Career Analysis of Cricketers in T20 Internationals and Prediction of Runs

Atul. A. Das[1], Brindha G R [2]*

School of Computing[1,2], SASTRA Deemed University, Thanjavur

*brindha.gr@ict.sastra.ac.in

ABSTRACT

The focus of this study is with the intent of analysing the careers of three prominent International Cricketers by taking the statistical data. The statistical data related to the matches which they have batted have been considered for the purpose of exploratory data analysis and prediction. The paper gives the reader the idea of the number of fours, the number of sixes and the average of the three players throughout the span of their careers. It has also described a Single Linear Regression predictive model and a Multiple Linear Regression model which predicts the total number of runs each player will score in the next year based on the data of the previous years. Three machine learning algorithms such as Logistic Regression, Support Vector Machines and Naive Bayes Classifier) have been used to classify whether the batsman gets out or not. To conclude, the three players considered for the analysis are Virat Kohli, Jos Buttler and David Warner.

Key Words: International Cricketers, Runs prediction, Linear Regression, Logistic Regression, Support Vector Machines, Naïve Bayes Classifier

1. INTRODUCTION

The format of T20 Internationals has its origins in the year 2005 when Australia and New Zealand played the first ever T20 International match. Over the years the format has become the most popular format of the game and due to this there has been an increase in viewership as well as an increase in profits for the game. The main reason for this was the fact that due to the shorter duration, the new generation of people liked this format and this had equally pressurizing situations as the two older formats of cricket-Tests and One Day Internationals (ODIs). The key problem with the two older formats was that the durations of the match lasted for one day in ODIs and in tests it lasted for almost five days. As far as T20Is are concerned the matches last only for a few hours of the day and it has only twenty overs per innings. The impact of T20Is is the fact that it has produced more aggressive batsmen and they play it with great panache. A really good example would be Chris Gayle. There have been a few classical players like Faf Du Plessis who anchor these matches. These players absorb the pressure and keep the runs coming despite the fact that many wickets have been lost in the other end. Then there are finishers like M S Dhoni and Rinku Singh who come to bat in the last minute when

1 **The paper presented at the International Mathematics Conclave (IMC-2023) in SASTRA, Thanjavur**

their team is under excessive pressure and win extremely difficult matches for their team. Seeing improvements in the batting, even the bowlers have improved in their skills and every bowler yearns to learn variations through which they get wickets and also increase the pressure in the batting team. The fielding has also improved considerably among teams. The impact that T20Is have had in various teams like India is immense as it has increased the morale and also allowed teams to play their cricket more confidently.

This topic has been a topic of immense study in the recent years. Quantifying the performance of cricketers is extremely essential for measuring player competencies [1][6]. It also indicates the level of achievements achieved by the players [8]. An example a statistic being to quantify the performance of cricketers is the Batting Strike Rate. Batting average is another such quantifier that is commonly used and it has a lot of versions suggested by mathematicians Kimber and Hansford (1993) and Damodaran (2006)[1]. Algorithms such as Naïve Bayes algorithm and K Nearest Neighbour [2], Support Vector Machines [11] and Logistic Regression [9] have been used for the prediction of scores and analysis of batting first[4]. In fact, the KNN and Naïve Bayes models have been used for analytics of Cricket games results and have given accuracies of 50% and 60% respectively. It has also been used for the selection of teams [7]. Correlation matrices help in the selection of proper features and help in the overall improvement of model training and performance [10]. Libraries like Numpy and Pandas are such libraries which have been used extensively for Data Analysis [12]. Descriptive and Predictive models can be used to predict the scores of players based upon factors like match locations and opponents [3]. These factors can be used to figure out the exactness of predictions made [13]. There are many other factors that also play a pivotal role like the statistics of lower-division matches played by them which will ensure their success in their International careers[5]. The best of all these factors have been considered and the analysis and prediction of players' careers have been done. To conclude, this paper does an analysis of their careers using graphs and pie charts and it also predicts and classifies data based on the players.

2. FEATURE EXTRACTION

The datasets of Virat Kohli, Jos Buttler and David Warner have been obtained from the ESPNCricinfo website. In the ESPNCricinfo website, the player profile was opened and after opening this, the batting innings list has been considered. After opening the batting innings list, the data has been scraped. Once the process of scraping has been done, the extracted data was stored in an Excel file. In this Excel file, there were many issues. Firstly, only numeric data has to be considered (in some numbers in the runs list there was a '*' appended after the number to denote the fact that the player was not out in the game) and secondly, there was existence of special characters. To resolve these issues, a not outs column was created and if the player was not out, it was represented with a '1' and if he was out, it was represented with a '0'. To resolve the other issues, the extra spacing and special characters were removed.

3. FEATURE DESCRIPTION

The data of the three players namely, Virat Kohli, Jos Buttler and David Warner have been taken from the ESPN Cricinfo website. These datasets are available for free. However, a few modifications are to be made to the dataset to make it fit for analysis. There have been many matches in which the three players were there in the team but they did not play it. These values have been removed from the datasets. In the datasets, the matches in which a player has remained not out have been marked with a star (*) next to the number. We have modified it in such a way that we have made runs to be numeric in nature and an extra column has been added which has binary values of 0 and 1. This will signify if the batsman is out or not (1 if not out 0 if out). From the data we observe the following Table 1.

Parameter: The parameter considered.
Virat Kohli, Jos Buttler, David Warner : The players
Min, Score: Minimum score of the player
Max Score: Maximum score of the player

Table 1. Scores of Players

Parameter	Virat Kohli	Jos Buttler	David Warner
Min Score	0	0	0
Max Score	122	101	100

3.1 Number of Matches Played against opposition

Among the three batsmen, Virat Kohli is the only batsman who has played matches against Zimbabwe, Scotland and Hong Kong. David Warner and Jos Buttler have never played against these three teams (Figure 1). With regards to matches played against Sri Lanka, David Warner tops the list with 17 matches played in his career against Sri Lanka. In comparison, Virat Kohli has played 7 matches against Sri Lanka and Jos Buttler has played 9 matches against Sri Lanka. Virat Kohli has played more matches against Australia (21 matches) as compared to Jos Buttler (15 matches).

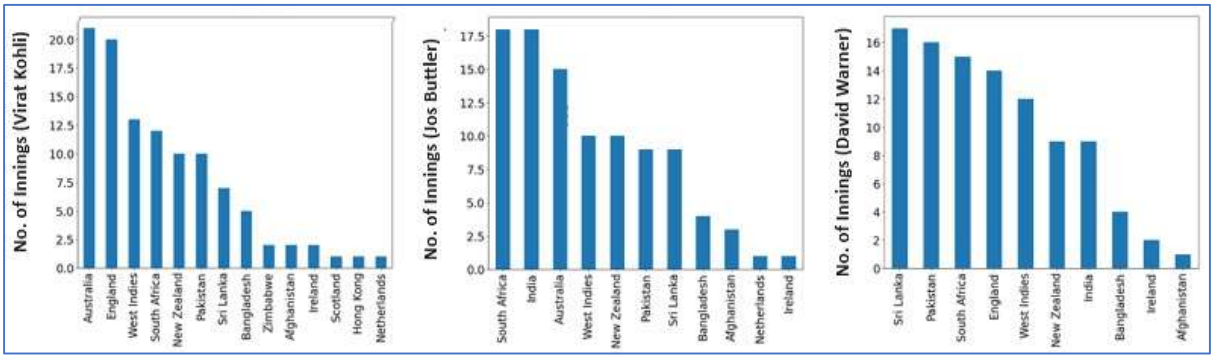


Fig 1. Country-wise Comparison of Innings

Number of Innings: Number of matches played

Country: The opposition the player played against

3.2 Number of Innings played per year

Among the three players, David Warner was the first one to start his T20I career (2009). Virat Kohli started his career in 2010 and Jos Buttler started his career in 2011 (Figure 2). Jos Buttler is the only player among the three to play matches in the year 2023. Except Virat Kohli, no other player among the three have played more than 14 matches in a particular year. Among the three players, Virat Kohli has played the maximum number of matches in two years (2022 and 2016)

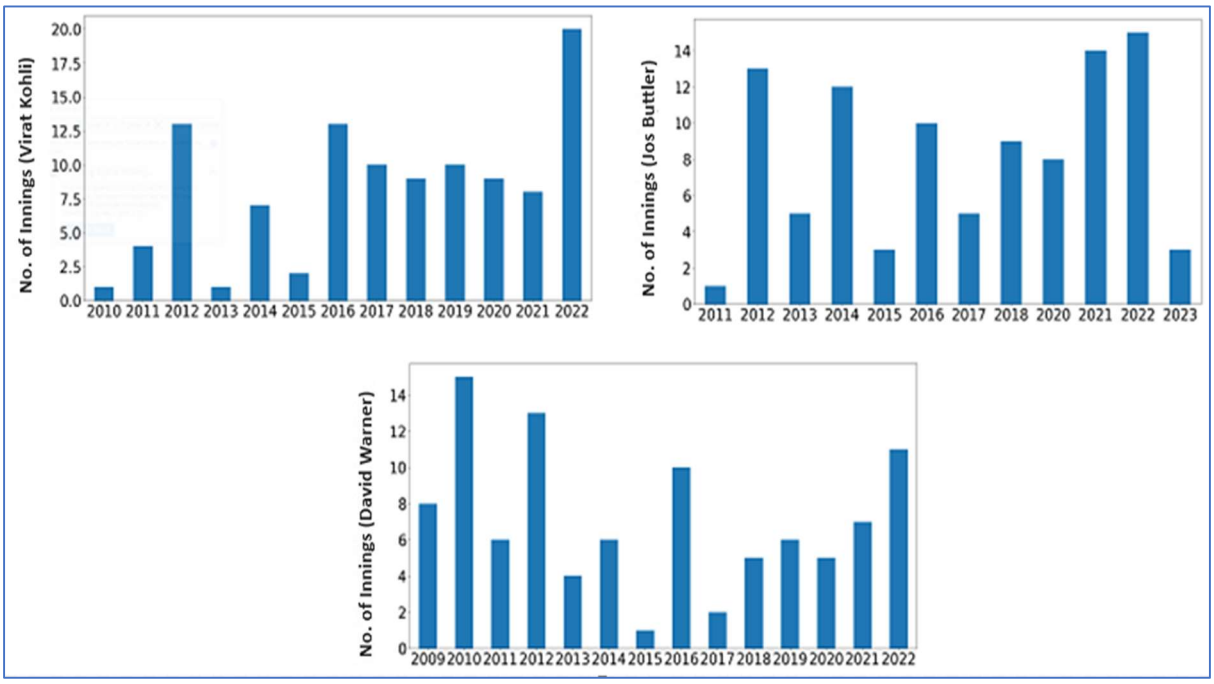


Fig 2. Year-wise Comparison of Innings

Number of Innings: Number of matches played

Year: The years in which the player has played

3.3 Number of Runs Scored Per Year

Jos Buttler is the only player to play T20 International matches in 2023. Both Virat Kohli and David Warner last played in the year 2022 (Figure 3). In the year 2022, Virat Kohli has scored the maximum number of runs among the three players. Second in the list is David Warner and third in the list is Jos Buttler. David Warner is the only player to play T20 International matches in the year 2009. We cannot say that these three players are inconsistent because in most of these years these three players have played different number of matches. Jos Buttler did not play T20 International matches in the year 2019.

In this year, Virat Kohli has scored more runs when compared to David Warner. David Warner scored the bulk of his runs in the year 2010. Whereas Virat Kohli and Jos Buttler scored the bulk number of runs in the year 2022 and 2021 respectively.

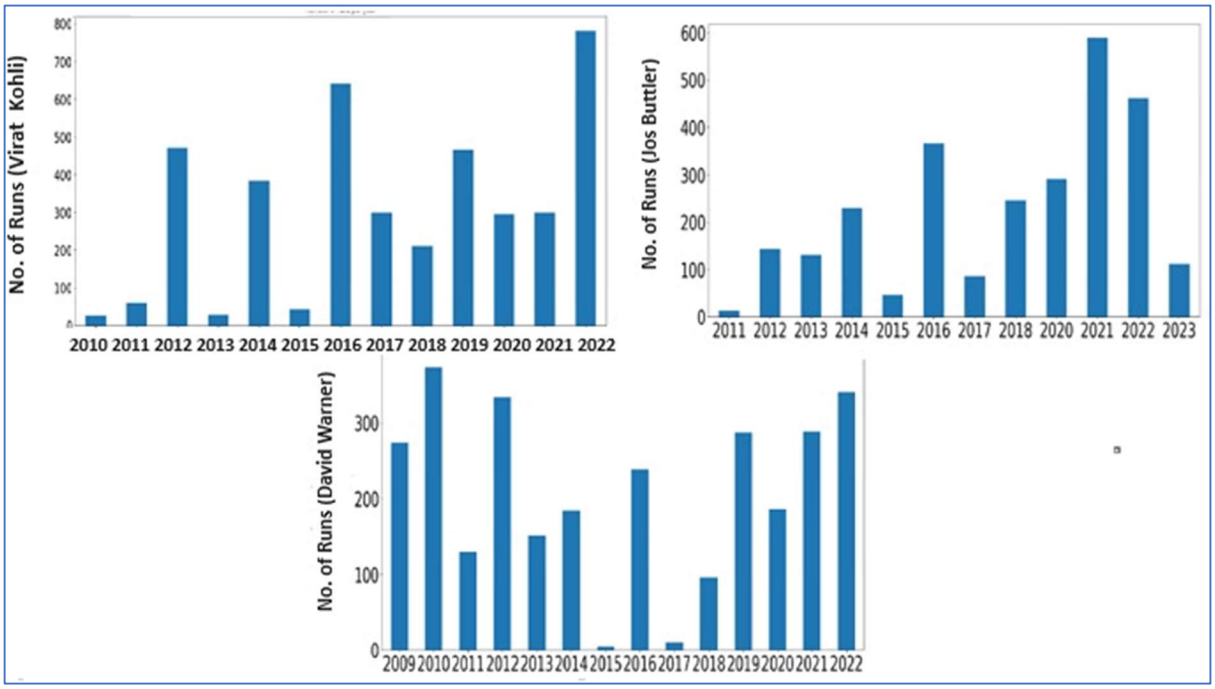


Fig 3. Year-wise Runs Comparison

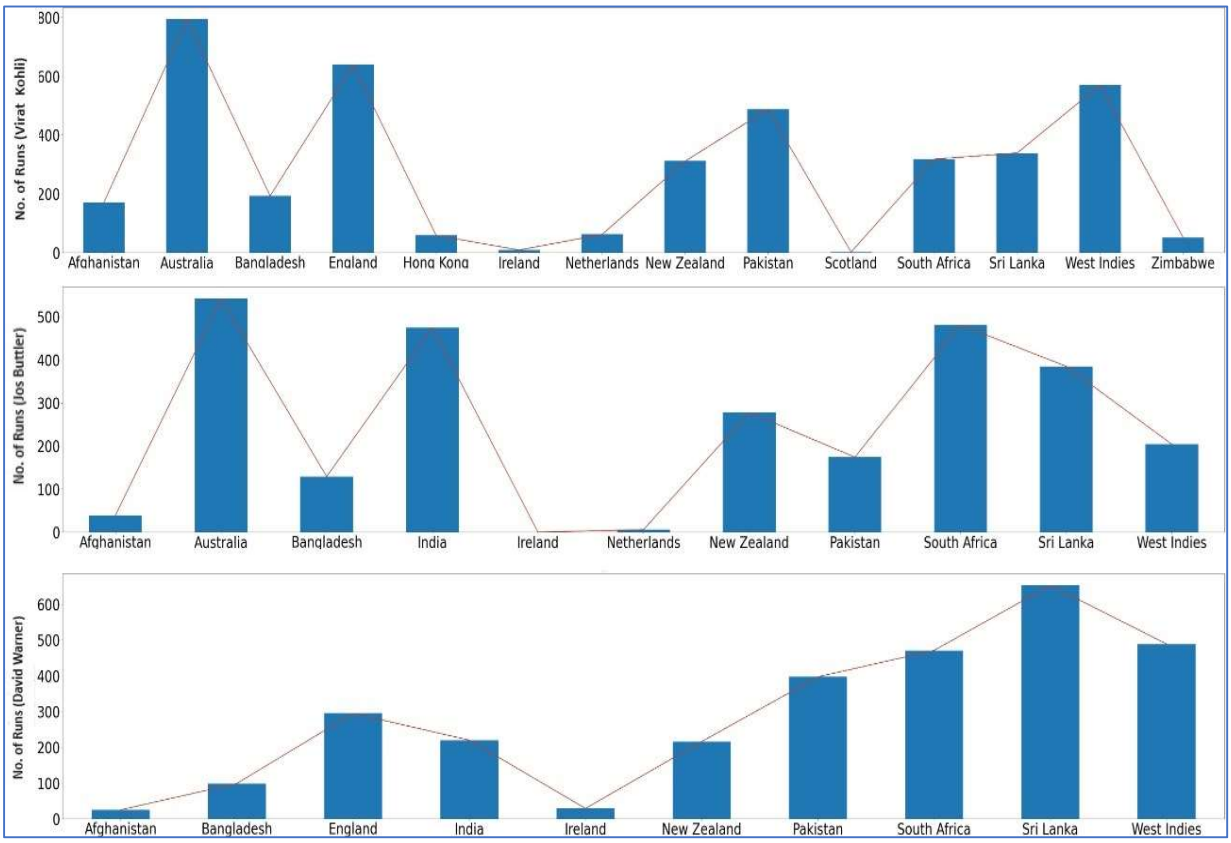
No of Runs: Number of Runs scored

Year: The years in which the player has played.

3.4 **Number of Runs against Opposition**

Compared to all the players only Virat Kohli neared the landmark of 800 runs against opposition. Jos Buttler has neared the maximum limit of 500 runs and David Warner has neared the maximum limit of 600 runs (Figure 4). Against Australia, between Virat Kohli and Jos Buttler, Virat Kohli is the highest run getter. David Warner has scored his bulk of runs against Sri Lanka. Whereas Virat Kohli and Jos Buttler have scored bulk of their runs against Australia. Compared to Jos Buttler, Virat Kohli has scored more runs against Netherlands. With regards to playing against New Zealand, Jos Buttler has scored the maximum number of runs. Second in the list is Virat Kohli and third is David Warner.

David Warner and Jos Buttler are the only ones to have played against Ireland and among these two players, David Warner has scored more runs. among the three players, Virat Kohli is the best while playing against most Asian nations (except Sri Lanka as David Warner is the better



player in this case).

Fig 4. Comparison of Runs against Opposition

No of Runs: Number of Runs scored

3.4 Correlation Analysis

Virat Kohli and Jos Buttler have strong correlation with Strike Rate and Runs. David Warner have strong correlation with Match Date and Runs.

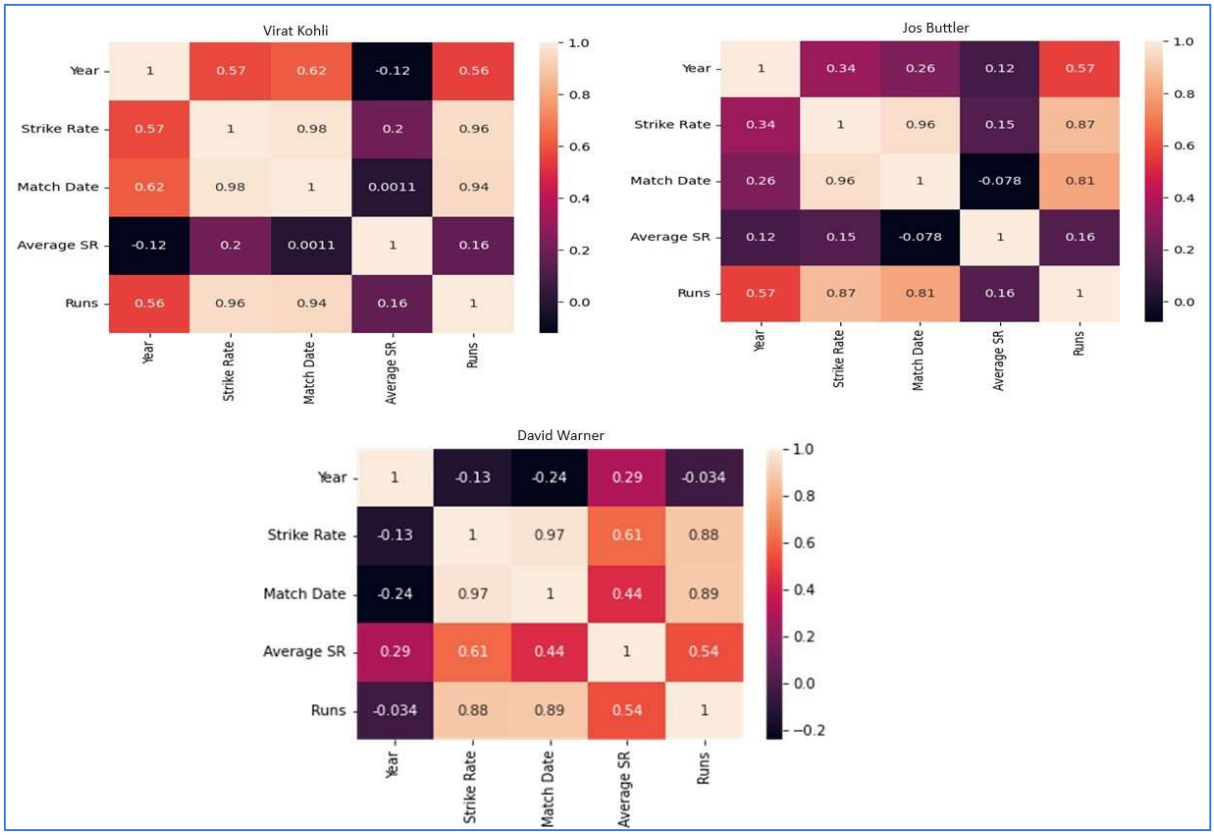


Fig 5.Correlation- Features Vs. Players

3.5 Total Number of Fours Hit in Career

Despite the fact that Virat Kohli has scored the bulk of his runs against Australia, he has hit the maximum number of fours against England. He has hit the second highest number of fours in his career against Australia (Figure 6). Jos Buttler has scored the bulk of his runs against Australia and he has hit the maximum number of fours against Australia as well. A similar case is present with respect to David Warner, he has scored most number of runs in his career against Sri Lanka and has hit the most number of fours against them as well. David Warner has hit fours against every opposition. However, the same cannot be told with regards to Virat Kohli and Jos Buttler. Virat Kohli has not hit a single four against Ireland and Scotland, and Jos Buttler has not hit a single four against Ireland and Netherlands. Among the three players, David Warner has hit maximum number of fours against any opposition as he has hit approximately 70 fours against Sri Lanka. Virat Kohli comes second with approximately 60 fours against England and Jos Buttler comes third in the list with approximately 50 fours

against Australia. Jos Buttler has hit more fours against India(40) as compared to David Warner(15). With regards to number of fours against England, Virat Kohli has hit more fours(60) as compared to David Warner(30).

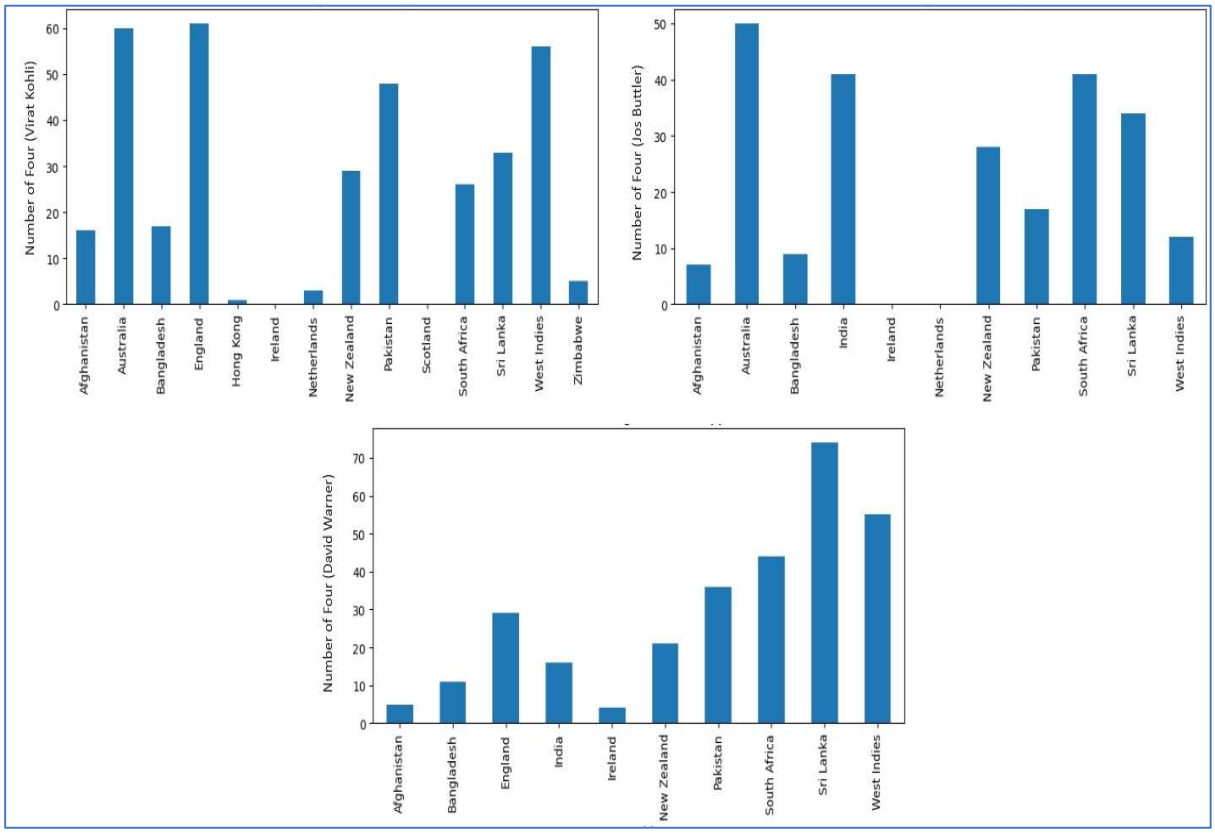


Fig 6. Comparison of fours against each opponent

Number of Fours: Number of Fours

Total Number of Sixes in Career

Virat Kohli is the only batsman to hit sixes against Afghanistan. Jos Buttler and David Warner have not hit a single six against Afghanistan (Figure 7). David Warner has hit almost 20 sixes against Pakistan, most number of sixes against this opposition among the three players. Virat Kohli has hit 12 sixes against Pakistan and Jos Buttler has hit 5 sixes against Pakistan. Virat Kohli is the only player to hit sixes against Hong-Kong. David Warner and Jos Buttler have not hit sixes against Hong-Kong. None of the three batsmen considered here have hit sixes against Ireland.

Jos Buttler is the only batsman in this list who has not hit a single six against West Indies. David Warner has hit 23 sixes against West Indies and Virat Kohli has hit 18 sixes against West Indies. Jos Buttler has not hit any six against Netherlands. David Warner has never played against Netherlands and Virat Kohli has hit 3 sixes against Netherlands.

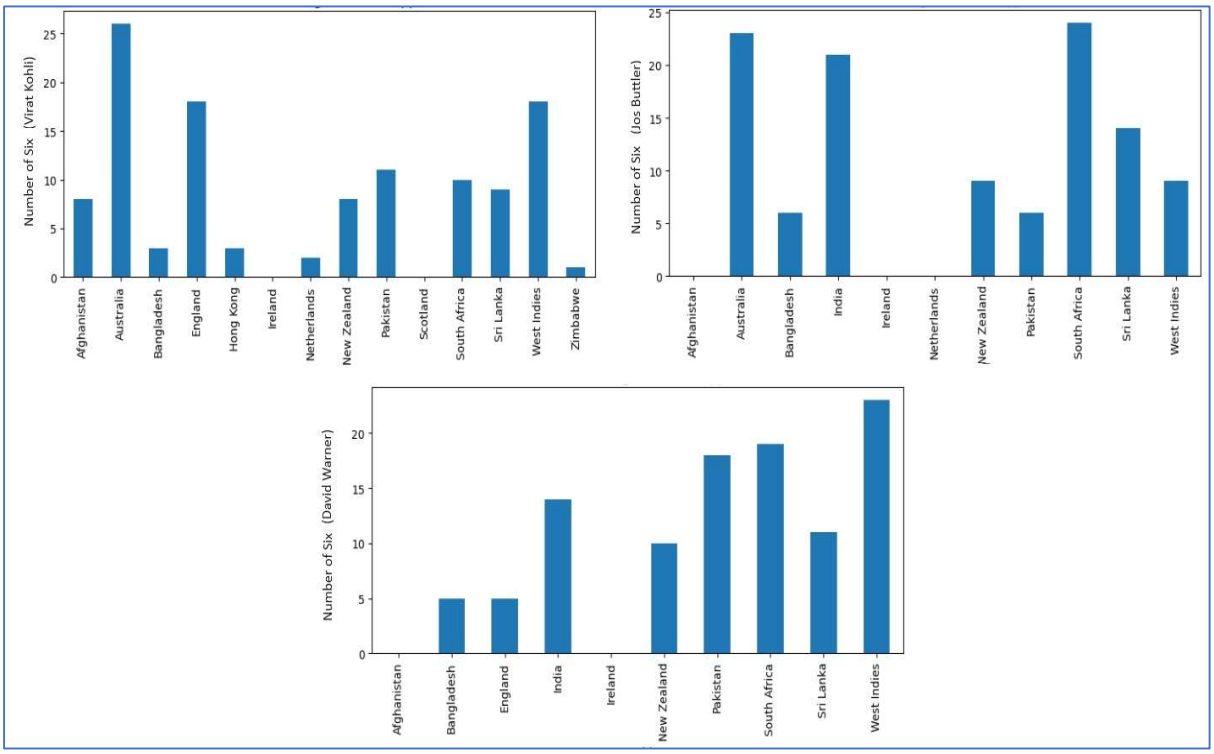


Fig 7. Comparison of sixes against each opponent

Number of Sixes: Number of Sixes

3.7 Pie Chart Showing Percentage of Boundaries

We can observe that approximately 75% of Virat Kohli's boundaries are fours and 25% of his boundaries are sixes.We can observe that approximately 68% of Jos Buttler's boundaries are fours and 32% of his boundaries are sixes. We can observe that approximately 74% of David Warner's boundaries are fours and 26% of his boundaries are sixes.

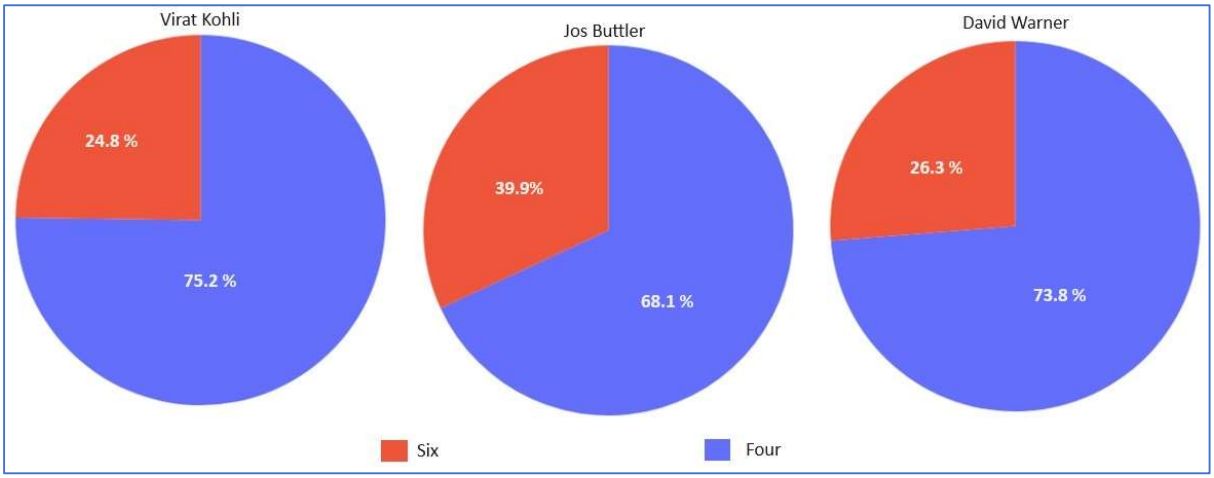


Fig 8. Comparison of fours and sixes

Six: Percentage of Sixes, Four: Percentage of Fours

3.6 Average Against Opposition

With regards to SENA nations (South Africa, England, Australia, New Zealand) Virat Kohli's average is above 50 only in Australia and in all the other nations his average is below 50 in all the other nations (Figure 9). Virat Kohli has the highest average among the three with an average of 175 against Afghanistan. Jos Buttler has the second highest average of 80 against Sri Lanka and David Warner's highest average is 50 against Sri Lanka. In all Asian nations with the exception of Sri Lanka, Jos Buttler's average has never crossed 50. His highest average in these countries has been achieved against Pakistan with an average of 40. Among the three players Jos Buttler has the highest average against New Zealand with an average of 40. Virat Kohli and David Warner have averages of 30 and 25 respectively. Jos Butter has the highest average against Sri Lanka with an average of 80. Virat Kohli comes second with an average of 75 and David Warner comes third with an average of 50. Among Virat Kohli and David Warner, Virat Kohli has a better average of 35 against England as compared to David Warner's 25. Virat Kohli has the highest average among the three players with regards to playing against West Indies. He has an average of 50 while playing against them. Second in the list is David Warner with an average of 45 and third in the list is Jos Buttler with an average of 30 against West Indies.

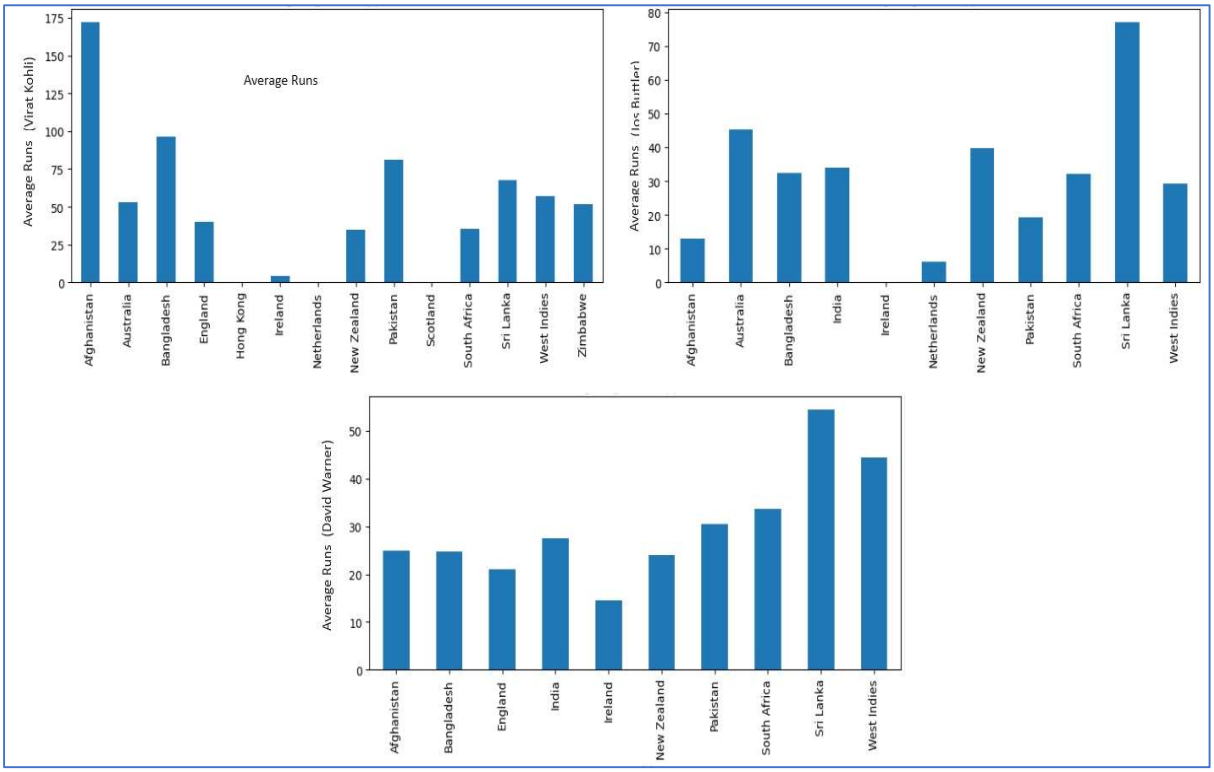


Fig 9. Comparison of Average Runs against Opponent

Average Runs: Average Number of Runs Scored

4 Predictive Analysis

This section consists of the different type of algorithms used in this research

4.1 Simple Linear Regression

It is a statistical method that determines the relationship between two types of continuous variables-the independent variable (the predictors) and the dependent variable, i.e. the variable to be predicted. This relationship between the two variable types is assumed to be a Linear Relationship, i.e. whenever the predictor value changes, the dependent variable value also changes. This al

The equation for the line is of the form

$$\boldsymbol{y = \beta_1 x + \beta_0 + \varepsilon}$$

where:

1. y is the dependent variable.
2. x is the independent/predictor variable.
3. β_0 is the intercept of the line and it is equal to y when x is zero.
4. β_1 is the slope of the line, i.e. it shows the change in value of y for every change in value for x.

5. ε is the error term. This is included to show the difference between the obtained y value and the original value, i.e. ý.

This type of Linear Regression will also estimate the value of β_0 and β_1 that minimize the Sum of Squared Errors (SSE) between the predictor variable and the dependent variable. Sum of Squared Errors can be represented by the following Cost function.

$$\textbf{Cost } (\boldsymbol{\beta}_0, \boldsymbol{\beta}_1) = \sum_{i=1}^{n} (\boldsymbol{yi} - (\boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \boldsymbol{xi}))^2$$

Multiple Linear Regression

This method is similar to Simple Linear Regression. The only difference is that in the case of MLR, there are multiple independent variables (known as predictor variables) and a dependent variable (also called the response variable). The assumption of a linear relationship holds true in this case as well, but instead of one independent variable, multiple independent variables are considered to influence the dependent variable.

The aim of Multiple Linear Regression is to find the best fit linear equation that represents the relationship between the dependent variable y and k independent variables $x_1, x_2, x_3 \ldots x_p$.

The equation for the line is of the form

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon$$

where:

1. y is the dependent variable.
2. $x_1, x_2 \ldots x_k$ are the independent/predictor variables.
3. β_0 is the intercept of the line and it is equal to y when x is zero.
4. $\beta_1, \beta_2 \ldots \beta_k$, is the slope of the line, i.e. it shows the change in value of y for every change in value for x.
5. ε is the error term. This is included to show the difference between the obtained value of y and the original value, i.e. ý.

In a similar form, there exists a cost function and this cost function can be represented as

$$\textbf{Cost } (\boldsymbol{\beta}_0, \boldsymbol{\beta}_1) = \Sigma_{i=1}^{n} (yi - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}))^2$$

Here:

1. n is the number of data points.
2. yi is actual observed value of the dependent variable.
3. $x_{i1}, x_{i2}, \ldots, x_{ip}$ are independent variable values.

4.2 Logistic Regression

This statistical method models the relationship between a dependent variable and independent variable (which can be one or more in nature). This type of regression is used to predict the probability of a categorical outcome such that it belongs to one of the two classes considered (e.g., Dog or Cat. 0 or 1).

The Logistic function, which is also called as the sigmoid function will transform a linear combination of the independent variables into a value between 0 and 1.

The function is defined in the following way:

$$\mathbf{f(z)=} \frac{\mathbf{1}}{\mathbf{1+}e^{-z}}$$

Where:

1. f(z) is output probability.
2. e is the base of the natural logarithm (14pprox.. 2.71828).
3. z is the linear combination of independent/predictor variables.

The linear combination z is calculated as

$$\mathbf{z=\beta_0+ \beta_1 x_1+ \beta_2 x_2+\ldots+ \beta_p x_p}$$

Where:

1. $\beta_0, \beta_1, \beta_{2\ldots}, \beta_p$ are the coefficients associated with predictor variables.
2. $x_1, x_2, \ldots, x_p$ are values of the predictor variables.

The outputs of this function, represents the likelihood of the binary outcome being 1.

This process also involves the estimation of coefficients $\beta_0, \beta_1, \beta_{2\ldots}, \beta_p$ to maximize the likelihood of the observed outcome given predictor variables.

Once the model is trained, it can be used to make predictions by calculating the probability of the binary outcome being 1 for new input data. A decision threshold can be chosen (often 0.5) to classify the predicted probabilities into the two classes.

There are a few assumptions such as the independence of errors, linearity of log odds and absence of multicollinearity among predictor variables.

4.3 Support Vector Machines

This supervised Machine Learning algorithm is used for both classification and regression purposes. With regards to Classification, it aims to find a hyperplane in a high-dimensional feature space to perfectly separate the data points from different classes. It maximizes the margin of distance of separation between the different classes. Margin of distance is defined as

the distance of the hyperplane from the nearest data points of each class. The data points are also called Support Vectors.

Consider a training dataset with n samples where each independent variable xi is associated with a class label yi (can be either +1 or -1 for a binary classification problem), the SVM seeks to find a hyperplane with the equation:

$$\mathbf{w.x + b=0}$$

Where:

1. w is the vector perpendicular to the hyperplane.
2. x is input feature vector
3. b is bias term

Classification decision is taken as follows:

1. If w.x + b>0 it is classified in the class with label +1.
2. If w.x + b<0 it is classified in the class with label – 1.

The goal of SVM is to maximize the margin and also ensure the correct classification of the data points. This can be considered as an optimization problem and can be mathematically represented in the following way:

$$\textbf{Minimize } \left(\frac{||\boldsymbol{w}||^2}{\mathbf{2}}\right)$$

$$\textbf{subject to constraints: } y_i\,(\mathbf{w}.\mathbf{x}_i+\mathbf{b})\geq=\mathbf{1 \text{ for all } i = 1, 2, \ldots, n}$$

where y_i is the class label of the i^{th} data point. The constraint ensures the correct classification of the data points and maximizes their distance from the decision boundary. Suppose the data is not linearly separable, the SVM can be extended to map the input data into a higher-dimensional space to ensure that the data becomes linearly separable in nature and no explicit calculation is involved while doing this. This process is commonly called as the Kernel Trick.

4.4 Naïve Bayes Classification

It is a Classification algorithm which is probabilistic in nature and it uses the concept of Conditional Probability in its functioning. The algorithm is called “naïve” because it considers the assumption of independence among features, which usually is not true in most cases. Albeit this assumption, it is surprisingly effective in the classification process.

The particular formula is used in the implementation of this algorithm:

$$\mathbf{P(Ci|x)} = \frac{\boldsymbol{P(x|Ci).P(Ci)}}{\boldsymbol{P(x)}}$$

Where:

1. P(Ci) is the Prior Probability and this is the probability of a class occurring without considering any features. Here i is the Class Index.
2. P(x|Ci) is the Likelihood and this is the probability of observing a given feature x for a specific class Ci.
3. P(x) is the Evidence and this is the probability of observing the feature x across all classes. It is used for normalization.
4. P(Ci|x) is the Posterior Probability and this represents the Probability of observing a Class, Ci given a particular feature x.

The Naïve condition can be written as follows:

$$\mathbf{P(x|Ci) = P(x_1|Ci).\ P\ (x_2|Ci)\ \ldots.\ P\ (x_n|Ci)}$$

where x_1, x_2, …., x_n are the individual features of input sample x.

The Classification is done based on the Posterior Probability value and the given feature x is classified to that Class Number Ci for which P(Ci|x) has the highest value.

5. RESULT ANALYSIS

In this analysis the 80:20 ratio has been considered between training and testing data (Figure 10). We can infer from Virat Kohli's data that the best algorithms are Logistic Regresion and Support Vector Machine (SVM). The worst algorithm is Naïve Bayes Classification. Logistic Regression and SVM will give the best results for this data if the accuracy of the algorithm is considered. These algorithms are used for classification purposes. We can observe that for Jos Buttler's data the best algorithms are Support Vector Machines, Naïve Bayes Classifier and K Nearest Neighbour. The worst algorithm for this data is the Logistic Regression. Therefore, it is recommended to use the Support Vector Machines, Naïve Bayes Classifier and K Nearest Neighbour for classification purposes. We can observe that for David Warner's data the best algorithms are Logistic Regression, Support Vector Machine (SVM) and K Nearest Neighbour. The worst algorithm for this data is Naïve Bayes Classifier. Therefore, it is recommended to use the Logistic Regression, Support Vector Machine (SVM) and K Nearest Neighbour for classification purposes.

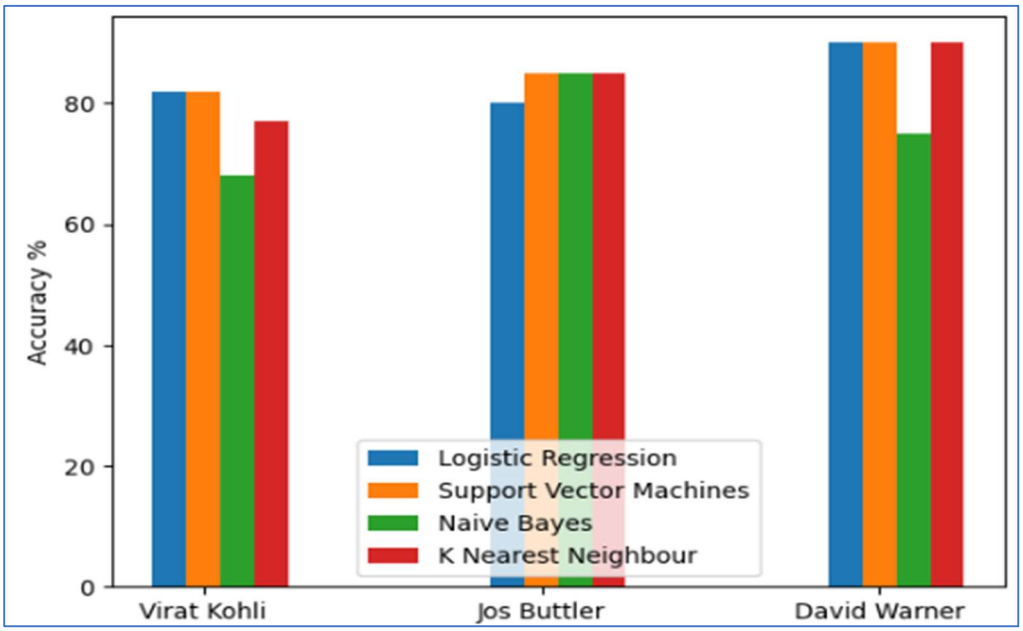


Fig 10. Performance of Predictive Models

Accuracy%=Percentage Accuracy

Virat Kohli, Jos Buttler, David Warner: Player names

Precision in machine learning refers to the measure of accuracy when identifying positive instances among those predicted as positive. It quantifies the proportion of true positive predictions relative to all positive predictions, providing insight into the model's ability to avoid false positives. Higher precision indicates a lower rate of false alarms and a more reliable positive prediction.

Precision= (True Positives+False Positives/True Positives)

Here:

True Positives (TP) are the number of correctly predicted positive instances.

False Positives (FP) are the number of instances that were incorrectly predicted as positive when they were actually negative.

Recall in machine learning measures the ability of a model to correctly identify all relevant instances among those that are actually positive. It is calculated as the ratio of True Positives to the sum of True Positives and False Negatives. Higher recall indicates a lower rate of missing relevant instances, making it crucial for tasks where identifying all positives is essential, like medical diagnoses.

Recall= (True Positives)/(True Positives+False Negatives)

ALGORITHM: Name of the Algorithm Implemented

VIRAT KOHLI, JOS BUTTLER, DAVID WARNER: Player names

PRECISION: Precision of the Algorithm

RECALL: Recall of the Algorithm

Table 2. Precision and Recall measure for the prediction of each player

	VIRAT KOHLI		JOS BUTTLER		DAVID WARNER	
ALGORITHM	PRECISION	RECALL	PRECISION	RECALL	PRECISION	RECALL
LOGISTIC REGRESSION	91	71	92	92	92	81
SVM	94	80	84	92	90	81
NAÏVE BAYES	93	70	94	92	88	81
K NEAREST NEIGHBOUR	94	80	83	100	90	100

Conclusion

The completion of data analysis marks a significant milestone in this research endeavor, revealing compelling insights into the efficacy of the novel idea under investigation. Through meticulous examination, accuracies, precisions, and recalls have been successfully determined, providing a comprehensive understanding of the model's performance. The obtained results not only validate the viability of the proposed concept but also highlight its potential for future advancements. The promising outcomes of this analysis underscore the importance of the innovative approach taken and pave the way for further exploration and development. The identified accuracies, precisions, and recalls serve as robust metrics, affirming the merit of this novel idea and positioning it as a valuable prospect for future research and applications.

STATEMENTS AND DECLARATIONS

Authorship:

The creation of graphs, tables and the overall code have been done by Atul.A.Das.

The Overall Structuring, Idea, and the Overall modules have been done by Brindha G R.

Competing Interests and Funding

Competing Interests: The authors declare that they have no competing interests.

REFERENCES

1.Saikia, Hemanta & Bhattacharjee, Dibyojyoti & Mukherjee, Diganta. (2019). Cricket Performance Management, Mathematical Formulation and Analytics. 10.1007/978-981-15-1354-1.

2.Kapadia, Kumash & Abdel-Jaber, Hussein & Thabtah, Fadi & Hadi, Wael. (2019). Sport Analytics for Cricket Game Results using Machine Learning: An Experimental Study. Applied Computing and Informatics. ahead-of-print. 10.1016/j.aci.2019.11.006.

3.Cricket Analytics and Predictor by Mr Suyash Mahajan, Ms Gunjan Kandhari, Ms Salma Shaikh, Ms Rutuja Pawar, Mr Jash Vora, Ms A. R Deshpande. (2019).Cricket Analytics and Predictor

4.Brydges, C. R. (2021). Analytics of batting first Indian Premier League twenty20 cricket matches. SportRxiv. https://doi.org/10.31236/osf.io/jq564

5. Vishal C V , Sathvik K B , Nischay N , Manoj Athreya H , Sagar N(2021)- Harnessing the Predictive Power of Lower-division Statistics of Cricketers to Predict Their Rates of Success at the International Level https://doi.org/10.22214/ijraset.2021.39354

6. Mukherjee, Satyam. (2012). Quantifying individual performance in Cricket $-$ A network analysis of Batsmen and Bowlers. Physical A Statistical Mechanics and its Applications. 393. 10.1016/j.physa.2013.09.027.

7. Vipul Punjabi, Rohit Chaudhari, Devendra Pal, Kunal Nhavi, Nikhil Shimpi, Harshal Joshi- A SURVEY ON TEAM SELECTION IN GAME OF CRICKET USING MACHINE LEARNING

8. Clarke, S. R. (2007). Studying Variability in Statistics via Performance Measures in Sport. 56th session,Kolkata: Indian Statistical Institute.

9. D. Böhning, Multinomial logistic regression algorithm, Ann. Inst. Stat. Math. 44 (1) (1992) 197–200

10. M.A. Hall, Correlation-based feature selection for machine, learning, 1999

11. Gunn, S.R., 1998. Support vector machines for classification and regression. ISIS technical report,14(1), 5–16.
12. W. McKinney, Python for data analysis: Data wrangling with Pandas, NumPy, and IPython, O'Reilly Media, Inc., 2012.
13. Mehvish Khan and Riddhi Shah, ―Role of External Factors on Outcome of a One Day International Cricket (ODI) Match and Predictive Analysis‖ International Journal of Advanced Research in Computer and Communication Engineering, vol. Vol. 4, no. Issue 6, pp. 192–197, Jun. 2015.