Statistical Analysis & Prediction of Player Performance Trends in Professional Cricket

A RESEARCH PROJECT SUBMITTED TO

KARMAVEER BHAURAO PATILUNIVERSITY, SATARA

FOR THE COURSE OF

M.Sc

IN

STATISTICS

BY

Raut Mayur Shivaji

Yashavantrao Chavan Institute of

Science, Satara (Autonomous)

Constituent college of Karmaveer Bhaurao Patil

University,Satara

2024-25

DECLARATION

I hereby declare that the Research Project entitled "**Statistical Analysis & Prediction of Player Performance Trends in Professional Cricket** " completed and written by me has not formed earlier the basis for the award of any degree or similar title of this or any other university or examination body. Further, I declare that I have not violated any of the provisions under the acts of copyright / Piracy / Cyber / IPR etc, amended from time to time.

The present research works original and the conclusions draw there in are based on the data collected by ourselves.

Place : Satara

Date:

Mayur Shivaji Raut

Research Investigator

Roll. No.	127

UID No.	202401200031

Rayat Shikshan Sanstha's

Yashavantrao Chavan Institute of Science, Satara (Autonomous)

(Constituent College of Karmaveer Bhaurao Patil University, Satara)

CERTIFICATE

Department of Statistics

Date:

This is to Certify that **Mr. Mayur Shivaji Raut**, UID No.:- 202401200031 partial fulfillment of curriculum of M.Sc. I (Sem-II) student has been successfully completed the Research project work in the statistics entitled ***"Statistical Analysis & Prediction of Player performance Trends in Professional Cricket"*** as prescribed by the Karmaveer Bhaurao Patil University, Satara during the academic year 2024-25.

Teacher-in-charge Examiner Head,
Department of Statistics

➢ ACKNOWLEDGEMENT :

In our efforts towards realization of our project work, we have drawn on the guidance of many people,whom we wish to acknowledge.

I am thankful to our **Principal Dr. B. T. Jadhav** and Head of the Department of Statistics **Dr. H. P. Umap** for their support and cooperation toward successful completion of this project also obliged to P.G. Co-ordinator of Department of Statistics **Dr. D.S. Jadhav** for continuous encouragement.

We are also extremely thankful to our beloved teachers **Dr.S.M.Nimbale, Mr.P.H.Mahadik, Ms.D.S. Tanpure, Mr.K.A.Ghorpade, Ms.S.S.Wagh, Ms.P.J.Thombare, Ms.D.S.Sabale, Ms.P.A.Kadam, Ms.A.M,Bagwan** for their support and guidance for this project.

Last but not the least we would like to thank all those who helped us directly and indirectly in our endeavour.

INDEX

Title :

"Statistical Analysis & Predictions of Player Performance Trends in Professional Cricket"

➢ ABSTRACT:

The Statistical analysis of Player performance over time is a crucial aspect of sports analytics, providing valuable insights into an athlete's consistency, growth and potential. This study focuses on Shubman Gill, a Prominent Cricketer, to evaluate his performance trends using Time Series analysis techniques in Python. I aim to identify key patterns, fluctuations and potential future projections and Prediction of his performance metrics, such as Batting Average, Strike Rate and Consistency in runs across ODI and TEST format of International Cricket. The study applies Augmented Dickey- Fuller Test (ADF) for checking the Stationarity, Holt's Linear trend model, Exponential Smoothing model to extract meaningful trends and provide actionable insights for Analysts, Coaches and Fans.

➢ INTRODUCTION:

Shubman Gill, an Emerging Star in Cricket has shown dynamic performances in international and domestic formats. This research applies time series analysis to evaluate his performance trends over multiple seasons. Using Python-based libraries like Numpy, Pandas, Matplotlib, Seaborn, Statsmodels, Exponential Smoothing. I analyze his Runs scored, Batting average, Strike rate, and form fluctuations over time.

➢ OBJECTIVES :

- To Collect and preprocess Shubman Gill's performance data across two international formats. (ODI & TEST)
- To apply Time Series Analysis techniques to identify Trends, Patterns and variations in his batting performance.
- To predict Future performance of his Batting Average, Strike rate and Runs based on historical data using forecasting model like Exponential Smoothing Model.

➢ PROBLEM STATEMENT :

In the field of professional cricket, evaluating a player's form and forecasting their future performance is essential for making informed decisions related to team strategy, selection, and training. This research specifically focuses on Shubman Gill, an emerging talent in Indian cricket, to explore and model his batting performance in the ODI and Test formats.

This study uses statistical tools and Python programming to turn raw cricket data into useful insights. This can help coaches, analysts, and fans better understand and expect a player's progress.

➢ LITERATURE REVIEW :

The application of Statistical models in sports analytics, particularly cricket, has grown rapidly in recent years. Here are some key insights from previous studies :-

- Hyndman & Athanasopoulos (2018) emphasized the efficacy of exponential smoothing methods for forecasting when data exhibits trends without seasonality.
- Studies like those by Bandulasiri (2008) and Lemmer (2011) explored the use of statistical models in evaluating batting and bowling performance using historical match data.

➢ METHODOLOGY :

DATA COLLECTION & PRE-PROCESSING :

The dataset contains Shubman Gill's Cricket performance statistics across two formats ODI from 2019 to 2025 & TEST from 2020 to 2024.

For this study on "*Statistical Analysis & Predictions of Player Performance Trends in Professional Cricket*" data on Shubman Gill's performance has been sourced from **ESPNcricinfo**, one of the most reliable platform for cricket Statistics. The dataset includes the following columns :-

Format	Date	Runs	Ball faced	Fours	Sixes	Strike rate	Position	Opposition	Fifty	Hundreds
ODI	31-Jan-19	9	21	1	0	42.85	3	NZ	0	0
ODI	3-Feb-19	7	11	1	0	63.63	3	NZ	0	0
ODI	2-Dec-20	33	39	3	1	84.61	2	AUS	0	0
ODI	22-Jul-22	64	53	6	2	120.75	2	WI	1	0
ODI	24-Jul-22	43	70	5	0	87.75	2	WI	0	0
ODI	27-Jul-22	98	98	7	2	100	2	WI	1	0
ODI	18-Aug-22	82	72	10	1	113.88	2	ZIM	1	0
ODI	20-Aug-22	33	34	6	0	97.05	3	ZIM	0	0
ODI	22-Aug-22	130	97	15	1	134.02	3	ZIM	0	1

➢ Data Analysis :-

1) Model Selection–

For this Research, I use the "EXPONENTIAL SMOOTHING MODEL" for Forecasting Runs, Batting Average and Strike rate of the Shubman Gill.

The Exponential Smoothing Model is one of the best method for Forecasting Cricket Statistics because,

1. Exponential smoothing captures short term fluctuations from our dataset.
2. This model can adjust the predictions based on recent performance rather than older data.

Exponential Smoothing equation –

The General Formula is,

$$y\hat{}t+1 = \alpha yt + (1-\alpha)\, y\hat{}t$$

where,

$y\hat{}\, t+1$: Forecast for the next period

yt : Actual value in the current period

$y\hat{}t$: Forecast for the current period

α : Smoothing constant ($0 < \alpha < 1$)

Holt's Linear trend model –

Equations are –

$$\text{Trend} : bt = \beta\,(lt-lt-1) + (1-\beta)\, bt-1$$

$$\text{Forecast} : y\hat{}t+h = lt + h \cdot bt$$

where,

β = Smoothing constant for the trend component ($0 < \beta < 1$)

bt = Smoothed trend at time t ; estimate of the slope or growth rate

h = Forecast horizon — how many steps into the future you're predicting

$y\hat{}t+h$ = Forecasted value for h periods ahead from time t (e.g., forecast for year $t+h$)

lt = Smoothed level of the series at time t; the estimate of the current base value.

Used Statistical Tests and Models :-

- Augmented Dickey Fuller test (ADF)
- Exponential Smoothing model
- Holt's Linear Trend Model

2) <u>Analyze the Data</u> –

Now I Start analyze the data, using Exponential Smoothing Model. Before applying this model, I should test for the following –

1. <u>Check the trend and seasonality</u> :

- If the trend is present, then apply Holts's Linear trend model
- If seasonality exists, use Holt-Winters exponential smoothing model.

After checking trend and seasonality in Python Software, I can apply the Holt's Linear trend model. Because, the data has the Trend but no Seasonality.

2. Stationarity Check :

After Checking the Trend & Seasonality, I Checked the Stationarity for Given Dataset. We can check stationarity for analysis, If and only if we have to Predict or Forecasts by using different models like ARIMA, LSTM, Exponential smoothing in Time Series method. For Checking the stationarity, I used the ADF test as follows –

- **Augmented Dickey Fuller Test :-**

The ADF test is used to check whether a time series is Stationary or Non-Stationary.

1) Apply ADF test on Data using Python software :

p- value = 0.4766 > 0.05, We fail to reject the null hypothesis. Hence, data is Non – Stationary. This means there is a trend in the runs scored over the years. Then I applied the first order differencing.

2) First order differencing :

p- value = 0.1573 > 0.05, We again fail to reject the null hypothesis. Hence, again data is Non-stationary. Even after taking a difference once, the trend is still present. Then I applied the second order differencing.

3) Second Order differencing :

p-value = 0.00321 < 0.05 We reject the null hypothesis (Ho) at 5% L.O.S. Hence The series is stationary and does not required again differencing.

3. Applying the Holt's Linear Trend Model :-

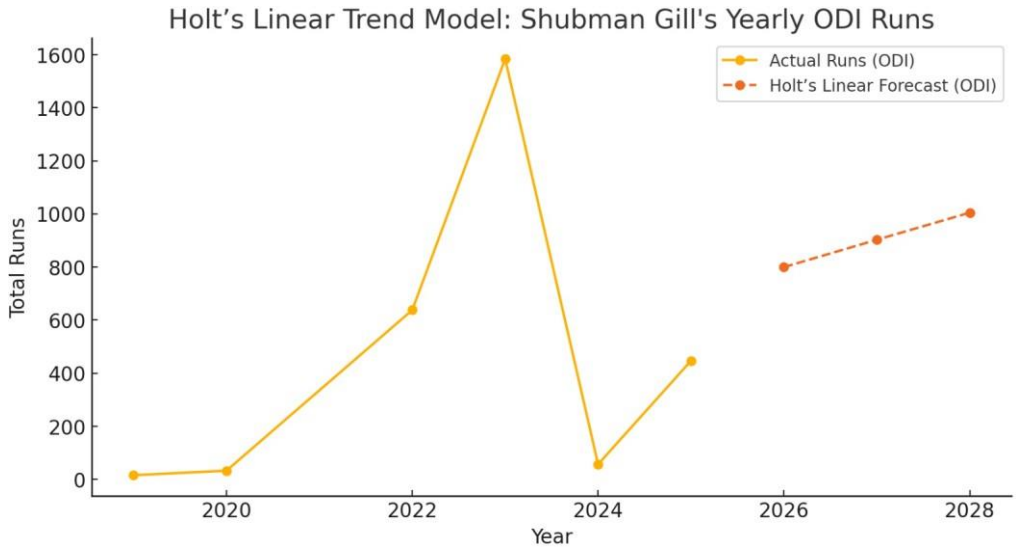
What this model does?

- Identify trend in past performance.
- Works well when data shows consistent increase or decrease.

➢ RUNS :-

For applying Holt's Linear trend model in Python software the steps are as follows :

1) Importing Libraries like Numpy, Pandas, Matplotlib, Statsmodels and load the given data.
2) Then apply Holt's Linear trend model.
3) After applying model, I plotted Actual and Forecasted Runs of Shubman Gill in **ODI Format** as follows –



Forecasted yearly ODI runs using Holt's Linear trend Model :

Year	Predicted Runs (ODI)
2026	800
2027	903
2028	1005

Conclusion : This Model suggests that Steady increase in yearly runs of ODI format over the next 3 years.

Forecasting Shubman Gill's **TEST Format** Runs in Python Software stepwise as follows –

- **Checking the Stationarity :**

For checking the stationarity of another dataset of Shubman gill's TEST format Performance, I used again the ADF test in Python software and the Result becomes –

p-value = 0.664 > 0.05

- **Interpretation:**

Since the p-value > 0.05, we fail to reject the null hypothesis, Means data is Non-stationary.

After applying, First and Second order differencing still the p-value > 0.05 and data becomes the Non-stationary. This suggests that the data has a Trend.

Still I can apply the Exponential smoothing model because this model also works well with Non-stationary data. If the data shows strong seasonality, Exponential smoothing may not be best choice, otherwise data shows Trends then we still apply the Exponential smoothing model.

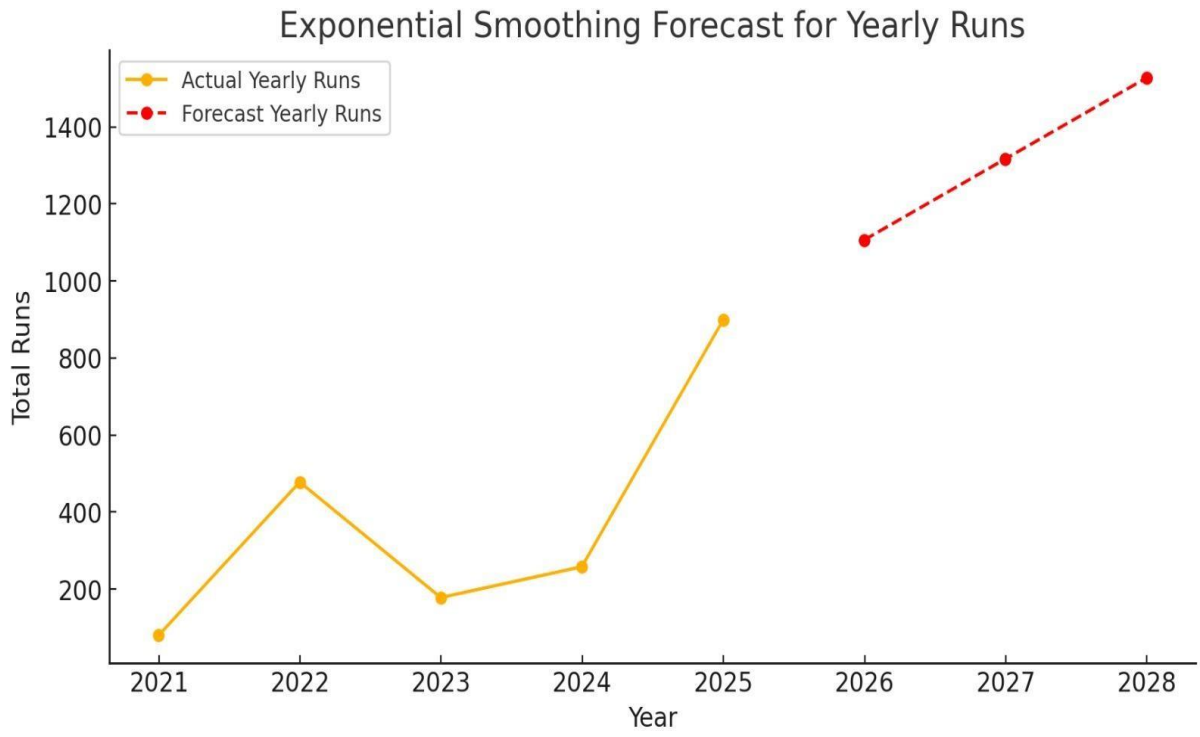
Then I applied Exponential smoothing model stepwise as follows –

1) Import the libraries Numpy, Pandas, Statsmodel and load the dataset.

2) Convert Date column to datetime format.

3) Then apply exponential smoothing model and forecast the next 3 years runs in **TEST** format on yearly basis.

4) Plot the results in form of Graph. (Line chart)



Here are the forecasted runs for the next 3 years in the **TEST** Format of Shubman Gill using Exponential smoothing model as :

Year	Forecasted Runs (TEST)
2026	1106.72 ≈1107
2027	1316.98 ≈ 1317
2028	1527.25 ≈ 1527

Conclusion :

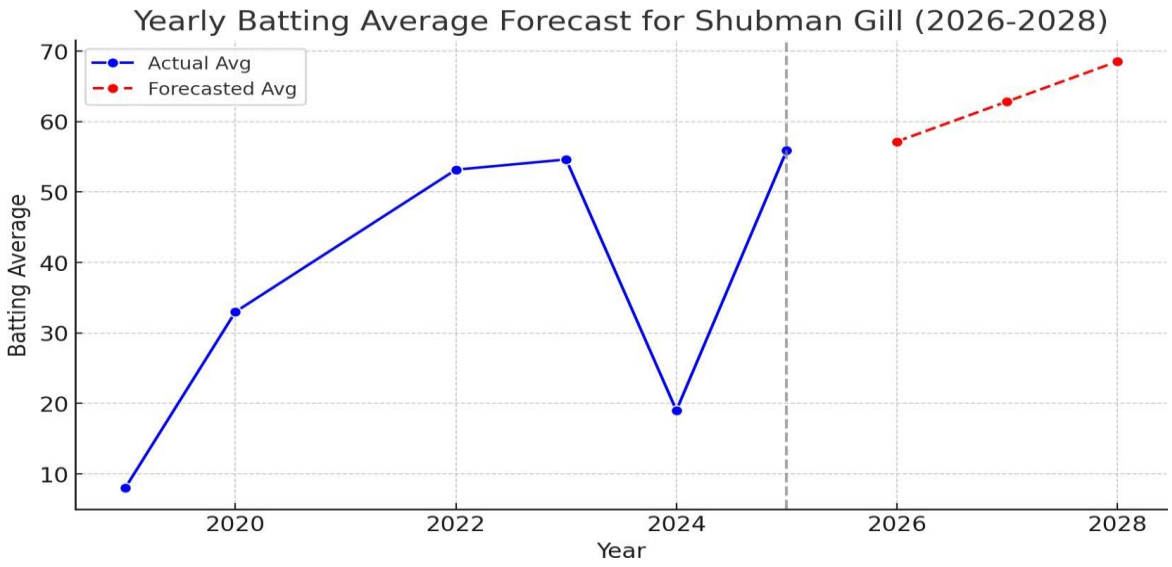
The model predicts a steady increase in runs for the next three years and the historical yearly runs along with the predicted values for the next three years. The increasing trend is captured well by the Exponential Smoothing model.

➢ BATTING AVERAGE :

Now, after forecasting Runs of Shubman Gill in ODI & TEST format I predicted the Batting average in both ODI & TEST format. (Yearwise)

1) ODI Format :-

I Repeated the same procedure of Forecasting Runs for Forecasting The Batting average of the ODI Format as :



Here is the Exponential smoothing forecast for Shubman Gill's yearly **ODI Batting Average** for the next 3 years as shown in above graph.

Year	Forecasted Batting avg. (ODI)
2026	53.94
2027	56.42
2028	58.40

Conclusion :

The above graph shows the gradual increase in his yearly Bating average. The model identifies a positive trend in Shubman gill's Bating average performance over the time.

➢ Strike Rate :-

Now, after forecasting Batting average of Shubhman Gill in ODI & TEST format I predicted the Strike Rate in both ODI & TEST format. (Yearwise)

1) ODI Format :-

Strike rate is computed by the formula as -

Strike Rate = (Total Runs / Total Ball Faced) / 100

I Repeated the same procedure of Forecasting Batting average for the Forecasting Strike Rate of the ODI Format as :

Start by checking the stationarity using ADF test –

After applying the ADF test in Python software for forecasting the Strike rate of next 3 years of Shubman Gill the interpretation becomes –

p-value = 0.112 > 0.05

This indicates that time series is Non – stationary. To make it Stationary, I applied the First order differencing and Recheck stationarity.

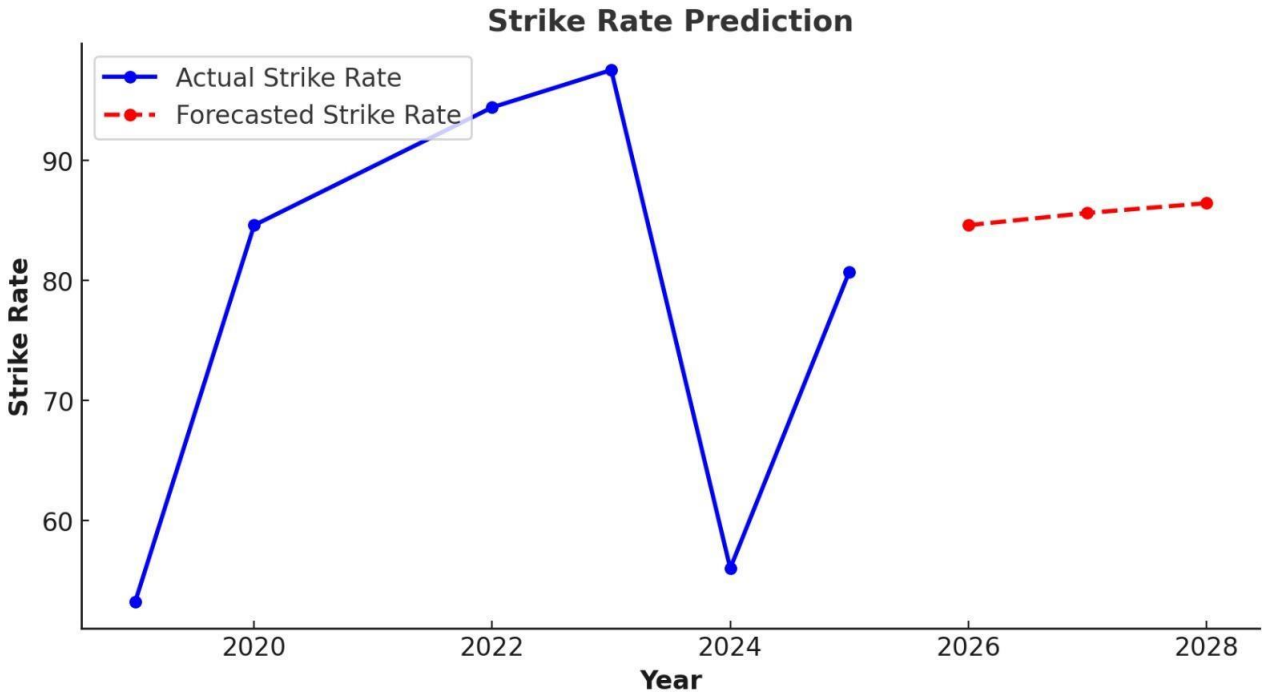
p-value = 0.260 > 0.05

After First order differencing the p-value still above 0.05 indicates that time series is Non-stationary. Then I applied the second order differencing and check again.

p-value = 0 < 0.05

Since after second order differencing p-value < 0.05 the data is Stationary. Now I can apply the Exponential Smoothing Model to the Dataset and forecast Strike rate for 2026 to 2028.

After applying the ADF test and Checking the Stationarity, I applied the Exponential Smoothing model for forecasting the **Strike Rate of ODI format** in Python software in form of Graphical Representation as –



Year	Forecasted Strike rate (ODI)
2026	84.61
2027	85.63
2028	86.44

After Ensuring Stationarity, the predicted Strike rate in ODI Format using Exponential Smoothing are :

Conclusion :-

The model predicts a steady increase in Strike Rate over the next three years and Plot shows the gradually increasing pattern.

2) TEST Format :-

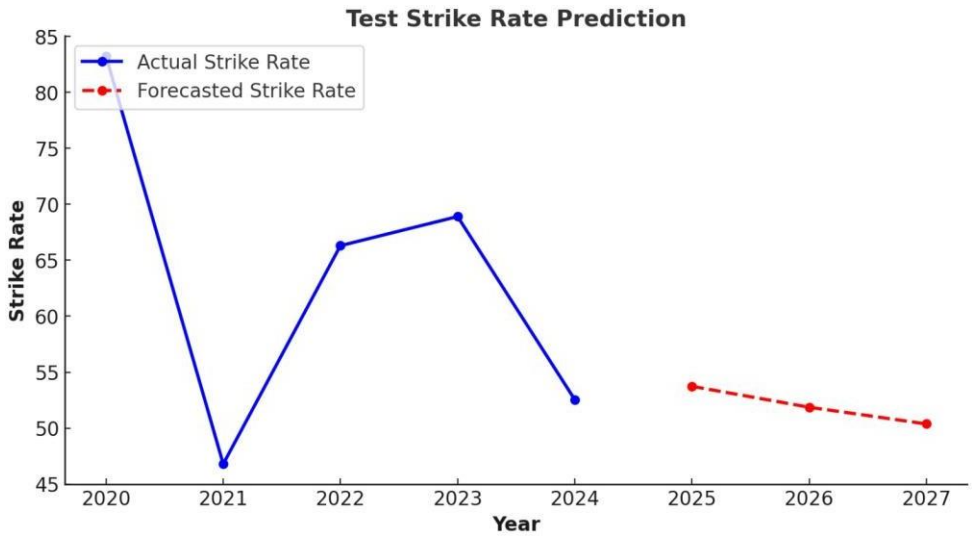
The dataset contains match-wise Test Performance data, including Strike rate, Runs, Balls faced, and Match dates. Now, I agrigated the rate yearly to prepare for Stationarity testing.

I checked if the time series is stationary using the Augmented Dickey-Fuller (ADF) test.

p-value = 0.00007 < 0.05

Since, p-value < 0.05 , We Reject the Null Hypothesis (Ho) at 5% L.O.S. This confirms that the time series is already stationary, so differencing is not required.

Now I applied the Exponential Smoothing Model and forecast the Strike rate in TEST format for the next 3 years. (2025-2027). I visualize the actual and forecasted strike rates using a Graph.



Year	Forecasted Strike rate (TEST)
2026	53.73
2027	51.87
2028	50.37

This suggests that if the current trend continues, Gill's Test strike rate may slightly decrease in the coming years.

- **Conclusion :** The fluctuations in past strike rates indicate inconsistency, which may be due to different playing conditions, opposition, or batting form. However, actual future performance can vary depending on factors such as playing style, team strategy, and match conditions.

➢ MAJOR FINDINGS :-

Based on Exponential Smoothing applied to the dataset, the key findings for Shubman Gill's projected performance in ODIs and Tests for 2026-2028 are :

1) Performance Trends in ODI's :

- The research suggests that **Steady Increase** in yearly **Runs** of ODI format over the next 3 years.
- The model identifies a **Positive Trend** in Shubman gill's **Bating average** performance over the time.
- The model predicts a **Steady Increase** in **Strike Rate** over the next three years and Plot shows the gradually increasing pattern.

2) Performance Trends in TEST :

- The research predicts a **Steady Increase** in **Runs** for the next three years and the historical yearly runs along with the predicted values for the next three years.
- The fluctuations in past **Strike Rates** indicate **inconsistency**, which may be due to different playing conditions, opposition, or batting form.

➢ RESULTS AND FORECASTING :

Forecasted Runs in ODI & TEST :

1. **ODI –**

2026 – 800 Runs

2027 – 903 Runs

2028 – 1003 Runs

2. **TEST –**

2026 – 1107 Runs

2027 – 1317 Runs

2028 – 1527 Runs

Forecasted Batting Avg. in ODI :

1. **ODI –**

2026 – 53.94

2027 – 56.42

2028 – 58.40

Forecasted Strike Rate in ODI & TEST :

1. **ODI –**

2026 – 84.61

2027 – 85.63

2028 – 86.44

2. **TEST –**

2025 – 53.73

2026 – 51.87

2027 – 50.27

➢ CONCLUSIONS :

The application of Exponential Smoothing and Holt's Linear Trend Model successfully captured trends in Gill's performance. The Analysis suggests **a Promising Future in ODIs, with consistent improvement in Runs, Batting Average and Strike rate**. However, the **Test format Strike Rate might slightly decline**, possibly requiring adjustments in playing style. This Research provides valuable insights for Cricket Analysts, Coaches and Fans in predicting future performance trends.

➢ LIMITATIONS :

- External factors : The model does not account for injuries, Team selection or Match Conditions which can significantly impact performance.

- This all predictions which helps in analyzing trend of player for team strategy, but they may not be directly useful for team selection.

- The dataset covers only few years which may not be sufficient for long- term analysis.

➢ REFERENCES :

- ESPNcricinfo : https://www.espncricinfo.com/cricketers/shubman-gill-1070173/bowling-batting-stats
- ICC Official Records (2025) : Player Rankings & Statistics Retrieved from www.icc-cricket.com.