ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 9, September 2014

Performance Evaluation of feature extraction model to identify student appraisals

Mamta Singh[1], Jyoti Singh[2], and Arpana Rawal[3]

Sai Mahavidyalaya, Sector-6, Bhilai, Chhattisgarh, India[1]

Directorate of Technical Education Raipur, Chhattisgarh, India[2]

Bhilai Institute of Technology Durg, Chhattisgarh, India[3]

Abstract: Educational Data mining techniques plays an important role in educational institution. It can be used to understand the difficulties arising in the teaching-learning professions. In machine learning, feature selection or Attribute analysis usually emerges as a pre-processing step. Feature selection is the problem of choosing a small subset of feature that ideally is necessary and sufficient for predictive / decision-making type of learning tasks. This study proposes a framework for identifying the most significant attributes towards academia, for the performance of second year students of computer science and application course. The authors realize that the some features are non-changeable and so do not contribute in upgraded academic performances of the students as they do not reveal any added academic effort. In this study, authors decided to work upon only external attributes of students by assigning weights that reflect their residual efforts put in for those attributes. The model is able to extract the fitness precedence relations of external efforts put up by student belonging to both 'above-risk' and 'at-risk' categories in their on-going course. The end-user can make use of these precedence relations to identify and resolve the most unfit governing factor for upgrading students' appraisals. The accuracy of these precedence relations is computed upon the most popular feature extraction (FE) algorithm 'RELIEF 'The model accuracy of 75% indicates the encouraging results in the direction of identifying graded precedence of the participating model attributes.

Keywords: Feature Extraction, Feature Selection, External Factors, Naïve Bayesian classification. Attribute Relevance, Precedence Relations, Relief Algorithm, Nearest-Hit (Z^+), Nearest-Miss (Z^-)

I. INTRODUCTION

Feature selection or subset selection is a process commonly used in machine learning. A subset of the features available from the data-sets is selected as input parameters to the learning algorithm. Feature selection usually acts as pre-processing step to machine learning. The elimination of irrelevant and redundant information is said to improve the quality of learning and also accuracy of model. In Academic field, usually, the input variables are selected beginning from the instant the student takes admission into the institution. The student data sets then range from his / her personal data, demographic data, and academic data to behavioural attributes. Most of the prediction tasks targeted to student performances were done taking all categories of attributes and using suitable machine learning algorithms. For instance:

- The web based model as formulated by B. Minaei-Bidgoli uses features like demonstrations, simulations, and individualized problems for use on homework assignments, quizzes, and examination in order to evaluate classification on an online course dataset.[1]
- Some other works have predicted drop-out feature of students using cumulative grade point average (CGPA) which was then handed over to teachers and management for direct or indirect intervention for their academic benefit [2].
- The above kind of machine learning task resembles the automated employee recruitment system along with his / her performance prediction done by Qasim A. AI-Radaideh and Emam AI Negi [3].
- In some other prediction tasks the social attributes were also considered apart from personal, demographic, academic and behavioral attributes. One such task was to predict student's exam scores by taking into account following feature-sets: feature extraction, topological feature and friends' grade features. These attributes helped in assessing varied levels of home assignment: Online individual assignments, Coding Assignment done in pairs and lastly, theoretical writing Assignment done in a groups of two to four students[4].

The proposed work follows the recent research trends that focus on the objective of utilizing the prediction results in gearing up the student's academic potential during their on-going period of studies. The salient feature of the proposed learning model, (Naïve Bayes), is that it works well both for prediction as well as graded feature-relevance tasks. The results obtained from the proposed model are compared to the popular RELIEF Feature Extraction model.

II. METHODOLOGY SURVEY

Feature selection step is traditionally categorized as wrapper and filter methods. In wrapper methods, the performance of a learning algorithm is used to evaluate the goodness of selected feature subsets, while in filter model criterion functions evaluate feature subsets by their information content, rather than directly. Among the popular feature extraction algorithm developed till date, the filter approach encompassing FOCUS and RELIEF algorithm with variants were found to give better combination of feature subsets as compared to the algorithm used in wrapper approach[5] [6] [7]. Above all RELIEF algorithms and its variants is said to achieve high success rates when it was invaded in practical application.

Irrespective of the methodologies used above in computing attribute relevance, some of the features can be nominated as totally irrelevant by manual effort owing to the criteria that they are non-changeable, this noble thought triggered formulating new kind of feature extraction model that only takes either partially relevant or strong relevant attributes. Such attributes can be suggested as parameters that involve academic effort and are changeable through student's counselling and result-oriented guidance.

Even the latest teaching-learning strategies make use of online educational system portals that assess the student performances based on the homework assignments, quizzes and online- examinations, one such system developed was LON-CAPA (Learning Online Network with Computer-Assisted Personalized Approach [8]. The system used minute level evaluation parameters to assess to the online student responses like correct answer count, correct answers in first attempt, total answering attempts, time taken by student, time taken for arriving at correct answers, study material support or group work in problem-solving.

During their in-depth study on feature selection methods, YijunSun and Depang Wu proved that RELIEF is the most successful algorithm that solves a convex optimization problem with a margin based objective function [9]. A system was observed that the relief model couldn't filter out redundant attributes as well as weakly relevant ones; this motivated the author to provide variant logistics to the approach.

In field of medicine Ying Liu proposed a method to evaluate and compare different feature selection method in the reduction of a high dimensional feature space in drug discovery. Two classifiers Naïve Bayesian and support vector machine [10.Hence, the innovative feature extraction model of providing graded feature selection was further embedded by the group into weight-updating step in 'RELIEF'. This graded evaluation can be seen as attribute-weight precedence.

III. THE WORK PRELUDE

The proposed investigation focuses on how the student's external factors govern his / her performance in academics. Initially, the test data set consisting of 20 tuples was classified normally using one of the robust machine learning tool, Naïve Bayesian classifier. The classification was done in accordance with the past performance shown by previous batches as a part of training data set. An attempt was further made to find the relative precedence illustrating the effect of each of the external attributes on the predicted 'at-risk / above-risk' class label of the students using naïve Bayesian approach.

A. *Working Model Parameters*

In Academic field, usually, the input variables are selected beginning from the instant the student takes into the institution. The student data sets then range from his / her personal data, demographic data, and academic data to behavioral attributes.

The logistics behind selecting input parameters was the dynamic nature of the attribute domain that resulted in formalizing two broad streams of attributes: inherent (static / non-changeable / past) and external (changeable with academic effort).The data set used in this study was gathered from different sources: Student personal data (Name, gender, cast, medium, living location, food habit…), demographic data (student's family background detail like father's occupation, mother's occupation, father mother qualification, family status and family income), past academic performance (10^{th} and 12th examination scores, previous exposure to programming, background stream).All the mentioned attributes were inherent attributes and hence, were found not to reflect the change in prediction-model accuracy.

Hence, there was a need to take up some additional attributes (changeable) that were used to upgrade student's academic performance to an optimal level. These attributes selected were students' attendance, internal assessment scores, assignment credit, and subject count (number of subjects in which the student appeared in internal examination).These attributes in turn act as an external factors, which if enhanced are sure to improve end-semester results of on-going batch students.

Table 1. External Attributes used for Feature Extraction

Sl. No.	Name of the Parameter	Description	Value in Binary
x_1	Attendance	Student's attendance	+7 to –7, (4 bits)
x_2	Assignment Credit	Out of 10 Marks	4 bits
x_3	Internal Score	Performance in 3 internal tests	2 bits
x_4	Subject count	Student appear in how many internal paper out of 10	4 bits

B. *Proposed Feature Extraction Model*

The training data sets of 88 tuples from three passed-out batches of BCA course were processed to compute prior probabilities of 'at risk' students. With the criterion that 'at-risk' students are prone to obtain less than 40% of aggregate score, the prior fit and unfit probabilities from the training data sets were found to be 0.76 and 0.23 for 'at-risk' and 'above risk' students respectively. As the nature of the problem involves the appearance of four independent experimental parameters (x_1 to x_4) enumerated in table 1, it was always appropriate to compute the individual conditional probabilities and compute the effect of these on an average.

The proposed model was carefully developed with two-phase functionality. The first phase helped in arriving at

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 9, September 2014

predicted values of class variable i.e. 'at-risk' and 'above-risk' values of the test data set (students of the on-going course). This was achieved with Naïve Bayesian posterior probability computations. The expressions for the same were formulated as mentioned in 3.2.1 and 3.2.2 respectively.

$$P(fit|\{x_1,x_2,x_3,x_4\}) = \frac{\sum_{i=1}^{4} p(fit).p\left(\frac{x_i}{fit}\right)}{\sum_{i=1}^{4} p(fit).p\left(\left(\frac{x_i}{fit}\right)+\sum_{i=1}^{4} p(unfit).p\left(\frac{x_i}{unfit}\right)\right.} \quad (3.2.1)$$

$$P(unfit|\{x_1,x_2,x_3,x_4\}) = \frac{\sum_{i=1}^{4} p(unfit).p\left(\frac{x_i}{unfit}\right)}{\sum_{i=1}^{4} p(fit).p\left(\frac{x_i}{fit}\right)+\sum_{i=1}^{4} p(unfit).p\left(\frac{x_i}{unfit}\right)} \quad (3.2.2)$$

File Edit Format View Help

PASS-FAIL.txt

tuple_id	p(fit|ti)	p(unfit|ti)	pass/fail
2	0.21	0.79	0
3	0.61	0.39	1
4	0.60	0.40	1
5	0.83	0.17	1
6	0.91	0.09	1
7	0.23	0.77	0
8	0.86	0.14	1
9	0.65	0.35	1
10	0.72	0.28	1
11	0.93	0.07	1
12	0.94	0.06	1
13	0.86	0.14	1
14	0.87	0.13	1
15	0.85	0.15	1
16	0.63	0.37	1
17	0.36	0.64	0
18	0.76	0.24	1
19	0.78	0.22	1
20	0.34	0.66	0
21	0.82	0.18	1

Fig.1. 'At-Risk' / 'Above-Risk' grades assigned to test-data (proposed FE model)

The higher of these posterior probabilities computed for each test tuple, pertaining to the current 2nd year batch: ($\boldsymbol{P(unfit\,|\{x_1,x_2,x_3,x_4\})}$) and ($\boldsymbol{P(fit\,|\{x_1,x_2,x_3,x_4\})}$) helped in deciding the predicted risk value of each test-instance as illustrated in figure 1. It may be noted that these predicted values further contribute in defining Z+ and Z- components of the weight update expression discussed in extended experiment using RELIEF algorithm.

The second phase was followed by relative fitness evaluation step among the participating attributes. The characteristic feature that the individual conditional probabilities upon each attribute, x1, x2, x3 and x4 together contribute for the classification task, their relative comparisons could be used to compare the degree of involvement in affecting the risk-category of the students.
In order to generate the precedence order of these external attributes, once again the components of the above formulae were revisited used for computing average fitness ($\boldsymbol{average\_fit(xi,tj)}$) and average unfitness ($\boldsymbol{average\_unfit(xi,tj)}$) of the students owing to each unfitness ($\boldsymbol{average\_unfit(xi,tj)}$) of the students owing to each attribute. Expression pairs for one such attribute x1 are shown in the equations 3.2.3 and 3.2.4 below.

$$average_{fit(x_i,t_j)} = \frac{\sum_{i=1}^{4} p(fit).p\left(\frac{x_i}{fit}\right)}{\sum_{i=1}^{4} p.(fit)p\left(\frac{x_i}{fit}\right)+\sum_{i=1}^{4} p(unfit).p\left(\frac{x_i}{unfit}\right)} \quad (3.2.3)$$

$$average_{unfit(x_i,t_j)} = \frac{\sum_{i=1}^{4} p(unfit).p\left(\frac{x_i}{unfit}\right)}{\sum_{i=1}^{4} p(fit).p\left(\frac{x_i}{fit}\right)+\sum_{i=1}^{4} p(unfit).p\left(\frac{x_i}{unfit}\right)} \quad (3.2.4)$$

C. The Output-Attribute-Precedence Relations

Once, the relative fitness probabilities namely average fit(x_1), average fit(x_2), average fit(x_3) and average fit(x_4), computed owing to the four attributes, they could be lay in increasing / decreasing order, the precedence relations shown in increasing order as in figure 2.

Output (1) - Notepad

File Edit Format View Help

OUTPUT.txt

Name of Stud	Tuple_id	p/f	precedence fitness	precedence unfitness
Adarsh Shrivast	1	0	0.00 0.00 0.00 0.00	0.00 0.24 0.24 0.24
Anshul Lanjawre	2	1	0.45 0.57 0.76 0.76	0.10 0.06 0.00 0.00
Chandan Kumar S	3	1	0.76 0.00 0.76 0.59	0.00 0.00 0.00 0.05
Deepesh Ganjir	4	1	0.00 0.00 0.74 0.00	0.00 0.00 0.01 0.00
Ku. Jigyasha	5	1	0.76 0.62 0.76 0.61	0.00 0.05 0.00 0.05
Manoj Kumar	6	1	0.76 0.62 0.59 0.76	0.00 0.05 0.05 0.00
Md. Majhar	7	1	0.45 0.57 0.76 0.00	0.10 0.06 0.00 0.00
Mona Mandavi	8	1	0.76 0.76 0.74 0.76	0.00 0.00 0.01 0.00
Nandita Sardar	9	1	0.76 0.76 0.76 0.76	0.00 0.00 0.00 0.00
Neha Khandelwal	10	1	0.76 0.76 0.74 0.76	0.00 0.00 0.01 0.00
Nidhi Khandelwa	11	1	0.76 0.76 0.59 0.76	0.00 0.00 0.05 0.00
Pratima Soni	12	1	0.76 0.62 0.59 0.76	0.00 0.05 0.05 0.00
Promod Kumar Sa	13	1	0.76 0.76 0.74 0.61	0.00 0.00 0.01 0.05
Ruchika Pandey	14	1	0.76 0.76 0.74 0.76	0.00 0.00 0.01 0.00
Saleya Khatoon	15	1	0.76 0.76 0.74 0.76	0.00 0.00 0.01 0.00
Sanju Patel	16	1	0.76 0.57 0.74 0.00	0.00 0.06 0.01 0.00
Santosh Kumar	17	1	0.00 0.00 0.76 0.00	0.00 0.00 0.00 0.00
Suraj Tripathi	18	1	0.00 0.76 0.76 0.61	0.00 0.00 0.00 0.05
T. anita Soni	19	1	0.76 0.76 0.76 0.61	0.00 0.00 0.00 0.05
Yogendra shyam	20	1	0.45 0.00 0.76 0.00	0.10 0.00 0.00 0.00

Fi.g. 2. Attribute Precedence of fitness / unfitness (proposed FE model)

Similarly, the precedence relations could be compiled for average unfit (x_i) probabilities for the above mentioned four attributes. These varying precedence relations, obtained from the above proposed FE modeling open the dimension of decision making tasks in the direction of individual student's counseling. As can be seen in figure 2, considering the attribute-relative precedence obtained for student enrolled with tuple-id: 3, it was observed that attribute x_4(subject count) exhibited 'NIL' fit and unfit values, while other three attributes exhibited values such that x_1 (students' attendance) was found most contributing to his academic appraisal, followed by attribute x_2 (internal assessment scores) and x_3 (assignment credit) respectively. This is justified from the maximum average-fit (0.51) and minimum average-unfit (0.0) values of attendance attribute (x_1). Similar kind of justifications can be made for other two attributes x_2 and x_3. Hence, for tuple-instance 3, the counseling should focus the most to the student's attendance, followed by motivation to complete his assignments and encouraging him to write the examinations for obtaining favorable internal scores. However, the precedence relations for other students are bound to vary with the fact that a class consists of students having diversified academic potentials.

D. Experiments with RELIEF

In order to arrive at performance evaluating of the proposed setup, it was decided to revisit the problem using one of the popular feature extraction model 'RELIEF'. The popularity of the model is due to its increased accuracy, reduced time complexity, usage of simple statistical approach and enormous success achieved in practical applications. Moreover, the traditional 'RELIEF' had been tailored by many other researchers to modulate their own problem objective by encoding RELIEF variants

E. Problem Mapping to RELIEF

The problem objective mentioned in section 1was mapped to 'RELIEF', as it readily fits into two-class classification problem of predicting 'at-risk' level of students and identifying the precedence levels of attributes contributing

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

to their performances. Given the training instances data:δ={1..88}; m=20 for given set of test instances so that every instance 'X' is denoted by 'p' dimension vector (x_1,x_2,x_3….,x_p) where p=4 for the current problem domain. Hereafter, the algorithm makes use of p-dimensional Euclidean distance to select 'near-hit' and 'near-miss' instances from the training data set. For the current domain,' Near-hit' and 'Near-miss' instances are defined as training instances (students from passed out batches) closest to the test instance but falling in 'pass' and 'fail' category, symbolized as Z+ and Z- respectively.

In order to compute updating of feature-weight vector of the test instances, the test feature vectors were extracted for 'Near-hit' and 'Near-miss' training instances and squares of the differences between respective attribute values for the above instance sets according to the expression mentioned below:

$$w_i=w_i\text{-diff}(x_i,\text{near-hit}_i)^2+\text{diff}(x_i,\text{near-miss}_i)^2 \qquad (4.1.1)$$

F. *Results of Comparisons*

Having the p-dimensional feature vector (p=4) for all test instances in sample size (m=20) updated, the logistics get deviated from original algorithm.

Here instead of comparing average value for each attribute x_i against set threshold to declare the relevancy status of the attribute, a precedence relation is established among all the participating attributes (p=4) according to the computed weight updates. Such observations were taken for all the test instances as illustrated in table 2, column 8.The initial weight values w_i were computed as discrete numeric values in normalized scale. For instance, w_1, is the student's attendance expressed as: (<attendance> + 8 /15); $w_2$is the assignment credit awarded out of 10; w_3 is the internal score computed out of '3' and w_4 is the subject count appeared in internal tests computed out of 10. It may be noted that Z^+ and Z^- components adheres to the heuristics defined in the original algorithm. If the test instance x_j is predicted as a positive instance (i.e. the student obtains 'above-risk' status) then near-hit instance (x_j) is assigned as Z^+ and near-miss instance (x_j) is assigned with Z^- value. The vice-versa happens if the test instance falls into 'at-risk' predicted status i.e. it is predicted as negative instance.

IV. CONCLUSION AND FUTURE SCOPE OF WORK

As the problem objective orients itself at identifying up to what level, the attributes affect the academic performance levels with the baseline fact that all the attributes bear partial or strong relevance. As tabulated in table 2, columns 8 and 9 compare the precedence relations showing computed graded attribute relevance due to both RELIEF and proposed FE logistics. Moreover, the attribute precedence deviated by one position hardly changes the relevance level of the attributes. The above opinion helped the work group in comparing partial or total match patterns of attribute-precedence by making use of RELIEF heuristics as benchmark. As can be seen in table 2, both the parameters:

- total precedence match patterns and partial precedence match patterns (having precedence deviations by 1 position) are taken into account to compute performance of the proposed FE model with respect to RELIEF approach. Upon accumulating the varied precedence match count from 1 to 4, the weighted accuracy of 75% from the proposed model was computed in ex-tracting attribute precedence as evaluated with reference to RELIEF heuristics. These results encouraged the authors to continue series of experiments upon varied combi-nation of contributing parameters.

Table 2. Precedence Relations showing graded Attribute Relevance using RELIEF

Sl. No.	Stu-dent Name (x)	Pre-dicted Risk	Wx1'	Wx2'	Wx3'	Wx4'	RELIEF method	Proposed Method
1	Adarsh Shrivastava	0	0.2	0.1	0.3	0.2	x2<x1=x4<x3	x2<x4<x3<x1
2	Anshul Lanjewar	1	0.1	-3	3.4	-15	x4<x2<x1<x3	x4<x3<x2<x1
3	Chandan Kumar Sahu	1	12	5	0.3	-15	x4<x3<x2<x1	x1<x4<x2<x3
4	Deepesh Kumar	1	0	9.3	3.2	9.2	x1<x3<x4<x2	x3<x1<x2<x4
5	Ku. Jigyasha	1	4.4	0.8	1.3	0.1	x4<x2<x3<x1	x4<x1<x2<x3
6	Manoj Kumar	0	75.3	7.2	-0.1	15	x3<x2<x4<x1	x2<x3<x4<x1
7	Md. Majhar	1	0	0.3	-0.7	3.2	x3<x1<x2<x4	x2<x1<x3<x4
8	Mona Mandavi	1	25.1	1.4	-0.3	1.3	x3<x4<x2<x1	x3<x4<x2<x1
9	Nandita Sardar	1	-1.7	1	0.2	-1.7	x1=x4<x3<x2	x4<x2<x3<x1
10	Neha Khandelwal	1	-3.6	-2.8	0.5	-8	x4<x1<x2<x3	x1<x2<x3<x4
11	Nidhi Khandelwal	1	0.4	0.3	0.4	0	x4<x2<x1=x3	x1<x2<x3<x4
12	Pramod	1	-5	3.3	0.2	4	x1<x3<x2<x4	x2<x1<x3<x4
13	Pratima Soni	1	-5	3.3	0.2	-1.2	x1<x4<x3<x2	x2<x1<x3<x4
14	Ruchika Pandey	1	4	-3.9	1.4	3.2	x2<x3<x4<x1	x2<x1<x3<x4
15	Saleya Khatoon	1	-2.9	0.1	0.3	-3	x4<x1<x2<x3	x2<x4<x3<x1
16	Sanju Patel	0	-12	5	1.3	12	x1<x3<x2<x4	x1<x3<x4<x2
17	Santosh Kumar	1	16	3.3	0.3	4.2	x3<x2<x4<x1	x1<x3<x2<x4
18	Suraj Tripathi	1	4	1.1	0.3	5.2	x3<x2<x1<x4	x3<x2<x1<x4
19	T. anita Soni	0	0.2	3	0.3	3	x1<x3<x2=x4	x3<x4<x1<x2
20	Yogendra shyam Kumar	1	-4	0	1.4	-3.8	x1<x4<x2<x3	x4<x2<x1<x3

Pre-dicted Risk Wx1' Wx2' Wx3' Wx4' RELIEF method

ACKNOWLEDGMENT

The authors sincerely thank the Management, Sai College, Sector-6, Bhilai for supporting the current on-going work in Educational Data Mining and Analytics and also to the Department of Computer Science for providing students' data of on-going batch of BCA second year, one of the approved courses at Sai College, affiliated to Pt. Ravi Shankar Shukla University, Chhattisgarh, India.

REFERENCES

[1] B. M. Bidgoli and W.F. Punch, "Predicting Student Performance:An Application of Data Mining Methods with the educational Web Based System LON_CAPA" ,G. A. Research and Application, Boulder,co, Nov 5-8,2003.
[2] M..N Quadril and N. V. Kalyankar, "Drop Out Feature of Student Data for Academic Performance Using Decision Tree Techniques", Global Journal of Computer Science and Technology, vol. 101, No. 2, pp. 2-5, April 2010.
[3] A. Qasim, AI-Radaideh and E.AlNegi"Using Data Mining Techniques to Build a Classification Model to Predicting Employees Performance", Int. J.Advanced Computer Science and Applications,vol. 3,No. 2, 2012.
[4] M. Fire, G. Katz, Y. Elovici, B. Shapira and L Rokach, "Predicting Student Exam's Scores by Analyzing Social Network Data", Telecom Innovation laboratories and information Engineering Department, Ben-Gurion,Israel, December 2012.
[5] R. Kohavi, H. John and Pfleger Karl, "Irrelevant Feature and the subset selection problem", In Machine Learning W.Cohen and HaymHiesh, eds. SanFrancisco, CA, [Proceedings 11th Int. Conf. Isreal, pp. 121-129, 1994.

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 9, September 2014

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

[6] K. Kira and L. A. Rendell, “A practical approach to feature selection,” Proc. 9th Int. Conf. Mach. Learn, pp. 249 , 1992.
[7] M Sewell, “Feature Selection”, 2007, http://machine-learning.martinsewell.com/feature-selection/feature-selection.pdfK. Kira and L. A. Rendell, “A practical approach to feature selection,” Proc. 9th Int. Conf. Mach. Learn, pp. 249, 1992M.
[8] B. M. Bidgoli and W.F. Punch, “Using Genetic Algorithms for Data Mining Optimization in an Educational Web-Based System”,G. A. Research and Application Group,Berlin, Heidelberg, pp 2252-2263,2003.
[9] Y.Sun and D.Wu, “A Relief Based Feature Extraction Algorithm”,Atlanta, Georgia, USA,[Proceedings Int. Conf. on Data Mining,,April 2008.
[10] Ying Liu “A Comparative Study on Feature Selection Methods for Drug Discovery” Georgia Institute of Technology, College of Computing, Atlanta, Georgia 30322, April 13, 2004.

BIOGRAPHIES

Mamta Singh received her MCA degree from Maharshi Dayanand University, Rohtak in 2005. She has also received her MPhil in Computer Science from Periyar University, Salem. She is working presently with Sai College as Assistant Professor and head in Computer Science Department.

Dr. (Prof.) Jyoti Singh received the MCA degree from Banasthali Vidyapith, Rajasthan in 1990. She has also done PhD in Computer Science and Application from Pt. Ravi Shankar Shukla University, Raipur in 2007.She worked as Dy. Registrar in Chhattisgarh Swami Vivekanand University(CSVTU) Bhilai since 2005. Act as Resource person in computer course under Canada-India Institute Industry Linkage Project. She is Life member of “The Indian Society for Technical Education. She is an approved PhD guide in CSVTU Bhilai, Dr. C V Raman University Bilaspur, Periyar University Salem, Vinayaka University Tamil Nadu.

Arpana Rawal completed her graduation degree in computer science from Nagpur University. She has awarded with doctoral degree in computer science in year 2013 from swami Vivekanand technical university. She has teaching experience of 14 years and research experience of 9 years.