over. In test cricket there are no limits to the amount of overs in an inning or the amount of balls a bowler may bowl. In the One Day International (ODI) format, each team bats once with each side having a limit of 50 overs. In T20 formats, there are 20 overs per side with each side batting once. T20 cricket matches last for an average of 3 hours and was introduced as a newer and faster-paced form of competition ("The Three Formats of Cricket").

When is Cricket Played?

Due to its popularity around the globe, cricket matches are played all around the world year-round. Cricket "seasons" generally line up with warmer and dryer weather, as this allows for optimal cricket playing conditions. In England, cricket is typically played between April and September. In India, Sri Lanka, and Australia, the height of the cricket season begins in September and lasts through the spring months (March/April). The Indian Premier League, one of the largest professional cricket leagues in the world, operates for approximately two months between March and May (Harris).

The largest global cricket event, the Cricket World Cup, is organized by the International Cricket Council. Occurring every 4 years, this tournament pits 10 teams in a month-long ODI competition that crowns a world champion. The time of year in which the World Cup is hosted depends on the host nation. For example, the 2023 edition of the Cricket World Cup will be held in India from October to November. The T20 World Cup, also hosted by the ICC, is a 20 team competition held every two years and also varies in its scheduling. The most recent T20 World Cup was held in Australia from October to November 2022 (ICC).

Capital Investments in Cricket

Although cricket falls outside of the top ten most watched sports in the US, cricket is the world's second most popular sport. According to the *Collector*, cricket has 2.5 billion followers around the world, largely stemming from India. A study from Forbes estimated that the average value of teams in the Indian Premier League rose from $67 million to $1.04 billion from 2009 to 2022. These figures reveal that the Indian Premier League has been growing in value faster than the NFL and the NBA over the past decade (Ray). Leveraging the league's outstanding expected growth, the IPL has negotiated broadcasting deals worth $6.2 billion for the next five years, up from $2.5 billion from the previous five years (Ray). As the sport continues to put up numbers on par with the world's biggest sports leagues, investment into cricket has been skyrocketing. According to the New York Times, some of America's largest sports investors, such as Alec Scheiner, former president of the Cleveland Browns and VP of the Dallas Cowboys, and Gerry Cardinale, partial owner of A.C. Milan and various other sports organizations, have begun investing in the IPL. The question remains whether cricket will be able to make such large strides in other countries around the world. The International Cricket Council hopes to break into larger markets like the US and China to increase global popularity (Jakeman).

Because of cricket's global popularity, sports analytics within the game of cricket have been around since the late 90s/ early 2000s and have become much more prevalent and in-depth ("Story of One of Cricket's First Analytics Pioneers"). With cricket's established prominence worldwide and potential for growth in untapped markets like America, it was a natural choice to perform a deep analysis on the way the game is analyzed and its role in global sports culture. In addition, cricket fundamentally has many similarities to baseball which is an established institution within the United States, and since baseball was the birthplace of sports analytics, we believe that there are many aspects of baseball analytics that can be incorporated to cricket in order to further improve the game.

Literature Review

Overview of Past Research in Cricket

Tim B. Swartz writes an excellent summary of cricket research in the Handbook of Statistical Methods and Analyses in Sports (2016). His chapter, titled Research Directions in Cricket, explores every major research question addressed by cricket analysts, giving a broad overview of the cricket analytics industry. According to Swartz, cricket analytics is expected to continue its tremendous growth because match data is easy to access via a number of online sources, and complex computer tracking systems, called Hawk-Eye, have already been established for reviewing plays, which give analysts access to spatial data. With this extensive data, cricket analysts have dove into a number of topics, including batting streakiness (Lemmer 2004, Borooah and Mangan 2010), home team advantage (Stefani 2008), and player chemistry (Valero and Swartz 2012). In terms of tactics, the strategy of declaration, when a team ends their own inning, has been studied extensively by Scarf and Akhtar (2011) and Perera et al. (2014). These studies concluded that teams generally do not declare as early as they should in the match. Batting order has also been studied, but conclusions are hard to reach, as Swartz points out that one team of 11 players has about 40 million possible lineup combinations. Finally, Swartz touches on fantasy cricket leagues, recognizing that research even exists on optimizing fantasy lineups (Brettenny et al. 2012). While relevant topics have been discussed in this overview, Swartz's chapter is extremely comprehensive. Given the limitations of the scope of this paper, additional topics from this chapter have been left out.

Evaluating Batsmen Performance

General to any sport is a coach's dilemma of how to evaluate player performance, and cricket is no different. For batsmen the batting average has long been used as the standard metric (Shah). However, as Uday Damodaran pointed out in his 2006 research, batting average is a flawed metric because of how it handles "not-out" innings (Damodaran). In cricket it's relatively common for a final inning to be prematurely terminated (ie before all 11 batsmen have gotten out) because the batting team has already surpassed the score of the other team. As a result, the number of runs a batsman is recorded as earning in this inning is lower than what it would have been had the inning been carried out to completion. Thus, when calculating batting average, a batsman who typically bats later in the line-up and has lots of "not-out" innings can have an artificially high or artificially low batting average, depending on how the calculation is carried out (Shah). Historically, to calculate batting average the numerator is the sum of all runs scored by a batsman over all innings played in a given period of time (ie a season, player's career, etc), and the denominator is the number of innings played excluding "not-out" innings in which the batsman never got out (Shah). As a result, if two batsmen of similar caliber were compared to one another, a batsmen who has more "not-out" innings would have a higher batting average because their denominator would be smaller. Thus, using batting average alone is not an ideal measure of a batsman's performance.

To try and resolve this issue of "not-out" innings and their impact on batting average, statistics researcher Uday Damodaran created a new methodology of evaluating batsmen. Under Damodaran's methodology a Bayesian approach is used to estimate the number of additional runs a batsman would have gone on to score had a "not-out" inning been played to completion and that player got out. Once these adjusted runs scored values have been determined, the Damodaran method involves creating stochastic dominance curves that visually illustrate a batsman's probability of scoring given the current score. As a result, different batsmen can quickly and easily be compared and ranked against one another by overlaying plots of their stochastic dominance curves in order to determine the stronger batsman, more consistent batsman, etc (Damodaran).

More recent research, done by Oliver George Stevenson and Brendon Brewer in 2017, has focused on how batting performance changes over the course of a game. Stevenson and Brewer created a Bayesian based survival analysis methodology that can be used to predict a batsman's performance. They found that typically a batsman's performance improves over the course of a game, and it takes time for a batsman to reach their optimal/peak performance (Stevenson). Interestingly, they were able to quantify this result and found that the typical batsman starts a game with somewhere between only 25% to 50% of their potential batting ability (Stevenson). This research has allowed for more complex analyses of batsmen, as the time it takes a batsman to reach their full potential in a game can be taken into consideration.

While it's clear that there's a wide array of different analytics that can be used to evaluate batsmen, there's far less research and analytics when it comes to evaluating bowlers. As it stands the only bowler based metric is the combined bowling rate, which is a combination of the average number of runs a bowler gives up as well as their average strikeout rate (Shah).

Cricket Decision Making: Who to Play

Another important decision for coaches is deciding which players on a roster to put in for a game. Currently, optimization techniques are being explored to tackle this problem and come up with the optimal subset of players. Some of these optimization techniques include integer programming methods and multi-objective algorithms that require relatively simple inputs such as basic statistics recorded by the box office (Shah). That being said, a major limitation of this area of data analytics is that these methods fail to account for the opposing team (Shah). The optimal team of players to put in for a game likely varies from game to game based on the strengths and weaknesses of the other team. Thus, future research should focus on algorithms that incorporate data on the opposing team to improve this decision-making process.

Artificial Intelligence- Generating Highlight Reels

Currently, data analysts have begun exploring how artificial intelligence and machine learning can be applied to the sport of cricket. One really cool application of AI is generating highlight reels of cricket games (Shah). As cricket games often last upwards of 8 hours, it's a very time-consuming process for analysts to go through all the film from a game and cut out only the clips that contain the most significant and interesting plays. Thus, researchers have begun exploring how to automate this process with semantic segmentation (Shah). Under semantic segmentation a computer groups together regions of an image and then categorizes each grouping. For a cricket match, the images analyzed are each of the still frames that make up the video. Since 2004 different methods of semantic segmentation have been utilized to classify cricket frames with the aim of determining if a frame is "highlight-worthy". One method is the Hidden Markov Model which analyzes each frame and predicts if that frame is part of a "ball movement", "fielding", or "wicket" shot (Shah). Then, a second algorithm is used to go through all the frames in each sub-category and determine which frames are "highlight-worthy."

A newer method that's been developed uses AI to count the proportion of pixels in a frame that are green, blue, and red (Shah). The computer algorithm groups together frames with similar pixel color proportions, and then classifies certain color proportions as being "highlight-worthy." In 2011 this method was performed by Hao Tang, Vivek Kwatra, and Mehmet Emre Sargin and was able to achieve a nearly 88% accuracy at categorizing a frame as highlight worthy or insignificant (Tang).

In addition to the pixel color method, another way to determine if a frame is highlight worthy is by analyzing the stance and pose of the umpire. Oftentimes an umpire's body language and hand symbols can reveal a lot of information about the current play. Thus, researchers Aravind Ravi, Harshwin Venugopal, and Sruthy Paul have generated a methodology called SNOW that is able to detect and categorize the posing of the umpire. In order to find and extract shots of the umpire from within hours of footage of a cricket game, this research group uses convolutional neural networks, CNNs. Then, the algorithm takes each frame and categorizes it based upon the umpire's stance. Ravi's group created 4 categories that frames can fall into: a sixer, a no ball, an out, or a wide. This methodology has been incredibly successful at generating highlight reels, as it achieved a 96% accuracy at classifying frames as highlight-worthy (Ravi).

<u>Artificial Intelligence- Automating Umpires</u>

In addition to using AI to generate highlight reels, there's growing interest in automating an umpire's determination of whether a bowl is legal or illegal (Shah). The hope is that utilizing AI to perform this process will help eliminate bias and make the process more standardized. Nikhil Batra, Harsh Gupta, Nakul Yadav, and Anshika Gupta pioneered this area of research in 2014 by creating the first automated system to decide if a bowl constitutes a "no-ball". More recently, in 2019 Md Nafee Al Islam, Tanzil Bin Hassan, and Siamul Karim Khan have expanded upon this idea and are working to classify bowls in greater detail, such as what type of bowl was thrown (ie googlie, yorker, bouncer, etc). This research group has done so using CNNs and has been able to achieve a 93% accuracy at classifying bowls (Al Islam).

In addition to classifying the bowl, it's also useful to classify the batsman's hit so that analyses can be performed to determine optimal swing types, predict the ball's trajectory, etc. Originally, to determine and classify a hit, analysts had to watch many different camera angles' worth of footage and manually make a decision. Now in the growing age of AI, researchers D Karmaker, A Chowdhury, and M Miah are training software to make this decision by measuring the angle of a shot, as well as the batsman's body movements (Karmaker).

Duckworth-Lewis Method

This paper would be remiss without discussing arguably the most well-known use of data science and analytics in cricket: The Duckworth-Lewis Method. This method was first created in 1988 and has been adopted by the ICC (Shah). The structure of cricket makes inclement weather a rather prominent challenge. Unlike baseball in which the teams batting and fielding frequently switch over the course of the game, in some forms of cricket, one team bats for the entire first half of the game and then switches to fielding for the entire second half. This presents concerns when it comes to inclement weather. If a game needs to be ended early due to weather, how do you account for one team having significantly less time batting? This is where the Duckworth-Lewis Method comes into play. This is a statistical prediction method that accounts for a team's remaining overs and wickets in order to calculate a new winning score that the team second to bat must score in order to win the game, rather than needing to beat the first team to bat's score (Shah).

Machine Learning: Predicting Match Outcomes

Many previous studies have attempted to use supervised and unsupervised machine learning to predict the total score of ODI cricket matches, measured as the total number of runs from each team in one match. Among these studies include Agarwal et al. (2017), Aburas et al. (2018), Vistro, Rasheed, and David (2019), and Awan et al. (2021). The methods of these studies range from simple linear regression to XGBoost classification. Awan et al. (2021) argues that of all previous methodologies, they have discovered the model that yields the highest accuracy percentage. They use the Spark ML linear regression package to model total score with ball-by-ball data, including runs, overs, runs in the last five overs, wickets in the last five overs, and more. Their model predicts a match's total score, as well as the team that will win the game, using live statistics during the match with around 95% accuracy. The success of this model is just one example of many representing the advancements of machine learning techniques in cricket analytics.

Gambling in Cricket

Match betting in cricket is rumored to have been around since the 1600s and grew wildly in popularity in correlation with cricket's rise as one of the most popular sports in the world. In recent years, cricket is estimated to bring India the equivalent of roughly 201 million U.S. dollars, despite the fact that gambling is illegal in the vast majority of the country ("How Cricket Became One of the Leading Sports to Bet On."). There are many types of bets a person can make on a cricket game, but some of the most popular include betting on the outcome of the match, outcome of each ball, who will win "man of the match", who is the best bowler, and which player will score the highest number of runs (Easton & Uylangco). While sports betting is illegal in many countries, inadequate legislation has allowed cricket betting to thrive. This is a big reason why many countries are being pushed to legalize betting in order to benefit (via taxes) from the large revenues that cricket gambling brings in.

Evaluating the Best Players and Teams in Cricket

The ICC has a list of the top 100 rated batters, bowlers, and all-rounders based on a complex algorithm for Test, ODI, and T20 matches that is updated after each game (Katewa). This algorithm has no human involvement or subjectivity and is composed by looking at many overall statistics for each player and results in a rating that can be anywhere from 0 to 1000. For batters in test matches, the ratings are calculated using the runs scored by the batsman, the rating of the opposition's bowlers, the level of run scoring (in short, what percentage of the team's runs did the batsman score), not-out innings (innings where the game was called short due to the team batting second scoring more runs than the other team but not all 11 outs were obtained), and the result of the match. For bowlers in test matches, they calculate the ratings using wickets taken and runs conceded, rating of the batsmen dismissed, levels of run scoring, heavy workload of bowler, and result of the match. All-rounders are rated based on (batting points x bowling points) / 1000. There are several other factors included for ODI and T20 matches as those have different constraints on the game as compared to test matches (Katewa). Currently, the best batter in test cricket is Marnus Labuschagne with a rating of 915, the best bowler in test cricket is Ravichandran Ashwin with a rating of 869, and the best all-rounder in test cricket is Ravindra Jadeja with a rating of 431.

The way teams are ranked in cricket differs significantly from many other sports, as the rating is calculated by taking the total number of points one team has (points not being the same as runs scored) and dividing by the total number of matches played within the last 4 years. Points are determined by an algorithm that includes the result of the series of games between two countries (1 point to a team for each game won, ½ point to each team for each game drawn, 1 point to the team that wins the series, ½ point to each team if the series is drawn) and then multiplying each team's number of points from the series by another number which is determined by how close in rating the two teams are at the commencement of the series (Katewa). Under this system, the top rated teams in test cricket are India with a rating of 121, Australia with a rating of 116, and England with a rating of 114. Upon looking at the rankings for Test, ODI, and T20I, the teams that are near the top in all three categories and are considered the best overall teams in the world are India, Australia, England, New Zealand, and Pakistan.

Future Work

Exit Velocity

As described in the literature review, evaluating batsmen performance in cricket is one of the most prevalent analytical issues. Currently, batsmen performance is traditionally evaluated using batting average, which is skewed by the batter's "not-out" innings. Batting average is a very situational metric and fails to consider the physical quality of the batsman's actual hit. One unique solution to this problem is applying hitting metrics from baseball to cricket.

Cricket and baseball are fundamentally similar in many ways, but baseball analytics are far more developed than that of cricket. Thus, we propose adopting some of baseball's analytical methods and modifying them to suit the needs of cricket.

In baseball, hitting performance is often evaluated using off-the-bat metrics like exit velocity and launch angle. Exit velocity describes the initial speed of the ball immediately after the batter makes contact with the ball. Launch angle is the vertical angle at which a batter hits the ball off the bat. Specific combinations of exit velocity and launch angle make a ball well-hit. In particular, baseball analytics uses the term "barrel" to describe a good hit. The minimum exit velocity for a barrel is 98 mph, and different ranges of launch angle are used to classify barrels based on different levels of exit velocity. Traditionally, exit velocity, launch angle, and the combination of the two, represented by barrels, are excellent measurements of a batter's hitting quality.
To apply these metrics to cricket, some adjustments must be made. Exit velocity in cricket should describe the initial speed of the ball off of the batsman's bat. Similarly, launch angle in cricket should describe the vertical angle of the ball off of the bat. Contrary to baseball, cricket would require an additional metric that represents the direction of the ball from the perspective of the batter. In baseball, batters can only hit the ball in play, which is mapped by two foul lines that intersect at home plate and form a 90 degree angle. Thus, a batter can hit the ball up to 45 degrees to the right and 45 degrees to the left. In cricket, a batsman can hit the ball anywhere, giving them a 360 degree angle to hit. Although the placement of the ball is important in both sports, the birds-eye angle at which the ball leaves the bat is much more critical to the quality of a hit in cricket than it is in baseball. For simplicity, we will refer to this metric as launch direction. As such, the quality of a hit in cricket should be measured by three metrics, exit velocity, launch angle, and launch direction, as opposed to just the two metrics, exit velocity and launch angle, that are used in baseball.

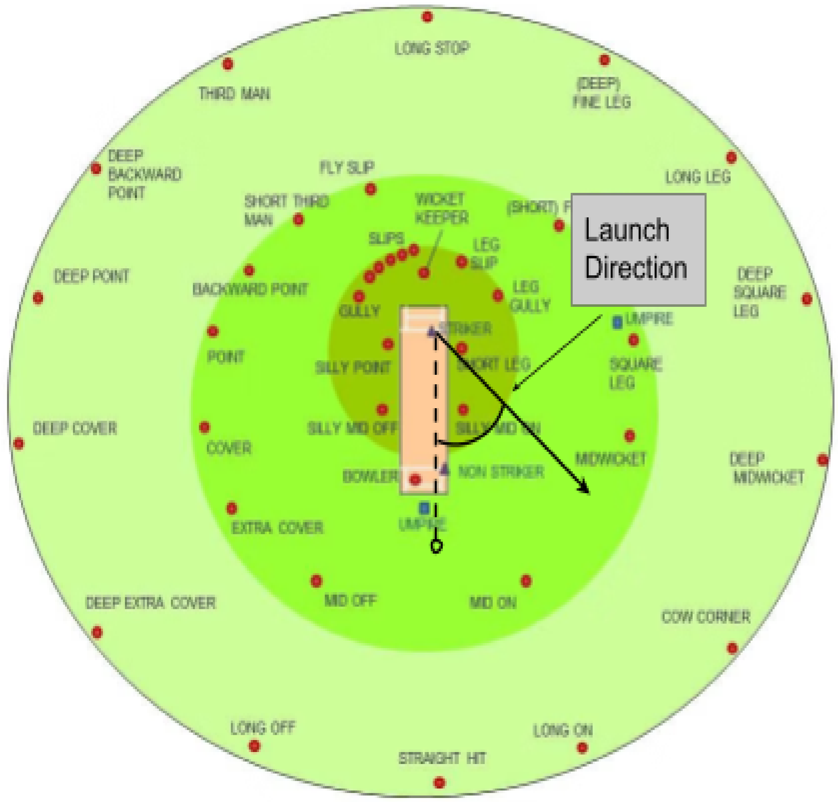

Figure 1. Visual depiction of launch direction metric.

From these three metrics, the quality of a hit can be summarized by certain combinations of these metrics, much like barrels in baseball. A better hit in cricket has a higher exit velocity, exactly like baseball. Accounting for launch angle, cricket follows the same pattern of baseball where a hit with higher velocity can have a wider range of launch angle to maintain the characterization of a good hit. This idea differs from baseball in how launch direction plays into the quality of a hit. In most types of cricket, the pitch is an oval shape, so the distance between the batsman and the boundary line is shorter if the launch direction is 90 or 270 degrees compared to the distance between the batsman and the boundary line if the launch direction is 0 or 180 degrees. Thus, a launch direction closer to 90 or 270 degrees would require lower exit velocity and greater range of launch angle to classify as a good hit. Conversely, a launch direction closer to 0 or 180 degrees would require a higher exit velocity and stricter range of launch angle to classify as a good hit. While batting quality can be approximated by the described trends, the precise combinations of measurements of exit velocity, launch angle, and launch direction to classify good hits and bad hits would require extensive research using thorough data and complex statistical methods.

Umpire Bias

From the literature review, it's apparent that AI models using machine learning and deeping learning have been theorized as a way to eliminate bias in cricket by automating the process of classifying delivery, such as classifying whether a bowl is legal or not. In addition, we know that in many sports there exists substantial referee bias. For example, in football there has been research on referee bias including that of Butler and Butler (2015) on big club favoritism and Dohmen (2008) on home team favoritism (Buraimo). Therefore, we are interested in whether cricket games and modeling is also subject to umpire bias.

In football, referee bias can be reflected by the stoppage time, while in cricket, we believe that umpire bias can be reflected by the determination of delivery type. For example, at the end of a game (assuming that the home team defends and their score is behind), if the umpire is more likely to classify the home team's ambiguous delivery as legal, then it somewhat indicates that there exists umpire bias. Umpire bias holds significance within games as it can influence audience reactions, thereby indirectly impacting the revenue of future games.

Therefore, our idea is creating a model which includes the number of ambiguous deliveries determined to be legal, the number of ambiguous deliveries determined to be illegal, the length of game time when the delivery decision is made, whether the team is home or away, whether one team has a significantly larger fan base than the other, the audience's satisfaction level, and the revenue difference between different games played by the same team, where audience satisfaction and game revenue are dependent variables. This model aims to determine the existence of umpire bias and assess its influence on audience satisfaction and game revenue.

Subsequently, the conclusions from this model could be utilized to modify the dataset, specifically by optimizing umpire decisions on the classification of delivery types. Finally, using this modified dataset, we intend to train the AI models such as Convolutional Neural Networks (CNN) to improve the classification process. Compared with the AI model in the literature review, rather than entirely eliminating bias, the focus shifts towards optimizing bias, especially for deliveries that pose challenges in determining their legality. Appropriately classifying such deliveries with intent is considered more favorable than striving for complete accuracy throughout the majority of cases.

Guiding a Batsman's Decision-Making

We propose that another future direction for cricket analytics could be the development of a model that provides batsmen with the optimized number of wicket runs they should attempt based upon how far the ball has been hit. After a hit, both batsmen must make a decision on how many scores they're going to attempt. In baseball this decision doesn’t necessarily need to be unanimous. A base runner on 2nd can run past 3rd and continue on to home while the hitter stops at 1st base. However, in cricket it’s imperative that the two base runners make a collective decision on when to stop running as there’s only 2 wickets and each wicket can have only 1 baserunner on it at one time. Thus, a baseman can only stop running if and when the other baseman does so. Therefore, one way to improve a cricket team’s performance is by developing a quantitative system to optimize this decision. Rather than batsmen simply following their gut, analytics can be used to guide batsmen to make the optimal decision of how many wickets to run, and ensure that both batsmen are following the same plan.

An initial generic model could first be created and implemented that determines the optimal number of wickets to run based on the distance the ball is hit and the fielding abilities of the average fielder. This simple model would estimate the average speed and distance of a ball thrown by an average fielder. From there, it could easily be calculated how long it will take average fielders to return a hit ball back to the wickets depending on the distance the ball is hit. In order to determine how many wickets the batsmen should run, the speed of the slower batsmen should be used, and it can be calculated how far they’re able to run during the time duration until the ball is returned to the wickets. This distance can then be converted into an optimal number of wickets attempted.

Once this simple model has been generated, it can be computed that batsmen should attempt 1 wicket if a hit ball falls within a distance of a particular range, 2 wickets if a hit ball falls within a slightly larger distance, and so on. These distances can then be used to create a ring-based model of the field with shells representing that if a ball is hit within shell X, the batsmen should attempt to run N wickets. We’ve created a mock-up template of what this model would look like. Note that the distances and values we’ve input as the optimal runs are arbitrary estimates in order to illustrate what the final product could look like.

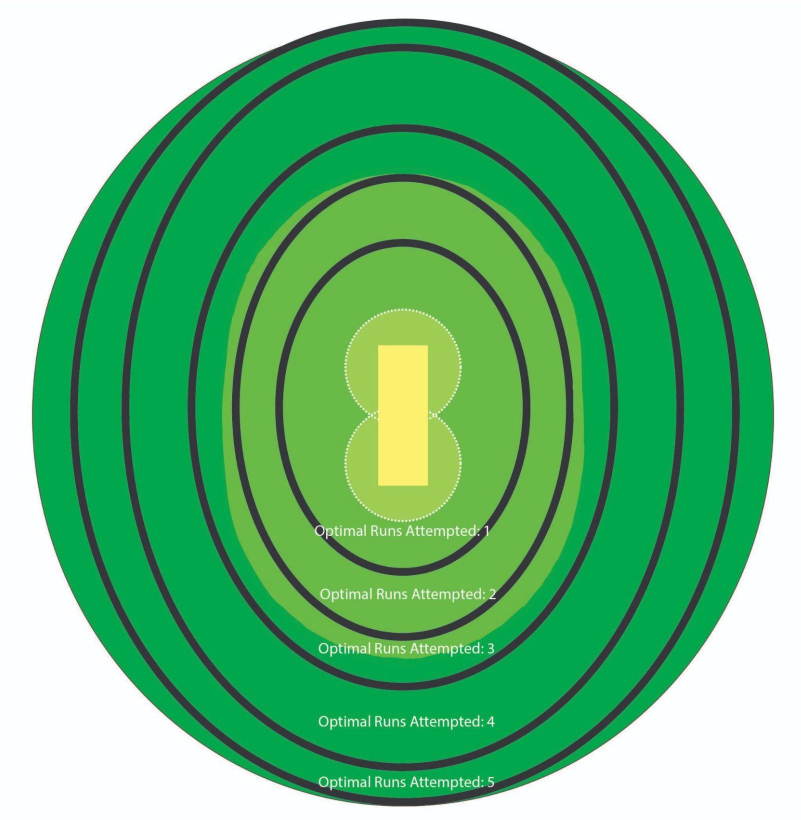

Figure 2. Mock-up sketch of hit distance zones. If a hit ball lands within a particular shell, the batsmen should attempt to run the optimal runs attempted value (determined via optimization modeling).

Once this model is generated based upon the speed of the slower batsmen (and estimations of the fielding abilities of the average fielder), the batsmen pair have a quantitative guide on how far they should attempt to run given the distance the ball is hit. From there, during training batsmen can practice incorporating this model by utilizing balls with built-in sensors. These sensors can give feedback on the exact distance from the wicket the ball is hit, such that batsmen can get practice gauging what a hit that falls within the different shells visually looks and feels like. Then when it comes time for a match batsmen are well-equipped to know how many wicket runs they should attempt based on where the hit ball lands.

To further improve the complexity of this model, the particular fielding abilities of the opposing team can be accounted for. Rather than simply estimating the fielding speed and distance of an average cricket player, analysts can generate more specific ring models that are particular to an upcoming game against a specific team. In this way, the radii of the rings will be shifted to account for the caliber of the opposing team's fielders in order to make the predictions for optimized runs attempted more accurate. Similarly, to expand the model even more so the positioning of the hit ball can be further accounted for. An opposing team's fielder in one spot on the field might be significantly worse of a fielder than a fielder in another spot. Thus, a batting team can account for the weak spots of their opponents by adjusting the size of the rings to be oblong, based upon the positions of various fielders on the opposing team. We've created a mock-up sketch of what this improved model could look like, accounting for a particularly strong fielder on one side of the field and a particularly weak fielder on another side.

(Note: Once again this is a rough mock-up, and the values expressed as well as the distances shown aren't to scale or based upon calculated optimized values. It's simply to illustrate what type of visual model could be created if optimization techniques are performed).

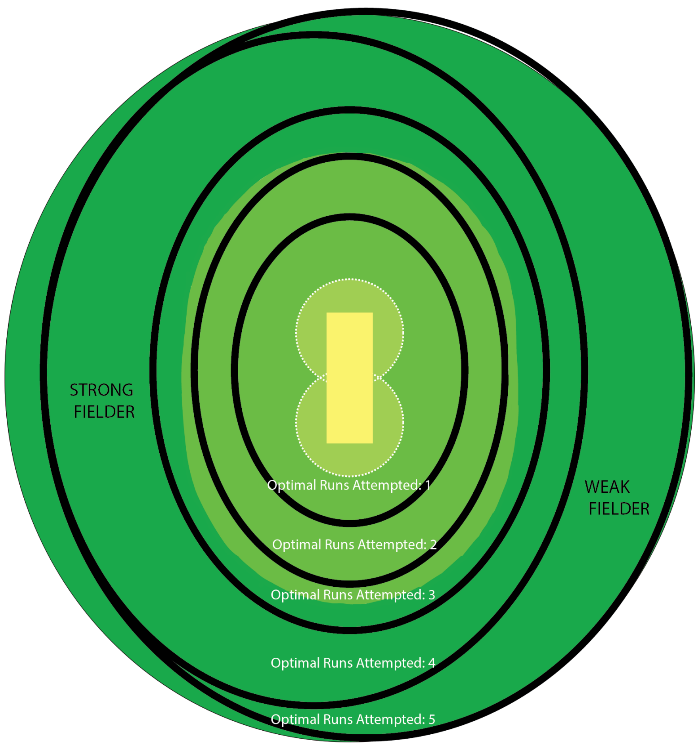

Figure 3. Adjusted mock-up accounting for a particularly strong fielder on the left side of the fielder and a particularly weak fielder on the right side. Since there's a strong fielder on the left side, if the ball is hit in this area, it's no longer optimal to attempt 5 runs, but rather attempting 4 runs becomes more strategic. Similarly, since there's a weak fielder on the right side of the field, if a ball is hit in this area the optimal number of runs attempted increases relative to Figure 1.

<u>Bowler Performance</u>

From the literature review, we found that the ICC ranked bowlers using multiple factors including wickets taken and runs conceded, the rating of batsmen dismissed, levels of run scoring, the workload of the bowler, and the result of the match. We can see that bowlers are generally rated by the level of batsmen they face and the result of the bowl. However, from our perspective, we think this ranking model could be improved by including some other factors such as the level of the defenders surrounding the bowler, errors committed by these players, and the combined effect of a bowler's and batman's preference/performance on types of deliveries.

If the level of the defense is high, then the performance of the bowler may be elevated due to the team's great defending. If the bowler gains runs from the batsman due to errors committed by the defense, then this score should not contribute to the bowler's ranting because it was likely not their fault. This concept is similar to the earned run in baseball. Bowler rating also does not take the combination of bowler and batsman preference for deliveries into consideration. If the batsman is skilled at hitting sliders and the slider is not a strength of the bowler delivering the ball to the hitter then it is not surprising the bowler will allow runs. In this scenario it is more likely to be regarded as a decision-making misplay rather than the bowler's poor performance. Therefore, in our model, we'll take these variables into consideration.

For the defenders in our model, assuming we have the ratings for defenders obtained from a considerate model, we subtract the rating of average defenders from that of the individual defenders to create a new variable representing the level of the defense above average. Then, we'll give a negative coefficient to this variable to decrease the influence of the performance of the defender on the bowler's rating. For defender errors, we regard the earned runs of the bowler to be 0 when the runs are caused by these errors. Similarly, if the runs are prevented by great plays of the defense, then we may consider adding earned runs to the bowler as a penalty. Creating criteria establishing when an earned run penalty may prove challenging and different situations would require different considerations. Substituting the established "runs conceded" metric with a new earned runs variable better represents bowler performance. Moreover, for the interaction between bowler and batsman preference in delivery types, we first rate batsmen and bowler's performances on different delivery types using a 0-10 scale based on the hitting percentages when facing or delivering different types of bowls. Then, we may subtract the hitter's rating on this scale from the bowler's rating to create a new variable representing the difference in preference/performance. If the absolute value of the preference difference is high, then give a penalty coefficient to a particular type of bowl. These additions may make the rankings of bowler performance better representative of the true skill of the players.

Conclusion

This project was a great way for us to dive deeper into the world of cricket and consider ways in which one of the world's most popular sports could still be improved. Due to the sport's fundamental similarities with baseball, we were able to look at cricket through the lens of a sport with a much more expansive history and emphasis on analytics and see what aspects could be translated to be of use in cricket. The explosion and improvement of AI has also influenced many of our ideas on how to push the sport forward. Our research showed promising trials of AI accurately and efficiently handling tasks like generating highlight reels to increase the popularity of the sport or officiating games far more accurately than any human ever could.

Before we began this project, only one member of our group knew the rules of cricket. Not understanding the rules of such a complex game was a significant barrier for us in terms of engagement with the sport. Learning the different rules, strategies, and formats of the game has certainly sparked a newfound interest in cricket. Throughout the research we did for this report, the passion for cricket that many have around the world was evident in the enormous amount of resources created to explore the sport and open it up for newcomers like us. Hopefully our contributions aiming to better evaluate the hitting ability of batsmen, create a better officiating system, provide strategy to optimize hitting strategy, and more accurately rate professional cricket bowlers may improve the way cricket is played in the future.

Works Cited

Al Islam, Md Nafee, et. al. "A CNN-based approach to classify cricket bowlers based on their bowling actions," 2019 IEEE International Conference on Signal Processing, Information, Communication & Systems (SPICSCON), Dhaka, Bangladesh, 2019, pp. 130-134, doi: 10.1109/SPICSCON48833.2019.9065090.

Awan, Mazhar Javed, et al. “Cricket Match Analytics Using the Big Data Approach.” MDPI, Multidisciplinary Digital Publishing Institute, 26 Sept. 2021, www.mdpi.com/2079-9292/10/19/2350. Accessed 08 May 2023.

Batra, Nikhil, et. al. "Implementation of augmented reality in cricket for ball tracking and automated decision making for no ball," 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Delhi, India, 2014, pp. 316-321, doi: 10.1109/ICACCI.2014.6968378.

Beyer, Greg. “A History of Cricket: The World’s Second-Most Popular Sport.” The Collector, 8 Apr. 2022, www.thecollector.com/history-of-cricket-worlds-second-most-popular-sport/.

Buraimo, Babatunde, et. al. Handbook of Statistical Methods and Analyses in Sports, by Jim Albert et al., CRC Press, 2016, pp. 401–424.

“Cricket in the United States.” *Wikipedia*, 28 Mar. 2023, en.wikipedia.org/wiki/Cricket_in_the_United_States#cite_note-mlb-47.

Damodaran, Uday. “Stochastic Dominance and Analysis of ODI Batting Performance: the Indian Cricket Team, 1989-2005.” Journal of Sports, Science, and Medicine, vol. 5, no. 4, 2006, pp. 503-508, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3861748/

Easton, Stephen Andrew, and Katherine Uylangco. “An Examination of In-Play Sports Betting Using One-Day Cricket Matches.” *SSRN Electronic Journal*, 2006, https://doi.org/10.2139/ssrn.948013.

Harris, Matt. “When Is Cricket Season and How Long Is It?” *Its Only Cricket*, 13 Jan. 2023, www.itsonlycricket.com/cricket-season.

“How Cricket Became One of the Leading Sports to Bet On.” Female Cricket, 3 Nov. 2021, femalecricket.com/women-cricket-news/20172-how-cricket-became-one-of-the-leading-sports-to-bet-on.html.

ICC. “ICC Men’s Cricket World Cup 2023.” *ICC Cricket*, 21 Nov. 2022, www.icc-cricket.com/news/2914163.

ICC. “ICC Men’s Test Player Rankings.” ICC, www.icc-cricket.com/rankings/mens/player-rankings/test. Accessed 8 May 2023.

Jakeman, Mike. “American Money Has Discovered Indian Cricket.” *The New York Times*, 1 Nov. 2022, www.nytimes.com/2022/11/01/sports/cricket/indian-cricket-us-investors.html.

Karmaker, Debajyoti, et. al. "Cricket shot classification using motion vector," 2015 Second International Conference on Computing Technology and Information Management (ICCTIM), Johor, Malaysia, 2015, pp. 125-129, doi: 10.1109/ICCTIM.2015.7224605.

Katewa, Shrot. “ICC Player Ranking System Explained – in under 5 Minutes!” Cricket Mastery, 16 Jan. 2023, cricketmastery.com/icc-player-rankings/.

Ozanian, Mike. “Indian Premier League Valuations: Cricket Now Has a Place among World’s Most Valuable Sports Teams.” *Forbes*, Forbes Magazine, 6 July 2022, www.forbes.com/sites/mikeozanian/2022/04/26/indian-premier-league-valuations-cricket-now-has-a-place-among-worlds-most-valuable-sports-teams/?sh=f01421e39514. Accessed 08 May 2023.

Ravi, Aravind, et. al. A Dataset and Preliminary Results for Umpire Pose Detection Using SVM Classification of Deep Features," 2018 IEEE Symposium Series on Computational Intelligence (SSCI), Bangalore, India, 2018, pp. 1396-1402, doi: 10.1109/SSCI.2018.8628877.

Ray, Siladitya. “Broadcast Rights for India’s Top Cricket League Sell for $6.2 Billion, Value per Game Second Only to NFL.” Forbes, 16 June 2022, www.forbes.com/sites/siladityaray/2022/06/14/broadcast-rights-for-indias-top-cricket-league-sell-for-62-billion-value-per-game-second-only-to-nfl/.

Shah, Kavya Mehul, and Ammaar Ahmed Saeed. “Innovating a Centuries-Old Sport: How Emerging Data Analytics Tools Are Redefining Cricket.” Harvard Data Science Review, 27 Apr. 2023, hdsr.mitpress.mit.edu/pub/g7fzhuzj/release/1?readingCollection=619bd457.

Smith, Thomas, and Michel Sakr. “How Popular Is Cricket around the World? [Statistics].” *SQaF*, 21 Apr. 2022, sqaf.club/how-popular-is-cricket/. Accessed 08 May 2023.

Stevenson, Oliver George and Brendon J. Brewer. "Bayesian survival analysis of batsmen in Test cricket" Journal of Quantitative Analysis in Sports, vol. 13, no. 1, 2017, pp. 25-36. https://doi.org/10.1515/jqas-2016-0090

“Story of One of Cricket’s First Analytics Pioneers.” *Wicky*, 6 May 2020, wicky.ai/content/analytics/story-of-one-of-crickets-first-analytics-pioneers/.

Swartz, Tim. Handbook of Statistical Methods and Analyses in Sports, by Jim Albert et al., CRC Press, 2016, pp. 445–458.

Tang, Hao, et. al. “Detecting highlights in sports videos: Cricket as a test case," 2011 IEEE International Conference on Multimedia and Expo, Barcelona, 2011, pp. 1-6, doi: 10.1109/ICME.2011.6012139.

“The Three Formats of Cricket.” *International Cricket Council*, 17 Nov. 2020, www.icc-cricket.com/about/cricket/game-formats/the-three-formats.

Wood, Robert, "Popularity of Cricket Around the World." Topend Sports Website, May 2010,https://www.topendsports.com/world/lists/popular-sport/sports/cricket.htm?ref=blog.duolingo.com