Multi-Criteria Decision Tree Approach to Classify All-Rounder in Indian Premier League

Pabitra Kumar Dey[1], Dipendra Nath Ghosh[2], Abhoy Chand Mondal[3]

[1]*Assistant Professor, Department of Computer Application,*
Dr.B.C.Roy Engineering College, India
[2]*Associate Professor, Department of Computer Science& Engineering*
Dr.B.C.Roy Engineering College, India
[3]*Associate Professor, Department of Computer Science,*
University of Burdwan, India

dey_pabitra@yahoo.co.in, ghoshdipen2003@yahoo.co.in,
abhoy_mondal@yahoo.co.in

Abstract

In this fast entertainment era Twenty-20 cricket becomes one of the most popular entertaining sports in all aged persons. Multi-criteria analysis plays a vital role to measure the performance of cricketers and Decision tree technique helps us to classify in very efficient manner. This paper makes use of Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) method to produce the overall performance of the all-rounder of Indian Premier League (IPL) T-20 session-III cricket tournament. The results of TOPSIS method are further used to classify all-rounder in four different categories by using Decision tree. Finally, this paper proposed a multi-criteria decision tree approach which provides accurate & efficient data classification upon the players' performance.

Keywords: *Multi-Criteria Analysis, Decision Tree, Data Classification, Performance Measure, Twenty-20 Cricket*

1. Introduction

Multi-Criteria Analysis (MCA), a promising and important field of study was introduced in the early 70's for supporting decision makers who are faced with making numerous and conflicting evaluations for both quantitative and qualitative evaluation criteria. Several MCA techniques like Simple Additive Weighting Model (SAW), Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) [1], Analytical Hierarchy Process (AHP) [2, 3], PROMETHEE, ELECTRE, *etc.* has been widely used for data analysis was described by Muralidharan [4]. TOPSIS is one of the most classical MCA methods, was first developed by Hwang and Yoon [5], and is based on the idea that the chosen alternative should have the shortest distance from the positive ideal solution and on the other side the farthest distance of the negative ideal solution.

A decision tree is a predictive modeling technique from the fields of machine learning and statistics that builds a simple tree-like structure to classify the data according their categories [6, 7, 8]. Decision trees are commonly used in operations research, specifically in decision analysis, to help identify a strategy most likely to reach a goal. The main advantage of decision trees [9, 10] over many other classification techniques is that they produce a set of rules that are transparent and easy to understand.

Among different forms of cricket played at the international level, Twenty-20 cricket [11, 12] is one of the most entertaining and favorite game in all around the world are becoming the most popular after started Indian Premier League (IPL) in India in the year 2008 by Board for Control of Cricket in India (BCCI) [13, 14]. Initially, IPL started with 8 franchises or teams but in IPL session-V 9 teams took participated. The franchise owner formed their teams by competitive bidding from a collection of Indian and international players and the best of Indian upcoming talents.

In cricket all-rounder means the player who can play in the both role as a batsman and a bowler and in twenty-20 cricket the all-rounder plays a vital role for success of any teams. IPL T-20 cricket tournament dataset [15] of session-III has been considered to classify the all-rounder. We here select such players who batted at least 11 innings and scored minimum of 20 runs, bowled for at least 21 over and got minimum of 3 wickets in IPL session-III. Different criterion such as Batting Innings(Inns-Bat), Batting Average(Avg-Bat), Batting Strike Rate(SR-Bat), Bowling Economy Rate (Econ-Bow), Average (Avg-Bow), Strike Rate (SR-Bow) have been considered for evaluating the all-rounder performances with the help of TOPSIS and according their performances classify them into several categories with the help of decision tree. This work proposed a multi-criteria decision tree approach which provides accurate and efficient data classification in the field of MCA on the basis of players ‟ performance. Finally, all-rounder is classified in four classes such as Performer, Under-Performer, Batting and Bowling All-rounder.

The paper is organized as follows: Section 2 discuss about the Multi-Criteria Analysis using TOPSIS. Section 3 focuses about the basic concepts of Decision Tree. Experiment and results are carried out on Section 4. Finally, Section 5 concludes the paper.

2. Multi-Criteria Analysis using TOPSIS

Multi-Criteria Analysis (MCA) is a decision-making tool developed for complex problems to handle a large amount of variables and alternatives assessed in various ways and consequently offer valuable assistance to the decision maker in mapping out the problem. The first complete exposition of MCA was given in 1976 by Keeney and Raiffa [16]. By using MCA the members don't have to agree on the relative importance of the Criteria or the rankings of the alternatives. Each member enters his or her own judgments, and makes a distinct, identifiable contribution to a jointly reached conclusion. Several MCA techniques like Simple Additive Weighting Model (SAW), Technique for Order Preference by Similarity to Ideal Solution (TOPSIS), Analytical Hierarchy Process (AHP), PROMETHEE, ELECTRE, etc. has been widely used for data analysis.

TOPSIS was first developed by Hwang and Yoon [5], is based on the idea that the chosen alternative should have the shortest distance from the positive ideal solution and on the other side the farthest distance of the negative ideal solution. TOPSIS works not only by the positive solution but also use the negative solution to find the relative closeness of the criteria. Due to this among several MCA techniques TOPSIS gives the better result in multi criteria analysis. Abo-sinna and Amer [17] extend TOPSIS approach to solve multi-objective nonlinear programming problems. Jahanshahloo, *et al.*, [18] extend the concept of TOPSIS to develop a methodology for solving multi-criteria analysis with interval data.

The steps of TOPSIS method are as follows:

Step 1: TOPSIS begins with a decision matrix having n criteria/attributes and m alternatives.

Step 2: Obtain the normalized decision matrix by using any normalization technique.

Step 3: Construct the weighted normalized matrix vij. This is calculated by multiplying each column of the matrix rij by the weight wj, which is calculated by AHP.

Step 4: Obtain the IDEAL "(best)" and Negative- IDEAL "(worst)" solutions. The "IDEAL" (best) and Negative- DEAL‟ (worst)" solutions can be expressed as

$$v_j^+ = \left\{ \left(\sum_i^{\max} v_{ij} \mid j \in J \right), \left(\sum_i^{\min} v_{ij} \mid j \in J^{'} \right) \mid i = 1, 2, ..., m \right\} = \left\{ v_1^+ \quad v_2^+ \quad ... \quad v_n^+ \right\} \tag{1}$$

$$v_j^- = \left\{ \left(\sum_i^{\min} v_{ij} \mid j \in J \right), \left(\sum_i^{\max} v_{ij} \mid j \in J^{'} \right) \mid i = 1, 2, ..., m \right\} = \left\{ v_1^- \quad v_2^- \quad ... \quad v_n^- \right\} \tag{2}$$

, where J= (j =1, 2, ..., n)/j is associated with the beneficial attributes and J' =(j =1, 2, . .., n)/j is associated with the non-beneficial attributes.

Step 5: Determine the separation distance between the alternatives.

The separation of each alternative from the 'IDEAL' solution is given by

$$S_i^+ = \sqrt{\sum_{j=1}^{n} \left(v_{ij} - v_j^+ \right)^2}, \quad i = 1, 2, ..., m \tag{3}$$

The separation from the 'Negative- IDEAL' solution is denoted by

$$S_i^- = \sqrt{\sum_{j=1}^{n} \left(v_{ij} - v_j^- \right)^2}, \quad i = 1, 2, ..., m \tag{4}$$

Step 6: Calculate the relative closeness to the ideal solution, which can be expressed as

$$C_i = \frac{S_i^-}{\left(S_i^+ + S_i^- \right)}, \qquad i = 1, 2, ..., m \tag{5}$$

3. Decision Tree

A decision tree, a predictive modeling technique from the fields of machine learning and statistics that builds a simple tree-like structure to classify the data according their categories described in [6, 7, 8]. Decision tree is a tree where the root and each internal node are labeled with a question; the arcs represent each possible answer to the associated question and each leaf node represents a prediction of a solution to the problem. The main advantage of decision trees [9, 10] over many other classification techniques is that they produce a set of rules that are transparent, easy to understand and the graphical representation of tree provides more realistic view to differentiate the different categories. In my previous work [19] IPL –T-20 players‟ classification was done with two criteria.

The decision tree algorithm is as follows:

T: Input Decision tree,

D: Input Database and

M: Output Model Prediction

For each t ε D do

n = root node of T

while n not leaf node do

Obtain answer to question on n applied t;

Identify arc from t which contains correct answer;

n = node at end of this arc;

Make prediction for t based on labeling of n;

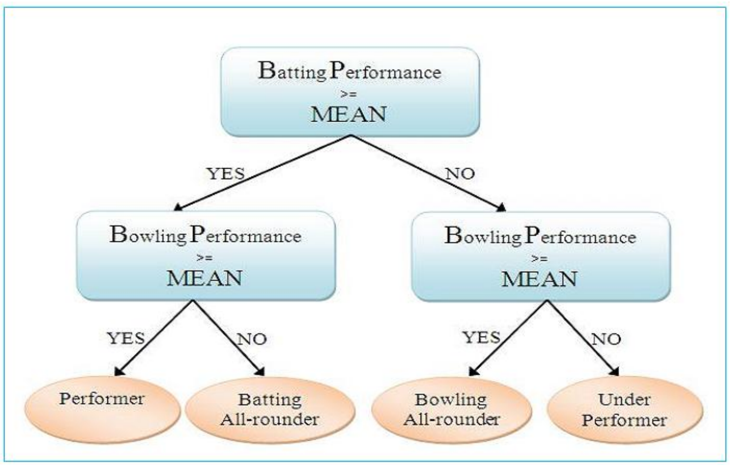

Figure 1. The prediction tree for classifying the all-rounder

4. Experiment & Result

In Twenty-20 cricket the all-rounder's (who can bat and ball both) performance plays vital role for team's success. As a batsman the batting Strike Rate, batting Average and the No. of Innings played by batsman are the most essential criteria for performance measure of a batsman although several other criteria like Runs Scored by a batsman and No. of Not-Out Innings of a batsman are there which is described in my previous work [20]. My previous work [21, 22] discussed about the bowling performance in IPL with multiple criteria like No. of Matches played by a bowler in a particular tournament, No. of Over bowled by a bowler and No. of Wicket taken by a bowler, Economy Rate of a bowler, Bowling Average and Bowling Strike Rate whereas the last three are the most essential criteria. For classification of the all-rounder IPL session-III dataset is considered and shown in the Table 1.

Table 1. IPL Session-III Dataset

Player Name	Team	Batting					Bowling				
		Inns	Runs	Avg	Balls	SR	Overs	Wkts	Avg	SR	Econ
Morkel, JA	CSK	20	338	30.7	233	145.1	88	30	24.2	17.6	8.26
Oram, JDP	CSK	11	106	15.1	105	101	31.3	8	35.3	23.6	8.95
Raina, SK	CSK	28	855	34.2	604	141.6	31	8	24.3	23.3	6.26
Sharma, RG	DC	28	766	31.9	588	130.3	27	12	15.5	13.5	6.89
Singh, RP	DC	15	35	4.38	41	85.37	111	38	22.6	17.5	7.74
Symonds, A	DC	11	410	45.6	273	150.2	30.5	7	37.3	26.4	8.46
Venugopal Rao, Y	DC	23	412	21.7	324	127.2	29	6	43.2	29	8.93
Smith, DR	DC+	11	260	23.6	172	151.2	25	6	34.5	25	8.28
Agarkar, AB	KKR	13	120	17.1	103	116.5	57	14	35.1	24.4	8.61
Ganguly, SC	KKR	24	538	23.4	514	104.7	32	8	28.1	24	7.03
Shukla, LR	KKR	18	246	18.9	207	118.8	26	9	24.9	17.3	8.62
Hopes, JR	KXIP	11	221	20.1	148	149.3	28	7	39.4	24	9.86
Pathan, IK	KXIP	22	327	20.4	274	119.3	103.2	32	23.1	19.4	7.15
Yuvraj Singh	KXIP	28	639	25.6	479	133.4	29	9	25	19.3	7.76
Bravo, DJ	MI	17	396	28.3	337	117.5	61.3	22	22.3	16.8	7.98
Harbhajan Singh	MI	13	113	11.3	70	161.4	54	17	19.9	19.1	6.26
Jayasuriya, ST	MI	26	735	30.6	501	146.7	40	11	29.3	21.8	8.05
Zaheer Khan	MI+	12	42	8.4	54	77.78	63	19	26.3	19.9	7.92
Kallis, JH	RCB	26	560	23.3	515	108.7	80.2	10	66.4	48.2	8.27
Kumar, P	RCB	20	177	11.8	144	122.9	94.2	24	31.7	23.6	8.06
Kumble, A	RCB	12	29	9.67	36	80.56	97.3	28	23.3	20.9	6.68
Vinay Kumar, R	RCB	12	72	9	66	109.1	55.2	14	32.4	23.7	8.19
Jadeja, RA	RR	22	430	23.9	369	116.5	25.3	6	28.7	25.5	6.75
Pathan, YK	RR	28	678	26.1	426	159.2	63.1	15	31.5	25.3	7.49
Warne, SK	RR	18	178	13.7	175	101.7	102	33	23.3	18.6	7.54
Watson, SR	RR	15	472	47.2	311	151.8	54.1	17	22.5	19.1	7.07

Methodology for classification of All-Rounder -

Step – 1: Taking the dataset as decision matrix and normalized the dataset.

Step – 2: The weights of the important criteria of the players are calculated with the help of pair-wise comparison method known as AHP [2, 3] and the corresponding weights of the criteria are described in the Table 2.

Table 2. Weights of the Criteria

Criteria	Weights
Inns-Bat	0.12
Avg-Bat	0.16
SR-Bat	0.22
Econ-Bow	0.20
Avg-Bow	0.17
SR-Bow	0.13

Step – 3: TOPSIS method is used separately for evaluation of batsmen and bowlers performance.

Step – 4: Calculate the mean of the both batting and bowling performance of the players.

Step – 5: With the help of our prediction Decision Tree classification of all-rounder is made into four different classes named as Batting All-rounder, Bowling All-rounder, Performer and Under Performer.

Step – 6: The players in several categories are described in the Table 3 and the corresponding graph is shown in the Figure 2.

Table 3. Players with Several Categories

Player Name	Batting				Bowling				All-rounder Status
	Inns	Avg	SR	Result	Avg	SR	Econ	Result	
Morkel, JA	20	30.73	145.06	0.7982	24.23	17.6	8.26	0.8351	Performer
Oram, JDP	11	15.14	100.95	0.3026	35.25	23.63	8.95	0.6398	Under Performer
Raina, SK	28	34.2	141.56	0.9112	24.25	23.25	6.26	0.8309	Performer
Sharma, RG	28	31.92	130.27	0.8344	15.5	13.5	6.89	1.0000	Performer
Singh, RP	15	4.38	85.37	0.0988	22.61	17.53	7.74	0.8743	Bowling
Symonds, A	11	45.56	150.18	0.9329	37.29	26.43	8.46	0.6015	Batting
Venugopal Rao, Y	23	21.68	127.16	0.5841	43.17	29	8.93	0.4948	Batting
Smith, DR	11	23.64	151.16	0.6074	34.5	25	8.28	0.6533	Batting
Agarkar, AB	13	17.14	116.5	0.3970	35.07	24.43	8.61	0.6435	Under Performer
Ganguly, SC	24	23.39	104.67	0.5717	28.13	24	7.03	0.7685	Performer
Shukla, LR	18	18.92	118.84	0.4698	24.89	17.33	8.62	0.8150	Bowling
Hopes, JR	11	20.09	149.32	0.5401	39.43	24	9.86	0.5662	Under Performer
Pathan, IK	22	20.44	119.34	0.5318	23.09	19.38	7.15	0.8730	Bowling
Yuvraj Singh	28	25.56	133.4	0.7080	25	19.33	7.76	0.8325	Performer
Bravo, DJ	17	28.29	117.51	0.6592	22.32	16.77	7.98	0.8722	Performer
Harbhajan Singh	13	11.3	161.43	0.4542	19.88	19.06	6.26	0.9272	Bowling
Jayasuriya, ST	26	30.63	146.71	0.8351	29.27	21.82	8.05	0.7517	Performer
Zaheer Khan	12	8.4	77.78	0.1046	26.26	19.89	7.92	0.8079	Bowling
Kallis, JH	26	23.33	108.74	0.5940	66.4	48.2	8.27	0.1056	Batting
Kumar, P	20	11.8	122.92	0.3702	31.67	23.58	8.06	0.7065	Under Performer
Kumble, A	12	9.67	80.56	0.1368	23.25	20.89	6.68	0.8650	Bowling
Vinay Kumar, R	12	9	109.09	0.2191	32.36	23.71	8.19	0.6936	Under Performer
Jadeja, RA	22	23.89	116.53	0.5966	28.67	25.5	6.75	0.7508	Performer
Pathan, YK	28	26.08	159.15	0.7660	31.53	25.27	7.49	0.7054	Batting
Warne, SK	18	13.69	101.71	0.3172	23.3	18.55	7.54	0.8661	Bowling
Watson, SR	15	47.2	151.77	1.0012	22.53	19.12	7.07	0.8842	Performer

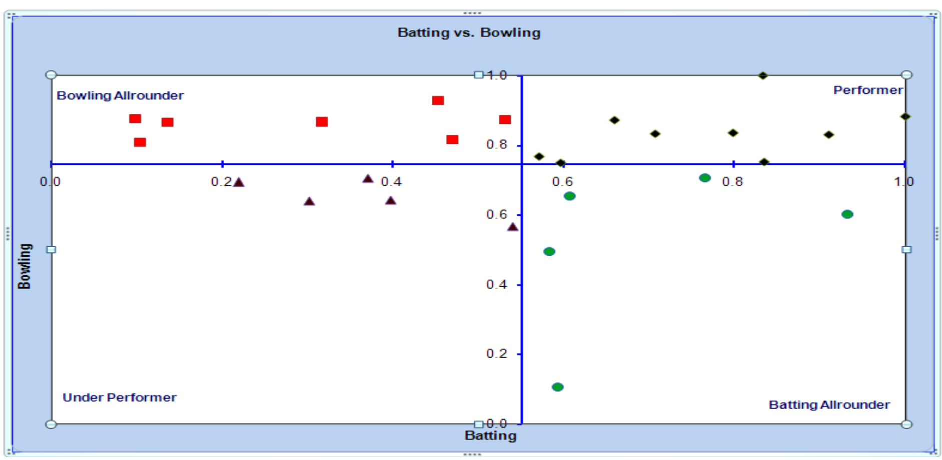

Figure 2. Graph of Several Categories of all-rounder

5. Conclusion

In IPL twenty-20 cricket tournament the franchisee owners selected their players through auction and the individual performances of the players reflected in team performances. For appropriate bidding price of each players depend on their individual performances and the franchisee owners get the actual picture and they decide their appropriate team with appropriate players and they get much more profit from IPL T-20 tournament. The role of the all-rounder in twenty-20 cricket is much more than other category players for team's better performance. TOPSIS used the relative closeness of positive ideal solution and negative ideal solution to produce the better result among several MCA techniques and use of decision tree is very easy to understand. Due to this reason our proposed methodology Multi-Criteria Decision Tree techniques gives the clear picture of the players several category like Performer, Batting All-rounder, Bowling All-rounder and Under Performer which helps the franchisee owners for selecting the all-rounder in auction and pay them according their performances. The players are also know their overall performances and also aware of their weakness of their playing where they really need the better performance and uplift themselves as a better player with this proposed. This proposed methodology helps decision makers to take their decision very easily according several criteria and the graphical representation of the classification is easy to understand and error free.

References

[1] K. Devi, S. P. Yadav and S. Kumar, "Extension of fuzzy TOPSIS method based on vague sets", International Journal of Cmputaional Cognition, vol. 7, no. 4, **(2009)**, pp. 58-62.
[2] T. L. Saaty, "The Analytic Hierarchy Process", McGraw-Hill, New York, **(1980)**.
[3] T. L. Saaty, "Priority Setting in Complex Problems", IEEE Transactions on Engineering Management, vol. 30, no. 3, **(1983)**, pp. 140-155.
[4] C. Muralidharan, "Application of Analytic Hierarchy Process (AHP) in materials and human resource management under group decision making environment", Unpublished Ph.D. Thesis, Annamalai University, **(2000)**.
[5] C. L. Hwang, K. Yoon, "Multiple Attribute Decision Making Methods and Applications", Springer, Berlin Heidelberg, **(1981)**.
[6] J. R. Quinlan, "Induction of decision trees", Machine Learning, vol. 1, **(1986)**, pp. 81–106.
[7] J. Gehrke, R. Ramakrishnan and V. Ganti, "RainForest - a Framework for Fast Decision Tree Construction of Large Datasets", Proceedings of the 24th VLDB conference, New York, USA, **(1998)**, pp. 416-427.
[8] P. Utgoff and C. Brodley, "An Incremental Method for Finding Multivariate Splits for Decision Trees", Machine Learning: Proceedings of the 7th International Conference, **(1990)**, pp. 58.
[9] J. R. Quinlan, "Programs for Machine Learning", Morgan Kaufmann, San Mateo, CA, **(1993)**.
[10] http://en.wikipedia.org/wiki/Decision_tree.
[11] Wikipedia, http://en.wikipedia.org/wiki/Twenty20.
[12] Cricinfo, http://www.cricinfo.com/.
[13] IPL Blog, http://premierleaguecricket.in/.
[14] Wikipedia, http://en.wikipedia.org/wiki/Indian_Premier_League.
[15] IPL, Indian Premier League, http://iplt20.com/.
[16] R. L. Keeney and H. Raiffa, "Decisions with Multiple Objectives: Preferences and Value Tradeoffs", John Wiley, New York, reprinted, Cambridge University Press, **(1976)**.
[17] M. A. Abo-Sinna and A. H. Amer, "Extensions of TOPSIS for multi-objective large-scale nonlinear programming problems", Applied Mathematics and Computation, vol. 162, **(2005)**, pp. 243–256.
[18] G. R. Jahanshahloo, F. Hosseinzadeh, L. fi and M. Izadikhah, "An algorithmic method to extend TOPSIS for decision-making problems with interval data", Applied Mathematics and Computation, **(2005)**.
[19] P. K. Dey, G. Chakraborty and A. C. Mondal , "Fuzzy Classification using Decision Tree Algorithms", Inc. in Proc. of 1st International Conference on Computing and Systems (ICCS- 2010), Burdwan, India, ISBN: 93-80813-01-5, **(2010)**, pp. 133-138.

[20] P. K. Dey, D. N. Ghosh and A. C. Mondal, "Statistical Based Multi-Criteria Decision Making Analysis for Performance Measurement of Batsmen in Indian Premier League", in International Journal of Advanced Research in Computer Science (IJARCS), ISSN 0976-5697, vol. 3, no. 4, **(2012)** July- August, http:// www.ijarcs.info.
[21] P. K. Dey, D. N. Ghosh and A. C. Mondal, "A MCDM Approach for Evaluating Bowlers Performance in IPL", in Journal of Emerging Trends in Computing and Information Sciences, ISSN 2079-8407, vol. 2, no. 11, **(2011)** November, pp. 563- 573, http://www.cisjournal.org.
[22] P. K. Dey and D. N. Ghosh, "Multi Criteria Fuzzy Classification of IPL3", in International Journal of Information and Computing Science (IJICS), ISSN 0972-1347, vol. 14, no. 1, (2012), pp. 43- 52.

Authors

Mr. Pabitra Kumar Dey is working as an Asst. Prof. in the Dept. of Computer Application, Dr. B.C.Roy Engineering College, Durgapur, India. He was born on 10/12/1978. He obtained B.Sc. (Math Hons.) in 2000, M.C.A. in 2004 & M.Tech.(CST) in 2011. He registered himself as a research scholar in the Dept. of Computer Science, Burdwan University. He has about more than of 8 years of Teaching Experience and 4 years of Research Experience. He has more than 15 research papers in reputed journals and conference proceedings. The broad area of his research interest is in "Soft Computing, Multi Criteria Analysis, Decision Theory, *etc.*"

Dr. Dipendra Nath Ghosh is currently Associate Professor in the department of Computer Science & Engineering, Dr. B.C.Roy Engineering College, Durgapur-713206, W.B., India. He obtained M.Sc. in Mathematics & M.C.A. from University of Burdwan and Ph.D. in Computer Science from that University in 2008. He has over 10 years of teaching experience and 07 years of research experience. He is guiding M.Tech. & Ph.D. students and has 20 research papers to his credit.

Dr. Abhoy Chand Mondal is currently Associate Professor of Department of Computer Science, Burdwan University, W.B, India. He received his B.Sc. (Mathematics Hons.) from The University of Burdwan in 1987, M.Sc. (Math) and M.C.A. from Jadavpur University, in 1989, 1992 respectively. He received his Ph.D. from Burdw an University in 2004. He has 1 year industry experience and 19 years of teaching and research experience. No. of journal paper is more than 20.