IPL Player Performance Prediction using Machine Learning

by

Doggala Ajay Kumar Reddy

A research submitted in partial fulfillment of the requirements for the degree of Master of Engineering in Information Management

Examination Committee:	Prof. Sumanta Guha (Chairperson) Dr. Chutiporn Anutariya (Co - chairperson) Assoc. Prof. Erik L.J. Bohez Mr. Olivier Nicole

Nationality:	Indian
Previous Degree:	Bachelor of Technology in Computer Science Jawaharlal Nehru Technological University Hyderabad, Telangana, India

Scholarship Donor:	AIT Fellowship

Asian Institute of Technology
School of Engineering and Technology
Thailand
May 2019

Acknowledgments

I, Ajay Kumar Reddy, would like you thank my advisor, Dr. Sumanta Guha for giving me guidelines and suggestions on my interesting topic Cricket. His support helped me a lot to learn more on the data mining techniques and get more focus on the data mining procedures.

Secondly, I would like to thank my committee members for advising me and suggest some valuable advices for the research study.

Last, I would like to thank my friends and parents for the continuous support. I would like to thank AIT which helped me to improve my knowledge in practical platform.

ABSTRACT

Selection Procedure for IPL has been very different procedure as the players are selected based on the previous performance in their matches in international cricket. Selecting a perfect team of eleven players with batsmen, bowler and wicket keeper is always a difficult task for franchises in IPL. The better the performance of the player, the more money he is sold for franchise. The research attempts to predict the runs and wickets taken in IPL. We classified both the results into different ranges. We used Naive Bayes, Decision Tree Classifiers, Random Forest and SVM to classify the prediction models.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

S. No	TERM	ABBREVATIONS
1	LBW	Leg before wicket
2	IPL	Indian Premier league
3	BCCI	Board of Control for Cricket in India
4	CRISP-DM	Cross Industry Standard Process for Data Mining
5	AI	Artificial intelligence
6	SVM	Support vector Machines
7	BPN	Back Propagation Network
8	RBF	Radial Basis function Network
9	SNA	Social Network Analysis
10	WEKA	Waikato Environment for Knowledge Analysis
11	SR	Strike Rate
12	HS	Highest Score
13	AHP	Analytic Hierarchy Process
14	SMOTE	Supervised Minority Oversampling Technique
15	ID3	Iterative Dichotomiser 3

CHAPTER 1
INTRODUCTION

1.1 Background

Cricket is a sport played in a ground by two teams. The batsman bats and a bowler bowls at a time in a session and is called as innings. The cricket team consists of eleven players each. The cricket is played where in center of the ground there is a 22 yard pitch.

The cricket ground consists of a boundary line where if the balls crosses the line, it is declared as a boundary or six. The players in field bats and bowls according to the toss. Partially, the cricket pitch consists of six wickets with bails on top of it, one at each ends of the pitch.

Figure.1.1 Cricket Pitch

Each team has a batsman, bowler and a Wicket keeper.

- For Runs, the batsman can score by running between the wickets as well as hitting the ball out of the boundary lines.
- For Wickets, the bowler can take by dismissing the batsmen in many ways.
- Every batting team consist of 10 wickets in an innings.
- The team needs to win a match by scoring runs in an innings or by defending the score by taking wickets.
- Any type of wickets except runout can be taken by bowler who bowled the ball.

Limited Overs cricket is a challenging task for all the players. Now a days, the mostly loved format of the games in the limited over cricket and 20 overs are bowled by a bowling team in an innings. In this we can predict how many wickets a bowler can take in match as well as the runs scored by a batsman.

1.1.1 Indian Premier League

The IPL is a limited Overs Twenty20 international cricket played in India. The IPL starts in month of April and ends in month of May every year. The IPL first season was started in 2008, by Lalit Modi who is also known as founder of IPL. The Board of Control for Cricket in India (BCCI) has started the season and till now IPL has contributed 11.5 billion dollars till now. Vivo mobile phones have been main sponsor for most seasons of IPL. The IPL has been the most successful franchise for many teams. Currently, the IPL title has mostly been successful for Chennai Super Kings as they won the tournament three times.

The IPL league starts with bidding of the players in the month of February, March where the players are auctioned based on their base prices. Every player have they base price, sometime the player who are unsold cannot participate in the IPL season. There are some chances of major injury in the middle of season and player can't play. The Total amount for base prices for franchises was around 400 million dollars and at the end of auction, the city names were announced. The cities where the IPL is conducted is Bangalore, Chennai and many other places in India. Most of the final and semifinals are conducted in Mumbai as the birth place of many legend players and the center city of the India.

In 2010, which is the fourth season of the IPL two new franchises where added, which are Pune Warrior and Kochi Tuskers Kerala. Where in 2012, the BCCI announced Deccan Charges will be terminated and Hyderabad Franchise will be further named as Sunrisers Hyderabad.

1.2 Statement of the Problem

The player performance has major role in the team selecting procedure. Especially in the IPL league, the performance of player is considered rather than a performance of a team. The players are auctioned every season where each franchise has a huge amount of money for bidding a player and own them. The unsold players are not allowed to play in the tournament. The players are selected based on various characteristics as players batting average, players batting Strike rate, players bowling economy rate, players bowling Strike rate, etc.

1.3 Objectives

A model for quantifying the performance of the batsman.
A model for quantifying the performance of the bowler.
Introducing new attributes based on the basic batting and bowling attributes as to increase performance.
Using Machine learning algorithms, to classify and predict the performance of the players.

1.4 Scope and Limitations

- The data used for modelling and training is from years 2008-2018.
- The models are built to predict the runs and wickets respectively and we also compare the accuracy by using different machine learning algorithms.

1.5 Research Outline

Chapter 2 Consists the data related to data mining procedures and various techniques which can be used to predict the performance of a player.
Chapter 3 includes the methodology process, to train and test the data based on the algorithms.
Chapter 4 Includes the evaluation process to select the best model.
Chapter 5 Conclusion and Future Work

CHAPTER 2
LITERATURE REVIEW

2.1 HISTORY OF CRICKET

The cricket has been started from 18th century as a national country sport for England. The first international match of test cricket started from 1844 and was further recognized in 1877. In South-east England, the origin of the game as in 16th Century as many believed game of cricket is for children at that time, as it has more fun in playing. In 1611, two men where prosecuted for playing cricket in the woods instead of going to church. In 17th Century no one had interest to play cricket as in England. In 18th Century, British members and the west indies were first introduced, unlike in summer they all used to play baseball. The laws in cricket were first introduced in 1744, as there are many recodifications and has been developed regularly. The laws mostly were like the height of the wickets, the weight of the bat, width between the wickets and bat width, especially the length and width of the pitch and batting, bowling crease. In 1760, the balling unit recognized the power of length ball, Yorker and bouncer. In 1772, Scoreboard has been introduced. In 18th century there was a smaller number of players playing and there was a lack of interest in them and no equipment for playing the game, investment was less. In 1817, A player named William Lambert was banned from playing cricket because of illegal match fixing with other team players. The first and foremost international game started in 1844 between United States and Canada. After the match, the Australians had a tour to England and had a picture clicked at Niagara Falls, the tour lead to victory for the Australians. In 1889, the South Africa was also introduced as cricket playing nation. There were many national and international championships between ashes, West Indies and South Africa in and around 1890s. Since then the game cricket kept growing and emerged as a number one sport entertainment till now.

Cricket as the starting was played four balls per over, later in 1920s eight balls per over has been played. After the World War in 1940 new laws of cricket were introduced and allowed only six balls per Over. The cricket started being in different versions in 1960s, as test series and ODI series. In 1970, the limited Overs cricket came into play and live coverage of the matches were also starting to develop. The first international World Cup was played in 1844, but the tournaments were started from 1975 between 20 various teams. Around 11 World Cups have been played in cricket till now.

2.2 DATA MINING

Data Mining does extract huge Volume of data and patterns which helps to gain knowledge from the existing data. Data mining does not only extract data but also helps to form an accessible structure which helps for further needs and use. Data mining procedure involves discovering of knowledge from databases (KDD).

The various stages for this process are mainly Data selection, Data preprocessing, Data Transformation, Data Mining, and Data evaluation. Mostly, Cross Industry Standard Process for Data Mining (CRISP-DM) is used, which is shown in figure.

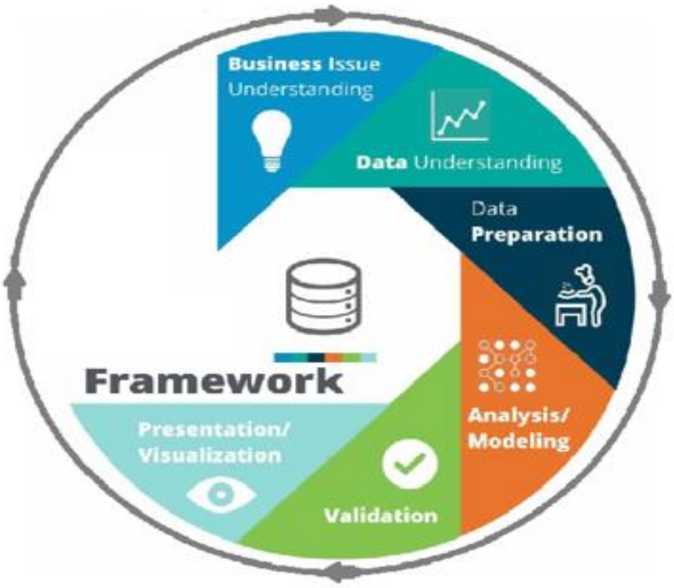

Figure.2.1 CRISP-DM PROCESS

The phases of the framework are iterative and flexible, it is required for subsequently solving business problems and focus more based on the data requirement as there are many changes during the process. The process also continues after a required solution has been deployed. The CRISP-DM is now mainly used as an industry tool as it is been a success.

1. The First phase, Business Understanding phase involves understanding of requirements from the user clearly, to perform an action of plan to gain knowledge and achieve goals.

2. Second Phase, Data Understanding involves the collection of all the raw data available and maintain quality of the knowledge gained by data analysis.

3. Third Phase, Data Preparation includes data cleaning, data preparation for model building, selection of attributes and many more tasks. In this phase, the dataset is ready and can be used for modelling.

4. Fourth Phase, Various data mining techniques can be used based on the required dataset. The data mining techniques can also be optimized and will be ready for evaluation.

5. Fifth Phase, Various data evaluation techniques are used to get an optimal model before deployment.

6. Last Phase, the end users have a good knowledge from the model.

2.3 DATA MINING TECHNIQUES

Many techniques are used for knowledge discovery through databases, some of them are: Classification, Regression, Clustering, Neural Networks, and Association rule.

2.3.1 Classification

The Classification technique is most commonly used for huge volume of data, which allows to apply the classified instances to a model. Data learning is the first and foremost step in this technique, which allows training dataset to be well analyzed by an algorithm. Whereas, in test data we will approximate the accuracy of the rules defined. If all the rules are accepted, they can be applied to data. On a record by record basis, both fraud detection and valid data, activities are determined by fraud detection technique. The set of parameters are then defined by classifier training algorithm.

Types of Models mainly used are:

- Bayesian Classification
- Neural Networks
- Support Vector Machines (SVM)

2.3.2 Regression

In Data Mining, there are two types of data mining variables. First, independent variables these are used as attributes. Second, Response variables which we want to be predicted. The model which used both variables and builds a relationship between them can be known as regression analysis. Most of the cases, the prediction is not a simple procedure where we have to many more complex techniques such as neural network, linear regression and many more. Neural networks help to create and solve complex procedures.

Types of regression methods used:

- Linear and Non Linear Regression
- Multivariate Linear and Non Linear Regression

2.3.3 Association Rule

This is mainly used to find frequent item sets in the dataset. Multilevel correlation and quantitative correlation can be used to help user to make useful decisions. Apriori algorithm can be used for obtaining the rules.

2.4 DATA MINING ALGORITHMS

The Data mining algorithms [15] and [16] takes data as input and produces patterns as an output. The training dataset is used to predict the pattern and is analyzed to identify rules. Some of the mainly used algorithms are:

2.4.1 Naive Bayes

It is the most supervised learning techniques used. The algorithms work very effectively for real world applications. The data can be in form of Category and continuous data. The technique is useful for independent variables, all the inputs variables and features are considered as independent. Very few training data sets are required for estimation of parameters so as they can be classified easily.

2.4.2 Decision Tree

The decision tree is built on their nodes and leaf. This algorithm is commonly used for prediction. The algorithm is used to predict the output based on the various input variables. Leaf node represents the predicted value. The technique is furthermore classified into algorithms such as REP tree, C4.5, ID3 and J45. All this algorithm has special characteristic as they select the best attribute for splitting the attribute table, creates multiple iterations and make proper decisions.

2.4.3 SVM

SVM performs both linear and nonlinear classification approach which is required. The technique is used mostly for unlabeled data. This is one of the mostly used industrial technique. The decision making helps data points to be classified based on the classes defined. There is a clear gap between each class known as hyperplane.

2.5 RELATED WORK

There are many articles and research papers related to prediction based on player performance in various sport games. Till date only few researchers had tried to predict the player performance rather than predicting team-based performance.

1. (Shanthi Muthuswamy and Sarah S. Lam) Indian Cricket team plays with eight to eleven international teams in various countries. The paper [2] discusses about prediction of bowler performance, to select the best performance team for wining cricket matches. The bowler main role is to take wickets with less economy, bowler need to take wickets in right time which will help his career for the selection for many upcoming tournaments. The paper discusses about a neural network approach which works effectively so that bowler concede less runs in a match. The two networks used in this paper are Backpropagation network (BPN) and radial basis function network (RBF).

2. (INDIKA PRADEEP WICKRAMASINGHE) The study [3] and [8], discusses about a method for prediction of performance of cricket test match based on previous five years of data. The rank of the team and the other team players performance also effect the performance of a batsman who is in form and out of form. For prediction a three-stage hierarchical linear model has been proposed for accuracy.

3. (G. D. I. Barr and B. S. Kantor) This paper [4] represents a graphical approach which happens as a criteria selection for batsman which also involves batsman strike rate and average runs scored.

4. (Iyer, S. R. and Sharda, R.) A neural network-based approach for predicting the performance of cricketer, the model [5] classifies the International cricketers into three categories – performer, moderate and failure. The model has been trained and tested based on the data collected from 1985 – 1007. The results provide a useful and supportive decision-making so as the cricket selection process can be carried out easily.

5. (Dibyojyoti Bhattacharjee, Darshan G Pahinkar) This paper [6] discusses about a limited over cricket where the bowlers have a difficult task. Various techniques were used to performance of bowler’s performance in relevant literature but in this paper Multiple linear Regression technique was used to identify the reasons that are responsible for performance of the bowlers. The bowler performance depends upon the runs he concede to a batsman this is related with economy of a bowler. It was very clearly observed that the experience of the bowler and economy is most important predictors which purely depends upon the performance of bowler in IPL (Limited Over Cricket).

6. (S. Mukherjee) This paper [7] discusses about the balanced team selection procedure thus by analysing the player performance in international cricket. The performance can be calculated by number of runs scored by batsmen as well as number of wickets taken by a bowler. The paper investigates the Social network analysis (SNA) through which a directed and weighted network has been generated.

7. (Parker, D., Burns, P. and Natarajan) This paper [9] describes a model for valuation for IPL. The Factors considered for selection of player is based on his experience, strike rate, economy, form etc. Like this [12] a model for integer optimization based on the performance measures.

8. (C. Deep Prakash, C. Patvardhan, C. Vasantha Lakshmi) The Mayo Prediction model has three major components and the models [10] are created using data analytics and machine learning. The training set can be increased for more improvement and can lead to increase in accuracy.

9. (Omkar, S.N. and Verma, R.) For selection of team fitness of player is calculated based on the performance. The Genetic algorithm [13] is used for evaluation and in this model each team is considered as string and players as string bit.

10. (Lewis, A. J.) Discusses a model [11] which includes a statistical analysis of player performance in cricket game, using the Duckworth-Lewis method.

CHAPTER 3
METHODLOGY

The methodology follows the industrial structure, the Model evaluation is done in the next chapter.

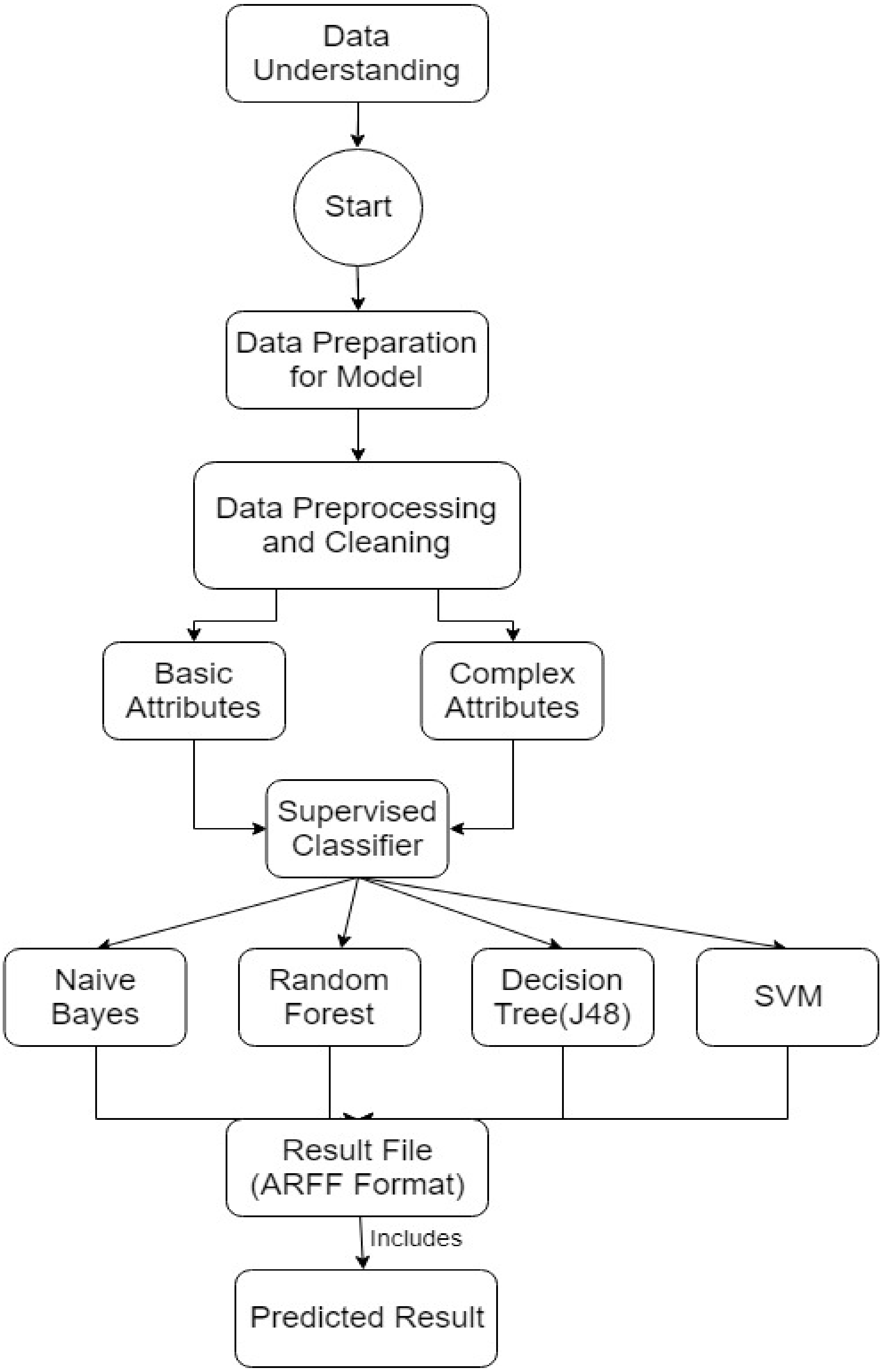

Figure.3.1 Methodology Flow Chart

3.1 Data Understanding

The data has been collected from www.ESPNCricinfo.com and www.HowStats.com [1]. The IPL is a professional limited over cricket tournament which is contested in April month of every year, the IPL season has started from 2008. There have been eleven seasons till date. The data consists of all the performance of the players in the IPL team. In 2011, two new franchise were added to IPL tournament. There were many changes in the changes in the team names in 2011 season. The IPL is a limited 20 overs cricket played in India every year except the Second IPL season was played in South Africa in 2009.

The data has every ball by ball where 470 players data, matches played in all seasons till 2018. We have around 136,590 ball by ball data it includes 16 attributes like batsmen_id, bowler_id, innings_id, Over_id, ball_id, etc. Every player data from starting season of the IPL has been collected. The batsmen data we have batsmen name, batting hand, his runs scored in a match ball by ball, the fielder or bowler with whom he got bowled out or caught. The bowler data consists of bowler name, bowling hand, his wickets taken in match, his economy rate based on the runs conceded in a match by bowler.

Every data is not available online like the Strike rate of the batsmen, the economy of the bowler, average runs scored by batsmen has been calculated by using simple formulas.

The performance of the player can be analyzed and predicted by using machine learning tools. For analysis of this data we used WEKA tool to obtain the results in the research study. These tools are very useful as they provide data mining features and provide preprocessing functionalities.

3.2 Data Preparation

Player data can be predicted by using several attributes. The main attributes we used for predicting and measuring the performance of batsmen are:

3.2.1 Batting Attributes

3.2.1.1 Innings: This attribute explains the number of innings played by a batsman till date. This depends on a batsman he played for a team in a season of the IPL. The attribute can be considered because it analyzes the experience of the batsmen based on the number of innings played.

3.2.1.2 Batting Average: The attribute measures the runs scored by a batsman in an innings. This can be calculated by using the formula:

$$Average = \frac{Runs\ Scored}{Number\ of\ Innings\ Played}$$

Calculation Formula for Batting Average

3.2.1.3 Batting Strike Rate (SR): IPL is a limited over cricket as a batsman need to score runs at a faster pace. The batsman can be out of form if the Strike rate is less depending on the matches he played with opponent team. The Strike rate can be calculated by using the formula:

$$Strike\ Rate = \frac{Runs\ Scored}{Number\ of\ Balls\ Faced} \times 100$$

Calculation Formula for Strike Rate of player

3.2.1.4 Centuries: If a batsman scored more than or equal to 100 runs in a venue, this helps the team performance and increases the capability of the player to score more runs in a venue.

3.2.1.5 Fifties: If a batsman scored more than or equal to 50 runs in a venue from a team, the player will get awarded if he wins a match because of his performance in match. The fifties are only counted if a player scores less than 100 runs and more than 50 runs in a match.

3.2.1.6 Zeroes: If a player gets out without scoring any kind of runs in a match, it is often called as a Duck Out.

3.2.2 Bowling Attributes:

3.2.2.1 Innings: At least one ball should be bowled by a bowler in a match, those matches will only be counted. The experience of the player is considered based on the number of innings he played for a team in a season.

3.2.2.2 Overs: The number of bowls bowled by a bowler in a match, this attribute also considered the experience of the bowler.

3.2.2.3 Bowling Strike Rate: The attribute depends by the bowler who is in form, and capability of the bowler to bowl out any kind of batsman in a match at any time. The Strike rate can be obtained as:

$$\text{Strike Rate} = \frac{\text{Number of Balls Bowled}}{\text{Number of Wickets Taken}}$$

Calculation Formula for Strike Rate of Bowler

3.2.2.4 Bowling Average: The attribute depends on the bowler if the bowler can bowl him out anytime and can concede less runs to a batsman in an over. The attribute also considers the experience and capability of the bowler based on the wickets taken. The bowling the more the bowling Average the lesser the capability of bowler. The bowling Average can be calculated by using the formula:

$$\text{Bowling Average} = \frac{\text{Number of Runs Conceded}}{\text{Number of Wickets Taken}}$$

Calculation Formula for Bowling Average of player

3.2.2.5 Economy Rate: The attribute depends upon the runs he conceded in overs bowled by a bowler in a match. The more the economy rate the bowler is not capable for getting a batsman out in fewer deliveries. The Economy rate is calculated by using the following formula:

$$\text{Economy Rate} = \frac{\text{Runs Conceded}}{\text{Overs Bowled}}$$

Calculation formula for economy rate of bowler

3.2.2.6 Four/More wicket Haul: The bowler capability to take more than three wickets in an innings, the performance measure can be considered as in the venue a bowler has performed well. The attribute considers the capability of the bowler.

3.3 Data Preprocessing

3.3.1 Calculating the weights

These above batting attributes and bowling attributes are mainly used for evaluating the performance of the player. Whereas, some features like the bowling player economy rate, the bowling player average, batting player average, and batting player strike rate can be most important in limited over cricket. We used weighted criteria to weight attributes upon other attributes to measure the performance accordingly. We will use Analytic Hierarchy Process (AHP) [14] to measure and determine the weights of the attributes. AHP is a complex decision making tool used for making decision and prioritize them accordingly. The AHP is mainly used to get the best results from the decision making process. For evaluation Criteria, the AHP will generate weights for each pairwise comparison according to the decision maker choice.

According to the decision maker pairwise comparison performance can be evaluated based on the score assigned for each criterion. Higher the score of criteria the better performance can be achieved. All the weights are then combined to obtain the final score for each criterion.

We are calculating these weights for the batting and bowling attributes to obtain new attributes which depends on player performance in matches epically in limited overs cricket.

Based on the experience and knowledge of cricket [17], we arrange the attributes in increasing order of importance as based on Fig.4:

Zeroes > Fifties > Centuries > Strike rate > Innings > Average

The Fundamental Scale for Pairwise Comparisons		
Intensity of Importance	**Definition**	**Explanation**
1	Equal importance	Two elements contribute equally to the objective
3	Moderate importance	Experience and judgment slightly favor one element over another
5	Strong importance	Experience and judgment strongly favor one element over another
7	Very strong importance	One element is favored very strongly over another; its dominance is demonstrated in practice
9	Extreme importance	The evidence favoring one element over another is of the highest possible order of affirmation
Intensities of 2, 4, 6, and 8 can be used to express intermediate values. Intensities 1.1, 1.2, 1.3, etc. can be used for elements that are very close in importance.		

Figure.3.2 The fundamental Scale of Pairwise comparisons

Table.3.1 Batting Pairwise Comparison

	Average	Innings	Strike Rate	Centuries	Fifties	Zeroes
Average	1	3	4	5	6	7
Innings	0.33	1	3	4	5	6
Strike Rate	0.25	0.333333333	1	3	4	5
Centuries	0.2	0.25	0.333333333	1	2	3
Fifties	0.166666667	0.2	0.25	0.5	1	3
Zeroes	0.142857143	0.166666667	0.2	0.333333333	0.333	1
SUM	**2.09285714**	**4.95**	**8.783333333**	**13.83333333**	**18.33**	**25**

For bowling, we arrange the attributes in order of importance as:

Innings > Overs > Strike Rate > Average > Economy Rate

Table.3.2 Bowling Pairwise Comparison

	Innings	Overs	Strike Rate	Average	Economy Rate
Innings	1	3	4	5	6
Overs	0.33	1	3	4	5
Strike Rate	0.25	0.333333333	1	3	4
Average	0.2	0.25	0.333333333	1	2
Economy Rate	0.166666667	0.2	0.25	0.5	1
SUM	**1.95**	**4.783333333**	**8.583333333**	**13.5**	**18**

For calculation of weights,

For a matrix of pairwise comparison,

$$\begin{vmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{vmatrix}$$

1) First, we need to sum the values in each column of the pairwise matrix.

$$P_{ij} = \sum_{i=1}^{n} P_{ij}$$

2) Divide each element in the matrix by its column total to get a normalized pairwise matrix.

$$X_{ij} = \frac{P_{ij}}{\sum_{i=1}^{n} P_{ij}} \begin{vmatrix} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \\ X_{31} & X_{32} & X_{33} \end{vmatrix}$$

3) Divide the sum of the normalized column of matrix by the number of criteria used to generate Weighted Matrix.

$$W_{ij} = \frac{\sum_{j=1}^{n} X_{ij}}{n} \qquad \begin{vmatrix} W_{11} \\ W_{12} \\ W_{13} \end{vmatrix}$$

We get the final weights of attributes as:
For Batting,
Average = 0.418
Innings = 0.250
Strike Rate = 0.155
Centuries = 0.080
Fifties = 0.059
Zeroes = 0.034

For Bowling,
Innings = 0.461
Overs = 0.260
Strike Rate = 0.151
Average = 0.075
Economy Rate = 0.049

The Three new attributes can be defined as the new weighted attributes [17].

These attributes are defined because the basic attributes cannot define the primary factors directly the new measures will quantify the player more accurately. The following three measures are Consistency, Venue and Opposition.

We calculate all the measures in MS-Excel file itself.

1) **Consistency:**

The player consistency depends on the batting experience in the cricket field. All the career is dependent on the consistency, which will help to evaluate whether a player scores runs in a match on average, bowler take wickets in a match, and many more measures.

For Batting Attributes,

Consistency = 0.418*Average + 0.250*Innings + 0.155* Strike Rate + 0.08* Centuries + 0.059* Fifties – 0.034* Zeroes

For Bowling Attributes,

Consistency = 0.461*Innings + 0.26*Overs + 0.151*Strike Rate + 0.075*Average + 0.049*Economy Rate

2) **Venue:**

This attribute also plays an important in performance evaluation of player.

For Batting Attributes,

Venue = 0.418*Average + 0.250*Innings + 0.155* Strike Rate + 0.08* Centuries + 0.059* Fifties – 0.034* Zeroes

For Bowling Attributes,

Venue = 0.461*Innings + 0.26*Overs + 0.151*Strike Rate + 0.075*Average + 0.049*Economy Rate

3) **Opposition:**

This attribute defines whether the player is capable of scoring runs against a team in a match. The more runs he scores the better the performance of the player will be.

For Batting Attributes,

Opposition = 0.418*Average + 0.250*Innings + 0.155* Strike Rate + 0.08* Centuries + 0.059* Fifties – 0.034* Zeroes

For Bowling Attributes,

Opposition = 0.461*Innings + 0.26*Overs + 0.151*Strike Rate + 0.075*Average + 0.049*Economy Rate

Rating the attributes:

The values in the tables of batting and bowling differ and have wide ranges, for instance the number of innings played by batsmen in IPL season will be 165 matches by a player, 150 matches by a player, so as we rated the attributes between 1 and 5 where minimum is 1 and maximum is 5. Based on the knowledge and the values available we rated the attributes. For instance, the batsmen batting average can be more than 70 and the number of overs bowled in seasons can be over 300, we used 5 if we have batting average more than 70 and we used 4 rating if the overs bowled are around 201 – 300. The rating we used are as follows:
For traditional Attributes (Basic attributes), runs and wickets respectively as:

Table.3.3 Rating basic attributes for Batting

No .of Innings	50 s	100 s	0s	Strike Rate	Average	Runs
1- 9 : 1 10 - 24 : 2 25- 49 : 3 50 - 69 : 4 >70 : 5	0 - 1 : 3 2 - 7 : 4 >8 : 5	0 - 1 : 3 2 : 4 > 2 : 5	0 - 2 : 3 3 - 5 : 4 >5 : 5	0 - 20 : 1 21 - 49 : 2 50 - 74 : 3 75 - 110 : 4 > 110 : 5	0 - 9.99 : 3 10 - 24.99 : 4 >25 : 5	0 - 500 : 2 501 - 1000 : 3 1001 - 1500 : 4 >1500 : 5

Table.3.4 Rating basic attributes for Bowling

No .of Innings	Overs	Average	Strike Rate	Economy Rate	4W	Wickets
>70 : 5 50 - 69 : 4 25 - 49 : 3 10 - 24 : 2 1- 9 : 1	>220 : 5 150 - 219.6 : 4 75 - 149.6 : 3 25 - 74.6 : 2 0 - 24.6 : 1	0 - 9.99 : 2 10 - 19.99 : 3 20 - 29. 99 : 4 >30 : 5	>25 : 2 15 - 24.99 : 3 10 - 14.99 : 4 0 - 9.99 : 5	1 - 2.2 : 5 2.3 - 5 : 4 5.01 - 8 : 3 >8 : 2	1 : 4 >1 : 5	0 - 20 : 2 21 : 40 : 3 41 : 60 : 4 >60 : 5

For Complex derived Attributes as:
Table.3.5 Rating complex attributes for No. of Innings

No. of Innings				
For Consistency	For Opposition	For Venue	For Batting Average	For Batting Strike Rate
1 - 15 : 1 16 - 40 : 2 41 - 60 : 3 61 - 71 : 4 >=71 : 5	1 - 13 :2 14 - 26 : 3 27 - 65 : 4 > 65 : 5	1 - 30 : 2 31 - 60 : 3 61 - 75 : 4 >75 : 5	0 - 9.99 : 3 10 - 24.99 : 4 >25 : 5	0.00 - 20.99 : 1 21.00 - 49.99 : 2 50.00 - 74.99 : 3 75.00 - 110.99 : 4 >111.00 : 5

Table.3.6 Rating complex attributes for Centuries

Centuries		
For Consistency	For Opposition	For Venue
0 - 2 : 4 >2 : 5	0 - 1 : 4 >1 : 5	0 - 1 : 4 >1 : 5

Table.3.7 Rating complex attributes for Fifties

Fifties		
For Consistency	For Opposition	For Venue
0 - 6 : 3 7 - 14 : 4 >15 : 5	0 - 5 : 4 >5 : 5	0 - 5 : 3 6 - 10 : 4 >10 : 5

Table.3.8 Rating complex attributes for Zeroes

Zeroes	
For Consistency	For Opposition & Venue
0 - 5 : 5 5 - 10 : 4 11 - 15 : 3 16 - 25 : 2 >25 : 1	0 - 5 : 5 >5 : 4

Table.3.9 Rating complex attributes for Overs

Overs						
For Consistency	For Opposition	For Venue	For Bowling Average	For Bowling Strike Rate	4/+ WH	For Bowling Economy Rate
0 - 24 : 1 25 - 74 : 2 75 - 149 : 3 150 - 219 : 4 >=220 : 5	0 - 20 :2 21 - 50 : 3 50 - 100 : 4 > 100 : 5	0 - 75 : 3 76 - 150 : 4 >150 : 5	>35 : 1 20 - 34.99 : 2 15 - 19.99 : 3 12 - 14.99 : 4 <11.99 : 5	10 - 14.99 : 3 15 - 19.99 : 4 >20 : 5	0 - 1 : 3 2 - 3 : 4 >3 : 5	0 - 4.99: 1 5 - 5.99 : 2 6 - 7.99 : 3 8 - 8.99 : 4 >8.99 : 5

3.4 Data Cleaning

We considered most of the missing values as zeroes and for attributes in Opposition and Venue like when a player plays a match and score zero runs, and when a player plays a match and bowl only one ball, these types of instances have many zeroes and are considered as missing values.

3.5 OUTPUTS

The Runs and wickets are classified into classes.

For Runs, 1-24: 1	For Wickets, 0-1: 2
25-50 :2	2: 3
51-75: 3	3: 4
75-99: 4	>3: 5
>=100: 5	

3.6 MODEL BUILDING AND TRAINING

3.6.1 Supervised Learning Algorithms

We will use three of the main learning algorithms through which a function is generated based on the training data set. We will use Naive Bayes, Decision Tree, SVM and Random Forest to generate the prediction models.

3.6.1.1 Naive Bayes Classifier

The classifier used in depends most on the naïve Bayes theorem. The classifier can be used to predict the difference features between the certain objects. The model provides good results based on the occurrence of events.

Bayes Theorem:

The features are independent to each other so as we can find the probability of the event y happening and X has occurred. Here, X and y are evidence and hypothesis respectively. The main important point is the features don’t affect each other they are independent.

We need to find the maximum probability of y as,

$$P(y|X) = \frac{P(X|y)P(y)}{P(X)}$$

Here, the X can be given as: X = (x_1, x_2, x_3, …., x_n);

By expanding the value of X, we get

$$P(y|x_1, ..., x_n) = \frac{P(x_1|y)P(x_2|y)...P(x_n|y)P(y)}{P(x_1)P(x_2)...P(x_n)}$$

This equation leads to a proportionality as the P(X) never changes as a denominator,

$$P(y|x_1, ..., x_n) \propto P(y) \prod_{i=1}^{n} P(x_i|y)$$

Hence, we need to the find the Maximum probability for y, we get the predictors using the class values.

$$y = argmax_y P(y) \prod_{i=1}^{n} P(x_i|y)$$

3.6.1.2 Decision Tree Induction

Decision tree induction is the process of creating decision trees for class-labeled training tuples. A decision tree is basically a tree structure like a flowchart. Each internal node of the tree represents a test on an attribute and each branch is the outcome of the test. Each leaf node is a class label. The first node at the top of the tree is the root node. Internal nodes of the tree are denoted by rectangles and leaf nodes are represented by ovals.

We are using j48 decision tree which is been developed by WEKA project team. This is usually an extension of ID3 algorithm.

3.6.1.3 Support Vector Machine

SVMs are highly accurate and less prone to overfitting. SVMs can be used for both numeric prediction and classification. SVM transforms the original data into a higher dimension using a nonlinear mapping. It then searches for a linear optimal hyperplane in this new dimension separating the tuples of one class from another. With an appropriate mapping to a sufficiently high dimension, tuples from two classes can always be separated by a hyperplane. The algorithm finds this hyperplane using support vectors and margins defined by the support vectors. The support vectors found by the algorithm provide a compact description of the learned prediction model. SVM takes different approaches to classify linearly separable and linearly non-separable data.

3.6.1.4 Random Forest

The algorithm is mainly used for best classification and regression problems. The Random forest is just combination of the decision trees as it adds randomness to model and the most important feature of the algorithm is picking the best subset when splitting a node based on the features selected so that the results generate a better classification model.

Random forest is very easy to measure the accuracy based on the importance of the features while prediction. The important feature of the algorithm is it reduces the impurities across the decision trees. Random forest also reduces the overfitting problem.

CHAPTER 4
Data Evaluation

We will use training and test sets to find the best combination that gives the most accuracy. We need to experiment by dividing the data set into five years of training data and five years of testing data.

We need to analyse and compare the performance of the algorithms in terms of several performance measures, which are described below in short:

Accuracy: The prediction accuracy of an algorithm is the ratio of the number of test instances correctly classified by the algorithm to the total number of test instances. The higher the accuracy, the better the performance.

$$Accuracy = \frac{Number\ of\ correctly\ classified\ instances}{Total\ number\ of\ instances}$$

Precision: Precision can be defined as the ratio of number of instances that were classified correctly to be in a class (True positives) to the total number of instances predicted in that class. Precision can be indicated as follows:

$$Precision = \frac{Number\ of\ instances\ of\ class\ x\ predicted\ correctly}{Total\ number\ of\ instances\ which\ were\ predicted\ to\ be\ in\ class\ x}$$

Recall: Recall can be defined as number of instances that are correctly predicted in a class to the total number of instances in the class. Recall can be indicated as follows:

$$Recall = \frac{Number\ of\ instances\ of\ class\ x\ predicted\ correctly}{Total\ number\ of\ instances\ in\ class\ x}$$

F1 Score: F1 score can be calculated as harmonic mean of both precision and recall in that class.

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Area under the curve: The curve depends upon the true positive rates and false positive rates. The accuracy can be calculated by the area under the curve. The range of the values is from 0.5 to 1 where, the values which are equal to 0.5 means the least accurate result is predicted and 1 means prefect accurate.

We simulated all the algorithms in WEKA machine learning software tool version 3.8 on Windows 10. The performance and accuracy are then compared as follows:

Basic Attributes:

Table.4.10 Classification values for runs using basic attributes

Basic (Runs)	Precision	Recall	F Score	ROC	Accuracy
Naïve Bayes	0.805	0.69	0.741	0.574	69%
j48	0.901	0.77	0.878	0.56	76.90%
Random Forest	0.814	0.834	0.824	0.576	83.38%
SVM	0.812	0.77	0.789	0.59	76.90%

Table.4.11 Classification values for wickets using basic attributes

Basic (Wickets)	Precision	Recall	F Score	ROC	Accuracy
Naïve Bayes	0.872	0.8	0.824	0.956	80%
j48	0.906	0.889	0.893	0.967	88.10%
Random Forest	0.702	0.818	0.78	0.906	81.70%
SVM	0.893	0.884	0.885	0.932	88.47%

From the above tables, we can say that the runs are predicted with highest accuracy 83.38% by using Random Forest method. While, the wickets are predicted with a highest accuracy 88.47% by using SVM.

Complex Attributes:

Table.4.12 Classification values for Runs using complex attributes

Complex (Runs)	Precision	Recall	F Score	ROC	Accuracy
Naïve Bayes	0.72	0.744	0.732	0.568	74.40%
j48	0.72	0.744	0.732	0.578	74.40%
Random Forest	0.749	0.799	0.763	0.512	79.80%
SVM	0.718	0.738	0.728	0.547	73.80%

Table.4.13 Classification values for wickets using complex attributes

Complex (Wickets)	Precision	Recall	F Score	ROC	Accuracy
Naïve Bayes	0.823	0.706	0.758	0.663	70.46%
j48	0.881	0.706	0.798	0.623	72%
Random Forest	0.806	0.726	0.763	0.603	72.63%
SVM	0.796	0.731	0.761	0.761	73.10%

From the above fig., we can say that the wickets are predicted with highest accuracy 79.8% by using Random Forest method. While, the wickets are predicted with a highest accuracy of 73.1% by using SVM.

By this we can that SVM and Random forest classifier works well for prediction based on the basic and complex attributes.

CHAPTER 5
CONCLUSION AND FUTURE WORK

5.1 Conclusion

In this paper, we used basic four models for prediction the performance of a batsman scoring runs and a bowler taking wickets. First, we used basic attributes to train and test the model and secondly, we used the complex attributes derived by using the basic attributes. The data is prepared for training and testing for model building and evaluation.

From this paper, we can conclude that the SVM and Random Forest are predicting with higher accuracy for both basic and complex attributes.

The model is used with five years of training and five years of testing dataset. The model predicts runs based on the attributes and prediction a margin. The approach of the whole research is to improve the performance of the machine learning and predict the result.

IPL plays a major role in international cricket for selection to various other tournaments and the World Cup which is held every four years. Every players dream is to win the world cup for their country, so this approach will be more helpful for the IPL season franchise which includes bidding of the players before the season starts and for future tournaments held.

5.2 Future works

- The weather data like the dry and wet pitch details, the sunny and daylight conditions can also be used to analyse the performance of the players.

- We can also take international cricket data, as this study can also be done for different formats of the game. For instance, the batsman should have capability to score more than and playing more balls in the test cricket whereas in limited overs his performance defers.

- The accuracy of the model can also be improved by making some more assumptions and data gathering.

REFERENCES

[ONLINE] www.espncricinfo.com

Muthusamy and S.S. Lam (2008), "Bowler Performance Prediction for One-day International Cricket Using Neural Networks," in Industrial Engineering Research Conference.

P. Wickremasinghe (2014), "Predicting the performance of batsmen in test cricket," Journal of Human Sport & Exercise, vol. 9, no. 4, pp. 744-751.

G. Barr and B. Kantor (2004), "A Criterion for Comparing and Selecting Batsmen in Limited Overs Cricket," Operational Research Society, vol. 55, no. 12, pp. 1266-1274.

S. R. Iyer and Sharda (2009), "Prediction of athletes performance using neural networks: An application in cricket team selection," Expert Systems with Applications, vol. 36, pp. 5510-5522.

M. G. Jhanwar and V. Pudi (2016), "Predicting the Outcome of ODI Cricket Matches: A Team Composition Based Approach," in European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases.

S. Mukherjee (2014), "Quantifying individual performance in Cricket - A network analysis of batsmen and bowlers," Physical A: Statistical Mechanics and its Applications, vol. 393, pp. 624-637.

P. Shah (2017), "New performance measure in Cricket," ISOR Journal of Sports and Physical Education, vol. 4, no. 3, pp. 28-30.

D. Parker, P. Burns and H. Natarajan (2008), "Player valuations in the Indian Premier League," Frontier Economics, vol. 116.

C. Prakash, C. Patvardhan and C. Lakshmi (2016), "Data Analytics based Deep Mayo Predictor for IPL-9," International Journal of Computer Applications, vol. 152, no. 6, pp. 6-10.

F. C. Duckworth and A. J. Lewis (1998), "A Fair Method for Resetting the Target in Interrupted One-Day Cricket Matches," The Journal of the Operational Research Society, vol. 49, no. 3, pp. 220-227.

G. Sharp, W. Bretteny, J. Gonsalves and M. Lourens (2011), "Integer optimisation for the selection of a Twenty20 cricket team," Journal of the Operational Research Society, pp. 1688-1694.

S. Omkar and R. Verma (2003), " Cricket team selection using genetic algorithm," in International congress on sports dynamics.

T. Saaty (1980), The Analytic Hierarchy Process, New York: McGrow Hill.

J. Han, M. Kamber and J. Pei (2012), Data Mining Concepts and Techniques, 3rd Edition ed., Waltham: Elsevier.

J. R. Quinlan (2015), C4.5: Programs for Machine Learning, Elsevier.

Kalpdrum Passi and Nivarkumar Pandey (2018), International journal of Data Mining and Knowledge Management Process Vol.8, No.2.

CHAPTER 6

APPENDIX – A Rating of Attributes

We rated the attributes based on the values in the batting and bowling instances. We used MySQL to rate the attributes between 1 to 5. We also used XAMPP for windows to start the MySQL server. Here, is the snippet of sample queries used in for rating:

QUERY PROCESSING.txt - Notepad

File Edit Format View Help

```
// No. of Innings
SELECT `No. of Innings`,
CASE WHEN `No. of Innings` BETWEEN 1 AND 9 THEN "1"
 WHEN `No. of Innings` BETWEEN 10 AND 24 THEN "2"
 WHEN `No. of Innings` BETWEEN 25 AND 49 THEN "3"
 WHEN `No. of Innings` BETWEEN 50 AND 69 THEN "4"
WHEN `No. of Innings` BETWEEN 70 AND 81 THEN "5"
ELSE "NO value"
END FROM `table 2`;

//50s
SELECT `50s`,
CASE WHEN `50s` BETWEEN 0 AND 1 THEN "3"
 WHEN `50s` BETWEEN 2 AND 7 THEN "4"
 WHEN `50s` BETWEEN 8 AND 16 THEN "5"
ELSE "NO value"
END FROM `table 2`;

//100s
SELECT `100s`,
CASE WHEN `100s` BETWEEN 0 AND 1 THEN "3"
 WHEN `100s` BETWEEN 1.01 AND 2 THEN "4"
 WHEN `100s` BETWEEN 2.01 AND 3 THEN "5"
ELSE "NO value"
END FROM `table 2`;

//0s
SELECT `0s`,
CASE WHEN `0s` BETWEEN 0 AND 2 THEN "5"
 WHEN `0s` BETWEEN 3 AND 5 THEN "4"
 WHEN `0s` BETWEEN 5 AND 7 THEN "3"
ELSE "NO value"
END FROM `table 2`;

//Strike Rate
SELECT `Strike Rate`,
CASE WHEN `Strike Rate` BETWEEN 0 AND 20.99 THEN "1"
WHEN `Strike Rate` BETWEEN 21 AND 49.99 THEN "2"
 WHEN `Strike Rate` BETWEEN 50 AND 74.99 THEN "3"
 WHEN `Strike Rate` BETWEEN 75 AND 110.99 THEN "4"
 WHEN `Strike Rate` BETWEEN 111 AND 166 THEN "5"
ELSE "NO value"
END FROM `table 2`;
```

CHAPTER 7

APPENDIX – B DATA PREPARATION

The player names are replaced with the Player_ID for data pre-processing.

Player_Id	Player_Name	DOB	Batting_Hand	Bowling_Skill	Country
1	Chandila , A	08-Jul-72	Left_Hand	Right-arm medium	India
2	Chopra , A	27-Sep-81	Right_Hand	Right-arm medium	New Zealand
3	Flintoff , A	19-Dec-74	Right_Hand	Right-arm medium	Australia
4	Kumble , A	15-Jul-77	Right_Hand	Right-arm offbreak	Australia
5	Mishra , A	17-Oct-80	Right_Hand	Right-arm offbreak	Pakistan
6	Mithun , A	11-Jan-73	Right_Hand	Right-arm offbreak	India
7	Mukund , A	16-Feb-78	Right_Hand	Right-arm offbreak	India
8	Nand Kishore , A	05-Nov-88	Right_Hand	Right-arm medium	India
9	Nehra , A	16-Oct-75	Right_Hand	Right-arm fast-medium	South Africa
10	Nel , A	18-Aug-83	Right_Hand	Legbreak googly	Australia
11	Singh , A	03-Dec-76	Right_Hand	Right-arm medium	South Africa
12	Symonds , A	07-Oct-77	Right_Hand	Right-arm medium-fast	India
13	Uniyal , A	30-Apr-77	Right_Hand	Right-arm fast-medium	Australia
14	Zampa , A	02-Oct-86	Right_Hand	Right-arm medium	India
15	Bilakhia , AA	07-Oct-78	Right_Hand	Left-arm fast-medium	India
16	Chavan , AA	06-Jun-70	Left_Hand	Slow left-arm orthodox	India
17	Jhunjhunwala , AA	09-Mar-85	Left_Hand	NULL	India
18	Kazi , AA	29-Oct-71	Left_Hand	Right-arm medium	Australia
19	Noffke , AA	27-May-75	Left_Hand	Right-arm medium	Australia
20	Agarkar , AB	07-Jul-81	Right_Hand	Right-arm medium	India
21	Barath , AB	27-Nov-86	Left_Hand	Right-arm offbreak	India
22	de Villiers , AB	28-Jul-78	Left_Hand	Right-arm fast-medium	New Zealand
23	Dinda , AB	30-Aug-80	Right_Hand	Right-arm offbreak	India
24	McDonald , AB	24-Dec-86	Left_Hand	Right-arm offbreak	India
25	Razzak , Abdur	24-Oct-78	Right_Hand	Right-arm medium	Australia
26	Blizzard , AC	27-Oct-77	Left_Hand	Right-arm offbreak	Sri Lanka
27	Gilchrist , AC	12-Dec-81	Left_Hand	Slow left-arm orthodox	India
28	Thomas , AC	21-Aug-75	Left_Hand	Slow left-arm chinaman	Australia
29	Voges , AC	27-Dec-84	Left_Hand	Left-arm medium-fast	India
30	Mascarenhas , AD	17-Dec-88	Right_Hand	Right-arm medium	India

The basic attributes for training dataset.

Table.7.14 Batting attributes basic_Training.csv

Player ID	No. of Innings	50s	100s	0s	Strike Rate	Average	Runs
274	3	4	3	3	4	5	3
285	3	5	3	3	4	5	4
298	5	5	3	3	5	5	5
437	5	5	3	4	5	5	5
299	3	3	3	3	4	3	2
190	5	4	3	4	5	4	3
393	5	5	3	4	5	5	4
327	4	4	3	5	4	4	3
451	2	3	3	3	2	3	2
410	1	3	3	3	2	3	2
196	2	3	3	3	2	3	2
260	1	3	3	3	1	3	2
205	2	3	3	3	3	3	2
106	1	3	3	3	1	3	2
389	2	3	3	3	5	4	2
7	1	3	3	3	1	3	2

259	4	3	3	3	3	3	2
242	4	3	3	4	3	3	2
317	1	3	3	3	1	3	2
262	3	4	4	4	4	4	4
160	1	3	3	3	2	3	2
203	1	3	3	3	2	3	2
477	1	3	3	3	2	3	2
345	3	3	3	4	3	3	2
122	3	3	3	3	2	3	2
409	2	3	3	3	1	3	2
312	1	3	3	3	2	3	2
418	4	3	3	3	2	3	2
91	1	3	3	3	1	3	2
510	3	3	3	3	5	4	2
124	4	4	3	3	5	4	3
547	1	3	3	3	1	3	2
480	1	3	3	3	1	3	2
420	2	3	3	3	4	3	2
233	1	3	3	3	2	3	2
149	2	4	3	3	2	3	2
358	4	3	3	3	4	4	3
90	1	3	3	3	1	3	2
506	2	3	3	3	1	3	2
3	1	3	3	3	2	3	2
12	3	4	3	3	4	5	3
365	5	5	3	4	5	5	5
515	4	4	3	3	4	4	3
27	4	5	4	5	5	5	5
505	2	3	3	4	4	4	2
248	1	3	3	3	2	3	2
170	3	4	3	4	4	4	3
121	3	3	3	3	4	4	2
135	1	3	3	3	2	3	2
428	2	3	3	3	2	3	2
417	2	3	3	3	2	3	2
511	2	3	3	4	3	3	2
373	4	3	3	4	3	3	2
114	1	3	3	3	1	3	2
165	1	3	3	3	1	3	2
339	4	3	3	4	2	3	2
137	1	3	3	3	1	3	2
337	1	3	3	3	1	3	2
140	2	3	3	3	4	4	2
482	3	4	3	3	4	4	3
366	3	3	3	3	4	4	2
15	1	3	3	3	2	3	2
150	1	3	3	3	1	3	2
431	1	3	3	3	1	3	2

167	2	3	3	3	2	3	2
191	1	3	3	3	1	3	2
72	1	3	3	3	1	3	2
49	1	3	3	3	1	3	2
272	2	3	3	3	3	3	2
296	1	3	3	3	2	3	2
218	1	3	3	3	2	3	2
354	3	3	3	4	2	3	2
396	4	5	3	3	5	4	5
224	4	5	3	4	5	4	5
109	3	4	3	3	4	4	3
67	2	3	3	3	3	3	2
406	2	4	3	4	3	3	2
206	3	5	3	3	4	4	3
143	2	3	3	3	2	3	2
5	4	3	3	5	4	4	2
182	1	3	3	3	2	3	2
145	4	3	3	4	3	3	2
281	2	3	3	4	2	3	2
174	1	3	3	3	2	3	2
48	1	3	3	3	1	3	2
226	1	3	3	3	1	3	2
175	3	3	3	3	3	3	2
126	1	3	3	3	2	3	2
328	1	3	3	3	2	3	2
17	2	3	3	3	3	3	2
336	1	3	3	3	2	3	2
97	1	3	3	3	1	3	2
50	1	3	3	3	2	3	2
495	1	3	3	3	1	3	2
76	1	3	3	3	1	3	2
474	1	3	3	3	1	3	2
486	1	3	3	3	1	3	2
152	5	5	3	4	5	5	5
496	4	5	3	4	5	5	5
284	4	4	3	4	4	4	3
225	5	4	3	4	5	5	4
275	2	3	3	3	4	4	2
22	4	5	3	4	5	5	4
483	3	5	3	4	5	5	4
346	4	3	3	3	4	3	2
433	1	3	3	3	2	3	2
158	2	3	3	3	1	3	2
291	1	3	3	3	1	3	2
335	3	3	3	3	2	3	2
257	3	3	3	5	4	4	2
68	1	3	3	3	1	3	2
134	3	3	3	3	4	4	2

118	3	4	4	4	4	4	4
514	3	3	3	3	3	4	2
24	2	3	3	3	4	3	2
136	2	3	3	3	1	3	2
9	3	3	3	3	3	3	2
45	1	3	3	3	1	3	2
332	1	4	3	3	2	3	2
232	2	3	3	3	2	3	2
270	1	3	3	3	3	3	2
400	1	3	3	3	1	3	2
493	3	3	3	3	2	3	2
236	2	4	3	3	4	4	3
138	4	5	3	3	5	4	4
322	1	3	3	3	2	3	2
37	1	3	3	3	3	3	2
178	5	3	3	4	5	4	3
33	1	3	3	3	2	3	2
315	4	4	3	4	4	3	3
250	3	4	3	4	5	4	3
490	1	3	3	3	2	3	2
161	1	3	3	3	1	3	2
258	3	3	3	4	3	3	2
501	2	3	3	3	1	3	2
466	1	3	3	3	1	3	2
401	2	3	3	3	1	3	2
119	1	3	3	3	2	3	2
363	2	3	3	3	3	3	2
20	3	3	3	3	4	3	2
245	2	3	3	3	3	5	2
445	2	4	3	3	3	4	2
427	3	5	3	3	4	5	5
508	2	3	3	3	2	3	2
521	4	4	3	3	4	4	4
338	5	3	3	5	4	4	2
208	2	4	3	3	3	4	2
380	1	3	3	3	1	3	2
71	3	3	3	3	5	4	2
211	2	3	3	3	3	3	2
502	2	3	3	3	2	3	2
407	3	3	3	3	2	3	2
319	1	3	3	3	1	3	2
377	3	3	3	3	4	3	2
154	1	3	3	3	1	3	2
488	1	3	3	3	1	3	2
75	3	4	3	4	5	4	4
422	4	4	3	4	4	4	4
128	3	4	3	4	5	4	3
415	1	3	3	3	2	3	2

249	3	3	3	3	5	3	2
120	2	3	3	3	2	4	2
491	1	3	3	3	2	3	2
476	1	3	3	3	2	3	2
255	3	3	3	3	4	3	2
384	1	3	3	3	1	3	2
292	1	3	3	3	1	3	2
2	1	3	3	3	2	3	2
84	3	4	3	4	5	4	4
23	3	3	3	4	2	3	2
432	1	3	3	3	1	3	2
181	3	3	3	3	2	3	2
74	1	3	3	3	1	3	2
463	3	4	3	3	4	3	3
387	5	4	3	3	5	4	5
522	2	3	3	3	2	3	2
455	4	5	3	4	5	4	5
42	3	3	3	3	5	3	3
131	1	3	3	3	1	3	2
446	2	3	3	3	2	3	2
166	4	3	3	4	5	4	2
462	4	4	3	3	5	4	3
265	1	3	3	3	2	3	2
423	2	3	3	3	4	3	2
341	2	3	3	3	3	3	2
243	1	3	3	3	2	3	2
43	3	4	3	4	4	4	3
504	1	3	3	3	1	3	2
283	3	4	3	5	4	3	3
10	1	3	3	3	1	3	2
112	2	3	3	3	3	3	2
378	2	3	3	3	2	3	2
141	3	3	3	3	3	3	2
386	1	3	3	3	1	3	2
310	1	3	3	3	2	4	2
157	3	4	3	4	4	4	3
456	3	5	3	3	5	4	4
412	2	4	3	4	4	3	2
518	5	4	3	4	5	4	4
219	1	3	3	3	2	3	2
115	1	3	3	3	2	3	2
254	3	3	3	3	4	3	2
440	4	3	3	5	3	3	2
450	2	3	3	3	2	3	2
142	1	3	3	3	1	3	2
30	2	3	3	3	4	3	2
261	2	3	3	3	2	3	2
439	4	3	3	3	3	3	2

519	1	3	3	3	1	3	2
286	4	3	3	4	2	3	2
329	2	3	3	3	2	3	2
302	3	3	3	3	4	3	2
349	5	4	3	4	5	4	5
452	2	3	3	3	4	3	2
507	1	3	3	3	2	3	2
200	5	5	3	5	4	4	5
278	1	3	3	3	2	3	2
65	2	3	3	3	4	3	2
494	5	5	3	3	5	4	5
342	4	3	3	5	4	3	2
19	1	3	3	3	1	3	2
523	4	3	3	4	4	3	2
394	1	3	3	3	1	3	2
357	4	3	3	3	4	3	2
184	1	3	3	3	1	3	2
144	1	3	3	3	1	3	2
4	3	3	3	3	2	3	2
419	1	3	3	3	1	3	2
25	1	3	3	3	1	3	2
381	2	4	3	3	2	3	2
375	1	3	3	3	1	3	2
516	2	3	3	3	1	3	2
503	1	3	3	3	2	3	2
289	1	3	3	3	2	3	2
100	3	5	5	4	5	5	5
31	2	3	3	3	3	3	2
517	1	3	3	3	1	3	2
461	1	3	3	3	1	3	2
47	1	3	3	3	1	3	2
107	1	3	3	3	1	3	2
268	1	3	3	3	1	3	2
460	1	3	3	3	1	3	2
51	1	3	3	3	1	3	2
162	1	3	3	3	1	3	2
290	1	3	3	3	1	3	2
441	4	3	3	4	3	3	2
94	1	3	3	3	1	3	2
359	1	3	3	3	1	3	2
185	2	3	3	3	3	4	2
244	1	3	3	3	2	3	2
57	2	3	3	3	3	3	2
368	1	3	3	3	1	3	2
472	1	3	3	3	1	3	2
220	1	3	3	3	1	3	2
331	2	4	3	3	2	3	2
11	2	3	3	4	1	3	2

444	1	3	3	3	1	3	2
194	3	4	3	4	4	3	3
113	1	3	3	3	1	3	2
347	1	3	3	3	2	3	2
239	1	3	3	3	1	3	2
6	2	3	3	3	3	3	2
235	2	3	3	3	1	3	2
297	2	4	3	3	3	3	3
21	1	3	3	3	1	3	2
435	2	3	3	3	2	3	2
241	1	3	3	3	1	3	2
82	2	3	3	3	4	3	2
383	1	3	3	3	2	3	2
316	2	4	3	3	4	4	2
98	2	3	3	3	3	3	2
382	1	3	3	3	1	3	2
267	3	4	3	3	3	3	2
426	1	3	3	3	1	3	2
269	1	3	3	3	1	3	2
195	2	3	3	3	1	3	2
59	3	5	3	4	4	4	4
217	3	4	3	3	4	3	3
350	2	3	3	3	2	3	2
353	2	3	3	3	2	3	2
92	1	3	3	3	1	3	2
54	1	3	3	3	2	3	2
464	2	3	3	3	4	4	2
35	1	3	3	3	1	3	2
46	1	3	3	3	2	3	2
29	1	3	3	3	2	3	2
151	2	3	3	3	2	3	2
139	1	3	3	3	1	3	2
52	1	3	3	3	2	3	2
320	2	3	3	3	2	3	2
470	2	3	3	3	2	3	2
402	1	3	3	3	2	3	2
13	1	3	3	3	1	3	2
36	1	3	3	3	1	3	2
421	1	3	3	3	1	3	2
408	1	3	3	3	2	3	2
146	2	3	3	3	2	3	2
305	1	3	3	3	1	3	2
487	1	3	3	3	2	3	2
96	1	3	3	3	1	3	2
300	1	3	3	3	1	3	2
467	1	3	3	3	2	3	2
73	2	3	3	3	3	3	2
309	1	3	3	3	1	3	2

513	1	3	3	3	2	3	2
256	1	3	3	3	1	3	2
385	2	3	3	4	1	3	2
324	1	3	3	3	1	3	2
323	1	3	3	3	1	3	2
372	1	3	3	3	2	4	2
83	1	3	3	3	2	3	2
429	2	3	3	3	3	3	2
199	2	3	3	3	2	4	2
26	1	3	3	3	2	3	2
129	1	3	3	3	1	3	2
388	1	3	3	3	1	3	2
311	1	3	3	3	2	3	2
103	1	3	3	3	2	3	2
403	1	3	3	3	2	3	2
509	2	3	3	3	2	3	2
479	1	3	3	3	1	3	2
198	1	3	3	3	1	3	2
459	1	3	3	3	1	3	2
28	2	3	3	3	2	3	2
207	1	3	3	3	2	3	2
69	2	3	3	3	2	3	2
40	3	3	3	3	2	3	2
123	2	3	3	3	3	3	2
16	2	3	3	3	2	3	2
264	2	3	3	3	3	3	2
202	1	3	3	3	2	3	2
60	1	3	3	3	3	3	2
222	1	3	3	3	1	3	2
351	1	3	3	3	1	3	2
18	1	3	3	3	1	3	2
188	1	3	3	3	4	3	2
390	2	3	3	3	3	3	2
117	1	3	3	3	2	3	2
164	1	3	3	3	2	3	2
64	2	3	3	3	2	3	2
304	2	3	3	3	1	3	2
318	2	3	3	3	1	3	2
453	2	3	3	3	2	3	2
252	1	3	3	3	1	3	2
362	1	3	3	3	2	3	2
367	1	3	3	3	2	3	2
105	1	3	3	3	1	3	2
413	1	3	3	3	1	3	2
355	1	3	3	3	1	3	2
454	2	3	3	3	2	3	2
53	1	3	3	3	2	3	2
288	1	3	3	3	1	3	2

177	1	3	3	3	2	3	2
279	1	3	3	3	1	3	2
246	1	3	3	3	1	3	2
216	1	3	3	3	2	3	2
229	1	3	3	3	2	3	2
156	1	3	3	3	1	3	2
1	1	3	3	3	1	3	2
173	2	3	3	3	1	3	2
374	1	3	3	3	1	3	2

The batting attributes consists of 361 players of data and bowling attributes consists of 314 players of data based on the five years of training data.

Table.7.15 Basic attributes_testing.csv

Player ID	No. of Innings	50s	100s	0s	Strike Rate	Average	Runs
274	3	5	3	3	3	4	4
437	5	5	3	4	5	5	5
298	5	5	3	3	5	5	5
358	5	3	3	4	5	5	5
510	4	4	3	3	5	5	4
262	4	5	3	3	4	4	4
393	2	3	3	3	1	3	4
190	2	3	3	3	4	5	2
124	4	3	3	3	5	4	2
345	4	3	3	3	4	4	2
101	3	4	3	4	5	5	2
70	1	3	3	3	1	3	2
287	5	3	3	4	4	3	2
389	1	3	3	3	1	3	2
56	2	3	3	3	1	3	2
204	2	3	3	3	3	3	2
136	1	3	3	3	1	3	2
493	4	3	3	3	5	4	2
85	1	3	3	3	2	3	2
118	5	5	3	4	5	5	5
496	3	4	3	3	4	4	4
138	2	4	3	3	2	3	2
110	2	3	3	3	3	3	2
232	4	4	3	4	5	4	2
271	1	3	3	3	1	3	3
178	3	3	3	3	3	4	2
490	2	3	3	4	2	3	3
67	1	3	3	3	1	3	4
20	1	3	3	3	1	3	4
258	3	3	3	3	5	3	3
363	1	3	3	3	1	3	2
315	4	3	3	4	5	4	2

86	1	3	3	3	1	3	2
185	2	3	3	3	3	3	2
33	3	4	3	3	4	4	2
401	3	3	3	4	2	3	2
9	3	3	3	4	2	3	2
515	1	3	3	3	2	3	2
398	2	3	3	3	2	3	2
322	3	3	3	3	4	4	4
117	4	5	3	3	5	5	2
427	3	4	3	3	5	5	2
245	1	3	3	3	1	3	2
27	2	3	3	3	2	3	2
128	2	3	3	3	2	4	3
353	2	3	3	3	2	3	5
263	3	4	3	3	5	4	2
267	3	4	3	5	4	4	5
64	2	4	3	3	2	3	3
548	2	3	3	4	2	3	2
299	1	3	3	3	2	3	2
338	4	3	3	4	4	3	2
342	4	3	3	4	4	3	3
331	1	3	3	3	1	3	2
167	1	3	3	3	1	3	2
366	1	3	3	3	1	3	2
318	2	3	3	3	1	3	2
73	1	3	3	3	1	3	3
82	2	3	3	3	4	4	2
416	4	3	3	3	3	3	3
30	1	3	3	3	1	3	2
277	1	3	3	3	1	3	2
518	5	4	3	4	5	5	2
372	2	3	3	4	4	4	2
146	3	4	3	3	4	4	2
152	5	5	3	5	5	5	2
402	1	3	3	3	2	3	3
71	1	3	3	3	2	3	4
200	2	3	3	4	2	3	2
297	2	4	3	3	3	3	4
284	3	4	3	4	4	4	2
346	3	3	3	3	4	4	2
120	2	3	3	3	1	3	2
447	1	3	3	3	1	3	2
242	2	3	3	3	2	3	2
453	4	3	3	4	4	3	2
75	4	5	3	3	5	4	2
293	3	3	3	3	3	3	2
320	1	3	3	3	1	3	2
350	1	3	3	3	1	3	2

249	1	3	3	3	2	3	2
400	1	3	3	3	1	3	2
181	2	3	3	3	4	3	2
335	1	3	3	3	3	3	5
217	5	5	3	3	5	5	2
365	5	5	3	4	5	5	2
140	4	5	3	5	5	5	5
225	5	5	3	4	5	5	2
54	3	3	3	4	4	4	2
455	2	3	3	3	2	3	5
166	5	3	3	5	5	4	2
59	4	4	3	5	5	5	2
161	4	4	3	5	5	4	2
348	2	3	3	3	5	4	2
384	1	3	3	3	1	3	2
276	3	3	3	3	4	3	2
313	3	3	3	3	4	3	2
441	4	3	3	3	4	3	2
396	5	5	3	4	5	5	2
46	1	3	3	3	5	3	2
201	3	3	3	3	2	3	2
520	4	3	3	3	3	3	2
286	1	3	3	3	1	3	2
196	1	3	3	3	1	3	2
141	3	3	3	3	4	3	2
326	1	3	3	3	1	3	5
454	4	4	3	3	5	5	2
37	4	5	3	4	5	5	2
387	5	5	3	4	5	5	2
31	3	3	3	3	4	4	2
283	4	4	3	4	5	5	2
246	1	3	3	3	2	3	2
296	2	3	3	3	2	3	2
521	4	4	3	4	5	5	2
240	2	3	3	3	2	3	2
177	3	3	3	3	1	3	2
250	1	3	3	3	2	3	2
42	2	3	3	3	2	3	2
482	1	3	3	3	1	3	2
69	5	3	3	4	4	3	2
489	1	3	3	3	1	3	2
509	2	3	3	3	2	3	2
35	1	3	3	3	2	3	2
261	1	3	3	3	1	3	2
330	2	3	3	3	3	3	2
23	3	3	3	4	2	3	2
288	1	3	3	3	2	3	2
354	2	3	3	4	1	3	2

74	1	3	3	3	1	3	3
385	1	3	3	3	4	3	2
257	2	3	3	3	2	3	2
216	1	3	3	3	1	3	2
84	2	3	3	3	3	4	2
456	4	4	4	3	5	5	2
43	5	5	3	5	5	5	3
464	4	3	3	4	5	4	2
349	2	4	3	3	2	3	2
468	4	4	3	4	5	5	2
123	2	3	3	3	2	3	2
316	1	3	3	3	2	3	2
207	4	3	3	3	5	4	2
229	2	3	3	3	3	3	2
273	1	3	3	3	2	3	2
414	2	3	3	3	3	3	2
439	2	3	3	3	1	3	2
40	1	3	3	3	1	3	2
470	1	3	3	3	1	3	2
407	1	3	3	3	1	3	4
156	2	3	3	3	2	3	3
392	2	3	3	3	1	3	2
503	1	3	3	3	1	3	2
1	1	3	3	3	1	3	2
16	1	3	3	3	1	3	2
343	3	3	3	3	2	3	2
355	1	3	3	3	2	3	2
357	3	3	3	3	3	4	2
100	4	5	4	4	5	5	2
494	5	5	5	4	5	5	2
270	4	4	3	3	5	5	2
22	4	5	4	4	5	5	2
462	3	4	3	3	4	4	2
98	2	3	3	3	2	4	2
7	1	3	3	3	1	3	2
483	1	3	3	3	1	3	2
222	2	3	3	3	2	3	2
264	3	4	3	3	5	4	2
231	3	4	3	3	4	4	2
352	2	3	3	4	2	3	2
187	1	3	3	3	1	3	2
230	4	5	3	3	5	4	2
143	2	3	3	3	2	3	2
6	1	3	3	3	1	3	2
255	2	3	3	3	1	3	2
24	1	3	3	3	1	3	2
195	3	3	3	3	2	3	2
523	3	3	3	3	2	3	2

259	2	3	3	3	1	3	2
373	2	3	3	3	1	3	2
132	2	3	3	3	3	4	2
145	3	3	3	3	4	4	2
76	1	3	3	3	2	3	2
312	3	3	3	3	4	4	2
327	4	4	3	4	5	4	2
239	4	3	3	3	5	4	2
339	2	3	3	3	4	4	2
109	2	3	3	3	2	3	2
159	2	3	3	3	2	3	2
5	4	3	3	4	3	4	2
121	1	3	3	3	1	3	2
224	1	3	3	3	1	3	2
328	1	3	3	3	1	3	2
512	1	3	3	3	1	3	2
344	3	4	3	3	4	4	2
48	1	3	3	3	1	3	2
50	2	3	3	3	4	4	2
175	3	3	3	3	1	3	2
149	3	4	3	3	4	4	2
209	1	3	3	3	1	3	2
90	1	3	3	3	1	3	2
497	1	3	3	3	2	4	2
206	3	4	3	3	4	5	2
236	2	3	3	3	2	4	2
197	1	3	3	3	1	3	2
180	3	3	3	3	2	3	2
160	3	4	3	3	3	4	2
55	4	3	3	4	4	4	2
430	1	3	3	3	2	3	2
79	1	3	3	3	1	3	2
221	1	3	3	3	2	3	2
413	4	3	3	3	4	4	2
429	3	4	3	3	4	4	2
97	2	4	3	3	4	4	2
334	2	3	3	3	2	3	2
247	3	5	3	3	4	5	2
102	3	4	3	3	4	4	2
89	1	3	3	3	2	3	2
397	1	3	3	3	2	3	2
252	1	3	3	3	1	3	2
213	1	3	3	3	1	3	2
356	1	3	3	3	3	3	2
78	1	3	3	3	4	4	2
480	2	3	3	3	5	3	2
403	1	3	3	3	2	3	2
522	1	3	3	3	2	3	2

499	1	3	3	3	2	3	2
266	3	3	3	3	2	3	3
379	1	3	3	3	2	3	2
308	1	3	3	3	2	3	2
173	2	3	3	3	3	3	2
418	1	3	3	3	1	3	2
501	3	3	3	3	3	3	2
364	1	3	3	3	1	3	2
458	3	4	3	4	3	4	2
130	1	3	3	3	1	3	2
163	1	3	3	3	1	3	2
546	1	3	3	3	1	3	2
189	2	3	3	3	1	3	2
164	2	3	3	3	2	3	2
51	2	3	3	3	1	3	2
449	2	3	3	4	2	3	2
223	2	3	3	3	2	3	2
186	2	3	3	3	2	4	2
171	3	3	3	3	4	4	2
282	3	3	3	4	4	3	2
127	3	3	3	3	4	4	2
188	1	3	3	3	1	3	2
77	2	3	3	3	1	3	2
116	2	3	3	3	3	3	2
448	2	3	3	3	3	4	2
411	1	3	3	3	2	3	2
390	2	3	3	3	1	3	2
237	2	4	3	3	3	4	2
381	1	3	3	3	2	3	2
478	2	3	3	3	1	3	2
376	2	4	3	4	3	4	2
469	2	4	3	3	3	3	2
111	2	3	3	3	3	3	2
147	1	3	3	3	2	3	2
183	2	3	3	3	2	3	2
32	1	3	3	3	1	3	2
399	2	3	3	3	1	3	2
294	2	3	3	3	2	3	2
172	2	4	4	3	3	4	2
325	1	3	3	3	2	3	2
228	1	3	3	3	2	3	2
314	1	3	3	3	1	3	2
148	1	3	3	3	2	3	2
471	1	3	3	3	1	3	2
93	1	3	3	3	1	3	2
238	2	3	3	3	1	3	2
227	3	3	3	3	3	4	5
303	2	4	3	3	3	4	2

192	2	3	3	3	3	4	2
280	2	3	3	3	3	3	2
492	1	3	3	3	2	3	2
14	2	3	3	3	1	3	2
251	2	3	3	3	1	3	2
133	1	3	3	3	2	3	2
443	1	3	3	3	1	3	2
484	2	3	3	3	3	4	2
104	2	3	3	3	1	3	2
475	1	3	3	3	1	3	2
34	1	3	3	3	1	3	2
301	2	3	3	3	1	3	2
545	1	3	3	3	1	3	2
549	1	3	3	3	1	3	2
550	1	3	3	3	2	3	2
544	1	3	3	3	2	3	2
542	1	3	3	3	1	3	2
551	1	3	3	3	1	3	2
552	1	3	3	3	1	3	2
543	2	3	3	3	1	3	2
541	1	3	3	3	1	3	2
539	1	3	3	3	1	3	2
540	1	3	3	3	1	3	2
182	1	3	3	3	1	3	2
553	1	3	3	3	2	3	2
537	2	3	3	4	2	3	2
538	2	3	3	3	1	3	2
554	1	3	3	3	2	3	2
535	2	3	3	3	2	3	2
532	2	4	3	3	2	3	2
533	1	3	3	3	1	3	2
534	1	3	3	3	2	3	2
536	2	3	3	3	1	3	2
530	1	3	3	3	2	3	2
555	1	3	3	3	1	3	2
529	1	3	3	3	1	3	2
528	1	3	3	3	1	3	2
531	1	3	3	3	1	3	3
526	2	3	3	3	2	3	2
524	1	3	3	3	1	3	2
525	1	3	3	3	1	3	2

These basic attributes are then used for calculating the derived or complex attributes. The complex attributes are Consistency, Venue and Opposition.

Table.7.16 The snippet for calculating runs with training dataset using complex attributes.

Player_ID	Consistency	Opposition	Venue	Runs
274	4	4	4	3
285	4	4	4	4
298	5	5	5	5
437	5	5	5	5
299	3	4	4	2
190	4	4	4	3
393	5	5	5	4
327	4	4	4	3
451	2	3	3	2
410	2	3	3	2
196	3	3	3	2
260	2	3	3	2
205	3	3	3	2
106	2	3	3	2
389	4	4	4	2
7	2	3	3	2
259	3	3	3	2
242	3	3	3	2
317	2	3	3	2
262	4	4	4	3
160	4	4	4	5
203	4	4	4	3
477	4	4	4	4
345	3	4	3	2
122	2	3	3	2
409	3	3	3	2
312	3	4	3	3
418	3	3	3	2
91	2	3	3	2
510	2	3	3	2

Table.7.17 The snippet for calculating runs with test dataset using complex attributes.

Player_ID	Consistency	Opposition	Venue	Runs
274	3	4	3	5
437	5	5	5	5
298	5	5	5	2
358	4	5	4	3
510	4	4	4	4
262	3	4	3	3
393	2	3	2	2
190	4	4	4	2
124	3	3	3	2
345	4	4	4	2
101	4	4	4	2
70	2	2	2	2
287	3	4	3	2
389	2	2	2	2
56	2	2	2	2
204	2	2	2	2
136	2	2	2	2
493	4	4	4	3
85	2	2	2	5
118	5	5	4	3
496	3	4	3	4
138	2	3	2	2
110	2	3	3	2
232	4	4	4	2
271	2	2	2	3
178	3	4	3	2
490	2	3	2	2
67	2	2	2	2
20	2	2	2	2
258	3	3	3	2

Table.7.18 The snippet for calculating wickets with training dataset using complex attributes.

Player_ID	Consistency	Opposition	Venue	Wickets
410	2	3	3	2
190	4	4	4	5
242	3	4	4	4
299	2	4	3	3
437	4	4	4	2
205	2	3	3	2
259	3	4	4	4
317	2	3	3	2
260	2	3	3	2
196	2	3	3	2
106	2	3	3	2
121	2	3	3	2
365	3	3	3	2
428	5	3	3	2
339	3	4	4	5
511	2	3	3	2
373	3	4	4	5
337	2	3	3	2
114	2	3	3	2
135	2	3	3	2
137	2	3	3	2
417	2	3	3	2
420	2	3	3	2
515	3	5	3	2
12	2	4	3	2
248	2	3	3	2
483	3	3	3	2
5	3	4	4	5
275	3	4	3	3
506	2	3	3	3

Table.7.19 The snippet for calculating wickets with test dataset using complex attributes.

Player_ID	Consistency	Opposition	Venue	Wickets
124	3	4	4	5
287	4	4	4	5
190	2	3	3	2
437	3	4	4	2
345	4	4	4	4
358	4	5	4	4
101	3	4	3	4
136	2	2	3	2
56	2	2	3	2
70	2	3	3	2
204	2	3	3	2
185	2	3	3	2
363	2	2	3	2
9	3	4	3	4
401	3	4	3	3
493	3	4	4	5
86	2	2	3	2
178	3	4	3	2
398	2	3	3	3
258	3	4	4	3
496	2	2	3	2
20	2	2	3	2
33	3	4	3	3
85	2	2	3	2
322	2	3	3	3
27	2	2	3	2
416	3	4	4	5
167	2	2	3	2
82	2	3	3	2
64	2	3	3	2

CHAPTER 8
APPENDIX – C RESULTS

These are some of the snippet of the result.arff file obtained after classification of the model.

For Prediction of Runs,

1) These is the Naïve Bayes Classifier output result obtained after classification from the basic attributes.

Viewer

Relation: Batting attributes_Training -weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last_predicted

No.	1: Player ID Nominal	2: No. of Innings Nominal	3: 50s Nominal	4: 100s Nominal	5: 0s Nominal	6: Strike Rate Nominal	7: Average Nominal	8: prediction margin Numeric	9: predicted Runs Nominal	10: **Runs** Nominal
1	274	3	5	3	3	3	4	-0.285714	5	4
2	437	5	5	3	4	5	5	0.285714	5	5
3	298	5	5	3	3	5	5	0.285714	5	5
4	358	5	3	3	4	5	5	-0.989933	2	5
5	510	4	4	3	3	5	5	-0.090909	3	4
6	262	4	5	3	3	4	4	-0.285714	5	4
7	393	2	3	3	3	1	3	-0.989933	2	4
8	190	2	3	3	3	4	5	0.979866	2	2
9	124	4	3	3	3	5	4	0.979866	2	2
10	345	4	3	3	3	4	4	0.979866	2	2

2) These is the decision tree classifier (j48) output result obtained after classification from the basic attributes.

Viewer

Relation: Batting attributes_Training -weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last_predicted

No.	1: Player ID Nominal	2: No. of Innings Nominal	3: 50s Nominal	4: 100s Nominal	5: 0s Nominal	6: Strike Rate Nominal	7: Average Nominal	8: prediction margin Numeric	9: predicted Runs Nominal	10: **Runs** Nominal
1	274	3	5	3	3	3	4	-0.285714	5	4
2	437	5	5	3	4	5	5	0.285714	5	5
3	298	5	5	3	3	5	5	0.285714	5	5
4	358	5	3	3	4	5	5	-0.989933	2	5
5	510	4	4	3	3	5	5	-0.090909	3	4
6	262	4	5	3	3	4	4	-0.285714	5	4
7	393	2	3	3	3	1	3	-0.989933	2	4
8	190	2	3	3	3	4	5	0.979866	2	2
9	124	4	3	3	3	5	4	0.979866	2	2
10	345	4	3	3	3	4	4	0.979866	2	2
11		3	4	3	4	5	5	-0.454545	3	2
12		1	3	3	3	1	3	0.979866	2	2
13		5	3	3	4	4	3	0.979866	2	2
14	389	1	3	3	3	1	3	0.979866	2	2
15		2	3	3	3	1	3	0.979866	2	2
16		2	3	3	3	3	3	0.979866	2	2
17	136	1	3	3	3	1	3	0.979866	2	2
18	493	4	3	3	3	5	4	0.979866	2	2
19		1	3	3	3	2	3	0.979866	2	2
20	118	5	5	3	4	5	5	0.285714	5	5
21	496	3	4	3	3	4	4	-0.7	3	4
22	138	2	4	3	3	2	3	1.0	2	2
23		2	3	3	3	3	3	0.979866	2	2
24	232	4	4	3	4	5	4	-0.454545	3	2
25		1	3	3	3	1	3	-0.979866	2	3
26	178	3	3	3	3	3	4	0.979866	2	2
27	490	2	3	3	4	2	3	-0.979866	2	3
28	67	1	3	3	3	1	3	-0.989933	2	4
29	20	1	3	3	3	1	3	-0.989933	2	4

3) These is the Random Forest classifier output result obtained after classification from the basic attributes.

Viewer

Relation: Batting attributes testing_predicted

No.	1: Player ID Numeric	2: No. of Innings Numeric	3: 50s Numeric	4: 100s Numeric	5: 0s Numeric	6: Strike Rate Numeric	7: Average Numeric	8: predictedRuns Numeric	9: **Runs** Numeric
1	274.0	3.0	5.0	3.0	3.0	3.0	4.0	3.17	4.0
2	437.0	5.0	5.0	3.0	4.0	5.0	5.0	4.8	5.0
3	298.0	5.0	5.0	3.0	3.0	5.0	5.0	4.88	5.0
4	358.0	5.0	3.0	3.0	4.0	5.0	5.0	3.09	5.0
5	510.0	4.0	4.0	3.0	3.0	5.0	5.0	3.61	4.0
6	262.0	4.0	5.0	3.0	3.0	4.0	4.0	4.03	4.0
7	393.0	2.0	3.0	3.0	3.0	1.0	3.0	2.0	4.0
8	190.0	2.0	3.0	3.0	3.0	4.0	5.0	2.11	2.0
9	124.0	4.0	3.0	3.0	3.0	5.0	4.0	2.36	2.0
10	345.0	4.0	3.0	3.0	3.0	4.0	4.0	2.41	2.0
11	101.0	3.0	4.0	3.0	4.0	5.0	5.0	3.87	2.0
12	70.0	1.0	3.0	3.0	3.0	1.0	3.0	2.0	2.0
13	287.0	5.0	3.0	3.0	4.0	4.0	3.0	2.18	2.0
14	389.0	1.0	3.0	3.0	3.0	1.0	3.0	2.0	2.0
15	56.0	2.0	3.0	3.0	3.0	1.0	3.0	2.0	2.0
16	204.0	2.0	3.0	3.0	3.0	3.0	3.0	2.0	2.0
17	136.0	1.0	3.0	3.0	3.0	1.0	3.0	2.0	2.0
18	493.0	4.0	3.0	3.0	3.0	5.0	4.0	2.53	2.0
19	85.0	1.0	3.0	3.0	3.0	2.0	3.0	2.0	2.0
20	118.0	5.0	5.0	3.0	4.0	5.0	5.0	4.6	5.0
21	496.0	3.0	4.0	3.0	3.0	4.0	4.0	3.04	4.0
22	138.0	2.0	4.0	3.0	3.0	2.0	3.0	2.1	2.0
23	110.0	2.0	3.0	3.0	3.0	3.0	3.0	2.0	2.0
24	232.0	4.0	4.0	3.0	4.0	5.0	4.0	3.5	2.0
25	271.0	1.0	3.0	3.0	3.0	1.0	3.0	2.0	3.0
26	178.0	3.0	3.0	3.0	3.0	3.0	4.0	2.02	2.0
27	490.0	2.0	3.0	3.0	4.0	2.0	3.0	2.0	3.0
28	67.0	1.0	3.0	3.0	3.0	1.0	3.0	2.0	4.0
29	20.0	1.0	3.0	3.0	3.0	1.0	3.0	2.0	4.0

4) These is the Random Forest output result after classification form the basic attributes.

Viewer

Relation: Batting Testing Runs_predicted

No.	1: Player_ID Numeric	2: Consistency Numeric	3: Opposition Numeric	4: Venue Numeric	5: predictedRuns Numeric	6: **Runs** Numeric
1	274.0	3.0	4.0	3.0	2.08	5.0
2	437.0	5.0	5.0	5.0	4.81	5.0
3	298.0	5.0	5.0	5.0	4.73	2.0
4	358.0	4.0	5.0	4.0	4.12	3.0
5	510.0	4.0	4.0	4.0	3.76	4.0
6	262.0	3.0	4.0	3.0	2.18	3.0
7	393.0	2.0	3.0	2.0	2.0	2.0
8	190.0	4.0	4.0	4.0	3.01	2.0
9	124.0	3.0	3.0	3.0	2.0	2.0
10	345.0	4.0	4.0	4.0	2.97	2.0

For prediction of Wickets,

5) These is the Naïve Bayes output result obtained from the classifier from the basic attributes.

Viewer

Relation: BOWLING ATTRIBUTE 5 YEARS DATA final-weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last_predicted

No.	1: Player ID	2: No. of Innings	3: Overs	4: Average	5: Strike Rate	6: Economy Rate	7: 4w	8: prediction margin	9: predicted W	10: **W**
	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Numeric	Nominal	Nominal
1	124		4	3	5	3	4	-0.793478	4	5
2			5	4	3	2	4	0.854603	5	5
3	190		2	4	3	3	4	-0.406122	3	2
4	437		2	5	2	2	4	-0.208981	3	2
5	345		4	4	3	3	4	0.391918	4	4
6	358		4	5	2	2	4	0.608241	4	4
7			4	3	3	3	4	0.562905	4	4
8	136		1	2	5	5	4	0.998655	2	2
9			2	3	4	3	4	0.686573	2	2
10			1	3	5	5	4	0.998172	2	2
11			2	5	3	3	4	-0.380228	3	2
12	185		2	4	3	4	4	0.361973	2	2
13	363		1	2	5	5	4	0.99965	2	2
14	9		4	4	3	2	4	-0.080049	5	4
15	401		3	5	3	3	4	0.663743	3	3

6) These is the Decision Tree (j48) output result obtained from the classifier from the basic attributes.

Viewer

Relation: BOWLING ATTRIBUTE 5 YEARS DATA final-weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last_predicted

No.	1: Player ID	2: No. of Innings	3: Overs	4: Average	5: Strike Rate	6: Economy Rate	7: 4w	8: prediction margin	9: predicted W	10: **W**
	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Numeric	Nominal	Nominal
1	124		4	3	5	3	4	-0.368421	4	5
2			5	4	3	2	4	0.666667	5	5
3	190		2	4	3	3	4	0.818182	2	2
4	437		2	5	2	2	4	0.818182	2	2
5	345		4	4	3	3	4	0.368421	4	4
6	358		4	5	2	2	4	0.368421	4	4
7			4	3	3	3	4	0.368421	4	4
8	136		1	2	5	5	4	1.0	2	2
9			2	3	4	3	4	0.818182	2	2
10			1	3	5	5	4	1.0	2	2
11			2	5	3	3	4	0.818182	2	2
12	185		2	4	3	4	4	0.818182	2	2
13	363		1	2	5	5	4	1.0	2	2
14	9		4	4	3	2	4	0.368421	4	4
15	401		3	5	3	3	4	0.555556	3	3
16	493		4	5	3	2	4	-0.368421	4	5
17			1	2	5	5	4	1.0	2	2
18	178		3	3	4	3	4	-0.555556	3	2
19			2	4	3	3	4	-0.818182	2	3
20	258		4	4	3	3	4	-0.526316	4	3
21	496		1	2	5	5	4	1.0	2	2
22	20		1	2	5	5	4	1.0	2	2
23	33		3	3	4	3	4	0.555556	3	3
24			1	2	5	5	4	1.0	2	2
25	322		3	4	3	2	4	0.555556	3	3
26			1	2	5	5	4	1.0	2	2
27			4	4	3	2	4	-0.368421	4	5
28	167		1	3	5	4	4	1.0	2	2
29	82		2	4	3	3	4	0.818182	2	2

7) These is the Random Forest output result obtained from the classifier from the basic attributes.

Viewer

Relation: Bowling attribute 10 years data _predicted

No.	1: Player ID Numeric	2: No.of Innings Numeric	3: Overs Numeric	4: Average Numeric	5: Strike Rate Numeric	6: Economy Rate Numeric	7: 4w Numeric	8: predictedW Numeric	9: W Numeric
1	124.0	4.0	4.0	3.0	5.0	3.0	4.0	3.693793	5.0
2	287.0	5.0	5.0	4.0	3.0	2.0	4.0	4.11668	5.0
3	190.0	2.0	2.0	4.0	3.0	3.0	4.0	2.089828	2.0
4	437.0	5.0	2.0	5.0	2.0	2.0	4.0	2.157403	2.0
5	345.0	4.0	4.0	4.0	3.0	3.0	4.0	3.929677	4.0
6	358.0	5.0	4.0	5.0	2.0	2.0	4.0	3.601704	4.0
7	101.0	3.0	4.0	3.0	3.0	3.0	4.0	3.596788	4.0
8	136.0	1.0	1.0	2.0	5.0	5.0	4.0	2.033972	2.0
9	56.0	2.0	2.0	3.0	4.0	3.0	4.0	2.043959	2.0
10	70.0	1.0	1.0	3.0	5.0	5.0	4.0	2.036294	2.0
11	204.0	2.0	2.0	5.0	3.0	3.0	4.0	2.08892	2.0
12	185.0	2.0	2.0	4.0	3.0	4.0	4.0	2.089827	2.0
13	363.0	1.0	1.0	2.0	5.0	5.0	4.0	2.026239	2.0
14	9.0	3.0	4.0	4.0	3.0	2.0	4.0	3.829752	4.0
15	401.0	3.0	3.0	5.0	3.0	3.0	4.0	2.731263	3.0
16	493.0	4.0	4.0	5.0	3.0	2.0	4.0	3.468862	5.0
17	86.0	1.0	1.0	2.0	5.0	5.0	4.0	2.027712	2.0
18	178.0	3.0	3.0	3.0	4.0	3.0	4.0	3.006092	2.0
19	398.0	2.0	2.0	4.0	3.0	3.0	4.0	2.097582	3.0
20	258.0	3.0	4.0	4.0	3.0	3.0	4.0	3.625865	3.0
21	496.0	2.0	1.0	2.0	5.0	5.0	4.0	2.01756	2.0
22	20.0	1.0	1.0	2.0	5.0	5.0	4.0	2.035324	2.0
23	33.0	3.0	3.0	3.0	4.0	3.0	4.0	2.890823	3.0
24	85.0	1.0	1.0	2.0	5.0	5.0	4.0	2.027712	2.0
25	322.0	3.0	3.0	4.0	3.0	2.0	4.0	2.834182	3.0
26	27.0	2.0	1.0	2.0	5.0	5.0	4.0	2.027712	2.0
27	416.0	4.0	4.0	4.0	3.0	2.0	4.0	3.70725	5.0
28	167.0	1.0	1.0	3.0	5.0	4.0	4.0	2.040648	2.0
29	82.0	2.0	2.0	4.0	3.0	3.0	4.0	2.093458	2.0

8) These is the SVM output result obtained from the classifier from the basic attributes.

Viewer

Relation: BOWLING ATTRIBUTE 5 YEARS DATA final-weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last_predicted

No.	1: Player ID Nominal	2: No. of Innings Nominal	3: Overs Nominal	4: Average Nominal	5: Strike Rate Nominal	6: Economy Rate Nominal	7: 4w Nominal	8: prediction margin Numeric	9: predicted W Nominal	10: W Nominal
1	124		4	3	5	3	4	-0.5	4	5
2			5	4	3	2	4	0.333333	5	5
3	190		2	4	3	3	4	0.166667	2	2
4	437		2	5	2	2	4	0.166667	2	2
5	345		4	4	3	3	4	0.166667	4	4
6	358		4	5	2	2	4	-0.166667	3	4
7			4	3	3	3	4	0.166667	4	4
8	136		1	2	5	5	4	0.166667	2	2
9			2	3	4	3	4	0.333333	2	2
10			1	3	5	5	4	0.166667	2	2
11			2	5	3	3	4	0.166667	2	2
12	185		2	4	3	4	4	0.166667	2	2
13	363		1	2	5	5	4	0.166667	2	2
14	9		4	4	3	2	4	0.166667	4	4
15	401		3	5	3	3	4	0.166667	3	3
16	493		4	5	3	2	4	-0.166667	4	5
17			1	2	5	5	4	0.166667	2	2
18	178		3	3	4	3	4	0.0	2	2
19			2	4	3	3	4	-0.166667	2	3
20	258		4	4	3	3	4	0.166667	3	3
21	496		1	2	5	5	4	0.166667	2	2
22	20		1	2	5	5	4	0.166667	2	2
23	33		3	3	4	3	4	0.0	2	3
24			1	2	5	5	4	0.166667	2	2
25	322		3	4	3	2	4	0.166667	3	3
26			1	2	5	5	4	0.166667	2	2
27			4	4	3	2	4	-0.166667	4	5
28	167		1	3	5	4	4	0.166667	2	2
29	82		2	4	3	3	4	0.166667	2	2

For Prediction of Runs,

9) These is the Naïve Bayes result obtained from the classifier from the derived attributes.

Viewer

Relation: Batting Training Runs-weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last_predicte

No.	1: Player_ID Nominal	2: Consistency Nominal	3: Opposition Nominal	4: Venue Nominal	5: prediction margin Numeric	6: predicted Runs Nominal	7: Runs Nominal
1	274	3	4	3	-0.655819	2	5
2	437	5	5	5	0.810885	5	5
3	298	5	5	5	-0.905223	5	2
4	358	4	5	4	-0.518197	4	3
5	510	4	4	4	-0.42658	3	4
6	262	3	4	3	-0.320555	2	3
7	393	2	3		0.997374	2	2
8	190	4	4	4	-0.819438	3	2
9	124	3	3	3	0.998617	2	2
10	345	4	4	4	-0.690944	3	2
11		4	4	4	-0.695999	3	2
12		2			0.981995	2	2
13		3	4	3	0.738006	2	2
14	389	2			0.984255	2	2
15		2			0.981995	2	2
16		2			0.981995	2	2
17	136	2			0.953364	2	2
18	493	4	4	4	0.662243	3	3
19		2			-0.982431	2	5
20	118	5	5	4	-0.524096	5	3
21	496	3	4	3	-0.784001	2	4
22	138	2	3		0.999043	2	2
23		2	3	3	0.999864	2	2
24	232	4	4	4	-0.690944	3	2
25		2			-0.981995	2	3
26	178	3	4	3	0.320555	2	2
27	490	2	3		0.999043	2	2
28	67	2			0.984255	2	2
29	20	2			0.984255	2	2

10) These is the Decision tree (j48) result obtained from the classifier from the derived attributes.

Viewer

Relation: Batting Training Runs-weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last_predicte

No.	1: Player_ID Nominal	2: Consistency Nominal	3: Opposition Nominal	4: Venue Nominal	5: prediction margin Numeric	6: predicted Runs Nominal	7: Runs Nominal
1	274	3	4	3	-0.764706	2	5
2	437	5	5	5	0.272727	5	5
3	298	5	5	5	-0.636364	5	2
4	358	4	5	4	-0.444444	4	3
5	510	4	4	4	-0.257143	3	4
6	262	3	4	3	-0.558824	2	3
7	393	2	3		1.0	2	2
8	190	4	4	4	-0.285714	3	2
9	124	3	3	3	1.0	2	2
10	345	4	4	4	-0.285714	3	2
11		4	4	4	-0.285714	3	2
12		2			0.774648	2	2
13		3	4	3	0.558824	2	2
14	389	2			0.774648	2	2
15		2			0.774648	2	2
16		2			0.774648	2	2
17	136	2			0.774648	2	2
18	493	4	4	4	0.257143	3	3
19		2			-0.811268	2	5
20	118	5	5	4	-0.636364	5	3
21	496	3	4	3	-0.735294	2	4
22	138	2	3		1.0	2	2
23		2	3	3	1.0	2	2
24	232	4	4	4	-0.285714	3	2
25		2			-0.774648	2	3
26	178	3	4	3	0.558824	2	2
27	490	2	3		1.0	2	2
28	67	2			0.774648	2	2
29	20	2			0.774648	2	2
30	258	3	3	3	1.0	2	2
31	363	2			0.774648	2	2
32	315	4	4	4	-0.285714	3	2

11) These is the Random Forest result obtained from the classifier from the derived attributes.

Viewer

Relation: Batting Testing Runs_predicted

No.	1: Player_ID Numeric	2: Consistency Numeric	3: Opposition Numeric	4: Venue Numeric	5: predictedRuns Numeric	6: **Runs** Numeric
1	274.0	3.0	4.0	3.0	2.08	5.0
2	437.0	5.0	5.0	5.0	4.81	5.0
3	298.0	5.0	5.0	5.0	4.73	2.0
4	358.0	4.0	5.0	4.0	4.12	3.0
5	510.0	4.0	4.0	4.0	3.76	4.0
6	262.0	3.0	4.0	3.0	2.18	3.0
7	393.0	2.0	3.0	2.0	2.0	2.0
8	190.0	4.0	4.0	4.0	3.01	2.0
9	124.0	3.0	3.0	3.0	2.0	2.0
10	345.0	4.0	4.0	4.0	2.97	2.0
11	101.0	4.0	4.0	4.0	3.88	2.0
12	70.0	2.0	2.0	2.0	2.0	2.0
13	287.0	3.0	4.0	3.0	2.05	2.0
14	389.0	2.0	2.0	2.0	2.0	2.0
15	56.0	2.0	2.0	2.0	2.0	2.0
16	204.0	2.0	2.0	2.0	2.0	2.0
17	136.0	2.0	2.0	2.0	2.0	2.0
18	493.0	4.0	4.0	4.0	4.6	3.0
19	85.0	2.0	2.0	2.0	2.0	5.0
20	118.0	5.0	5.0	4.0	3.84	3.0
21	496.0	3.0	4.0	3.0	2.55	4.0
22	138.0	2.0	3.0	2.0	2.0	2.0
23	110.0	2.0	3.0	3.0	2.0	2.0
24	232.0	4.0	4.0	4.0	3.04	2.0
25	271.0	2.0	2.0	2.0	2.0	3.0
26	178.0	3.0	4.0	3.0	2.65	2.0
27	490.0	2.0	3.0	2.0	2.0	2.0
28	67.0	2.0	2.0	2.0	2.0	2.0
29	20.0	2.0	2.0	2.0	2.0	2.0

12) These is the SVM result obtained from the classifier from the derived attributes.

Viewer

Relation: Batting Training Runs-weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last_predicte

No.	1: Player_ID Nominal	2: Consistency Nominal	3: Opposition Nominal	4: Venue Nominal	5: prediction margin Numeric	6: predicted Runs Nominal	7: **Runs** Nominal
1	274	3	4	3	-0.5	2	5
2	437	5	5	5	0.166667	5	5
3	298	5	5	5	-0.333333	5	2
4	358	4	5	4	-0.333333	4	3
5	510	4	4	4	-0.166667	3	4
6	262	3	4	3	-0.166667	2	3
7	393	2	3		0.166667	2	2
8	190	4	4	4	-0.333333	3	2
9	124	3	3	3	0.166667	2	2
10	345	4	4	4	-0.333333	3	2
11		4	4	4	-0.333333	3	2
12		2			0.166667	2	2
13		3	4	3	0.166667	2	2
14	389	2			0.166667	2	2
15		2			0.166667	2	2
16		2			0.166667	2	2
17	136	2			0.166667	2	2
18	493	4	4	4	0.166667	3	3
19		2			-0.5	2	5
20	118	5	5	4	-0.5	4	3
21	496	3	4	3	-0.5	2	4
22	138	2	3		0.166667	2	2
23		2	3	3	0.166667	2	2
24	232	4	4	4	-0.333333	3	2
25		2			-0.166667	2	3
26	178	3	4	3	0.166667	2	2
27	490	2	3		0.166667	2	2
28	67	2			0.166667	2	2
29	20	2			0.166667	2	2

For Predicting Wickets,

13) These is the Naïve Bayes output obtained from the classifier from the derived attributes.

Viewer

Relation: Bowling Training Wickets-weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last_predicted

No.	1: Player_ID	2: Consistency	3: Opposition	4: Venue	5: prediction margin	6: predicted Wickets	7: **Wickets**
	Nominal	Nominal	Nominal	Nominal	Numeric	Nominal	Nominal
1	124	3	4	4	-0.301914	4	5
2		4	4	4	0.4717	5	5
3	190	2	3	3	0.892537	2	2
4	437	3	4	4	-0.382955	4	2
5	345	4	4	4	-0.224428	5	4
6	358	4		4	-0.356635	5	4
7		3	4	3	-0.3717	3	4
8	136	2	2	3	0.548645	2	2
9		2	2	3	0.824999	2	2
10		2	3	3	0.931619	2	2
11		2	3	3	0.931619	2	2
12	185	2	3	3	0.944975	2	2
13	363	2	2	3	0.857511	2	2
14	9	3	4	3	-0.189618	3	4
15	401	3	4	3	0.188533	3	3
16	493	3	4	4	-0.005574	4	5
17		2	2	3	0.824999	2	2
18	178	3	4	3	-0.363802	3	2
19		2	3	3	-0.931619	2	3
20	258	3	4	4	-0.120768	4	3
21	496	2	2	3	0.857511	2	2
22	20	2	2	3	0.548645	2	2
23	33	3	4	3	0.188533	3	3
24		2	2	3	0.824999	2	2
25	322	2	3	3	-0.944975	2	3
26		2	2	3	0.824999	2	2
27		3	4	4	-0.014261	4	5
28	167	2	2	3	0.548645	2	2
29	82	2	3	3	0.944975	2	2

14) These is the Decision Tree(j48) output obtained from the classifier from the derived attributes.

Viewer

Relation: Bowling Training Wickets-weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last_predicted

No.	1: Player_ID	2: Consistency	3: Opposition	4: Venue	5: prediction margin	6: predicted Wickets	7: **Wickets**
	Nominal	Nominal	Nominal	Nominal	Numeric	Nominal	Nominal
1	124	3	4	4	-0.737705	2	5
2		4	4	4	-0.737705	2	5
3	190	2	3	3	0.659836	2	2
4	437	3	4	4	0.659836	2	2
5	345	4	4	4	-0.717213	2	4
6	358	4		4	-0.717213	2	4
7		3	4	3	-0.717213	2	4
8	136	2	2	3	0.659836	2	2
9		2	2	3	0.659836	2	2
10		2	3	3	0.659836	2	2
11		2	3	3	0.659836	2	2
12	185	2	3	3	0.659836	2	2
13	363	2	2	3	0.659836	2	2
14	9	3	4	3	-0.717213	2	4
15	401	3	4	3	-0.659836	2	3
16	493	3	4	4	-0.737705	2	5
17		2	2	3	0.659836	2	2
18	178	3	4	3	0.659836	2	2
19		2	3	3	-0.659836	2	3
20	258	3	4	4	-0.659836	2	3
21	496	2	2	3	0.659836	2	2
22	20	2	2	3	0.659836	2	2
23	33	3	4	3	-0.659836	2	3
24		2	2	3	0.659836	2	2
25	322	2	3	3	-0.659836	2	3
26		2	2	3	0.659836	2	2
27		3	4	4	-0.737705	2	5
28	167	2	2	3	0.659836	2	2
29	82	2	3	3	0.659836	2	2

15) These is the Random Forest output result obtained from the classifier from the derived attributes.

Viewer

Relation: Bowling Testing Wickets_predicted

No.	1: Player_ID Numeric	2: Consistency Numeric	3: Opposition Numeric	4: Venue Numeric	5: predictedWickets Numeric	6: **Wickets** Numeric
1	124.0	3.0	4.0	4.0	3.74	5.0
2	287.0	4.0	4.0	4.0	3.85	5.0
3	190.0	2.0	3.0	3.0	2.0	2.0
4	437.0	3.0	4.0	4.0	3.88	2.0
5	345.0	4.0	4.0	4.0	3.98	4.0
6	358.0	4.0	5.0	4.0	2.86	4.0
7	101.0	3.0	4.0	3.0	2.35	4.0
8	136.0	2.0	2.0	3.0	2.02	2.0
9	56.0	2.0	2.0	3.0	2.0	2.0
10	70.0	2.0	3.0	3.0	2.01	2.0
11	204.0	2.0	3.0	3.0	2.02	2.0
12	185.0	2.0	3.0	3.0	2.0	2.0
13	363.0	2.0	2.0	3.0	3.63	2.0
14	9.0	3.0	4.0	3.0	2.81	4.0
15	401.0	3.0	4.0	3.0	2.7	3.0

16) These is the SVM output result obtained from the classifier from the derived attributes.

Viewer

Relation: Bowling Training Wickets-weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last_predicted

No.	1: Player_ID Nominal	2: Consistency Nominal	3: Opposition Nominal	4: Venue Nominal	5: prediction margin Numeric	6: predicted Wickets Nominal	7: **Wickets** Nominal
1	124	3	4	4	-0.333333	4	5
2		4	4	4	-0.166667	2	5
3	190	2	3	3	0.166667	2	2
4	437	3	4	4	0.166667	2	2
5	345	4	4	4	0.0	2	4
6	358	4		4	-0.166667	2	4
7		3	4	3	-0.333333	2	4
8	136	2	2	3	-0.166667	3	2
9		2	2	3	0.166667	2	2
10		2	3	3	0.166667	2	2
11		2	3	3	0.166667	2	2
12	185	2	3	3	0.166667	2	2
13	363	2	2	3	0.166667	2	2
14	9	3	4	3	0.0	2	4
15	401	3	4	3	-0.166667	2	3
16	493	3	4	4	-0.333333	4	5
17		2	2	3	0.166667	2	2
18	178	3	4	3	0.0	2	2
19		2	3	3	-0.166667	2	3
20	258	3	4	4	-0.333333	4	3
21	496	2	2	3	0.166667	2	2
22	20	2	2	3	-0.166667	3	2
23	33	3	4	3	-0.166667	2	3
24		2	2	3	0.166667	2	2
25	322	2	3	3	-0.166667	2	3
26		2	2	3	0.166667	2	2
27		3	4	4	-0.333333	2	5
28	167	2	2	3	-0.166667	3	2
29	82	2	3	3	0.166667	2	2