Yakshita Pratha Model (YPM): An Alternative to Duckworth Lewis Stern (DLS) Method

A Dissertation

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE AWARD OF THE DEGREE

OF

MASTER OF SCIENCE

IN

APPLIED MATHEMATICS

Submitted by:

Pratha Chaudhary (24/MSCMAT/43)

Yakshita (24/MSCMAT/46)

Under the supervision of

Prof.(Dr.) Anjana Gupta

DEPARTMENT OF APPLIED MATHEMATICS

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

MAY 2026

DEPARTMENT OF APPLIED MATHEMATICS

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

We, Pratha Chaudhary (24/MSCMAT/43) & Yakshita (24/MSCMAT/46) students of MSc (Mathematics), hereby declare that the Project Dissertation titled — **"Yakshita Pratha Model: An Alternative to DLS Method"** which is submitted by us to the Department of Applied Mathematics, DTU, Delhi in fulfillment of the requirement for awarding of the Master of Science degree, is not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma, Fellowship or other similar title or recognition.

Place: New Delhi	**Pratha Chaudhary**	**Yakshita**
Date: 23/05/2026	**(24/MSCMAT/43)**	**(24/MSCMAT/46)**

DEPARTMENT OF APPLIED MATHEMATICS

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

CERTIFICATE

I hereby certify that the Project titled **"Yakshita Pratha Model: An Alternative to DLS Method"** which is submitted by Pratha Chaudhary (24/MSCMAT/43) & Yakshita (24/MSCMAT/46) for fulfillment of the requirements for awarding of the degree of Master Of Science (Msc) is a record of the project work carried out by the students under my guidance & supervision. To the best of my knowledge, this work has not been submitted in any part or fulfillment for any Degree or Diploma to this University or elsewhere.

Place : New Delhi

Date : 23/05/2026

Prof.(Dr.) Anjana Gupta

(SUPERVISOR)

Professor

Department of Applied Mathematics

Delhi Technological University

ABSTRACT

Keywords - DLS Method, T20 Cricket,ML in sports, Run Prediction Model

The Yakshita Pratha model (YPM) is an innovative machine learning model designed to set accurate, revised target scores for limited over cricket matches that are interrupted by inclement weather, floodlight failure or some crowd issues. While the former Duckworth-Lewis- Stern (DLS) method operates on static mathematical framework centered primarily on resource preservation (wickets left and overs remaining), the YPM captures modern cricketing dynamics . This model is grounded in a series of robusts data-driven assumptions regarding a team's evolving behavioural patterns, scoring potentials and a highly non-linear impact of losing wickets as the match progresses.

The core philosophy of this research is the transition away from the static models towards a highly adaptive , observe scoring paradigm. In contrast to the traditional method that rely heavily on generalized baseline decay curves, the YP model is based on modern shifts in playing style . It accounts for contemporary strategic changes, including deep batting lineup, aggressive data driven strategies and multi phased power hitting . This model treats match scoring as interconnected, path dependent process, the proposed model conditions, predictions based on how a team has used their resources in previous matches and real time execution context , rather than theoretical remaining capacity.

ACKNOWLEDGEMENT

The successful completion of any task is incomplete and meaningless without giving any due credit to the people who made it possible without which the project would not have been successful and would have existed in theory.

First and foremost, we are grateful to **Dr. R.Srivastava**, HOD, Department of Applied Mathematics, Delhi Technological University, and all other faculty members of our department for their constant guidance and support, constant motivation and sincere support and gratitude for this project work. We owe a lot of thanks to our supervisor, **Dr. Anjana Gupta**, Professor, Department of Applied Mathematics, Delhi Technological University for igniting and constantly motivating us and guiding us in the idea of a creatively and amazingly performed Major Project in undertaking this endeavor and challenge and also for being there whenever we needed her guidance or assistance.

We would also like to take this moment to show our thanks and gratitude to one and all, who indirectly or directly have given us their hand in this challenging task. We feel happy and joyful and content in expressing our vote of thanks to all those who have helped us and guided us in presenting this project work for our Major project. Last, but never least, we thank our well-wishers and parents for always being with us, in every sense and constantly supporting us in every possible sense whenever possible.

Pratha Chaudhary
(24/MSCMAT/43)

Yakshita
(24/MSCMAT/46)

Contents

List of Tables

List of Figures

Chapter 1

INTRODUCTION

1.1 The evolving game of cricket

Cricket has historically been rich , strategically complex team sport played with a bat and a ball played on a 22 yard pitch . It originated in england during the sixteenth century . This sport has undergone an immense global migration, evolving from a localized game into an internationally celebrated sport across multiple continents. Cricket was traditionally all about test matches- a long format played for over 5 days where endurance, defensive technique, and prolonged physical and mental stamina decided the outcome . However , the late 20th and early 21st centuries marked a huge shift within the sports global economy and competitive architecture . Driven by the demands of the public and fans shifting their engagement patterns, cricket progressively changed into faster and highly structured limited over formats most notably one day internationals and Twenty20 internationals. These limited formats requiring definite results within a fixed time. Therefore , these shortened formats changed how players approach run accumulation, field placings and resource risk management.

Cricket being played in open stadiums remains vulnerable to environmental conditions. Matches are regularly halted and reduced by multiple factors.

• **Inclement weather** : sudden rain , thunderstorms , wet outfields , fog etc

• **Flood light failure**

• **Some crowd issues** When these kind of incidents occur mid match, a major problem arises: some overs are permanently lost from the games original allocation. Because the team, which is batting second, will now no longer have the same access to the total number of allocated overs as that of the team batting first had, a fair and mathematically correct framework is required. Therefore, to maintain sporting integrity and produce authentic outcome, cricket requires advanced methods to adjust and revise target scores flawlessly. Thus declining, a winner based on fair distribution of opportunity.

1.2 The Traditional DLS Method and its Practical Limitations

For a very long time for a very long time.The standard for adjusting targets in limited over cricket has been the Duckworth- Lewis -Stern (DLS) method. Structurally, the DLS method operates on resource based model. It treats a team's scoring capacity as a function of two different resource remaining at any given moment: the number of **overs left to face** and the **number of wickets in hand**. However, now scientists and sports analysts argue and highlight, it's rigid and real world limitations - mainly, it's failure to adapt to the explosive batting philosophies of modern day cricket players. The DLS method was adapted during 1990's. During that era, Cricket was characterised by conservative highly predictable scoring trends. Teams played cautiously during the opening power play overs, focused heavily on preserving their wickets, playing with low risk in the middle overs and consolidated their explosive acceleration, almost exclusively for the final 10 overs of their innings. Individual strike rates were lower and players played long and comparatively slower innings with minimal risks.

In stark contrast, the modern cricketing Players feature an entirely different mindset. The global rise of domestic and international T20 leagues has embedded a permanent " T20 mindset" even within standard 50 over ODI matches. Modern cricket completely challenges the traditional resource to run assumptions:

• **360 degree and unorthodox hitting** : the routine deployment of switch hits, ram shots , reverse sweeps and different ways of clearing boundary ropes across the entire ground

•**Deep batting lineups**: teams regularly construct accomplished power heaters extending all the way down to 8th ,9th or 10th position. This allows the top order batsman to play with full freedom, minimising the fear of sudden collapse.

• **Continuous and multi phase acceleration**: teams no longer accelerate in the final death overs instead they target explosive scoring bursts across different phases of the match.

Consequently the core limitation of DLS method lies in its static nature generalised assumption of resource behaviour. It assumes that every wicket lost impact a team's scoring capability uniformly, regardless of that teams specific batting line up, current momentum, or specialized tactical aggression. While and Elite team with exceptionally deep batting lineup loses early wicket due to high aggressive intent, the DLS Framework penalizes their remaining resource percentage severely thereby dropping their projected score. In reality, however, modern teams frequently maintain and even accelerate their run rate is losing multiple wickets, relying on their deep batting lineups and power hitting capabilities. DLS fails to incorporate these real time acceleration and Momentum changes, it's revised target often seems to be disconnected from the actual scoring Trends of modern Limited over cricket. This establishes and urgent need for an updated and data driven alternative framework that is capable of tracking live momentum and which adapts to the current playing trends of the players.

1.3 The Need for a Data Driven Machine Learning Model

Bridge this gap between static mathematical frameworks and the modern playing patterns cricket analytics must embrace of fundamental shift: from a resource- based model to a **live observed modelling**. Rather than evaluating a match by a generalized mathematical formula, modern predictive systems can consider huge amount of historical ball by ball International Cricket data available today. This will allow the algorithm to learn how modern teams play through Complex match situations under varying pressure.

Key indicators such as Run rates, historical acceleration curves, boundary frequencies, dot balls percentage provides a much broader picture of a match's true direction then just simply considering remaining overs and wickets in hand.

By analysing these features across thousands of historical match progressions a machine learning model will learn to recognise and incorporate shifting momentums and acceleration of any team. Implementing a data driven model ensures that when a match is interrupted the targets are genuinely aligned with the real world tempos of modern cricket.This approach eliminates outdated assumptions replacing them with a highly responsive data driven model that reads the game as a living continuous process. It is within this modern context that YP model is introduced.

Chapter 2

Literature Review

The mathematical resolution of interrupted cricket matches represents a basic challenge in sports economics, computational analytics, and applied mathematics.Over the course of several decades, the regulatory bodies have moved away from primitive airthmetic solutions to sophisticated mathematical ones.
This chapter reviews the mathematical frameworks of the historical target predictive methods, evaluates the foundational equations and structural vulnaribilities of the DLS method, and explores the shift towards path-dependent machine learning architectures.

2.1 The Mathematics behind Historical Target-Predictive Frameworks

Before the adoption of mathematically derived mathematical modeling, interrupted cricket matches used to rely on empirical airthmetic rules. These early systems failed because there was a lack of rigorous formulation of multi-variable resource constraints.

2.1.1 The Average Run Rate(ARR) Method

Let Team A had played the total overs and innings of Team B was interrupted and overs reduced then
This Method used to determine the revised targets as:

$$\text{Revised target} = \frac{Total\ runs\ scored\ by\ team\ A}{Total\ overs} \times (overs\ reduced)\ +\ 1$$

Mathematically, the ARR method assumes only linear run-accumulation function that is seperated from the wicket preservation metrics.It treats the marginal value of a delivery constant throughout the entire innings:

$$\frac{\partial(Expected\ Runs)}{\partial(Overs\ Faced)} = C$$

This linear model ignores the reality that a chasing team already knowing that the overs have reduced can optimize their score by accelerating their scoring rate and consuming the wicktes left.
ARR rewards the chasing team heavily by fail to penalize them for accelerated wicket consumption.

2.1.2 The Most Productive Overs(MPO) Method

To overcome the biasness of ARR method towards the team batting on number 2, leading bodies introduced the MPO method, which was used during the 1992 ICC Cricket World Cup. If Team B's innings is reduced to overs 0_2 overs, their revised target is calculated based on the cumulative runs scored by Team A in their highest-scoring 0_2 overs.
Let $R_{1,j}$ be the run output of Team A's j-th over.
The target is formulated as:

$$Target\ Score = \sum_{j=1}^{O_2} R_{1,j} \ + \ 1$$

ARR penalized the team batting first,MPO introduced an aggressive mathematical penalty against the chasing team.This method completely disregards the contextual conditions of Team A's lowest-scoring overs, which typically include defensive maiden overs.
Case Study Analysis: The 1992 World Cup Semifinal
The problems in MPO was clearly seen during the 1992 World Cup Semi Final between South Africa and England.England scored 252 runs in 45 overs($O_1 = 45$).South Africa, chasing the target, scored 231 runs with the loss of 6 wickets in 42.5 overs and then the match was delayed due to rain and overs were reduced by 2, compressing South Africa's remaining 2.1 overs(13 balls) to 1 ball ($0_2 = 43$).
Using MPO formula,the match officials calculated the revised target for South Africa based on England's 43 highest-scoring overs. This meant removing the least-productive

overs from England's total score:

$$T_2 = 252 - (0+0) = 252\ runs$$

Which lead South Africa to score 21 runs in one ball from the target to score 22 runs in 13 balls. The Mathematical Framework of MPO penalizes the chasing team for unexpected bowling performances delivered during their own first innings, revealing a big flaw in the calculation of resource value.

2.2 Foundational Mathematics of the Duckworth-Lewis(DL) Method

To resolve these problems,statisticians Frank Duckworth and Tony Lewis formulated a multi-variable resource-based model known as Duckworth-Lewis method.
The model states that a batting team's capacity to accumulate runs at any given state of match is basically bounded by a combined resource vector which consists of two primiary components:**overs remaining(u) and wicktets lost(w).**

2.2.1 Modeling of Resource Functions

This follows exponential decay modeling and the basic assumption of DL method is that the percentage of scoring capacity remaining,$P(u,w)$,asymptotically decays the number of remaining overs u which approaches to zero.This relationship is modeled using a single-parameter exponential decay curve:

$$P(u,w) = P_0(w) \times [1 - \exp(-b(w) \times u)] \tag{2.1}$$

where:

- u $\in [0, O_{max}] representsthenumberofoversremainingintheinnings.$
- w $\in [0, 10] representsthediscretestateofwicketslost.$
- $P_0(w)$ is the asymptotic total scoring capacity coefficient for a team with w wickets lost, acting as a normalization scaling factor.

- $b(w)$ is the characteristic exponential decay constant corresponding to the wicket state w

The derivatives of this resource function tell how the value of remaining deliveries changes over time:

$$\frac{\partial P}{\partial u} = P_0(w) \times b(w) \times \exp(-b(w).u) > 0 \tag{2.2}$$

$$\frac{\partial^2 P}{\partial u^2} = -P_0(w) \times [b(w)]^2 \times \exp(-b(w).u) < 0 \tag{2.3}$$

This formula satisfies the marginal laws of cricket: the resource function is monotonically increasing with respect to remaining overs, while demonstrating decreasing maginal returns as the available over window expands.

2.3 Machine Learning Applications in Sports Analytics

To overcome the assumptions of traditional static mathematical models,contemporary research has turned toward data -driven,live-observed modeling.This shifted the analytical focus from general resource calculations to real time,context-aware run prediction.

2.3.1 Run Prediction Frameworks

Early applications of predictive machine learning in cricket relied on ordinary least sqaures(OLS) multi linear regression and generalized additive models(GAMS).They modeled total runs as a linear combination of static features:

$$\hat{Y} = \beta_0 + \beta_1(overs) + \beta_2(wickets) + \sum_i \beta_i X_i \tag{2.4}$$

These models provided baseline interpretability but they failed to capture the complex and non-linear interactions between the variables.For example,the impact of a wicket falling is highly dependent on the stage of the innings and the active scoring pattern. Recent research addressed these limitations by adopting ensemble tree-based methodologies,such as Random Forests and Gradient Boosted Decision trees. These frameworks construct non-linear decision spaces capable of tracking the interactions of complex features - such as evaluating how a dot-ball percentage alters scoring potential when paired with a batting order under scoreboard pressure.

2.3.2 Path-Dependency and Momentum Tracking

A key drawback of existing cricket models is their dependency on Markovian assumptions,which presumes that future run-accumulation depends on the current discrete state entirely:

$$P(X_{t+1}|X_t, X_{t-1}, ..., X_0) = P(X_{t+1}|X_t) \tag{2.5}$$

In modern limited overs format, match progression is deeply path-dependent.Two teams arrives at an identical resource state- for example,90 runs for the loss of 4 wickets at the 10-over mark - can possess entirely different momentum profiles.

Path A : 0/0 (Powerplay Acceleration) $\rightarrow RapidCollapse \rightarrow 90/4$ *at Over* $10(NegativeMomentum)$

Path B : $20/3(EarlyDefensiveShell) \rightarrow SteadyRecovery \rightarrow 90/4$ *at Over* $10(PositiveMomentum)$

Static models like DLS treats both scenarios identically because their resource values are equal.While machine learning models address this gap by incorporating rolling window features.By evaluating these indicators accross large historical datasets,ensemble algorithms can learn to the shifiting momentums directly, offering a responsive,data-driven framework for a fair prediction.

Chapter 3

Yakshita Pratha Model

The core contribution of this research is creation of an advanced machine learning model that is designed to predict the expected future runs of a batting team when a match is halted by inclement weather, infrastructure failure or crowd disruptions. This model considers Run scoring process as a dynamic and path dependent process. This model conditions it's on the basis of how a team has historically and currently used it' s resources within a match. This approach allows the model to capture strategic batting styles, incorporates multi-phase powerplay acceleration and deep batting lineups comma with the DLS often penalize or miscalculate.

3.1 Theoretical and Operation Highlights

This model is built on four fundamental operational pillars:

• **Multivariable real time prediction:** the model predicts the expected total of a team based on overs it has played runs already scored wickets fallen and current scoring trend of that team.

• **Path dependency:** The model captures path dependency which means it's runs predicted depend on how the team has actually scored so far in the match and previously. Momentum and scoring acceleration is incorporated. For example, a score of 90 /4 in 10 overs achieved by early explosive hitting followed by a collapse is treated differently than a steady and uninterrupted score of 90/4 as the two scenarios reflect distinct momentum states and resource usage patterns.

• **Incorporation of momentum and acceleration indicators:** The model adapts to the modern aggressive batting styles including higher strike rates, power hitting in different phases of a match and deep batting line ups.

• **Real time scoring behaviour:** This model replaces the static resource based DLS as it records real time scoring behaviour and learns from historical match patterns.

3.2 Data Formulation and Feature Engineering

To build and alternative to the resource tables the model utilises over by over structure data set compiled from historical T20 International matches taken as a sample full stop this data set records the precise match state at the conclusion of every single over providing a look at batting Momentum and maths progression. The primary feature Matrix input into the model includes both cumulative match features and short term window indicators:

Feature Category	**Specific Dataset Variables**	**Analytical Purpose**
Static & Situational Context	venue, innings, batting_team, bowling_team	Captures ground dimensions, pitch history, and team capabilities
Cumulative Match State	over, runs_till_now, wickets_till_now, wickets_remaining, overs_remaining	Establishes the baseline macro-state of the innings
Short-Term Window Dynamics	runs_last_over, wickets_last_over, runs_last_3_overs, wickets_last_3_overs	Measures immediate momentum, defensive consolidation, or mounting bowling pressure
Aggression & Efficiency Metrics	current_run_rate, boundaries_till_now, dot_ball_percentage	Quantifies batting intent, boundary-hitting frequency, and situational control
Phase-Specific Performance	powerplay_runs, powerplay_wickets	Anchors early-game performance to predict late-game acceleration

Table 3.1: Feature categories, dataset variables, and their analytical purpose in cricket match modeling

3.3 Machine Learning implementation via XGBoost:Methodology

The prediction of the YP model is done by XG boost (Extreme Gradient Boosting), an advanced implementation of gradient-boosted decision trees. The model takes the over by over data feature matrix as input during training. XGBoost is selected for its performance on structured tabular data and capacity to handle multi-collinear variables.Why sequentially building decision trees that minimizing the residual errors of the preceding trees, the algorithm maps Complex dependency across features such as how a high dot ball percentage interacts with a deep wicket remaining count during different stages of the innings.

Pseudo code

1. **Import Libraries and Load Dataset**

   Import pandas, numpy, matplotlib Import sklearn modules: GroupShuffleSplit, cross_val_score, GroupkFold, StandardScale Import xgboost Import sklearn matrics: MAE, MSE, R^2

   Load dataset from CSV file into dataframe

2. **Define TargetEncoder Class**

   Class TargetEncoder:

   Initialize with categorical columns and smoothing factor

   fit(x,y):

   Copy x

   Add target column = y

   Compute global mean of target

   For each categorical column:

   Group by column – calculate mean and count of target

   Apply smoothing formula

   Store encoding map

   transform(x):

   Copy x

   Replace categorical values with smoothed encodings

   Replace the missing values with global mean fit_transform(x,y):

Call fit()

Call transform()

3. **Train-Test Split with GroupShuffleSplit**

   Define categorical columns= [“batting_team”,”bowling_team”,”venue”]

   Define target column = “final_total”

   Initialize GroupShuffleSplit with test_size=0.2

   Now split the dataset in train and test ensure that the groups by match_id are preserved

   Create train_df and test_df data frames

   Separate features (x) and target (y) for train and test sets

4. **Apply Target Encoding**

   Initialize TargetEncoder with categorical columns

   Fit encoder on the training data and transform to x_encoded And test data to x_test_encoded

5. **Defining the Model Training Function**

   Function train_model(x_train_encoded, y_train, model_type)

   Then use the models random_forest and xgboost accordingly

   If model is random forest initialize RandomForestRegressor with parameters

   If model is xgboost initialize XGBRegressor with parameters.

6. **Model Evaluation Function**

   Function evaluate_model(model,x_test_encoded,y_test):

   Predict on test set

   compute MAE, RMSE,R^2

   Print results

   Return Matrix

7. **Train and Evaluate moddel**

   Train XGBoost model using train_model()

   Evaluate model using evaluate_model()

8. **Cricket Prediction Functions**

Function dot_ball(total_balls, dot_ball):

Return percentage of dot balls

Function current_run_rate(runs_till_now,overs):

Return runs_till_now overs 100

Function predictor(innings,over,batting_team,bowling_team,venue,runs_till_now etc use all features included in the table)

Create dataframe with all input features

Encode dataframe using trained encoder

Predict score using trained model

return prediction

3.4 Mathematical Formulation and Evaluation Metrics

Step 1: Quantifying First-Innings Performance (R)

The model first establishes how efficiently the team batting first (Team A) converted its match situation into actual runs compared to a statistical baseline. The algorithm ingests the completed over-by-over features of the first innings to output a final predicted score. The performance ratio (R) is defined mathematically as:

$$\mathrm{R} = \frac{Actual\ score\ placed\ by\ Team\ A}{Predicted\ Score\ by\ Model\ for\ Team\ B} \tag{3.1}$$

This performance ratio (R) functions as an index of competitive output.

An $R > 1.0$ indicates that the first batting team overperformed historical baseline trends, while an $R < 1.0$ indicates they fell short of the expected run potential given their match progression

Step 2: Calculating the Revised Target

Once the performance ratio is established, it is used to scale the chasing team's (Team B) expected scoring potential within their newly

adjusted over allocation. When an interruption cuts the second innings short, the model recalculates Team B's baseline potential using their active over mark and the remaining overs left to face.

The revised target total for the chasing team is determined using the formula:

$$\text{Revised Target for Team B } = \mathrm{R} * (Predicted\ Runs\ for\ Team\ B\ in\ Revised\ over\ Allocation) \tag{3.2}$$

To determine the exact number of runs Team B needs to score when play resumes or concludes,
the model subtracts their current cumulative score from the revised target:

$$\text{Runs left to Score} = (\textit{Revised target}) - (\textit{Runs Scored Till Interruption}) \tag{3.3}$$

3.5 Practical application illustration

To demonstrate the Framework in practise consider a match between Team A and Team B under normal conditions where team A bats first and sets an initial target:

• **Team A actual score :**180 runs

• **Model predicted score for team A (by the YP model) :** 200 runs (indicating team A slightly under performed expectation)

Using these values the performance ratio 'R' is computed:

$$\text{R} = \frac{\textit{Actual Score}}{\textit{Predicted Score}} \tag{3.4}$$

$$\text{R} = \frac{180}{200} = 0.9 \tag{3.5}$$

This means that team B's target will be adjusted to 90percent of their predicted score. Now suppose an interruption occurs Midway through the second innings. Team B has played 16 overs scoring 130 runs when they play is halted. The match officials reduce the total match length by 2 overs leaving 18 overs in total for the innings. The YP model evaluates team B's performance at the 16-over mark, accounts for the 2 remaining overs and outputs a prediction:

•**Model predicted runs for team B (scaled to 18 overs):**160 runs

The revised target for team B is then calculated using the performance ratio :

$$
\begin{aligned}
\text{Revised Target} &= \text{R} \times 160 \\
&= 0.9 \times 160 \\
&= 144
\end{aligned}
$$

Finally, the remaining runs required for team B to win in the final 2 overs are determined:

$$\text{Runs left score} = 144 - 130 = 14 \text{ Runs} \tag{3.6}$$

Team B must score **14 runs in remaining 2 overs** to secure a victory under the YP model's target revision.

3.6 VISUAL ANALYSIS

To visually demonstrate the mathematical stability, predictive alignment, and feature dependency structure of the developed machine learning framework, this section presents a comprehensive graphical evaluation of the test results. Evaluating a predictive engine on an extensive validation partition requires analyzing not just localized accuracy metrics, but also visualizing the overall error distribution and the continuous trajectory of predictions across a standard match progression. The following figures and accompanying analytical commentaries provide a detailed look into how the trained XGBoost algorithm handles real-time cricket dynamics, manages multi-variable momentum tracking, and identifies the structural thresholds where minor scoring lags begin to manifest. .

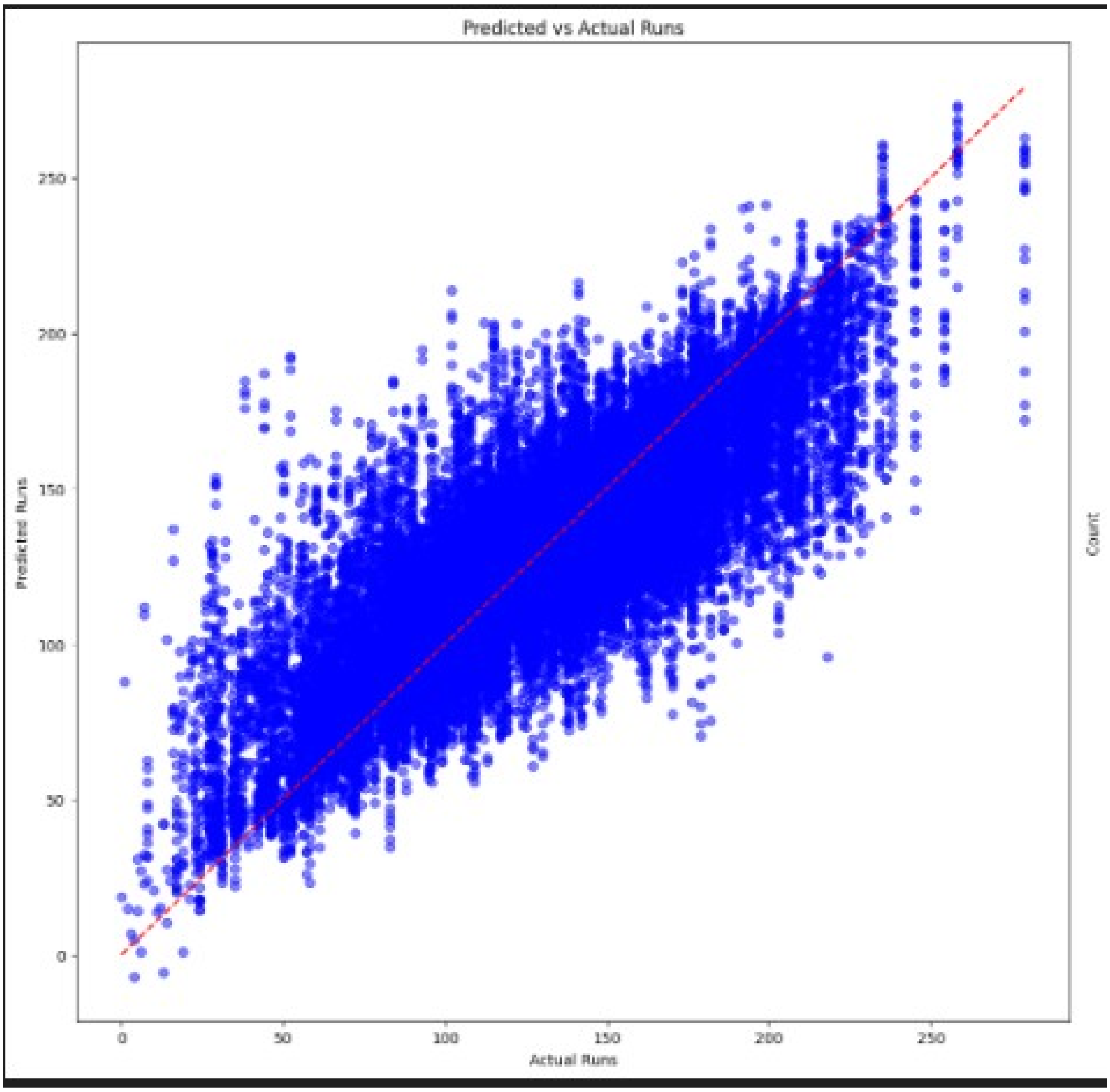

Figure 3.1: Scatter Plot(Predicted vs Actual)

Figure 2.1 place the relationship between the actual match outcomes and the values predicted by the model the data points form a tight linear cluster along the 45 degree identity line, indicating a strong correlation between the feature inputs and the final scores . The uniform distribution of data points across both low scoring and high scoring match contacts shows that the XG boost algorithm effectively adapts to different team profile and scoring without over fitting or under fitting.

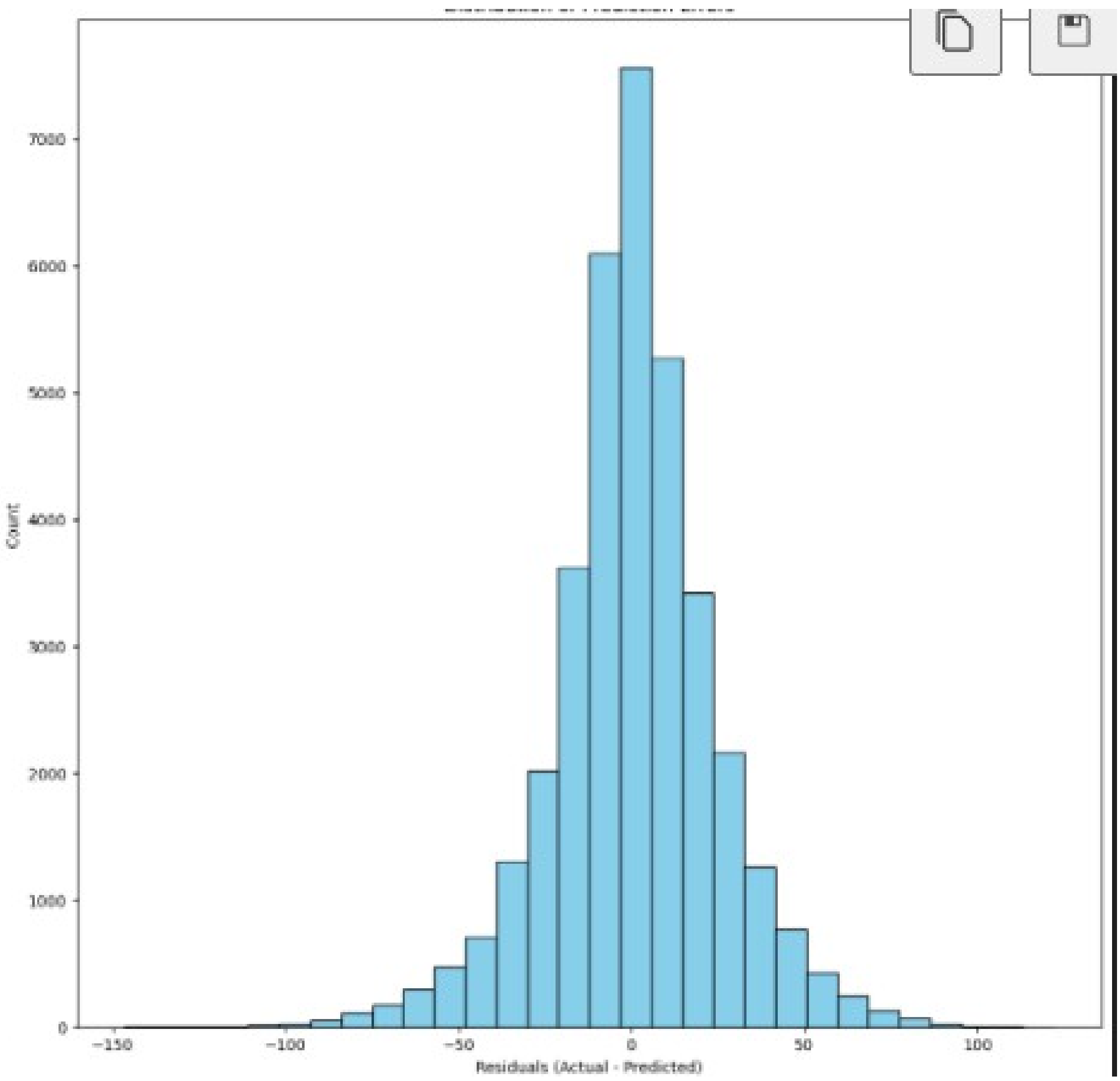

Figure 3.2: Residual Error

The residual error distribution shown in figure 2.2 serves as a validation of the models statistical consistency full stop the residuals are centred around 0 and so the model does not show any bias. It does not show any systematic tendency to over estimate or underestimate a batting sides target.

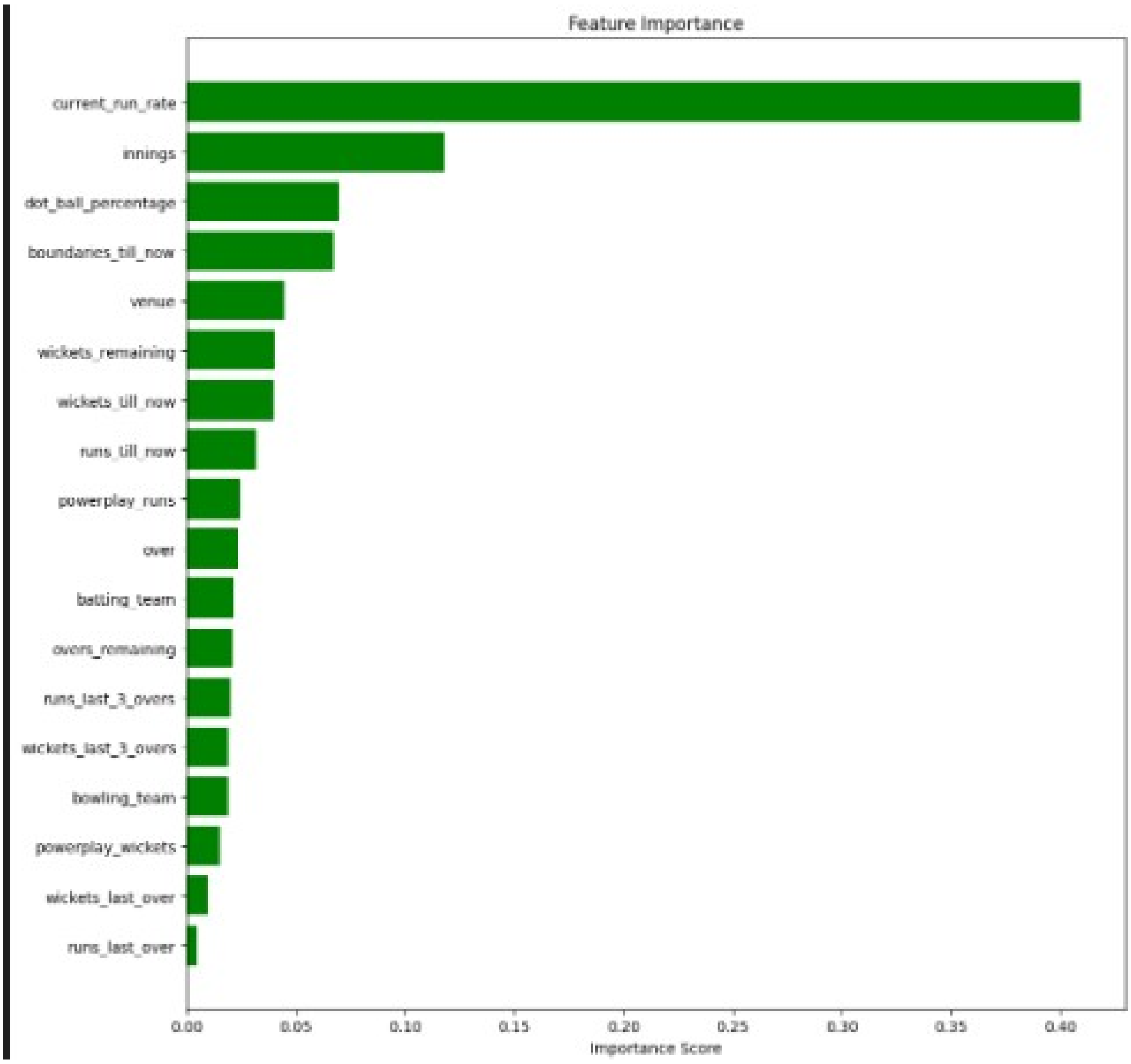

Figure 3.3: Feature Importance

Figure 2.3 outlines the structural hierarchy of features driving the model's decision nodes. The current run rate registered as the most important features followed by innings, dot ball percentage, cumulative boundary count.

This shows that the system prioritizes real time performance tempo and bowling pressure over static resource variables, allowing it to dynamically capture a team's active momentum

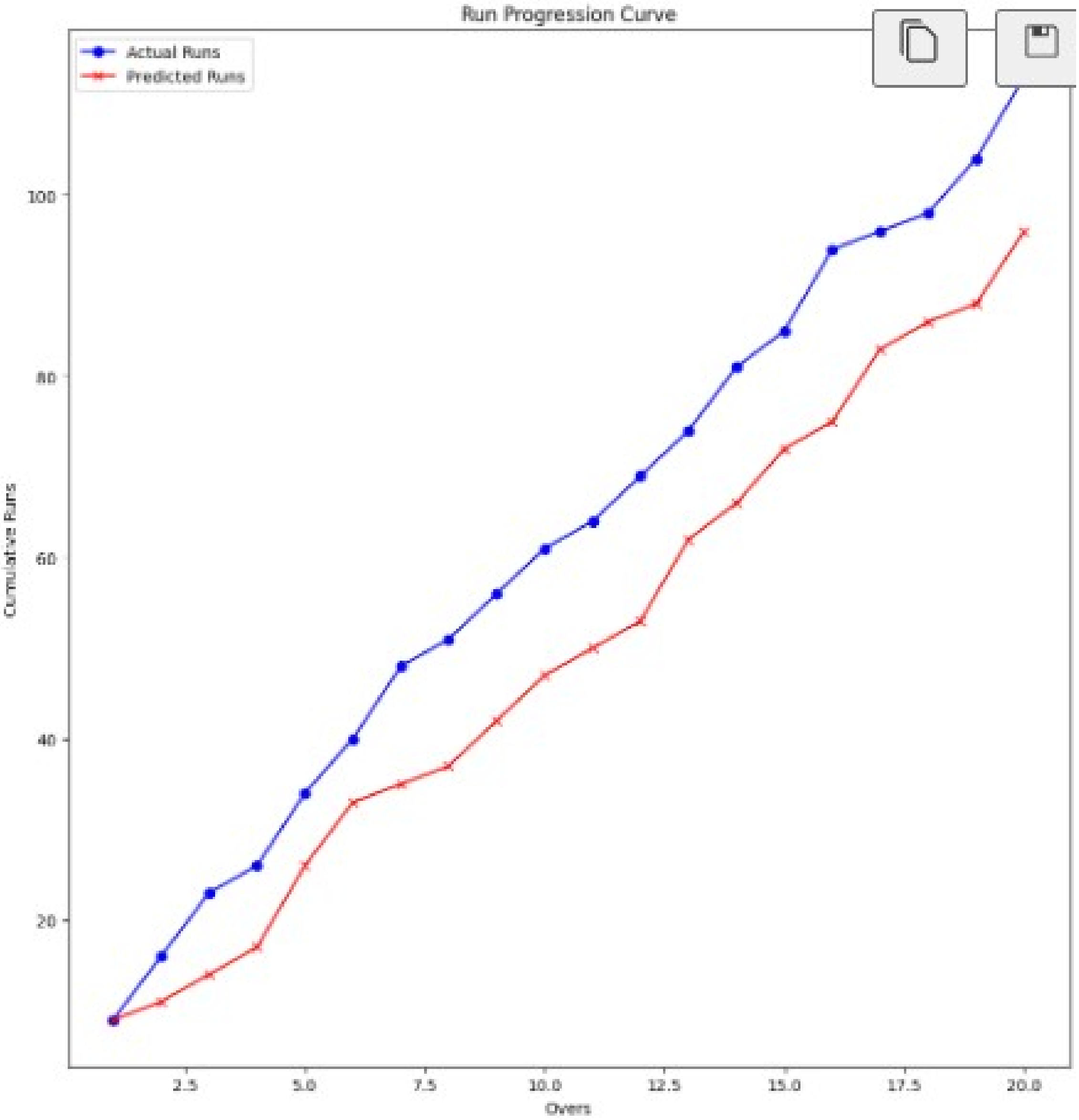

Figure 3.4: Actual vs Predicted Trace

This figure tracks step by step scoring progression of a typical innings against the live, evolving projections generated by the YP model. The graph highlights that the model successfully captures match momentum, its trajectory smoothly moves during the initial power play and Middle overs. However, as the timeline nears the final overs a slide under prediction gap becomes visible.

Chapter 4

CONCLUSION

4.1 EMPIRICAL ANALYSIS

To evaluate the predictive reliability of the model the trained XG boost Framework was validated against an independent unseen data set comprising 145,305 over by over international matches. The empirical results established a mean absolute error (MAE) of 17.58 runs indicating that the model's prediction of a team's final innings deviates by 17 runs on an average at any given over. The root means square error (RMSE) calculated at 24.19 runs. The Framework achieved a coefficient of determination (R^2) of 0.678. This indicates that the engineered feature Matrix and the sequential gradient boosted trees successfully account for approximately 68% of the total variance in match score. This represents a robust predictive capability for a sport heavily influenced by real time human variables and situational volatility.

4.2 Identified Model Limitations and Future Scope:

While the model tracks run progression at every phase of match in Limited overs cricket, the validations revealed a different behavioural limitation during highly volatile stages of the match.
The model exhibits slight **underestimation of scoring acceleration during the final 4 to 5 overs of an innings**. Death overs phase in cricket is characterized by extreme power hitting and non linearly scores surging with increased boundaries.
Because the gradient boosting trees operate on a sequence of historical splits, the model's

prediction can lack slightly behind these rapid bursts of late game acceleration when a team retains a deep batting card.

Bibliography

[1] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

[2] Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press.

[3] Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.

[4] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. *Nature*, 521(7553), 436–444.

[5] Shah, H., Sampat, J., Savla, R., & Bhowmick, K. (2015). Review of Duckworth Lewis Method. *Procedia Computer Science*.

[6] Jha, A. (2023). *International Cricket Dataset*. Retrieved February 23, 2026, from `https://www.kaggle.com`

[7] XGBoost Documentation. Retrieved February 2026, from `https://xgboost.readthedocs.io`

[8] Python Software Foundation. *Python Documentation*. Retrieved from `https://www.python.org`

[9] Scikit-learn Developers. *Scikit-learn: Machine Learning in Python*. Retrieved from `https://scikit-learn.org`

[10] Pandas Development Team. *Pandas Documentation*. Retrieved from `https://pandas.pydata.org`

12% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text
- Cited Text
- Small Matches (less than 10 words)

Match Groups

19 Not Cited or Quoted 12%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

9% Internet sources
2% Publications
11% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

*% detected as AI

AI detection includes the possibility of false positives. Although some text in this submission is likely AI generated, scores below the 20% threshold are not surfaced because they have a higher likelihood of false positives.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

PAPER ACCEPTANCE PROOF

Reference No.-ICAORFDI2026_02 - Abstract approved Inbox

ICAORFDI-2026 13 Feb

to me, deepmala, Dr....

Dear Authors,

Congratulations! Your abstract, Ref. No. **ICAORFDI2026_02**, "*ALTERNATIVE OF DUCKWORTH LEWIS STERN (DLS) METHODOLOGY USING PRINCIPAL COMPONENT ANALYSIS AND MACHINE LEARNING*", has been **accepted for oral presentation** at the International Conference on Applications of Operations Research in Finance, Defence, and Industry **(ICAORFDI-2026)** & 58th Annual Convention of Operational Research Society of India (ORSI-2025) to be held at Govt. Holkar Science College, Indore, Madhya Pradesh, India during **06-08 March 2026.**

Although your abstract has been accepted, presenting authors must register **on or before 20/02/2026** in order to ensure inclusion in the final program.

REGISTRATION PROOF

Conference Registration Form

1. Full Name: YAKSHITA

2. Designation: STUDENT,

3. Department / Affiliation: Dept. of Applied Mathematics

4. Institution / Organization: Delhi technolo ical University

5. Email ID: yakshitadangi2511@gmail.com

6. Mobile Number: 9871086875

7. Address: 431, lig flats, Hastsal, Uttam Nagar

8. City: New Delhi State: Delhi

9. Country: INDIA PIN Code: 110059

10. Category (Tick one):

☑ Student □ Research Scholar □ Faculty □ Industry □ Other

11. Mode of Participation:

□ Oral Presentation ☑ Presentation + Publication □ Listener

12. Title of Paper (if applicable): YP method – a DLS alternative

13. Author(s) Name(s): PRATHA CHAUDHARY and YAKSHITA*

14. Payment Details:

Amount Paid: 3500 Date: 18/02/2026

Transaction ID / Reference No.: Y2M12399949085788606464

15. Accommodation Required:

☑ Yes □ No

16. Date of Arrival: 5/03/2026 Date of Departure: 8/03/2026

Signature of Participant: yakshita

Date: 18/02/2026

Conference Registration Form

1. Full Name: PRATHA CHAUDHARY

2. Designation: STUDENT,

3. Department / Affiliation: Dept. of Applied Mathematics

4. Institution / Organization: Delhi technological University

5. Email ID: prathachaudhary4@gmail.com

6. Mobile Number: 9354141629

7. Address: 1113, Baoli, patti Dehu

8. City: Baghpat State: Uttar Pradesh

9. Country: INDIA PIN Code: 250621

10. Category (Tick one):

☑ Student ☐ Research Scholar ☐ Faculty ☐ Industry ☐ Other

11. Mode of Participation:

☐ Oral Presentation ☑ Presentation + Publication ☐ Listener

12. Title of Paper (if applicable): YP method - a DLS alternative

13. Author(s) Name(s): PRATHA CHAUDHARY and YAKSHITA*

14. Payment Details:

Amount Paid: 3500 Date: 18/02/2026

Transaction ID / Reference No.: T2602181615033412536418

15. Accommodation Required:

☑ Yes ☐ No

16. Date of Arrival: 5/03/2026 Date of Departure: 8/03/2026

Signature of Participant: Chaudhary

Date: 18/02/2026

CONFERENCE PROOF

PARTICIPATION CERTIFICATES

International Conference on Applications of Operations Research in Finance, Defence, and Industry (ICAORFDI-2026)

&

58th Annual Convention of
Operational Research Society of India (ORSI-2025)

CERTIFICATE

This is to certify that

PRATHA CHAUDHARY

Dept. of Applied Mathematics, Delhi Technological University

has presented a paper entitled "YP method — a DLS alternative" in the *International Conference on Applications of Operations Research in Finance, Defence, and Industry (ICAORFDI-2026)* & *58th Annual Convention of Operational Research Society of India (ORSI-2025)* held from March 06 - 08, 2026, organized by ORSI Indore Chapter at Government Holkar Science College, Indore, India.

Dr. Chitaranjan Sharma
Conference Chair

Dr. Deepmala
Conference Chair

International Conference on Applications of Operations Research in Finance, Defence, and Industry (ICAORFDI-2026)

&

58th Annual Convention of Operational Research Society of India (ORSI-2025)

CERTIFICATE

This is to certify that

YAKSHITA

Dept. of Applied Mathematics, Delhi Technological University

has presented a paper entitled "YP method — a DLS alternative" in the *International Conference on Applications of Operations Research in Finance, Defence, and Industry (ICAORFDI-2026) & 58th Annual Convention of Operational Research Society of India (ORSI-2025)* held from March 06 - 08, 2026, organized by ORSI Indore Chapter at Government Holkar Science College, Indore, India.

Dr. Chitaranjan Sharma
Conference Chair

Dr. Deepmala
Conference Chair