Analytics and scope (of mathematics) in the sports industry

(MTRP Camp for seniors, at ISI Kolkata)

Soudeep Deb

Indian Institute of Management Bangalore
Bannerghatta Road, Bengaluru 560069, India.
Email: soudeep@iimb.ac.in.

March 28, 2026

Outline

Who I am

Career Path

- **B.Stat. & M.Stat.**
  Indian Statistical Institute, Kolkata
- **Ph.D. in Statistics**
  University of Chicago
- **Senior Lead Data Scientist**
  NBCUniversal Media
- **Faculty**
  Indian Institute of Management Bangalore

Research Interests

- Time Series Analysis
- Bayesian Methods
- Statistical Machine Learning
- Nonparametric Methods
- Sports Analytics

Academia–Industry Synergy

We are trying to bridge rigorous statistical theory with real-world data science applications in sports.

Outline

Seminar Objectives

By the end of this session, you will:

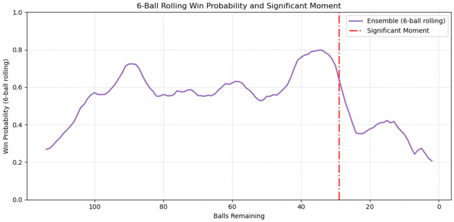
- Understand how advanced mathematics powers sports analytics.
- Apply mathematical reasoning to real-world problems.
- Analyze strategy, gameplay, and betting markets mathematically.
- Explore connections with algebra, combinatorics, geometry, etc. Then, extend it to advances in statistics and machine learning.

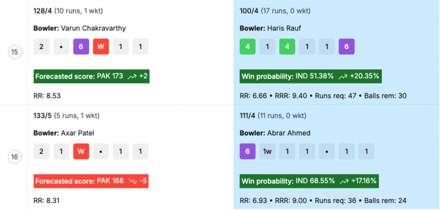
What is Sports Analytics?

- The science of extracting insights from sports data.
- Used by teams, analysts, and betting markets.
- Helps answer questions such as:
  - What is the optimal strategy?
  - Which player is most valuable?
  - What is the true probability of winning?

A sports analytics illustration from ongoing project

We employ a machine learning algorithm to forecast cricket match outcome when the second innings is in play; and accordingly determine he most significant moment of the innings.





A sports analytics illustration from basketball

Let us watch a sequence of a play from a basketball match, created using player tracking data, available in an article from Harvard Data Science Review.

Mathematical Foundations

Field	Applications
Combinatorics	Scheduling, lineups, tournaments
Probability	Predictions, betting markets
Algebra	Ranking systems, regression models
Geometry	Spatial analysis and shot selection
Graph Theory	Passing networks
Number Theory	Tournament scheduling
Optimization	Strategy and decision-making

Outline

Combinatorics: Counting Strategies

- Determines the number of possible outcomes.
- Essential in team selection and tournament design.
- Helps evaluate feasibility and fairness.

Problem 1: Optimal Batting Order

A cricket team has 6 specialist batters and 5 all-rounders.

- Exactly 4 specialist batters must be in the top 5.
- The captain must bat in position 3 or 4.

How many valid batting orders exist?

Problem 2: Round-Robin Scheduling

In a round-robin tournament with n teams, how many matches will be played in total?

Problem 2: Round-Robin Scheduling

In a round-robin tournament with n teams, how many matches will be played in total?

$$\text{Number of Matches} = \binom{n}{2}$$

Problem 2: Round-Robin Scheduling

In a round-robin tournament with n teams, how many matches will be played in total?

$$\text{Number of Matches} = \binom{n}{2}$$

Prove that if n is even, the tournament can be completed in $n - 1$ rounds with each team playing once per round.

(Hint: Use cyclic permutations.)

Problem 3: Knockout Tournaments

In how many ways can 8 teams be paired in the first round?

Problem 3: Knockout Tournaments

In how many ways can 8 teams be paired in the first round?

$$\frac{8!}{2^4 \cdot 4!} = 105$$

Problem 3: Knockout Tournaments

In how many ways can 8 teams be paired in the first round?

$$\frac{8!}{2^4 \cdot 4!} = 105$$

Can you generalize this for 2^n teams?

How many matches are required to find the best team of the tournament?

Statistics and Machine Learning in Scheduling Problems

Motivation

- Scheduling is crucial in sports leagues, tournaments, and broadcasting.
- Objectives include fairness, efficiency, logistics optimization, and revenue maximization.
- Examples: IPL, NBA, Club Football, Olympic timetables.

Key Challenges

- Balanced home and away matches
- Minimizing travel distance and fatigue
- Avoiding venue and time conflicts
- Maximizing audience engagement and broadcast value
- Ensuring fairness and competitive integrity

Statistics and Machine Learning in Scheduling Problems

- **Optimization Techniques:**
  - Integer Programming and Mixed-Integer Linear Programming (MILP)
  - Constraint Programming and Network Optimization
- **Statistical Methods:**
  - Time Series Forecasting for demand and viewership
  - Bayesian Models for uncertainty quantification
- **Machine Learning:**
  - Reinforcement Learning for adaptive scheduling
  - Predictive Models for attendance and revenue
  - Clustering for geographic grouping of teams

Statistics and Machine Learning in Scheduling Problems

Mathematical Formulation

$$\max_{x_{ijt}} \sum_{i,j,t} w_{ijt} x_{ijt}$$

subject to

$$\sum_{j} x_{ijt} = 1, \quad x_{ijt} \in \{0,1\},$$

where $x_{ijt} = 1$ if team i plays team j at time t.

Sports Analytics Insight

Modern leagues use data-driven optimization combining statistics, machine learning, and operations research to design fair, efficient, and commercially successful schedules.

Probability in Sports Analytics

- Models uncertainty and randomness.
- Determines outcome predictions and betting odds.
- Quantifies risk and expected returns.

Problem 4: Penalty Shootout

Two equally skilled teams take penalties alternately. What is the probability that the team shooting first wins?

Problem 5: Binomial Model

A basketball player has a scoring probability of 0.6. What is the probability that the player scores at least 7 out of 10 shots?

Problem 5: Binomial Model

A basketball player has a scoring probability of 0.6. What is the probability that the player scores at least 7 out of 10 shots?

$$P(X \geqslant 7) = \sum_{k=7}^{10} \binom{10}{k} (0.6)^k (0.4)^{10-k}$$

Who will you bet for?

International

World Cup Betting Odds

OVERVIEW FREE BETS MATCHES TABLE **OUTRIGHTS** TIPS

World Cup 2026 Winner

Compare odds (25+ Bookmakers) ›

Spain	5/1	England	6/1	France	8/1
Brazil	17/2	Argentina	17/2	Portugal	12/1
Germany	14/1	Netherlands	22/1	Italy	33/1
Belgium	50/1	Mexico	80/1	USA	80/1
Uruguay	80/1	Croatia	100/1	Ecuador	100/1
Japan	100/1	Senegal	125/1	Denmark	125/1

Problem 6: Expected Value in Betting

Decimal Odds and Implied Probability: $p_{\text{implied}} = \frac{1}{O}$

For a unit stake:

- Profit if the bet wins: $O - 1$
- Loss if the bet loses: 1

$$EV = p(O-1) - (1-p) = pO - 1$$

Interpretation

$EV > 0:$ Profitable bet, $EV = 0:$ Fair bet, $EV < 0:$ Losing bet.

A bet is attractive only if your estimated probability exceeds the bookmaker's implied probability.

Advanced Probability and Statistics in Betting Markets

- Betting markets use sophisticated probabilistic models to set accurate and profitable odds.
- The goal is to manage risk, ensure profitability, and reflect real-time information.
- Modern sportsbooks operate similarly to financial markets.
- Bookmakers introduce a margin (overround):

$$\sum_i \frac{1}{O_i} > 1.$$

- This guarantees long-term profit regardless of the outcome.

Advanced Probability and Statistics in Betting Markets

- **Bayesian Inference:** Updates odds as new information becomes available.
- **Poisson Models:** Predict goals and scores in football and hockey.
- **Time Series Analysis:** Tracks player form and team performance.
- **Monte Carlo Simulation:** Estimates outcome probabilities through repeated simulations.
- **Regression and Machine Learning:** Logistic regression, random forests, and gradient boosting.
- **Elo and Bradley–Terry Models:** Estimate team strengths and win probabilities.

Advanced Probability and Statistics in Betting Markets

Risk Management and Market Efficiency

- Odds are adjusted dynamically to balance betting volumes and minimize exposure.
- Concepts from financial mathematics and arbitrage theory are employed.
- Efficient market principles ensure that prices reflect collective information.

Sports Analytics Insight

Betting markets represent one of the most advanced real-world applications of probability, statistics, machine learning, and quantitative finance.

Geometry in Sports

- Used in football, basketball, and hockey.
- Evaluates angles, distances, and spatial dominance.

Problem 7: Shooting Angle

A goal is 7.32 m wide. A player stands 10 m from the midpoint. Find the angle subtended by the goal.

Problem 7: Shooting Angle

A goal is 7.32 m wide. A player stands 10 m from the midpoint. Find the angle subtended by the goal.

$$\theta = 2\tan^{-1}\left(\frac{3.66}{10}\right)$$

Expected Goals (xG) in Football

$$xG = f(\text{distance}, \text{angle}, \text{defender pressure})$$

Closer and wider angles yield higher scoring probabilities.

Expected Goals (xG) in Football

- xG measures the probability that a shot results in a goal.
- Each shot is assigned a value between 0 and 1 based on data.
- Widely used in modern football analytics.

Key Factors Influencing xG

- Distance from the goal
- Shooting angle
- Type of assist (cross, through-ball, set-piece)
- Body part used (foot, head)
- Defensive pressure and goalkeeper positioning
- Match context (open play, counterattack, penalty)

Expected Goals (xG) in Football

Mathematical Model xG is typically estimated using logistic regression:

$$\text{xG} = P(\text{Goal} \mid \mathbf{x}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)}},$$

where $\mathbf{x}$ represents shot characteristics.

Sports Analytics Insight

- Total xG evaluates team performance beyond the final score.
- Helps identify finishing efficiency and defensive strength.
- Used by clubs, analysts, and platforms such as Opta and StatsBomb.

Problem 8: Optimal Passing

Three attacking players are located at:

$$A(0,0), \quad B(6,0), \quad C(3,4).$$

A defender is positioned at:

$$D(3,1).$$

Tasks:

1. Determine the shortest passing distance.
2. Identify which passes are safe from interception.
3. Compute a risk-adjusted optimal pass.
4. Optimize player positioning along the line $x = 3$.

Solution: Optimal Passing Strategy

Given Positions:

$$A(0,0), \quad B(6,0), \quad C(3,4), \quad D(3,1)$$

Objectives:

1. Determine the shortest passing distance.
2. Identify passes safe from interception.
3. Compute a risk-adjusted optimal pass.
4. Optimize player positioning along the line $x = 3$.

Mathematical Tools

Distance Formula, Coordinate Geometry, Optimization, and Sports Analytics.

Step 1: Shortest Passing Distance

Using the distance formula

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2},$$

$$AB = \sqrt{(6-0)^2 + (0-0)^2} = 6,$$
$$AC = \sqrt{(3-0)^2 + (4-0)^2} = 5,$$
$$BC = \sqrt{(6-3)^2 + (0-4)^2} = 5.$$

Conclusion

Shortest passes are AC and BC, each of length 5.

Insight: Triangle ABC is a 3-4-5 triangle.

Step 2: Defensive Interception Analysis

A pass is interceptable if the defender lies within a perpendicular distance of 1 unit from the passing line and the projection lies on the segment.

Pass AB: Line $y = 0$

$$d_{AB} = 1.$$

Pass AC: Line $4x - 3y = 0$

$$d_{AC} = \frac{|4(3) - 3(1)|}{5} = \frac{9}{5} = 1.8.$$

Pass BC: Line $4x + 3y - 24 = 0$

$$d_{BC} = \frac{|4(3) + 3(1) - 24|}{5} = \frac{9}{5} = 1.8.$$

Step 2: Safety of Passes

To confirm interception, check whether the defender's projection lies on the segment.

Pass	Distance from Defender	Safe?
AB	1	Unsafe (borderline interceptable)
AC	1.8	Safe
BC	1.8	Safe

Conclusion

Passes AC and BC are safe from interception.

Step 3: Risk-Adjusted Optimal Pass

Define a risk-adjusted cost function:

$$\text{Cost} = d + \frac{2}{\delta},$$

where d is the pass length and δ is the defender's distance from the passing line.

$$\text{Cost}_{AB} = 6 + \frac{2}{1} = 8,$$
$$\text{Cost}_{AC} = 5 + \frac{2}{1.8} \approx 6.11,$$
$$\text{Cost}_{BC} = 5 + \frac{2}{1.8} \approx 6.11.$$

Optimal Decision

The optimal passes are AC and BC.

Step 4: Optimization of Player Position

Suppose player C moves along the line $x = 3$ to minimize the total passing distance from A to B via C.
Let $C = (3, y)$. Then

$$AC = \sqrt{9 + y^2}, \quad BC = \sqrt{9 + y^2}.$$

Total distance:

$$f(y) = 2\sqrt{9 + y^2}.$$

Differentiate:

$$f'(y) = \frac{2y}{\sqrt{9 + y^2}}.$$

Setting $f'(y) = 0$ gives:

$$y = 0.$$

Optimal Position of Player C

$$f''(y) = \frac{18}{(9+y^2)^{3/2}} > 0,$$

confirming a minimum at $y = 0$.

Optimal Position

$$\boxed{C = (3, 0)}$$

Minimum total passing distance:

$$f_{\text{min}} = 2\sqrt{9} = 6.$$

Geometric Insight:

- The optimal point lies on line segment AB.
- This is consistent with symmetry and reflection principles.

Final Summary

Task	**Result**
Shortest Passing Distance	$AC = BC = 5$
Safe Passes	AC and BC
Risk-Adjusted Optimal Pass	AC or BC
Optimal Position of C	$(3, 0)$
Minimum Two-Pass Distance	6

Sports Analytics Insight

Mathematics enables teams to optimize passing strategies using geometry, optimization, and risk modeling.

Algebra in Sports

- Used to build predictive models and rankings.
- Forms the basis of performance metrics.

Problem 9: Deduce rating

Given score differences:

A beats B by 10 points
B beats C by 8 points
C beats A by 2 points

How would you define team ratings?

Least Squares Rating Method

Let the team ratings be r_A, r_B, r_C.

$$r_A - r_B = 10,$$
$$r_B - r_C = 8,$$
$$r_C - r_A = 2.$$

We estimate ratings by minimizing:

$$(r_A - r_B - 10)^2 + (r_B - r_C - 8)^2 + (r_C - r_A - 2)^2.$$

To ensure uniqueness:

$$r_A + r_B + r_C = 0.$$

Matrix Formulation

$$X = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 10 \\ 8 \\ 2 \end{bmatrix}.$$

Solve:

$$\min \|X\mathbf{r} - \mathbf{y}\|^2.$$

Normal equations:

$$X^\top X\mathbf{r} = X^\top \mathbf{y}.$$

Final Team Ratings

With the constraint $r_A + r_B + r_C = 0$:

$$\boxed{r_A = \frac{8}{3}, \quad r_B = -\frac{2}{3}, \quad r_C = -2}$$

Ranking

$$A > B > C$$

Sports Analytics Insight

This method forms the basis of Massey ratings, Elo systems, and modern ranking algorithms.

Elo Rating System

- Developed by Arpad Elo to rank chess players.
- Widely used in sports analytics, including football, cricket, tennis.
- Provides a dynamic measure of player or team strength.

Elo Rating System

Expected Score If players (or teams) A and B have ratings R_A and R_B, the expected probability that A wins is:

$$E_A = \frac{1}{1 + 10^{(R_B - R_A)/400}}, \quad E_B = 1 - E_A.$$

Rating Update Rule After a match, ratings are updated as:

$$R'_A = R_A + K(S_A - E_A),$$

where

- $S_A = 1$ (win), 0.5 (draw), 0 (loss),
- E_A is the expected score,
- K is the learning rate or adjustment factor.

Bradley–Terry Model

- A probabilistic model for paired comparisons.
- Introduced by Bradley and Terry (1952).
- Widely used in sports rankings, psychology, and machine learning.

Bradley–Terry Model

Model Formulation Let each team i have a positive strength parameter π_i. The probability that team i defeats team j is:

$$P(i \text{ beats } j) = \frac{\pi_i}{\pi_i + \pi_j}.$$

Logistic Representation Setting $\pi_i = e^{\beta_i}$ gives:

$$P(i \text{ beats } j) = \frac{e^{\beta_i}}{e^{\beta_i} + e^{\beta_j}} = \frac{1}{1 + e^{-(\beta_i - \beta_j)}}.$$

Parameter Estimation

- Estimated using Maximum Likelihood Estimation (MLE).
- Can incorporate covariates such as home advantage.

Passing Networks

- Players are nodes; passes are edges.
- Used to measure team coordination.

Problem 10: Complete Passing Network

If each of 11 players passes to every other player once, how many directed passes occur?

Problem 10: Complete Passing Network

If each of 11 players passes to every other player once, how many directed passes occur?

$$11 \times 10 = 110$$

Generalize for n players.

Problem 11: Training drill

Can a training drill be designed for n players so that each passing link is used exactly once?

Solution: Eulerian Strategy in Training Drills

Graph-Theoretic Interpretation:

- Represent each player as a **vertex**.
- Represent each passing link between two players as an **edge**.
- Drill corresponds to traversing the graph without repeating any pass.

Eulerian Concepts:

- An **Eulerian Path** uses every edge exactly once.
- An **Eulerian Circuit** uses every edge exactly once and returns to the starting player.
- A connected graph has an **Eulerian Circuit** if every vertex has an even degree.
- A connected graph has an **Eulerian Path** if exactly two vertices have odd degrees.

Sports Analytics Insights from Eulerian Paths

- If all players make an even number of passes, the drill forms a closed loop.
- If exactly two players make an odd number of passes, the drill starts and ends with them.
- Otherwise, such a drill is impossible without repeating a pass.

Outline

Optimization in Sports

- Determines optimal strategies.
- Maximizes expected returns.

Problem 12: Optimal Shot Selection

A basketball player:

- Scores 2-pointers with probability 0.55.
- Scores 3-pointers with probability 0.38.

Which strategy maximizes expected points?

Advanced Machine Learning for Optimal Shot Selection

Motivation

- Teams seek to maximize scoring by selecting high-quality shots.
- Data-driven decision-making has transformed sports such as basketball and football.
- Examples include the rise of three-point shooting in the NBA and Expected Goals (xG) in football.

Key Data Inputs

- Player and ball tracking data (spatial coordinates and movement)
- Distance and angle to goal or basket
- Defensive pressure and positioning
- Shot type and player skill
- Time remaining and match context

Advanced Machine Learning for Optimal Shot Selection

Machine Learning Models

- **Logistic Regression:** Estimates scoring probabilities
- **Tree-Based Methods:** Random Forests and Gradient Boosting
- **Deep Learning:** Convolutional and Recurrent Neural Networks
- **Reinforcement Learning:** Learns optimal shot policies through sequential decision-making.

Mathematical Formulation Let $\mathbf{x}$ denote contextual features. The probability of scoring is:

$$P(\text{Score} \mid \mathbf{x}) = \frac{1}{1 + e^{-(\beta_0 + \boldsymbol{\beta}^\top \mathbf{x})}}.$$

Optimal shot selection maximizes expected value:

$$\text{Optimal Shot} = \arg\max_{s \in \mathcal{S}} \left[P(\text{Score} \mid s) \times \text{Point Value}(s) \right].$$

Advanced Machine Learning for Optimal Shot Selection

Applications

- Basketball: Optimal three-point vs. mid-range shot selection.
- Football: Expected Goals (xG) analysis.
- Cricket: Boundary-hitting strategies in T20 formats.

Sports Analytics Insight

Advanced machine learning converts tracking data into actionable strategies, enabling teams to identify and exploit the highest-value scoring opportunities.

Optimization in IPL Auction Strategy

- The Indian Premier League (IPL) auction requires franchises to build a competitive team under financial and structural constraints.
- Decisions must balance player performance, team composition, and budget efficiency.
- Data-driven optimization plays a crucial role in strategic bidding.

Optimization in IPL Auction Strategy

Let

$$x_i = \begin{cases} 1, & \text{if player } i \text{ is selected,} \\ 0, & \text{otherwise.} \end{cases}$$

$$\max \sum_{i=1}^{n} v_i x_i$$

subject to:

$$\sum_{i=1}^{n} c_i x_i \leqslant B \quad \text{(Budget Constraint)}$$

$$\sum_{i \in \mathcal{R}_k} x_i \geqslant r_k \quad \text{(Role Requirements)}$$

$$\sum_{i \in \mathcal{F}} x_i \leqslant F \quad \text{(Overseas Player Limit)}$$

Optimization in IPL Auction Strategy

Methodologies

- Mixed-Integer Linear Programming (MILP)
- Dynamic Programming and Knapsack Optimization
- Bayesian Valuation Models
- Reinforcement Learning for Sequential Bidding
- Monte Carlo Simulation for uncertainty analysis

Sports Analytics Insight

IPL franchises leverage optimization, statistical modeling, and machine learning to construct high-performing squads while maximizing return on investment.

Outline

Moneyball

- Used statistical analysis in baseball.
- Focused on on-base percentage.
- Revolutionized team-building strategies.

Moneyball

- Used statistical analysis in baseball.
- Focused on on-base percentage.
- Revolutionized team-building strategies.

Watch the movie (starring Brad Pitt).

Duckworth–Lewis–Stern (DLS) Method

- The Duckworth–Lewis–Stern (DLS) method is used to reset targets in rain-interrupted limited-overs cricket matches.
- Developed by statisticians Frank Duckworth and Tony Lewis; later refined by Steven Stern.
- Officially adopted by the ICC for One Day Internationals (ODIs) and T20 matches.

Duckworth–Lewis–Stern (DLS) Method

Core Principle

- A team's scoring ability depends on two key resources:
  1. Overs remaining
  2. Wickets in hand
- These resources are quantified using empirically derived statistical tables.

Duckworth–Lewis–Stern (DLS) Method

Mathematical Model: $Z(u,w) = Z_0(w)\left(1 - e^{-b(w)u}\right)$, where:

- u = overs remaining, w = wickets lost,
- $Z(u,w)$ = expected runs obtainable,
- $Z_0(w)$ and $b(w)$ are parameters estimated from historical data.

Target Adjustment Formula If Team 1 scores S_1 runs using R_1% of resources and Team 2 has R_2% remaining, then:

$$\text{Revised Target} = \begin{cases} S_1 \times \dfrac{R_2}{R_1}, & \text{if Team 2 has fewer resources,} \\ S_1 + S_1\left(\dfrac{R_2 - R_1}{100}\right), & \text{if Team 2 has more resources.} \end{cases}$$

Outline

Challenge 1

Design a ranking system for a school football league using linear algebra.

Challenge 2

Create a mathematical model to detect undervalued players in an auction-based league.

Challenge 3

Develop a profitable betting strategy using expected value and probability theory.

Outline

Careers in Sports Analytics

- Academician / Researcher
- Data Scientist
- Quantitative Analyst
- Performance Analyst
- Sports Strategist

Tools to Learn

- **Statistical models, Optimization**
- Python, R
- Excel, Tableau, etc.
- Machine Learning and AI

Useful Reading

- *Soccermatics* – David Sumpter
- *Mathletics* – Wayne Winston
- *The Numbers Game* – Anderson and Sally
- *Moneyball* – Michael Lewis

Research papers from my group

- Deb, S. (2022). A goal based index to analyze the competitive balance of a football league. Journal of Quantitative Analysis in Sports, 18(3), 171-186.
- Deb, S., Das, S. (2026) Optimal selection of the starting lineup for a football team. IIMB Management Review, 100642.
- Deb, S., Dey, D. (2019). Spatial Modeling of Shot Conversion in Soccer to Single out Goalscoring Ability. Journal of Sports Analytics, 5(4), 281-297.
- Divekar, C., Deb, S., & Roy, R. (2024) Real-time forecasting within soccer matches through a Bayesian lens. Journal of the Royal Statistical Society Series A: Statistics in Society, 187(2), 513–540.
- Gupta, K., Krishnamurthy, V., & Deb, S. (2024). What elements of the opening set influence the outcome of a tennis match? An in-depth analysis of Wimbledon data. IIMB Management Review.
- Nahata, S., Deb, S. (2021) A machine learning approach to analyze the effect of situational and player-dependent features on converting freekicks in soccer. In Conference Proceedings 2021 Asia-Singapore Conference on Sport Science (p. 19).

Thank you!

Contact: soudeep@iimb.ac.in
Webpage: https://soudeepd.github.io/