A Bayesian Classification Model for Predicting the Performance of All-rounders in the Indian Premier League

Hemanta Saikia

Department of Business Administration,

Assam University, Silchar – 788011

E-mail: h.saikia456@gmail.com

and

Dibyojyoti Bhattacharjee

Department of Business Administration,

Assam University, Silchar – 788011

E-mail: dibyojyoti.bhattacharjee@gmail.com

Abstract

The game of cricket got a new dimension, when the Indian Premier League (IPL), a competition of twenty over-a-side featuring eight teams named after various Indian cities/states started in 2008. The teams were franchisee driven and the players were selected via competitive bidding from a pool of available players. All-rounders i.e. players with the ability to both bat and bowl play a noteworthy part in cricket, whatever is the version of the game. The study measures the performance of all-rounders in Indian Premier Leagues (IPL) based on their strike rate and economy rate. The all-rounders are divided into four non-overlapping class viz. performer, batting all-rounder, bowling all-rounder and under performer. Stepwise multinomial logistic regression is used to determine the significant predictors responsible for such categorization. A Naïve Bayesian classification model is developed that can use the significant predictors to forecast the class in which an incumbent all-rounder is expected to lie. The classifier is build based on the performance of all-rounders who participated in IPL-I and II, and the validity of the classifier is subsequently tested over the incumbent all-rounders of IPL-III. The classifier though moderately successful in predicting the appropriate class of the incumbent all-rounders in IPL III, is expected to perform better in future with increase in the size of training sample. This classification would be useful for the participating teams' management while deciding about which all-rounder to be bided for and to what amount in the next addition of the league.

Keywords: Naïve Bayesian classification, Multinomial logistic regression, Cricket, Performance measurement, Data mining, Decision making.

1. Introduction

Cricket in India has become the only sports that have captured the attention of the mass. The media also attended its developments regularly. Cricket in India is run by an autonomous body, Board of Control of Cricket in India (BCCI). It is well managed compared to the governing bodies of other sports which lack adequate institutional support and planning (Ramani, 2008). When India won the inaugural Twenty-20 world cup in 2007, Indians had several reasons to be happy about. First, it was her major success after the cricket world cup victory in 1983. Secondly, in the same year India performed poorly in the 50 over version of world cup cricket, being eliminated in the first round itself. The deep wounds of the fans, fresh in their memories were more than just been healed when M. S. Dhoni and his men clichéd the title.

Though the first international Twenty-20 cricket was played in 13th June, 2005, between England and Australia, yet this version of cricket was not very popular in India till the world cup win in 2007. Taking this opportunity, Zee Telefilms started the Indian Cricket League (ICL), a

Twenty-20 cricket tournament, after they failed to obtain official broadcast rights for Indian cricket tournaments. But for a cricket league to become official a league needs to be sanctioned by the national governing authority and the International Cricket Council (ICC). The ICL was declared unofficial by BCCI before its very commencement. This meant that the players participating in ICL are officially banned from participating in world cricket. However, the well attended Twenty-20 matches of ICL reflected a sense of opportunity that the Twenty-20 format of the game is a probable platform for making money from cricket.

In April 2008, BCCI initiated the Indian Premier League, a Twenty-20 cricket tournament to be played among eight domestic teams, named after eight Indian states or cities but owned by franchise. The franchise formed their teams by competitive bidding from a collection of Indian and international players and the best of Indian up-and-coming talent. Team owners bid for the services of cricketers for a total of US $42 million. Each team can purchase a maximum of eight overseas players; though, only four can be considered in a playing eleven. The franchisees bid for the salaries that they are ready to offer to the players. Each player has a base price fixed by the IPL authorities and there is no upper limit. However, the salary offer is valid for three years only. As three seasons of IPL are already complete and from the next version of IPL two other teams joins the league so the salary offer to the available cricketers are supposed to undergo substantial change. Such change should be related to the performance of cricketers in the yester seasons of IPL and in domestic and international tournaments.

Of the different dexterities required to become a cricketer, batting and bowling are undoubtedly the prime skills. Though a balanced cricket team has specialist batsmen and bowlers, yet players with reasonably good performance both with the bat and the ball are always assets to their teams. Such players termed as all-rounders, are of real utility in the team for which they play, whatever be the version of cricket. The paper evaluates the performance of the all-rounders who figured in the first two versions of IPL and the factors responsible for their performance are identified. These factors can be used to estimate the performance of an incumbent all-rounder in the coming IPL using Naïve Bayesian classifier. The classification model is developed, based on a training sample of all-rounders and their performance in IPL I and II. The external validity of the model is tested by predicting the performance of all-rounders outside the training sample who participated in IPL III. The predicted performance is than compared with the actual performance of those incumbent all-rounders. Such a model can help a franchisee to decide which all-rounder is to be bided for prior to the fourth season of IPL when fresh agreements are to be signed with the players related to their salary.

2. Review of Literature

Each game of cricket generates a huge amount of numerical information and often termed by various academicians as a game of statisticians delight. The studies of academic significance that centered on cricket can broadly be classified into two groups. The first group comprises of the studies related to physical properties, technological advances, use of advanced equipment in professional cricket etc. while the other group is related to decision making in cricket using stochastic tools. Some studies that fall under the former category are Morris (1976) who examined the performance of catching for different coloured balls against white and black coloured background, Koslow (1985) and Kingsbury *et al*. (2001) examined the slip-catching performance of professional cricketers when ball colour and luminance level differed. Campbell *et al*. (1987) investigated visual reaction times using computer based pointer movement task and attempt to generalize findings to the game of cricket. Mehta (2005) provides a scientific explanation for the various aerodynamic phenomena that effect cricket ball swing on a cricket ground. Sayers, Koumbarakis and Sobey (2005) study the effect of ‘Knock-in’ on surface hardness and impact of performance of the bat. Maman, Balasaheb and Sandhu (2008) examined the impact of visual skills training programs on batting performance in cricketers.

In the second category, mainly deals with analytic research on uncertainties, performance measurement and optimal decision making related to the game. Elderton (1945) seems to be the pioneer who used the first statistical analysis of cricket data to demonstrate some of the fundamental aspects of Statistics. Wood (1945), examined the performance of consistency in cricket and applied the geometrical distribution to model cricket scores based on results from test cricket. Clarke (1988) gives a study about optimal batting strategies using dynamic programming model. Wood (1945) defined a batting average by considering not-out scores as complete innings. Kimber and Hansford (1993) and Damodaran (2006) proposed alternative batting averages when batsman remains not-out in one-day cricket. Some other authors working on the same issue are Clark and Norman (1999), Lewis (2005), Maini and Narayanan (2007), Borooah (2007) and Lammer (2008). Norman and Clark (2004), Ovens and Bukeit (2006) applied mathematical modeling approach to optimize the batting order of a team. Another area where several analytic works has been done is the rescheduling of the target for a rain truncated match, for the team batting second. Though the works of Duckworth-Lewis (1998) and Jayadevan (2002) are popular, a number of other authors like Rogo (1995), Gurram and Narayanan (2004), O’Riley and Ovens (2004) etc. had made remarkable contribution towards this issue.

Though Twenty-20 is the latest version of cricket, but IPL has immerged as a focus of discussion and a hot-spot for research workers of various disciplines viz. Economics, Management, Decision Science, Finance, Human resource and so on. This is mainly because the players’ salaries are decided by auction which has lead to a rare opportunity of obtaining the values of the players in monetary terms. The valuation of players obtained through auction and availability of players' performance has allowed researchers to infer on different aspects of the game. Consequently, considerable amount of literature on Twenty-20 cricket has evolved considering its age and compared to the other versions of cricket. Some of such works are Lemmer (2008) who discussed the performance of cricketers in the first T–20 world cup, Vig (2008) studied the implications of having two cricket leagues in India viz. Indian Cricket League (ICL) and Indian Premier League (IPL), Ramani (2008) reported IPL as a "…distorted form of commodity and consumer excess". Staden (2009) developed a performance measure for cricketers in Twenty-20 cricket considering data from IPL-I. Several other research activities relates to players performance in IPL and their valuation in auction, some of such works are Parker, Burns and Natarajan (2008), Rastogi and Deodhar (2009) and Depken and Rajasekhar (2010) etc.

3. Data and Variables

3.1 Data

Two variables viz. strike rate of the batsman, which is the number of runs scored per 100 balls faced and economy rate of the bowler, which is the average number of runs scored per over are used as performance measures of the all-rounders. Ideally, an all-rounder is supposed to have a high strike rate and a low economy rate to be effective in the Twenty-20 format of cricket. Based on the performance of the all-rounders in the two IPLs, they can be divided into four classes' viz. performer, batting all-rounder, bowling all-rounder and under performer. The four classes can be as:

(i) An all-rounder with strike rate above average and economy rate below average is a performer which is the ideal situation.

(ii) An all-rounder with strike rate above average and economy rate above average is a batting all-rounder.

(iii) An all-rounder with strike rate below average and economy rate below average is a bowling all-rounder.

(iv) An all-rounder with strike rate below average and economy rate above average is termed as under–performer, which is the worst case.

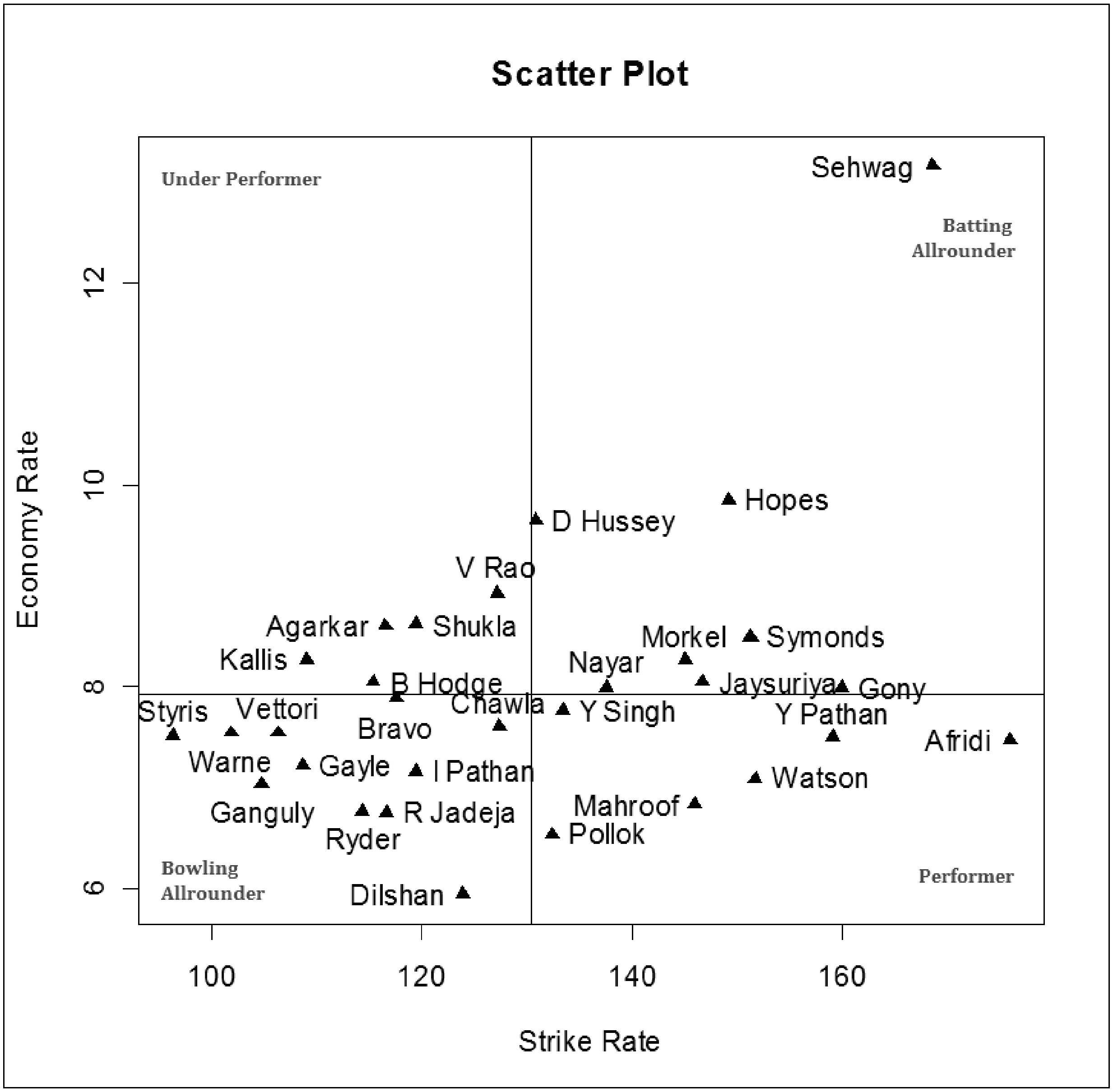
Thus, the dependent variable is categorical in nature i.e. the type of all-rounder. The training sample (Appendix–A) considered here is populated with the all-rounders, who had performed in the two IPLs. However, players who satisfied all the following conditions were spaced in the training sample:

- The player was in the playing eleven in atleast 5 matches of IPL.
- The player has bowled atleast 10 overs in IPL.
- The player has faced atleast 100 balls in the IPL.

As discussed earlier, an all-rounder will fall in any one of the categories viz. performer, batting all-rounder, bowling all-rounder and under performer based on his strike rate and economy rate. The classification is shown in the following graphical depiction.

The training sample initially consisted of 30 all-rounders, but in case of some of the all-rounders the values of some independent variables were not available. For example, Y. Venugopal Rao has not bowled in any ODI and hence his economy rate in ODI was not available. Manpreet Singh Gony has not batted in any ODI and so his Strike rate in ODI is not available, at the time when the paper was prepared. Ultimately the training sample size gets restricted to 27. The information about the all-rounders, for the variables considered earlier was collected from the websites *www.cricinfo.org* and *www.cricketnirvana.com.*

Figure 1: Graphical display of categorizing all-rounders of the training sample



Data Source: The graph is drawn in R with data from Appendix-B.

3.2 Variables

Several independent variables that are supposed to affect the said classification were considered. The information about the variables were collected on 10th March, 2010 from website *www.cricinfo.org* prior to the start of IPL-III. Some of the variables are nominal (e.g. Batting hand either left or right); some are discrete (e.g. IPL team, country etc.) while others were continuous (e.g. Age, strike rate in Twenty-20 etc.) in nature. A brief description of the different independent variables are as follows and their values for different cricketers can be seen in Appendix-B.

Age: Demographic variable measuring the age of the player. It is the number of years completed at the end of IPL 2.

Batting hand: Nominal variable i.e. either left or right. It is the dominating hand of the player while batting. The binary code '0' represents left hand batsmen and '1' represents right hand batsmen.

Bowling hand: Nominal variable i.e. either left or right. It is the throwing arm of the bowler when he comes to bowl. The binary code '0' represents left hand bowler and '1' represents right hand bowler.

Type of bowler: The type of bowling is characterized into two groups either fast or spin. Medium fast bowlers are considered under the fast bowler category. The code '1' used for spin bowler and '2' for fast bowler.

ODI matches played: Measures the experience of the player in terms of number of international one day matches in which he was in the playing eleven.

Twenty-20 matches played: Measures the experience of the player in T20 cricket. It measures the number of international T20 matches in which he was in the playing eleven.

International cricket experience: This is also a measure of experience. It is measured in years and counts the international career of the player in terms of years.

Bid amount in IPL: It is the amount of money in Dollars for which a given player was auctioned in IPL.

IPL team: The team for which the cricketer has played in IPL. There are a few players who changed their team in IPL 2, but the training sample does not contain any of those players.

Country: The country to which the player belongs.

Average batting position in IPL: IPL matches have seen huge changes in the batting order and hence the average batting positions of a player for both the IPLs taken together were considered.

Strike rate in ODI: Career Strike rate of the player in one day international matches.

Economy rate in ODI: Career Economy rate of the player in one day international matches.

Strike rate in Twenty-20: Career Strike rate of the player in international T20 matches. In case the player has not batted in any T20 international matches his strike rate at various other domestic T20 matches were considered.

Economy rate in Twenty-20: Career Economy rate of the player in one day international matches. In case the player has not bowled in any T20 international matches his economy rate at various other domestic T20 matches were considered.

4. Analysis and Results

This study attempt to identify the various factors that are associated with the grouping of the cricketers into the four classes of all-rounders discussed earlier. To determine the variables that are influential in shaping the class to which an all-rounder may belong, stepwise multinomial logistic regression is used. Multinomial logistic regression is used whenever the dependent variable is to be classified into more than two groups, as the case here. In stepwise multinomial logistic regression, the way in which independent variables are considered in the model combines both forward and backward procedures. Due to the existence of intercorrelations between variables, the variance explained by certain variables may change as and when new variable or variables are entered in the equation. If this takes place then stepwise method would remove the weakened variables.

Here, the dependent variable is the type of all-rounder. As discussed the incumbent all-rounder will fall in any one of the class of all-rounder. The players are classified as, Performer ($Y=1$), Batting all-rounder ($Y=2$), Bowling all-rounder ($Y=3$) and under performer ($Y=4$). Also let X_{i1}, be the value of the first independent variable when Y takes the value i, obviously $i=1, 2, 3$ and 4. Similarly, we can have $X_{i2},\ldots, X_{ik}$ as the other independent variables. Thus, we can have the multinomial logistic regression equation as,

$$Prob(Y = i) = \frac{\exp(a + b_1 X_{i1} + b_2 X_{i2} + \ldots + b_k X_{ik})}{1 + \exp(a + b_1 X_{i1} + b_2 X_{i2} + \ldots + b_k X_{ik})} \text{ for } i = 1, 2, 3 \text{ and } 4 \qquad (4.1)$$

Where a, b_1, b_2, …. are parameters of the model to be estimated from data (Appendix-A).

Table 1: Information about the fitted model

Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	71.129			
Final	10.849	60.280	15	.000

Nagelkerke's R-Square = 0.962

Considering data from the training sample the model is fitted. The information about the fitted model provided in Table 1 above shows that the independent variables explain the said classification in a significant manner. It is important to measure the amount of uncertainty that is been explained by the model. One such measure is $\rho^2 = 1 - \frac{LL_F}{LL_0}$, where LL_F is the log–likelihood function of the full model and LL_0 is the log–likelihood function of the intercept only model. A model that adds nothing to the intercept only model produces $\rho^2 = 0$ and a model with complete explanatory power has $\rho^2 = 1$. The above table indicates, how well the model fits the data where a smaller -2loglikelihood value implies that the model fits the data better. In addition to this the Nagelkerke's value of Pseudo R-square is 0.962 which implies that 96.2% variation in the implying that data provides a good fit to multinomial logistic regression analysis.

Table 2: Results of Likelihood Ratio Tests to identify the significant variables

Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	10.849	.000	0	.
Strike rate in ODI	24.126	13.277	3	.004
Economy rate in Twenty-20	26.264	15.415	3	.001
Strike rate in Twenty-20	40.367	29.518	3	.000
Economy rate in ODI	37.251	26.401	3	.000
Bowling type	32.793	21.944	3	.000

The likelihood ratio-tests identified those variables which are found to be significant in stepwise multinomial logistic regression analysis. From the above table, it has been seen that the *p*-value corresponding to Strike rate in ODI cricket, Economy rate in ODI cricket, Strike rate in Twenty-20 cricket, Economy rate in Twenty-20 cricket and Bowling type are significant in determining the class to which an incumbent all-rounder is supposed to plummet. The calculations were done using SPSS 17.0.

4.2 Naïve Bayesian Classification

Bayesian classifiers are probabilistic classifiers which are based on the Bayes theorem. It can predict class membership probabilities such as the probability that a given subject belongs to a particular class. The naïve Bayes classifier assumes that the effect of a variable value on a given class is independent of the values of the other variable. This assumption is called as class conditional independence (Flach and Lachiche, 2004). It ignores the interaction between variables within individuals of the same class. An important advantage of the naïve Bayes classifier is that it requires a small size of training data to estimate the parameters necessary for classification. The naïve Bayesian classification works in the following way:

(i) Each data sample is considered as an n-dimensional feature vector, $\boldsymbol{X} = (X_1, X_2, ..., X_n)$. This actually is the values of the n variables measured from the same sample member (subject) having n attributes $A_1, A_2, ..., A_n$ respectively.

(ii) Suppose there are m classes, $C_1, C_2, ..., C_m$ to which the subjects are to be classified. Given an unknown sample vector $\boldsymbol{X}$, the classifier will predict to which class $\boldsymbol{X}$ belongs using the highest posterior probability, conditioned by $\boldsymbol{X}$. Thus, in this classifier an unknown sample vector $\boldsymbol{X}$ is classified to the class C_i if and only if,

$$P(C_i|\boldsymbol{X}) > P(C_j|\boldsymbol{X}) \text{ for } 1 \leq j \leq m \text{ and } j \neq i \qquad (4.2)$$

The expression $P(C_i|\boldsymbol{X})$ is called as the posterior probability of the class C_i given the sample vector $\boldsymbol{X}$. Thus, one has to find C_i that maximizes the posterior probability $P(C_i|\boldsymbol{X})$. From Bayes Theorem,

$$P(C_i \mid \mathbf{X}) = \frac{P(\mathbf{X} \mid C_i)P(C_i)}{\sum P(\mathbf{X} \mid C_i)P(C_i)} \qquad (4.3)$$

(iii) As the term $\sum P(\mathbf{X} \mid C_i)P(C_i)$ is constant irrespective of the value of i, so maximizing of $P(C_i|\boldsymbol{X})$ is equivalent to maximization of only $P(\mathbf{X} \mid C_i)P(C_i)$. $P(C_i)$ is called as the class prior probability of the i^{th} class. The class prior probability would be estimated by

$$P(C_i) = \frac{s_i}{s} \qquad (4.4)$$

where s_i is the number of training samples of class C_i, and s is the total number of training samples.

(iv) The attributes are assumed to be independent of each other. However, the criterion of independence does not seem to be abnormal for the problem under consideration. In such a case,

$$P(\mathbf{X} \mid C_i)P(C_i) = \prod_{k=1}^{n} P(X_k \mid C_i)P(C_i) \tag{4.5}$$

Now, the probabilities $P(X_k|C_i)$ $(k = 1, 2,\ldots,n)$ can be estimated from the training sample in the following manner:

(a) If the k^{th} attribute A_k, is categorical, then $P(X_k|C_i) = \frac{s_{ik}}{s_i}$, where s_{ik} in the number of training sample points of class C_i for which the k^{th} attribute attains the value X_k.

(b) However, if the k^{th} attribute A_k, is continuous, then $P(X_k|C_i) \sim N(\mu_{c_i}, \sigma_{c_i})$, so that

$$p(x_k \mid c_i) = \frac{1}{\sqrt{2\pi}\sigma_{c_i}} e^{-\frac{(x_k - \mu_{c_i})^2}{2\sigma_{c_i}^2}} \tag{4.6}$$

where μ_{c_i}, σ_{c_i} are the mean and standard deviation of the values of the attribute A_k belonging to the i^{th} class. However before setting up the distribution it is obligatory to test such a claim.

(v) Once the probability functions are determined, the values of $P(\mathbf{X} \mid C_i)P(C_i)$ be determined for each value of i. Sample $\boldsymbol{X}$ is then assigned a class C_i if $P(\mathbf{X} \mid C_i)P(C_i)$ is maximum.

The success of this classifier depends on the accuracy of the assumptions i.e. independence of the attributes and the normality assumption of the attributes that are continuous in nature.

4.3 Computation of Prior and Posterior Probabilities

The calculations are based on the values of the significant independent variables viz. Strike rate of ODI cricket (X_1), Strike rate of Twenty-20 cricket (X_2), Economy rate of ODI cricket (X_3), Economy rate of Twenty-20 cricket (X_4) and Bowling type (X_5) either spin or fast, obtained from Appendix-A. The paper computes the prior probabilities and distributions of the significant variables under different classes as discussed in (iii) and (iv) of the above section. The results of such probabilities are shown in the Appendix-D and Appendix-E.

However, before considering the continuous variables to be normally distributed it is essential to check this normality assumption through appropriate normality test. Here the Kolmogorov-Smirnov test is used to check the normality assumption of the variables with parameters estimated from data (See Appendix-C). Based on the probabilities in Appendix-D and Appendix-E, and using (4.5), (4.3) one can determine the class (C_i) which has the maximum posterior probability for a given sample vector $\boldsymbol{X}$. Simply speaking, with the knowledge of the The model would help to classify several such subjects into appropriate class viz. performer, batting all-rounder, bowling all-rounder and under performer.

5. External Validation of the Model

To test the external validity of the model a fresh sample of all-rounders are considered. The considered players did not participated in the two previous versions of IPL but have played in the third season only. To get sufficient information about their performance in IPL- III, the players who has played at least 5 innings and bowled at least 10 overs are considered. Six all-rounders were found to satisfy the above criteria they are Pollard, McLarean, Mathews, Voges, Bond and White. Information about the players for the significant variables viz. Strike rate in ODI cricket, Economy rate in ODI cricket, Strike rate in Twenty-20 cricket, Economy rate in Twenty-20 cricket and Type of bowling were collected from www.crickinfo.com (See Appendix-F). For each all-rounder, the values of the independent variables are replaced in (4.5) and using the appropriate distribution under each class, the posterior probability that the player would plummet in a given class is computed using (4.3). This is repeated for each of the classes for the same player. The class with maximum probability would be the expected one for that player. Also, the actual performances of the six all-rounders in IPL III are recorded and their actual category was determined. The expected results from the classification model and the actual class based on performance from IPL-III are provided in Table 3. The model can predict four of the six cases correctly.

Thus, a franchisee depending on the team's requirement can decide which players to bid for. If the team requires a bowling all-rounder Bond and Mathews are the two options available. Team requiring a batting all-rounder can prefer to auction in favour of White. Again, if the team requires a performer then Pollard is the best choice and definitely all the franchise would prefer not to bid for McLarean, as per the findings of the aforesaid model.

Table 3: Estimated class of six incumbent all-rounders in Indian Premier League

Sl. No	Players Name	Class of all-rounders	Posterior Probability	Expected Class	Performance in IPL-III
1	KA Pollard*	Performer	**0.3341**	Performer	Performer
		Batting all–rounder	0.2691		
		Bowling all–rounder	0.1648		
		Under–performer	0.2317		
2	R McLarean*	Performer	0.0002	Under Performer	Under Performer
		Batting all–rounder	3.62×10^{-16}		
		Bowling all–rounder	0.0056		
		Under–performer	**0.9941**		
3	A Mathews	Performer	0.1239	Bowling All-rounder	Under Performer
		Batting all–rounder	0.0010		
		Bowling all–rounder	**0.4467**		
		Under–performer	0.4283		
4	A Voges	Performer	6.97×10^{-8}	Batting All-rounder	Under Performer
		Batting all–rounder	**0.9972**		
		Bowling all–rounder	1.96×10^{-7}		
		Under–performer	6.06×10^{-25}		
5	S. Bond*	Performer	0.0016	Bowling All-rounder	Bowling All-rounder
		Batting all–rounder	0.00001		
		Bowling all–rounder	**0.9982**		
		under–performer	0.00007		
6	CL White*	Performer	0.1238	Batting All-rounder	Batting All-rounder
		Batting all–rounder	**0.4608**		
		Bowling all–rounder	0.4153		
		Under–performer	3.48×10^{-17}		

5. Conclusion

IPL is a young professional league, yet it has planned its player distribution procedure in the same lines as of the other professional sports leagues around the world. The salaries of players that are decided through auction are a way of quantifying players' performance in monetary terms. Thus, it is a matter of decision making from the part of the franchise to decide about

* The predicted and actual performance/categories of these players are similar by applying naive Bayesian model.

which player to be bided for and at what cost. Therefore, such a model can help a franchisee to take a decision. The salary offered to the players is valid for three years. At the beginning of the fourth season of IPL, new agreements are to be made and fresh bidding is probable. In such a context such studies may be used to predict the probable performance of all-rounders and the franchisee might decide which all-rounders they should target for their team and who are not to be considered.

The model when applied for external validity was found to be only 66.7 percent accurate. This has mainly resulted as the training sample used to develop the model was loaded with only twenty-seven all-rounders. As the age of the league increases, more number of all-rounders can be considered in the training sample and much better prediction are expected from the model. However, the form of the player when the player actually plays the tournament is a deterministic factor for the actual classification. But the model, once rich in data, is supposed to work well provided the performance of the player is not much variant to what he had been performing in the yester matches. The IPL appears likely to be around for many years, providing opportunities for further research (Parker *et. al*, 2008).

References

Campbell, F W; Rothwell, S E and Perry, M J (1987). "Bad light stops play," *Ophthalmic and Physiological Optics*, 7, 165 – 167.

Clarke, S R (1988). "Dynamic programming in one-day cricket- optimal scoring rates," *Journal of the Operational Research Society*, 50, 536 – 545.

Duckworth, F C and Lewis, A J (1998). "A fair method of resetting the target in interrupted one-day cricket matches," *Journal of Operational Research Society*, 49, 220 – 227.

Damodaran, U (2006). "Stochastic Dominance and Analysis of ODI Batting Performance: The

Indian Cricket Team, 1989-2005," *Journal of Sports Science and Medicine*, 5, 503 – 508.

Depken, C. A. and Rajasekhar, R., (2010) Open Market Valuation of Player Performance in Cricket: Evidence from the Indian Premier League (April 1, 2010). Available at SSRN: http://ssrn.com/abstract= 1593196

Elderton, W P (1945). "Cricket scores and some skew correlation distribution," *Journal of Royal Statistical Society* (Series A), 108, 1 – 11.

Flack, P A and Lachiche, N (2004). "Naïve Bayesian classification of structured data," *Machine Learning*, Boson: *Kluwer Academic Publishers*, 1 – 37.

Gurram, M and Narayanan, S (2004). "Comparison of the methods to reset targets for interrupted One-Day Cricket Matches," *Proceedings of 7th Australasian Conference on Mathematics and Computers in Sport*, 165 – 171.

Jayadevan, V (2002). "A new method for the computation of targets scores in interrupted, limited over cricket matches," *Current Science*, 83, 577 – 586.

Koslow, R E (1985). "Peripheral reaction time and depth perception as related to ball colour," *Journal of Human Movement Studies*, 11, 125 – 143.

Kimber, A C and Hansford, A R (1993). "A Statistical Analysis of Batting in Cricket," *Journal of Royal Statistical Society*, 156, 443 – 455.

Kingsbury, D; Scott, K; Bennett, S; Davids, K and Langley, M (2000). "Effects of cricket ball colour and illuminance levels on catching behaviour in professional cricketers," *Ergonomics*, 10, 1681-1688.

Lemmer, H (2008). "An analysis of players' performances in the first cricket Twenty20 world cup series," *South African Journal for Research in Sport*, 30, 71 -77.

Morris, D G S (1976). "Effects of ball and background colours have upon the catching performance of elementary school children," *The Research Quarterly*, 47, 409 – 416.

Mamman, P; Balashaheb, T and Sandhu, J S (2008). "The impact of visual skills training program on batting performance in cricketers," *Serbian Journal of Sports Sciences*, 2, 17 – 23.

Mehta, R D (2005). "An overview of cricket ball swing," *Sports Engineering*, 8, 181– 192.

Norman, J and Clarke, S R (2004). "Dynamic programming in cricket: Batting on sticky wicket," *Proceedings of the 7th Australasian Conference on Mathematics and Computers in Sport*, 226 – 232.

Ovens, M and Bukeit, B (2006). "A mathematical modeling approach to one day cricket batting orders," *Journal of Sports Science and Medicine*, 5, 495-502.

O'Riley, B and Ovens, M (2006). "Impress your friends and predict final score: An analysis of the psychic ability of four target resetting methods used in One-dayInternational cricket," *Journal of Sports Science and Medicine*, 5, 488 – 494.

Parker, D., Burns, P and Natarajan, H., (2008) "Player valuations in the Indian Premier League," *Frontier Economics,* October 2008, 1-17.

Ramani, S. (2008) "Cricket, Excesses and Market Mania," *Economic and Political Weekly*, March 8, 13-15.

Rastogi, S K and Deodhar, S Y (2009). "Player pricing and valuation of cricketing attributes: Exploring the IPL Twenty-20 vision," *Vikalpa*, 34 (2), 15 – 23.

Sayers, A T; Koumbarakis, M and Sobey, S (2005). "Surface hardness of cricket bat after 'Knock-In'," *Sports Engineering*, 8, 233 – 240.

Wood, G H (1945). "Cricket scores and geometric progression," *Journal of Royal Statistical Society*, Series A (Statistics in Society) 108, 12 – 22.

Vig, A (2008). “Efficiency of sports league – The economic implications of having two in the Indian cricket market,” Masters Project: The University of Nottingham.

Staden, P J (2009). “Comparison of cricketers’ bowling and batting performances using graphical displays,” *Current Science*, 96, 764-766.

Appendix-A

Players name	Age	L/R bat (1-R, 0-L)	Bowl type (1–spin, 2–fast)	No. of TWENTY-20 matches played	No of ODI matches played	SR One Day	SR in TWENTY-20 Cricket	Econ Rate in ODI	Econ Rate TWENTY-20	Years of International Cricket	Country	Avg. Batting Position in IPL	L/R bowl (1-R, 0-L)	Bidding Amount
S.B. Styris	33	1	2	30	157	78.89	109.34	4.75	7.57	10	NZ	4.85	1	175000
D.J. Bravo	25	1	2	35	99	81.19	125.4	5.3	8.35	5	WI	4.24	1	100000
M.F. Mahroof	24	1	2	20	91	85.5	133.07	4.75	6.96	5	SL	7.13	1	225000
S. Afridi	29	1	1	34	282	110.78	151.66	4.63	6.51	13	PAK	4.22	1	675000
J. Kallis	33	1	2	35	292	71.92	113.02	4.84	8.31	14	SA	2.44	1	900000
J. Hopes	31	1	2	11	61	90.12	107.14	7.64	4.34	4	AUS	2.00	1	300000
V. Rao	27	1	1	32*	16	60.05	120.14	_____	8.39	1	IND	5.62	1	30000
S. C. Ganguly	36	0	2	44*	311	73.71	105.9	5.06	7.58	13	IND	2.87	1	1092500
D.Hussey	31	1	1	11	23	88.46	140.47	5.36	6.36	1	AUS	4.34	1	625000
L.R. Shukla	27	1	2	29*	3	94.73	123.59	4.94	8.11	1	IND	5.71	1	30000
J.A. Morkel	24	0	2	47	41	100.39	141.42	5.32	7.94	5	SA	5.90	1	675000
S. Pollock	36	1	2	12	303	86.69	122.85	3.67	7.62	13	SA	5.87	1	550000
S.T. Jaysuriya	39	0	1	47	439	91.28	145.41	4.76	7.8	20	SL	1.63	0	975000
V. Sehwag	30	1	1	37	205	101.85	161.52	5.28	13.7	10	IND	2.25	1	833750
Y.K. Pathan	26	1	1	9	29	100.8	171.83	5.76	7.95	1	IND	3.97	1	475000

S. Watson	28	1	2	22	85	80.91	145.31	4.82	7.36	8	AUS	3.80	1	125000
S. Warne	39	1	1	31	194	72.04	100.52	4.25	7.4	15	AUS	7.55	1	450000
Yuvraj Singh	27	0	1	15	239	89.46	160.85	5.16	9.3	9	IND	4.00	0	1063750
P.chawla	20	0	1	34	21	65.11	121.27	4.96	7.33	3	IND	7.88	1	400000
I. Pathan	24	0	2	44	107	77.68	121.05	5.25	7.35	5	IND	6.12	0	925000
A. Nayar	25	0	2	28*	0	------	137.58	-----	8.09	0	IND	5.45	1	30000
R. Jadeja	20	0	1	4	3	74.44	87.5	5.73	5.66	1	IND	5.68	0	30000
A. Agarkar	31	1	2	4	191	80.62	136.36	5.07	8.09	11	IND	7.41	1	350000
M. Gony	25	1	2	23*	2	--	160	5.84	7.98	1	IND	8.38	1	30000
A. Symonds	33	1	1	26	198	92.46	158.94	5.01	8.74	10	AUS	4.13	1	1350000
J. D. Ryder	24	0	2	9	20	90.08	132.75	6.6	6.8	1	NZ	1.00	1	160000
D. Vettori	30	0	1	27	239	79.26	106.32	4.17	6.12	12	NZ	7.00	0	625000
T. Dilshan	32	1	1	42	164	82.85	128.02	4.73	6.61	10	SL	4.27	1	250000
Chris Gayle	29	0	1	19	205	82.99	139.2	5.02	7.04	9	WI	1.67	1	800000
B Hodge	34	1	1	8	25	87.5	122.07	4.63	10	4	AUS	3.22	1	30000

* Did not play international Twenty-20 matches till August 2009. The data from other Twenty-20 matches were used.

Source: *www.cricinfo.org* and *www.cricketnirvana.com.*

Appendix-B: IPL performance of the all–rounders

Serial No.	Players name	IPL Career			Country	IPL Team
		Matches	Strike rate	Economy Rate		
1	S.B. Styris	10	96.18	7.51	NZ	DC
2	D.J. Bravo	20	117.51	7.88	WI	MI
3	M.F. Maharoof	13	146	6.83	SL	DD
4	S. Afridi	10	176.09	7.47	PAK	DC
5	J. Kallis	26	108.93	8.27	SA	RCB
6	J. Hopes	11	149.32	9.86	AUS	KP
7	V. Rao	27	127.16	8.93	IND	DC
8	S. C. Ganguly	26	104.67	7.03	IND	KKR
9	D.Hussey	17	130.82	9.65	AUS	KKR
10	L.R. Shukla	22	119.32	8.62	IND	KKR
11	J.A. Morkel	25	145.06	8.26	SA	CSK
12	S. Pollock	13	132.43	6.54	SA	MI
13	S.T. Jaysuriya	26	146.71	8.05	SL	MI
14	V. Sehwag	25	168.72	13.17	IND	DD
15	Y.K. Pathan	29	159.15	7.49	IND	RR
16	S. Watson	15	151.76	7.09	AUS	RR
17	S. Warne	28	101.71	7.54	AUS	RR
18	Y. Singh	29	133.4	7.76	IND	KP
19	P. Chawla	29	127.37	7.6	IND	KP
20	I. Pathan	28	119.34	7.16	IND	KP
21	A. Nayar	27	137.59	8	IND	MI
22	R. Jadeja	27	116.53	6.75	IND	RR
23	A. Agarkar	20	116.5	8.61	IND	KKR
24	M. Gony	23	160	7.99	IND	CSK
25	A. Symonds	12	151.29	8.5	AUS	DC
26	J. D. Ryder	5	114.29	6.76	NZ	RCB
27	D. Vettori	9	106.25	7.54	NZ	DD
28	T. Dilshan	21	123.81	5.94	SL	DD
29	Chris Gayle	6	108.55	7.22	WI	KKR
30	B Hodge	15	115.29	8.05	AUS	KKR

Source: *www.cricinfo.org* and *www.cricketnirvana.com*

Appendix-C: Normality check for four continuous variables

One-Sample Kolmogorov-Smirnov Test

		Strike rate of ODI	Strike rate of TWENTY-20	Economy rate of ODI	Economy rate of TWENTY-20
Most Extreme Differences	Absolute	.069	.079	.194	.200
	Positive	.069	.074	.194	.200
	Negative	-.059	-.079	-.150	-.104
Kolmogorov-Smirnov Z		.366	.435	1.028	1.093
Asymp. Sig. (2-tailed)		.999	.992	.241	.183
The test distribution is		Normal	Normal	Normal	Normal

Appendix-D: Prior Class probabilities derived from the data in Appendix-A

Class	Notation of Class Probability	Class Probabilities
Performer	$P(C_1)$	0.222
Batting All-rounder	$P(C_2)$	0.222
Bowling All-rounder	$P(C_3)$	0.407
Under Performer	$P(C_4)$	0.148

Appendix-E: Distribution of the significant variables under different classes

Variable / Class	SR of ODI $N(\mu,\sigma^2)$	SR of TWENTY-20 $N(\mu,\sigma^2)$	ER of ODI $N(\mu,\sigma^2)$	ER of TWENTY-20 $N(\mu,\sigma^2)$	Bowl Type	
					Spin	Fast
Performer	μ = 92.256 σ = 10.24	μ = 147.595 σ = 16.371	μ = 4.798 σ = 0.627	μ = 7.61 σ = 0.881	0.5	0.5
Batting All-rounder	μ = 94.093 σ = 5.619	μ = 144.06 σ = 17.756	μ = 5.601 σ = 0.958	μ = 8.118 σ = 2.647	0.666	0.333
Bowling All-rounder	μ = 78.022 σ = 6.631	μ = 116.115 σ = 15.481	μ = 5.074 σ = 0.676	μ = 7.073 σ = 0.746	0.545	0.455
Under Performer	μ = 78.964 σ = 13.523	μ = 123.036 σ = 8.477	μ = 4.87 σ = 0.185	μ = 8.58 σ = 0.804	0.25	0.75

Appendix-F: Data related to the six incumbent all-rounders

Sl. no	Players name	SR in ODI	SR in TWENTY-20	ER in ODI	ER in TWENTY-20	Bowling Type
1	KA Pollard	90.78	122.76	5.38	8.52	Fast
2	Ryan McLaren	45.83	100	4.79	7.37	Fast
3	A Mathews	77.8	135.02	4.8	7.64	Fast
4	S Bond	70.39	100	4.24	6.8	Fast
5	Adam Voges	87.15	121.15	6.36	2.5	Spin
6	CL White	82.48	148.45	6.36	6.25	Spin

Source: *www.cricinfo.org*