Modern Techniques for Analysing Performance of Cricketers

Debopriya Dey[1], Ankesh Samanta[2], Ankan Sadhukhan[3] andAniket Pd Singh[4]

Assistant Professor, Basic Science and Humanities, Narula Institute of Technology, Kolkata, India[1]

Assistant Professor, Mechanical Engineering, Narula Institute of Technology, Kolkata, India[2]

Student, Computer Science and Engineering, Narula Institute of Technology, Kolkata, India[3]

Student, Computer Science and Business System, Narula Institute of Technology, Kolkata, India[4]

Abstract: Maths is a promising prospect in any field right now. It is involved in overwhelming number of sectors. In this paper we intend to provide deeper insights into how the entire sports industry has been affected in a multifarious way, we provide example of data driven analysis and also field level analysis based on statistics focused on the training, selection and playing of players and in their management. Mathematics is everywhere from daily lives to sports. This article will give a brief understanding into the world of mathematics and how it influences sports across the world. The continuous development of technology, maths and statistics in recent years has played an important role in sport. As use of technology to compute, analyse and improve performance gains momentum, sport is not an exception in this regard. From amateur athletic training to premium sporting teams and clubs, parallel technology is used to give athletes feedback on their training and also predict injuries. Statistical data is taken into equal consideration by certain football coaches when the time comes to get ready for the next tournament. Analysing a team's performance can help coaches and managers carve out strategic decisions based more on data and less on the coach's instinct. It's a complex process, but it makes all the difference.

In this paper we are going discuss elaborately about the different parameters used to assess data e.g., heart-rate, average speed, calories burnt, distance covered etc. Overall, this paper will bring forth the role of statistics in the field of sports and also throw light on the development of new technology or devices to ease such data collection and thus ultimately the benefit the players in the world of sports. Here we use multiple regression model shows the trend of performance and predicts the future performance of the players by analysing some important parameters.This method has proven itself to be requisite as it enables athletes to have a scientific awarenessof their performance and spot areas for improvement.

Keywords: Strategic Decisions, Multiple regression model, Trend of performance, Future performance prediction.

I.   INTRODUCTION

Cricket, or the gentleman's game is a very old, widespread, and uncomplicated pastime game. In the late 16th century, the sport of cricket has originated in the south-east England. It became the country's national sport in the 18th century and has developed globally in the 19th and 20th centuries and yet the one of the most popular games in the planet. It is a game of uncertainty, hard to predict the outcome of the game up to the last moment though the possible results are known to all, therefore, an appropriate statistical model can be applied to predict the performance and outcome of the game which we have tried in this paper.

Sports analytics and Data Visualization play an essential role for Player selection, team managers, and boosting their on-field performance. Decision making and analysis, is the process of applying different algorithms on data to gain insights into prediction of the future. This data is made to undergo several algorithms, tools, and visualization techniques to make way for suggestion of the players to create the team. To build predictive models various machine learning techniques are applied.

As we all know that cricket has his own huge fan following and is a very popular sport in the India and some of the other parts of the world also, which makes it an interesting game to have research on. Much research has been going and had been done in this game like and will be going on.

This paper discusses about the use of statistical use of cricketer's data and using them for the following purpose:

- Analysing and data.
- Performance prediction
- Age as factor of decrease in cricketer's performance.

To do these statistics and calculations we have used the below mentioned statistical methods:
- Multiple linear regression model
- Piecewise liner regression model
- Data analysing
- Data collection

The above-mentioned methods have been explained in the paper with the use of it in the predictions and calculations.
By the end of the paper, we tried to achieve the following goals:
- Performance prediction old given data.
- Effect of aging in players performance.

II.   METHODOLOGY

From the introduction part it must have been cleared that the paper consists of research on prediction of performance using statistical methods. For this three of the popular methods have been used in this paper.

    a)   Multiple regression model
    b)   Segmented regression model/ Piecewise linear model
    c)   Data analysis

The multiple regression model and segmented regression model are mainly used for getting the impact of age in an athlete's performance, while the other method (Data analysis) is used to show and predict the performance of a player with time and seasons.

Multiple regression model
Multiple regression model is a statistical method which is used to analyse the relationship between one dependent variable and several independent variables.
In this case an athlete's performance is based on not only one factor but many factors like physical fitness, skills, training, age etc.
So, the use of this model is most with the help of which we can get to predict the performance and rank them.

Segmented regression model/ piecewise regression model
Segmented regression, also known as broken-stick regression, is a method in regression analysis in which the independent variable is divided into parts or interval and a separate line segment is fit to each interval which helps to fit a model linearly.

Data analysis
Statistical analysis is the process of collecting and analysing data to discern patterns and trends. This technique is useful for collecting the interpretations of research, developing statistical models, and planning surveys and studies.
In this paper we have taken data of 5 Indian cricketers and 5 footballers from premiere league to analyse and predict their performance for future.
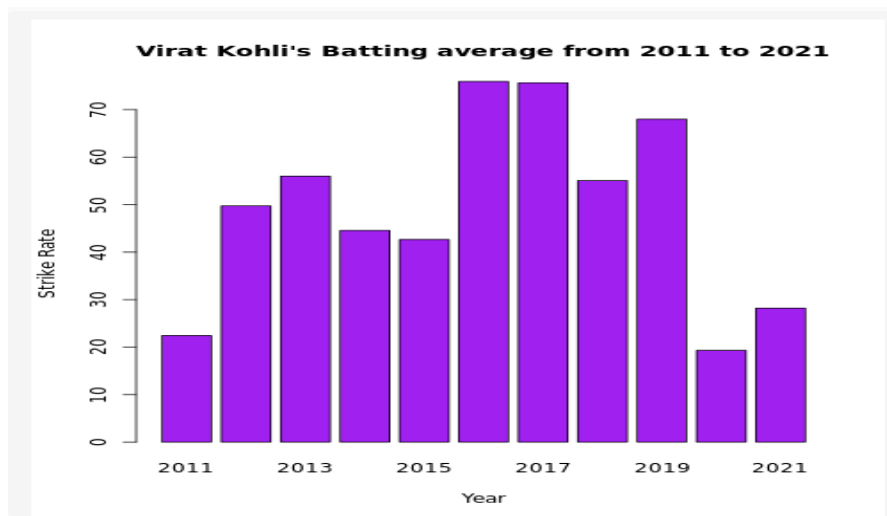

Fig. 1 Virat Kohli's Batting Average from 2011-2021

III. CASE STUDY OF CRICKETERS

To do the research on prediction of performance for cricket athletes we gathered performance data of their previous matches with the help of internet and used some specific factors only which tells about the performance of a cricketer. We gathered data of 3 Indian cricketers Rohit Sharma, Virat Kohli and Ravi Chandra Ashwin. The raw data gathered for all the three players has been given below in table 1,2 and 3 respectively.

TABLE 1 DATA COLLECTION OF ROHIT SHARMA

YEAR	HIGH SCORE	AVERAGE	STRIKE RATE
2013	333	66.6	62.83
2014	237	44.54	44.54
2015	326	48.65	48.65
2016	288	64.14	64.14
2017	217	69.7	69.77
2018	184	47.30	47.30
2019	556	75.95	75.95
2020	906	48.19	48.19

TABLE 2 DATA COLLECTION OF VIRAT KOHLI

YEAR	HIGH SCORE	AVERAGE	STRIKE RATE
2011	202	22.44	42.7
2012	689	49.21	46.74
2013	616	56	54.65
2014	847	60.57	60.54
2015	1000	69.66	74.05
2016	1215	75.93	76.41
2017	1359	75.64	80.24
2018	1022	55.08	69.33
2019	612	68	53.28
2020	516	40	44.07
2021	116	28	35.45

TABLE 3 DATA COLLECTION OF RAVI CHANDRA ASHWIN

YEAR	WICKETS TAKEN	AVERAGE	ECONOMY
2011	26	24.8	2.91
2012	37	37.75	3.15
2013	41	22.51	2.59
2014	10	43.8	3.10
2015	62	17.20	2.83
2016	72	23.9	2.94
2017	56	27.58	2.78
2018	38	23.56	2.49
2019	20	24.15	2.78
2020	13	21.23	2.41
2021	54	16.64	2.32

The above given data has been used in this paper for the prediction, by using the statistical method multivariate linear regression model for getting a graph of which shows the trend of performance and predicts the future performance. So, after using the data for multiple linear regression model we get a table which have the following useful data. Table number 4,5 and 6 shows the result of using the given table data for multiple linear regression model.

The general form of Multiple linear regression model is given by:

$$Y = \text{intercepts} + b_1(x_1) + b_2(x_2) + \ldots + b_n(x_n)$$

Where: b_1, b_2…b_n are the coefficients of independent variables. X_1, x_2, x_3… x_n are the independent variables. Intercepts denotes intercept coefficient.

TABLE 4 ROHIT SHARMA'S DATA AFTER USING MLR

Column1	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	1.103750542	2.935056887	0.376057632	0.722302139	-6.441053378	8.648554462	-6.441053378	8.648554462
HIGH SCORE	0.000390147	0.002380891	0.163866096	0.876254519	-0.005730127	0.006510422	-0.005730127	0.006510422
AVERAGE	0.970501681	0.046857588	20.71172953	4.85776E-06	0.850050418	1.090952945	0.850050418	1.090952945

The above data is obtained with the help of Microsoft Excel using the Data Analysis tool which is being used for getting the required values like the coefficients and P-values as shown above.

Using TABLE 4 we have the required coefficients to get the multiple linear regression line as:

$$Y = 1.10375054210861 + 0.000390147268911624(x_1) + 0.97050168132702(x_2)$$

Here, x_1 & x_2 indicates high score and batting average respectively and Y is the strike rate for the batsman Rohit Sharma.

TABLE 5 VIRAT KOHLI'S DATA AFTER USING MLR

Column1	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	28.68173891	5.390828945	5.320469116	0.000710635	16.25046507	41.11301275	16.25046507	41.11301275
HIGH SCORE	0.034836926	0.008469486	4.113227844	0.003375572	0.015306256	0.054367597	0.015306256	0.054367597
AVERAGE	0.060791462	0.181682587	0.334602576	0.746528648	-0.358169336	0.479752259	-0.358169336	0.479752259

Similarly, the above table is what we get from MS Excel with the help of the same Data Analysis tool. And from the table we can get another equation for Virat Kohli which is mentioned below.

$$Y = 28.6817389118202 + 0.0348369264675218\ (x_3) + 0.0607914617006704\ (x_4)$$

Here, x_3 & x_4 indicates high score and batting average respectively and Y is the strike rate for the batsman Virat Kohli.

TABLE 6 RAVI CHANDRA ASHWIN'S DATA AFTER USING MLR

Column1	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	1.800736896	0.286046513	6.295259029	0.00023398	1.141112453	2.460361339	1.141112453	2.460361339
WICKETS TAKEN	0.00474871	0.00315708	1.504146471	0.170954853	-0.002531528	0.012028949	-0.002531528	0.012028949
AVERAGE	0.029862594	0.007882274	3.788575697	0.005322628	0.011686036	0.048039151	0.011686036	0.048039151

Similarly, the above table is what we get from MS Excel with the help of the same Data Analysis tool. And from the table we can get another equation Ravi Chandra Ashwin which is mentioned below.

$$Y = 1.80073689590561 + 0.00474871012800812\ (x_5) + 0.0298625936003207\ (x_6)$$

Here, x_5 & x_6 indicates wickets taken and bowling average respectively and y is the economy for the bowler Ravi Chandra Ashwin. In Cricket strike rate and economy of the cricketer which are one of the crucial parameters used for depicting a player's performance. So using data analysis and Multiple regression model we can easily calculate these parameters for any batsman or bowler and by help of these we can predict the performance of the players.

IV.   AGE FACTOR

The natural process of aging has a significant impact on everyone, and so the effects of aging on sportspersons are no exception. As per the demand of sports like hockey, soccer, basketball, and many more the physical aspect of players dominates their performance very much in the field. It has been widely reported as relative age effect (RAE) in sports. The presence of age effect on the performance of the cricketers is also conveyed continuously in different newspapers and magazines. In cricket, physical as well as the mental ability and technical skills separate the best-performing players from the rest.

A case study of Sachin Tendulkar

We all know about Sachin Tendulkar and his contribution in the game of cricket, he is an inspiration for all the cricketers and fans of cricket not only in India but also outside the country too. The famous cricketer is well known all over the world as the "God of Cricket". He started his career as a 16-year-old kid in 1989 playing against the country's biggest rivals Pakistan where he faced the mighty baller of Pakistan Wasim Akram, Imran khan and Waqar Younis. He was hit on the face by Waqar Younis but continued to bat in a blood-soaked shirt.

The way Sachin played the game so beautifully during his active playing years (1989-2013) was worth of watching. If we look in his career's performance, then by using data analysis we can tell the duration of his peak performance.

TABLE 7 SACHIN TENDULKAR'S CAREER RECORD

1989	215	35.83
1990	373	41.44
1991	78	19.5
1992	419	41.9
1993	640	91.42
1994	700	70
1995	58	29
1996	623	41.53
1997	1000	62.53
1998	647	80.87
1999	1088	68
2000	575	63.88
2001	1003	62.68
2002	1392	56.68
2003	153	17
2004	915	91.5
2005	444	44.4
2006	267	24.27
2007	776	55.47
2008	1063	48.31
2009	541	67.62
2010	1562	78.1
2011	756	47.25
2012	357	23.8
2013	276	34.5

Now from the above data, if we make a line graph showing the above-mentioned credentials then it will look like the below graph only.
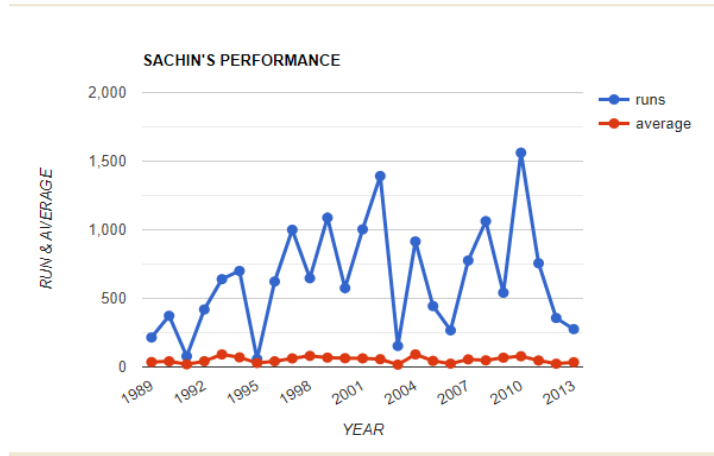
IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

5ᵗʰ National Conference on Science, Technology and Communication Skills – NCSTCS 2K22

Narula Institute of Technology, Agarpara, Kolkata, India

Vol. 9, Special Issue 2, September 2022

Fig. 2 Sachin Tendulkar's Performance Graph From1989-2013

Result

Now, from the above graph we can see that Sachin was in his peak or the best in his career during the years of 1997-2004 as during these periods he scored most of his runs consistently. If we try to calculate his age during that period, then he would be around the age of 24 in 1997 and 31 in 2004.

Case study of M.S.Dhoni

another name that comes in our mind when we talk about cricket is M.S.Dhoni the great captain of Indian cricket team under whose captainship India won the ICC World cup 2011.he played the international games from 2004 to 2019 and we get to talk about his playing data then we can get the following result.

TABLE 8M. S. DHONI'S PERFORMANCE TABLE

2004	19	9.5
2005	895	49.72
2006	821	41.05
2007	1103	44.52
2008	1097	57.74
2009	1198	70.47
2010	600	46.15
2011	764	58.77
2012	524	65.5
2013	753	62.75
2014	418	52.25
2015	640	45.71
2016	278	27.8
2017	788	60.62
2018	275	25
2019	600	60

If the above-mentioned data is derived in a line graph, then the following result can be obtained using the data analysis method for depicting the best performing age for him.
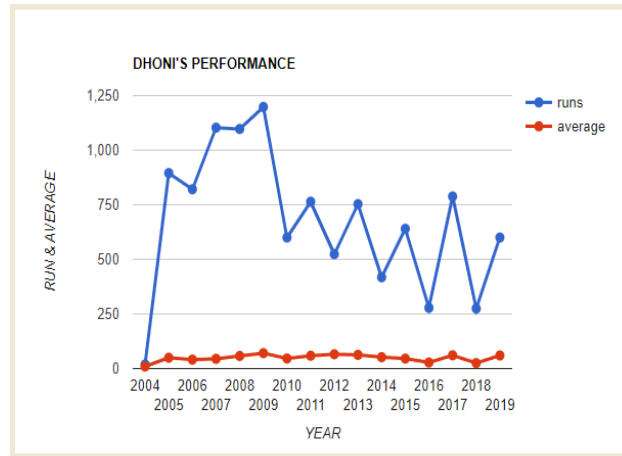
Fig. 3 Line graph for M. S. Dhoni's performance

RESULT

From the above graph we can see that from the years of 2005 to 2010 M.S.Dhoni was at the peak of his career and scoring runs consistently. If we calculate his age during this period, then we can say that during the time when he was of 24-29, he was at peak if his career.

V. CONCLUSION

In this paper, we have provided a comprehensive review of how we can (1) predict the performance of a player by obtaining a simple equation using the statistical methods and talked about the (2) impact of age in the performance in a player's career, (3) the point of his life in which he is at the best of career.
 However, the analysis of the literature on sports data has led to the conclusion that although researchers have proposed some methods to resolve the problems in sports data area, the solutions of some crucial issues remain unknown, such as predicting the athletes' performance in the knowledge graph, finding a rising star of sports, unified sports big data platform, open sports big data, and privacy protections.

REFERENCES

[1] Weihs, C., &Ickstadt, K. (2018). Data science: the impact of statistics. *International Journal of Data Science and Analytics*, *6*, 189-194.
[2] Bai, Z., & Bai, X. (2021). Sports big data: management, analysis, applications, and challenges. *Complexity*, *2021*, 1-11.
[3] Saikia, H., Bhattacharjee, D., & Mukherjee, D. (2019). *Cricket Performance Management: Mathematical Formulation and Analytics*. Springer Singapore.
[4] www. https://www.espncricinfo.com/player/ms-dhoni-28081/bowling-batting-stats
[5] www. http://www.howstat.com/cricket/statistics/players/PlayerYears_ODI.asp?PlayerID=3243
[6] www. https://www.rapidtables.com/tools/line-graph.html
[7] www. https://www.statskingdom.com/410multi_linear_regression.html
[8] www. https://stats.blue/Stats_Suite/multiple_linear_regression_calculator.html.
[9] Ahmed, F., Deb, K., & Jindal, A. (2013). Multi-objective optimization and decision making approaches to cricket team selection. *Applied Soft Computing*, *13*(1), 402-414.
[10] Hughes, M. D., & Bartlett, R. M. (2002). The use of performance indicators in performance analysis. *Journal of sports sciences*, *20*(10), 739-754.