Applicability of Traditional Classification Techniques on Educational Data

Balwinder Kaur, Anu Gupta, R. K. Singla

Abstract: Student performance prediction and analysis is an essential part of higher educational institutions, which helps in overall betterment of the educational system. Various traditional Data Mining (DM) techniques like Regression, Classification, etc. are prominently utilized for analyzing the data coming from educational settings. The usage of DM in the area of academics is called Educational Data Mining (EDM). The current pilot study aims to determine the applicability of these standalone classification techniques namely; Decision Tree, BayesNet, Nearest Neighbor, Rule-Based, and Random Forest (RF). The present pilot study uses the WEKA tool to implement traditional classification techniques on a standard dataset containing student academic information and background. The paper also implements feature selection to identify the high influential features from the dataset. It helps in reducing the dimensionality of the dataset as well as enhancing the accuracy of the classifier. The results of classifiers are compared on basis of standard statistical measures like Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Kappa, etc. The results show the applicability of classification algorithms for student performance prediction which will help under-achievers and struggling students to improve. It is found the output that, J48 algorithm of the Decision tree gave the best results. Further, it is deduced from the comparative analysis that individual classifiers give different accuracy on the same dataset due to class imbalance in a multiclass dataset.

Keywords: Performance Prediction, Data Mining Techniques, Classification, Decision Tree, BayesNet, Random Forest, WEKA.

I. INTRODUCTION

The merger of computing with communication has produced a community that feeds on information [1]. The unrestricted growth of databases in previous years has resulted in a huge amount of data stored in different repositories in every field[1][2]. There is a need to analyze data stored in these repositories for finding hidden knowledge, which is difficult to do manually. Data mining(DM) sector deals with discovering unique knowledge and beneficial learning from large volumes of data stored in various repositories [3]. In the past few years, there is an engrossing interest in the usage of DM for research and scientific inquiry within the educational system.

Balwinder Kaur*, Defense Security Cooperation Agency, Panjab University, Chandigarh, India. Email: balwinder@pu.ac.in.
Aun Gupta, Defense Security Cooperation Agency, Panjab University, Chandigarh, India. Email: anugupta@pu.ac.in
Ravinder K. Singla, Defense Security Cooperation Agency, Panjab University, Chandigarh, India. Email: rksingla@pu.ac.in

The usage of DM used in the field of academics is known as Educational Data Mining ("EDM")[3]. Educational Data Mining has evolved as an individualistic and self-sufficient research area, starting from research in Intelligent Tutoring System, learning management system, etc [3].
The origin of EDM lies in workshop series organized into related conferences started in 2000. [6],[7],[8].
The 1st workshop, introduced 'Educational Data Mining', was held in the year 2005
EDM is a field that focuses on predicting the performance of students related to academics based on personal, socio-economic, demographic and other environmental factors. The knowledge discovered is potentially useful to the teachers, course planners, parents, and above all students. The areas of application of EDM are related to student performance prediction, giving recommendations to students, providing feedback to instructors, grouping students, planning and scheduling, etc. [3].
Various traditional techniques of Data Mining like Classification, Regression, Association rule mining, Clustering, and Feature selection techniques are being prominently used on data coming from the educational environment by various researchers [2], [5], [6], [7], [8]. The focus of the current pilot study is on the application of different methods of selecting features as well as the applicability of traditional classification techniques namely; J48, BayesNet, PART, RepTree and Random Forest on the student academic dataset. Further, the performance of the individual classifier is measured based on percentage accuracy as well as other statistical measures.
The paper is organized into the following sections namely; Section 2: related-work, Section 3: methodology of research, Section 4: Experiment and Results, Section 5 is Discussion and Section 6 presents the Future work and Conclusion.

II. RELATED WORK

A survey on the usage of DM in various education systems for a period from 1995 to 2005 is presented in [6]. The survey shows the application of DM in different systems of education like traditional education systems, web-based courses, adaptive systems, etc. The survey finds that every educational system has specific requirements and the main need is considered different instructional and academic features of the learner as well as the system. M. Ramaswami et al., in their study[7] investigated the selection of most relevant attributes for attaining high-performance prediction.

The comparative study is carried on 6 different filter feature section algorithms to find the best method. The outcomes of the present study adequately support the fact of the increase in the accuracy of prediction with the least number of features. The results show reduction in constructional cost as well as computational time in training and classification phases of the performance model. Researchers in [8] aims to bring forward performance analysis of different algorithms of feature selection applied on dataset related to students. In this study, various feature selection algorithms such as CfsSubsetEval, ChiSquaredAttributeEval, Filtered-Attribute-Evaluator, Gain-Ratio-Attribute-Eval, Principal Component, and Relief-Attribute-Eval have been assessed and analyzed. The results of experiment has shown that out of different methods of feature selection used, principal components shows superior results when used with RF classifier. The research has shown that Multi-Layer Preceptor is slightly better in comparison to various classifiers used on data set. The study shows that, fine tuning of parameters related to feature selection methods is required, for achieving better performance. M. Chalaris et al., in their study[9], represent the potential abilities of data mining with reference to the Higher Educational Institute (HEI). They use techniques like Correlation analysis, Clustering, and Association rules on data set of student and course obtained Technological Educational Institute located in Athens. The study finds that DM techniques can be used in HEI decision making. Kabakchieva [10], focuses on implementing data mining techniques on a dataset collected by universities. The study aims at performance analysis of following classification algorithms namely; OneR, JRip, J48, NaiveBayes, BayesNet, and Nearest Neighbour on the provided dataset. The researcher finds that features associated with student university admission marks as well as no. of failures in 1st year exams of university are the important elements in the process of classification. AU Khasanah[11] et al., have conducted study with an aim to find and select most influential features from student performance data of Industrial Engineering Department of an University in Indonesia. For this purpose 2 classification algorithms; Decision tree and Bayesian Network are applied on the dataset. The outcome shows that GPA and attendance are the most influential attributes and Bayesian Network out-performs DT in terms of accuracy. Further in studies [12] and [13], the researchers have implemented various Decision Tree algorithms for student performance prediction at the end of the semester based on the previous semester and in semester results. They used academic attributes to compare the performance accuracy of algorithms. Hussain et al., in their research[14], aims to assess the performance of college students related to academics depended upon educational and personal individual data collected. The researcher implements different classifiers namely; RF, Bayes Network, J48, and PART. The outcomes showed that Random Forest is better than the rest of classifiers implemented based upon the accuracy of classification as well as classification errors. Further, to discover the association rule between different features the Apriori algorithm is used. Parneet Kaur et al., are

focusing on recognizing the slow learning students in [15]and showing it by a predictive DM model based on different algorithms of classification namely; J48. MLP, SMO, REPTree, and Naïve Bayes. These algorithms are implemented of the dataset using cross-validation. The classification accuracy of classifiers is compared. The result of the comparison shows that the MLP method has 75% accuracy highest among all the classifiers implemented.

III. RESEARCH METHODOLOGY

A. Research Aim

Predicting and analyzing student academic performance is an essential part of educational institutions. This pilot study aims to analyze the applicability and performance of traditional classification techniques namely; J48, BayesNet, Naïve Bayes, PART, RepTree, 1- Nearest Neighbor and Random Forest in predicting the student performance based on their academic as well as demographic attributes. Different feature selection algorithms will be applied to the student performance dataset with an aim to extract the most relevant features.

B. Data Set

The standard academic dataset for the study is available at the well as domain theories, brought into use by the researcher for UCI. It is an open-access Machine Learning (ML) Repository. It stores widely used a compilation of different databases, data generators as the experimental study and analysis of ML algorithms [19]. This archive was initiated in the year 1987 by David Aha along with his fellows at UC Irvine [16]. This pilot study uses a dataset that contains 131 instances with 22 attributes. The attributes store both the academic data related to marks of class 10th, 12th, backlog, pre-university, attendance, etc. and demographic information like family size, lived in village or town, family income and so forth, of the students[15], [18]. All the attributes are converted into nominal values and stored in the CSV file and later converted into WEKA compatible Attribute-Relation File Format (.arff) [3].

C. Tools and Techniques

- Tools: In this work, the WEKA tool is used for analysis purposes. WEKA is an assemblage of different ML algorithms used for different tasks of data mining which are developed in Java by "University of Waikato" located in New Zealand [2]. It's open-source software that provides extensive support for the whole process of experimental data mining. It includes options for different tasks like pre-processing of data, filters for selecting features, classification, regression, clustering, association rules, as well as visualization. The algorithms can either be embedded in java code or can be applied on a dataset directly.

1673

The WEKA tool provides four interfaces namely; The Explorer, The Experimenter, The Knowledge Flow and Simple CLI [2], [3].

- Techniques for Feature Selection (FS): FS process is also named as variable or attribute selection. In the feature selection process, subsets of relevant features are selected from the entire set of features related to a dataset [8], [9]. It helps to minimize the dimensions of the dataset and also reduces the computational cost of the classification process. The Feature Selection methods are categorized into 3 types namely; Wrapper Method, Filter Methods and Hybrid Methods[20], [21]. Using WEKA, feature selection methods namely; CfsSubset Evaluator, Correlation-Attribute-Eval, Gain-Ratio-Attribute- Eval, Info-Gain-Attribute-Eval, ReliefF-Attribute-Eval, Symmetrical-Uncert-Attribute-Eval are applied and attributes selected by these methods are shown in table 1. The description of data for the selected attributes is presented in table 5 Appendix.

- Classification Techniques: Classification techniques come under the category of supervised learning techniques [3]. These techniques classify the data items into the predefined class labels[2]. The data classification process is a 2 step process that involves model construction and model usage. In model construction, first training data is used to construct a model and during model usage, it is used for predicting the unknown value of data [2], [22]. Data mining encompasses various techniques of classification like SVM, Decision tree, Rule-Based, K-Nearest Neighbour, Bayesian classification and so on [21]. The classification techniques based on single classifiers are traditional classification approach. In this approach, a number of classification algorithms like Decision Tree (DT), BayesNet, Rule-based, and Random Forest are experimented independently for investigating the behavior of these core classifiers on the student dataset. The algorithms are compared and evaluated using different evaluation measures.

Table I: Attributes selected by Feature Selection Methods

Method of Feature Selection	Highly Influential Attributes/ Features
CfsSubsetEval	TNP,TWP,IAP, ARR, AS, FS, SH, ME, ATD
CorrelationAttributeEval	TNP, TWP, AS, IAP, ME, FS, ARR, ATD, SH, NF, TT, SS, LS
GainRatioAttributeEval	TNP, TWP, IAP, ARR, ME, AS, ATD, SH
InfoGainAttributeEval	TNP, TWP,IAP, ATD, ME, SH, FQ, ARR, AS, NF, MQ
ReliefFAttributeEval	TNP,IAP,TWP, ME, AS, NF, ARR, ATD, SH
SymmetricalUncertAttributeEval	TNP, TWP, IAP, ME, ATD, ARR,SH,AS

IV. EXPERIMENT AND RESULTS

Classification technique is one of the frequently used prediction techniques used by various educational organizations for student performance prediction.

In the experiment, after applying different feature selection algorithms, the classification algos. were applied for predicting the academic performance of learners using WEKA. 7 different classification methods namely; J48, RepTree, BayesNet, Random Forest, Naïve Bayes (NB), 1-Nearest Neighbor and PART are applied on the dataset. In the experiment, student end semester percentage (ESP) is chosen as predicted variable and grouped into 5 classes namely: 'a' as 'Best', 'b' as 'Very Good', 'c' as 'Good', 'd' as 'Pass', and 'e' as 'Fail'.

According to the algorithms specification provided by WEKA [22]; J48 is a decision tree classifier, the structure starts from the topmost root node and moves to end at leaf attributes. It implements a C4.5 algorithm [23]. NB classifier which is based on Bayes theorem is a simple probability-based classifier, it makes the assumption that all the attributes are independent of one another [2]. Bayes Net [13],[17] is a directed acyclic graph that shows the dependencies between random variables. 1Bk is k-NN based technique. This method is based on classifying objects which are nearest to training examples feature space. It is an instance-based learning method [10]. PART is a rule-based classifier that divides and conquers approach[24]. Random forest (RF) is a group of trees that are unpruned and these unpruned trees are based on regression or classification trees build by applying bootstrap, reduce over-fitting, bias, and variance[14]. REPtree builds a regression tree or decision tree by applying information gain as split criteria and further apply reduced error pruning to trim the unwanted branches.

There are 131 student records with 9 selected attributes. All the above classification methods used for the experiment are trained using 10-folds cross-validation.

A. Measures of Performance

Various performance measuring scales are used to gauge the performance of the classifiers [9], [13], [24]. The performance comparison helps to select the most optimal classifier.

- Classification Accuracy: Accuracy of a classifier is measured as its ability to correctly predict the value of unseen data of an instance divided by the total no. of instances multiplied by 100 [1], [2], [14].
- Mean Absolute Error (MAE): MAE is a statistical term brought into usage to calculate the average difference between predicted and real value [1], [2], [14].
- Root Mean Square Error (RMSE): RMSE is a measure to estimate dissimilarities among actual values perceived and values calculated by the model. It is calculated by taking the square root of the MSE [1], [2], [13].
- Confusion Matrix: It comprises the information about real and predicted groupings done by a classification technique [14].

The classification accuracy percentage is shown in table 2 and Figure 1.

The performance compared on the basis of MAE, RMSE, kappa of all the classifiers implemented in the pilot study are shown in Table 3.

Table III: Performance on the basis of accuracy

Classifier	Correctly Classified Instances	Incorrectly Classified Instances	Accuracy
J48	88	43	67.17%
RepTree	82	49	62.59%
Random Forest	84	47	64.12 %
BayesNet	81	50	61.83%
Naïve Bayes	81	50	61.83%
PART	83	48	63.35%
1-Nearest Neighbor	69	62	52.67%
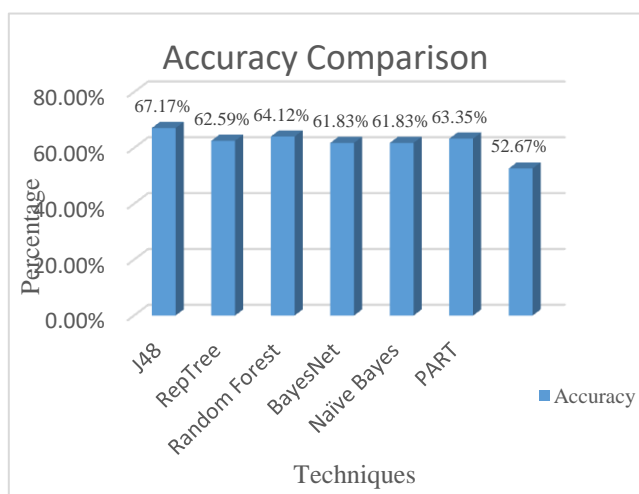
Figure I: displays percentage accuracy comparison between J48, Bayes Net, Part, RF, and REPTree classifiers

Table III: Performance comparison on the basis of kappa statics, MAE and RMSE

Classifier	Kappa	MAE	RMSE
J48	0.5005	0.1857	0.3241
RepTree	0.4266	0.1974	0.3293
Random Forest	0.4541	0.1972	0.3273
BayesNet	0.426	0.1863	0.3344
Naïve Bayes	0.426	0.1866	0.3296
PART	0.45	0.1902	0.341
1-Nearest Neighbor	0.2886	0.2034	0.3922

Further, table 4 shows the comparison among all the implemented classifiers on the basis of the confusion matrix of all the classifiers.

Table IV: Classifier comparison based on the Confusion matrix

Classifiers	a	b	c	d	e	Class
J48	0	7	1	0	0	a = Best
	0	29	13	0	0	b = Vg
	0	6	41	7	0	c = Good
	0	1	8	18	0	d = Pass
	0	0	0	0	0	e = Fail
RepTree	0	7	1	0	0	a = Best
	1	29	12	0	0	b = Vg
	0	9	40	5	0	c = Good
	0	3	11	13	0	d = Pass
	0	0	0	0	0	e = Fail
Random Forest	1	6	1	0	0	a = Best
	0	30	12	0	0	b = Vg
	0	9	38	7	0	c = Good
	0	1	11	15	0	d = Pass
	0	0	0	0	0	e = Fail
BayesNet	0	7	1	0	0	a = Best
	1	30	11	0	0	b = Vg
	0	11	34	9	0	c = Good
	0	0	10	17	0	d = Pass
	0	0	0	0	0	e = Fail
Naïve Bayes	0	7	1	0	0	a = Best
	1	30	11	0	0	b = Vg
	0	11	34	9	0	c = Good
	0	0	10	17	0	d = Pass
	0	0	0	0	0	e = Fail
PART	0	6	2	0	0	a = Best
	0	28	13	1	0	b = Vg
	1	8	36	9	0	c = Good
	0	1	7	19	0	d = Pass
	0	0	0	0	0	e = Fail
1-Nearest Neighbor	1	7	0	0	0	a = Best
	1	28	11	2	0	b = Vg
	0	16	29	9	0	c = Good
	0	0	16	11	0	d = Pass
	0	0	0	0	0	e = Fail

V. DISCUSSION

After measuring the performance of the student, the educator, and the student can be informed. Various constructive steps can be taken by the educator as well as students to improve the performance in the case of a student with weak academics performance and positive feedback will be provided to students with good academics performance for motivation which will help to enhance overall academic performance.

From the above study, it is also found that the most influential features are the academic attributes like the tenth percentage, 12th percentage etc. as compared to demographic and socio-economic attributes related to students for academic performance prediction. Further, the J48 technique displayed a highly accurate classification out of all the classifiers implemented. The statistical comparison also shows that the J48 algorithm is the most suitable among all the classifiers. Some issues related to the data mining techniques were also identified during the study. These issues are:

- The size of data-set: the data-set utilized in experimentation is of small size, therefore not many records are available for training and testing. The accuracy can be improved by using a larger dataset.
- Use of multiclass dataset: In multiclass datasets, the data may be imbalanced as the number of instances for all the classes is not equally represented. The majority of instances may belong to fewer classes. Due to imbalanced data, the learning algorithm oversees classes which are less frequent and focus on frequent classes. Therefore, the classifier obtained classify data instances incorrectly, hence reducing the accuracy. This issue can be resolved by applying sampling or class balancing or rebalancing algorithm during data pre-processing.
- Multiple Classification methods: Application of individual traditional classification methods produces different accuracies on the same dataset. A combination of different but complimenting algorithms can be applied to increase the prediction accuracy.

VI. FUTURE WORK AND CONCLUSION

The tremendous quantity of data associated with students is stored by the educational institutes at different levels. This data contains concealed knowledge that can be utilized to enhance the educational/ pedagogical achievement of the learners as well as the institute. Many prediction models using diverse approaches to predict student performance are reported by the researchers; mostly these models are based on traditional classifications techniques. In this pilot study, 7 different classification algorithms namely; REPTree, PART, Random Forest, J48, 1-Bk, Naïve Bayes and BayesNet are implemented using WEKA. The results show the applicability of classification algorithms to predict student performance which will help under-achieving and struggling students to improve. The results obtained are on the basis of academic and some personal attributes of students which are further reduced by feature selection. Based on the comparative analysis performed, it is concluded that J48 method gives better results among all the implemented algorithms and it is the best algorithm for student dataset used. The study helps in gaining insight into the working of the algorithms i.e. how prediction is done. From the discussion, it is concluded that different factors like the size of a dataset, usage of multi-class data set which can lead to class imbalance affect the performance accuracy of the technique. Taking into consideration various classification methods applied, there exist not a single classifier that claims to be uniformly superior to the other, a hybrid multi-classifier algorithm can be proposed as future work which may enhance the accuracy of classification in case of multi-class datasets.

APPENDIX

Table V: Data Description of selected features [15, 18]

Attribute	Description
TNP	10th Percentage
TWP	12th Percentage
IAP	Internal Assessment Percentage
ESP	End Semester Percentage
ARR	Student backlog or arrear papers
LS	Living status (Town/ Village)
AS	Category of Admission
FS	Size of Family
FQ	Qualification of Father
MQ	Qualification of Mother
NF	Total Friends
SH	No. of Hours of study
SS	Secondary School attended
ME	Medium
TT	Travel Time(home to college)
ATD	Attendance Percentage

REFERENCES

1. Witten I. H., L. Trigg, Eible F., Hall M., & Holmes G., "WEKA: Practical Machine Learning Tools and Techniques with JAVA Implementations," 1999.
2. Jiawei H., Micheline K., & Pei J., "*Data Mining: Concepts and Techniques*", Morgan Kaufmann Publishers.
3. Romero C. & Ventura S., "Data Mining in Education", *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Volume. 3, Issue no. 1, pages. 12–27, 2013.
4. S. Ventura and C. Romero, "Data mining in education," vol. 7, no. October 2015, pp. 43–48, 2012.
5. Ryan S. J. D. Baker and Kalina Y., "The State of Educational Data Mining in 2009 : A Review and Future Visions," *Journal of Educational Data Mining*, volume. 1, Isuue no. 1, pages. 3–16, Fall 2009.
6. Cristobal Romero and Sebastian Ventura, "Educational data mining: A survey from 1995 to 2005", *Expert Systems with Application*, volume. 33, Issue 1, pages. 135–146, July 2007, Elsevier.
7. M. Ramaswami and R. Bhaskaran, "A Study on Feature Selection Techniques in Educational Data Mining", *Journal of Computing*, Volume. 1, Issue. 1, pages. 7–11, December 2009.
8. Maryam Z., Manzoor A. Hashmani and K.S. Savita, "Performance Analysis of Feature Selection Algorithm for Educational Data Mining", IEEE conference of Big Data and Analyticsm, pages. 7–12, IEEE, 2017.
9. Manolis C., Stefanos G., Manolis M., Cleo S., and Anastasios T., "Improving Quality of Educational Processes Providing New Knowledge Using Data Mining Techniques," *Procedia - Social Behavioral Sciences*, Volume. 147, pages. 390–397, 25 August 2014.
10. Kabakchieva D., "Predicting Student Performance by Using Data Mining Methods for Classification", Cybermetics and Information Technologies, Volume. 13, no. 1, pages. 61–72, 2013.
11. Annisa U. Khasanah and Harwati, "A Comparative Study to Predict Student's Performance Using Educational Data Mining Techniques," *IOP Conference Series: Material Sciences and Engineering*, Volume. 215, no. 1, IOP publishing 2017.
12. Brijesh Baradwaj and Saurabh Pal, "Mining educational data to analyze student's performance," *International Journal of Advanced Comput. Sci. Application*, Volume 2, no. 6, pages. 63–69, 2012.
13. Yadav S.K., "Data Mining : A Prediction for Performance Improvement of Engineering Students using Classification", World of Computer Science and Information Technology Journal Volume. 2, Issue . 2, pp. 51–56, 2012.

14. Sadiq Hussain, Neama A. Dahan, F. M. Ba-Alwib, and N. Ribata, "Educational data mining and analysis of students' academic performance using WEKA," *Indonesian Journal of Electrical Engineering and Computer Science*, Volume. 9, no. 2, pages. 447–459, Feb 2018.

15. Parneet Kaur, Manpreet Singh, and Gurpreet. S. Josan, "Classification and Prediction Based Data Mining Algorithms to Predict Slow Learners in Education Sector", *Procedia Computer Science*, Volume. 57, pages. 500–508, 2015, Elsevier.

16. "UCI info." Available [Online]: https://archive.ics.uci.edu/ml/index.php. [Accessed on : 14-May-2019].

17. Sadiq Hussain, "STANDARD DATA STUDENT PERFORMANCE." Available [Online]: https://archive.ics.uci.edu/ml/datasets/Student+Academics+Performance . [Accessed: 05-Feb-2019].

18. Sanmay Das, "Filters, wrappers and a boosting-based hybrid for feature selection," ICML-2001, Volume 1, pages. 74–81.

19. Eric P. Xing, Michael I. Jordan, and Richard M. Karp, "Feature selection for high-dimensional genomic microarray data," *Proceedings 18th International Conf. Machine Learning*, Volume 1, pages. 601–608, 2001.

20. Mrinal Pandey and S. Taruna, "Towards the integration of multiple classifiers pertaining to the Student ' s performance prediction," *Perspectives in Science*, Volume 8, pages. 364–366, Sept 2016.

21. Ryan S. J. D. Baker, "Data mining for education," *International Encyclopedia Educ.*, volume. 7, pages. 112–118, 2010.

22. "weka documentation.", Available [Online]: "https://www.cs.waikato.ac.nz/ml/weka/documentation.html.", [Accessed: 04-May-2019].

23. Hilal Almarabeh, "Analysis of Students ' Performance by Using Different Data Mining Classifiers", International Journal of Modern Edu. and Comp. Sci., Volume. 9, pages. 8–15, August 2017.

24. Pang-Ning Tan, Michael. Steinbach, and V. Kumar, "*Introduction to data mining*.", Pearson Education India, 2005.

AUTHORS PROFILE

Mrs. Balwinder Kaur, is Research Scholar in Computer Sc. and Applications, Panjab University, Chandigarh where she has been working as Assistant Professor since 2011. She received her MCA degree from Kurukshetra University. Her research interest is in Educational Data Mining.

Dr. (Mrs.) Anu Gupta, has been working as Professor in Computer Sc. and Applications, Panjab University, Chandigarh since July 2015 where she has been working as a faculty member since 1998. She has the experience of working on several platforms using a variety of development tools and technologies. Her research interests include Open Source Software, Software Engineering, Cloud Computing, and Data Mining. She is a life member of the 'Computer Society of India' and 'Indian Academy of Science'. She has published 25 research papers in various journals and conf.

Dr. Ravinder K.., Singla has been Professor of Computer Sc. and Applications, Panjab University, Chandigarh since July 2004 where he has been a faculty member since 1988. Currently, he is also holding the position of Chairperson, DCSA and Dean Research, PU. He held many IT-related positions including Director, Computer Center; Coordinator, TIFAC-DST; Project Leader; and Programmer/Analyst. His research interests include scientific computing, Linux networking, mobile computing, open-source software and software cost estimation. He is also a member of the editorial board of the 'Panjab University Research Journal (Science)-New Series'. He is a life member of the 'Computer Society of India'. He has published 70 research papers in various journals and conf.