Prediction of Student's Performance based on Incremental Learning

Pallavi Kulkarni
Dept. of Computer Engineering
Dr .D. Y. Patil SOET
Pune, Maharashtra, India

Roshani Ade
Dept. of Computer Engineering
Dr .D. Y. Patil SOET
Pune, Maharashtra, India

ABSTRACT
It is necessary to use Student dataset in order to analyze student's performance for future improvements in study methods and overall curricular. Incremental learning methods are becoming popular nowadays since amount of data and information is rising day by day. There is need to update classifier in order to scale up learning to manage more training data. Incremental learning technique is a way in which data is processed in chunks and the results are merged so as to possess less memory. For this reason, in this paper, four classifiers that can run incrementally: the Naive Bayes, KStar, IBK and Nearest neighbor (KNN) have been compared. It is observed that nearest neighbor algorithm gives better accuracy compared to others if applied on Student Evaluation dataset which has been used.

General Terms
Incremental learning, Classification

Keywords
Student prediction; KStar; NNGe; IBK; Naïve Bayes

1. INTRODUCTION
The ability to predict a student's performance is of great significance in educational field. Students' academic performance is based on diverse factors such as personal, social, psychological and other environmental aspects. Student's performance can be evaluated by making use of this factors.

There are different data mining techniques that can be used for analysis of factors like mentioned above and finding out scope of improvement for the same. Supervised learning techniques use instances, which have already been pre-classified in some manner. That means each instance has a label, which recognizes the class to which it belongs. Classification is a supervised data mining technique, makes prediction about values of data using known results found from different data. Classification maps data into predefined groups of classes [4][5].

Almost all traditional classification algorithms work in batch mode. That means they require to have input as their training data before building the actual model of the system. This approach is suitable for systems having small amount of data or, the systems where data comes in batch format. Means once collected, there is no need to update it frequently. It may be a kind of historical data where analysis of past performances can be done.

But today's educational field is very vast, and various enhancements are being done in it by researchers in that field to improve quality of education thereby producing a skilled manpower from respective fields. So it is all about iterative data mining process. Incremental technique is an approach where data is being continuously added [7][8]

Incremental learning ability is too essential to machine learning strategies designed for solving real-world problems because of two reasons:

1) It is difficult to gather all useful training instances before the trained system executes. Hence when the new instances are given, the learning algorithm must have capability of doing some revisions on the trained system so that the knowledge which is not learned yet, can be encoded in those newly arrived examples can encompass in the trained system.

2) Updating a trained system can be less costly in terms of time than creating a new system from scratch, which is of great importance in real-world applications[20][21].

2. BACKGROUND
Data Mining deals with recognizing new patterns from huge data. It has numerous application in educational filed, for example to enlist which alumni can donate in grand amount to the institution.

2.1 Educational Data Mining
The process of tracing and mining student data for the purpose of improvement in teaching and learning is comparatively new still there are various works before which attempt to do so and researchers are starting to combine their ideas .The usefulness of mining such data is auspicious but still require to show beyond doubt and oversimplified analysis to be done[4][5][7].

Decision making in traditional classroom system encompasses perceiving a student's way in which he/she conduct, examining their past records, and roughly calculating the virtue of pedagogical policies. Nevertheless, when students are in electronic environments, this informal supervising is difficult; educators should find alternatives to obtain such knowledge. Many organizations work in distance education scenario and therefore they gather massive amount of data, most of the times automatically generated online environmental devices such as servers and recorded in server access logs. Online learning environments can track data of most learning behaviors of the students therefore able to collect numerous learning profile. Nowadays, people started working with more enthusiasm in online environment to automatically analyze student and his/her interactions and learning profiles[27].

Taking out knowledge with effort from information and evaluating it in area of educational system can be applied to enhance learning process and it is having a profound influence on student's development. A technique known as formative evaluation in education environment is the assessment of student's accomplishment while it is in progress and ongoing refinement of a course and student's conduct [28].

To enhance study material which is used by tutor he/she can examine in what way student use the system. By doing this instructions for the course can be prepared in formative style [29].

In pedagogical perspective data mining techniques may help for the formative assessment of students so that tutors can upgrade teaching methods and overall approach towards learning process.

Distance education have great importance and there are various issues regarding student's performance in it. There is really a need to apply some intelligent techniques to examine the difficulties which can be arise in such environment and find solutions on order to improve the quality of distance education.[20]

There are several data mining techniques which can be used in educational systems like
1) Statistics and visualization
2) Web mining
3) Classification, clustering and outlier detection
4) Association rule mining and sequential pattern mining
5) Text mining [38-40].

2.2 Incremental Learning

Several algorithms have been recommended for incremental learning, where incremental learning implied different problems. Some of the studies use the term incremental learning to talk about growing or pruning of classifier architectures some of them discuss choosing most elucidative training samples. In some cases, some form of controlled updation of classifier weights are proposed, this is advised by retraining with tuples which are misclassified. [20]-[23].

Incremental learning ability is very vital to machine learning approaches designed for solving real-world problems due to two reasons. Firstly, it is very difficult to gather all useful training samples before the trained system is put into use. Therefore when new instance are fed, the learning approach should have the ability of doing some revisions on the trained system so that unlearned knowledge encoded in those new examples can be included. Secondly, modifying a trained system may be cheaper in time cost than building a new system from scratch, which is valuable mainly in real-time applications [15][16].

Online learning is a way where one can capture knowledge from training instances which are already labelled and continuously updated. This type of learning is of importance in case of many analysis oriented applications [1].

It happens that customer changes proclivity as new services and products arrive. For instance, smart phones are preferred by people instead of commonly used phones.

Recommendable characteristics for incremental learning systems in online environments are:

1 .Capability of detecting modification in data without providing any explicit information about its change in the system.

2. Capability to recuperate from change in the state of instances already present and change hypothesis accordingly

3. Capability of reusing the foregoing experiences in circumstances whenever same problem arises again [30-33].

There is no requirement of storing each and every training sample as they arrive and process it again and again in online training environment. Here, instances arrive over time and new ones are processed, if they are identical to previous one in the dataset then no need to reprocess it. So "reusability" saves time and storage space too. So these algorithms are efficient and works well specially for huge dataset where number of instances are massive. If same thing is to be done by traditional batch algorithms then they will take number of passes to go through dataset and therefore will be costly. Various surveys are performed by researchers to find out applications where incremental ensembles algorithms can be used[24-25].

A classifier alone can perform according to particular steps, it can have its own advantages and some shortcomings too. So combining classifiers can be the new way which is also called as ensemble technique in which two or more classifiers are merged together to work in cooperation. At final step, decisions of all of them gets combined, before that they may work parallelly and independently. For combing the decisions also there are various schemes like majority voting, minimum probability, maximum probability and so on. In other words multiple new weak classifiers can be constructed for the part of sample space which is yet not seen. This can be alternative to approach of generating new node for each and every unknown feature. This allows to change fundamental idea of incremental algorithm which can build incremental algorithm which is not sensitive to the sequence of training instances or few adjustments of the parameters that can be set in algorithms [15][33].

By using idea of boosting in which a weak learner can be converted into strong learner for binary class problem where weak learner is simply a technique which is slightly better than random guessing in terms of performance can achieve low error rate, Freund et al proposed AdaBoost which works for multiclass problem and regression too. Concept of combining weak classifier came into picture, which gets benefited from instability of the weak classifier. With AdaBoost, it is possible to use even weak features for creating a pattern classifier, assuming you have a sufficient number of such features and assuming you just want to carry out binary classifications [22-24].

Learn++ is an incremental learning algorithm was inspired by the AdaBoost algorithm, originally developed to improve the classification performance of weak classifiers implementation and addresses the problem of incremental learning by creating diverse classifiers and combines their decisions through a weighted majority voting process. Each classifier's weight is based on the performance of a classifier on the entire training data[25].They use majority voting scheme at final stage to combine results of classifiers and get their final classification output [26].

3. METHODOLOGY

So what has been done here is, application of 4 well-known incremental algorithms on the student dataset created by us.

The algorithms which are used here are:

1) NaïveBayesUpdatable
2) IBK
3) KStar
4) K Nearest neighbor (NNGe)

By analyzing the results and performance of these algorithms, it is found that nearest neighbor algorithms (NNGe) works well among all of them. It has been proven in this paper by calculating accuracies of each algorithm.

3.1 Naïve Bayes

Bayesian classifiers are statistical classifiers. They can predict class membership probabilities, such as the probability that a given tuple belongs to a particular class. Naïve Bayes is the simplest form of Bayesian network, it assumes class conditional independence. That is, every attribute is independent of all the all other attributes, given the state of the class feature.

Naïve Bayes algorithm traditionally used in batch mode. In other words, naïve bayes algorithm used to demand all the training instances first and then perform final computations on it i.e. it requires "batch" (group) of instances at once. This is the algorithm that can work incrementally. It's incremental version is called NaiveBayesUpdateable. It works in steps by scanning training instances in each pass.

Learning method of Naïve Bayes consists of obtaining probabilities required to compute values of Bayes rule and predict class of that particular instance by analyzing probability.

Assume it makes one pass through entire training set. Let us have an example, it initializes some variables like count to zero in first step and then scans training instance one at a time. For each training tuple some features are assigned. So in next step it will increment the count and update values accordingly and in last step it may compute probabilities (conditional and prior) [15-17].

3.2 Nearest neighbor classification algorithms

Nearest neighbor classifiers are popular kind of lazy learning algorithms which are based on learning by analogy, that is, by comparing a given test instance with training instances that are similar to it. [6]

Lazy learners stores the training tuples and do nothing on training instances until it is fed with a test tuple. As soon as it observes that test tuple is provided, it perform generalization to classify the instance based upon its similarity to the instances which are fed during training i.e. instances which it has already stored. These methods do less work when training tuple is given and do more work when classifying data or while numeric prediction. Here, most of the learning is based on instances, these techniques are also called as "Instance based learners" [14].

The main benefit of applying lazy learning method is that the target function will be approximated locally such as in the k-nearest neighbor algorithm. As a result of this, lazy learners can solve more than one problems simultaneously and are able to deal with updates in the problem arena [13]. One more advantage of instance-based learners is that they are able to learn quickly from a very small dataset.

In nearest neighbor algorithms, number of training instances are described by n attributes. Each instance is a representation of a point in n-dimensional space forming pattern space of training tuples. When unknown instance comes, a KNN algorithm looks the pattern space for the k training tuples that are closest to new (unknown) instance [6].

3.2.1 IBK

IBk implements k-NN. It uses normalized distances for all attributes so that attributes on different scales have the same impact on the distance function. It may return more than k neighbors if there are ties in the distance. Neighbors are voted to form the final classification.

Euclidian distance metric is used in IBK for finding nearest neighbor. The number of nearest neighbors can be specified explicitly in the object editor or decided automatically using leave-one-out cross-validation focus to an upper limit given by the specified value. Different search algorithms can be used to increase the speed of the task of finding the nearest neighbors

3.2.2 KStar

Neighbors K* is an instance-based classifier, that is the class of a test instance is based upon the class of those training instances similar to it, an as determined by some similarity function. It differs from other instance-based learners in that it uses entropy-based distance function.

The approach in this case is to compute the distance between two instances is motivated by information theory. The intuition is that distance between instances can be determined by the complexity of transforming one instance into another. Complexity calculation is done in two steps. Mapping of one instance into another is done first and for finite set of transformations is done in next step. [8]

In many datasets there is common problem of missing values. It may require preprocessing to deal with missing values. KStar algorithm can handle such dataset which consists of missing values of one or more attribute. KStar manages this by assuming that missing values can be served as if they were, randomly drawn among the instances in the database. The technique used in this algorithm of summing probabilities over all possible paths solves the smoothness problem and thus leads to better performance [41]-[44].

3.2.2 NNGe

In instance based learning method a main issue is that if more instances are added to memory, classification time increases. To deal with such a situation generalized examples can be the choice. Instead of storing all examples exactly as they are, they can be integrated together so that number of instances can be minimized. If new instances arrives they can be treated in two ways:

1) Consolidated into already present generalized examples
2) Drop completely if covered already by a generalized example

Generalised exemplars are the one which is representation of more than one of the actual instances in the training set. There are generalization methods such as nested, overlapped etc.

Non-Nested Generalised Exemplars (NNGE) is an algorithm that generalises exemplars without nesting or overlap. NNGE is an extension of NGE which performs generalisation by combining exemplars which in turn forms hyper rectangles in sample space that represent conjunctive rules with internal disjunction. Generalisation is formed by NNGE each time a new example is added to the database, by joining it to its nearest neighbour of the same class.

NNGE learns incrementally by first classifying, then generalising each new instance.

It uses a modified Euclidean distance function that handles hyper rectangles, symbolic features, and exemplar and feature weights. Numeric feature values are normalised by dividing each value by the range of values observed. The class predicted is that of the single nearest neighbour. NNGE uses dynamic feedback to adjust exemplar and feature weights after each new example is classified. When classifying an example, one or more hyperrectangles may be found that the new example is a member of, but which are of the wrong class. NNGE prunes these so that the new example is no longer a member.[34]-[37]

4. DATASET DESCRIPON

Name of relation is student-evaluation.

Data collection method can be described as follows:

This dataset is created at our institute known as Dr. D.Y.Patil School Of Engineering and Technology Pune, Department of Computer engineering, for study purpose. For evaluation of students, test has been conducted in our institute. Among 29 attributes in the dataset, each one represents possible criteria for evaluating student's performance. They have been structured in the form of questionnaires that is Questions has been designed for assessment of students and values has been filled accordingly.

This dataset divides the data into 3 classes which in turn represents performance levels of student.

There are total 34 attributes and 3000 records (instances) in the dataset. This dataset contains numeric values. That means all the values in the Student evaluation dataset are numbers (integers). This data is an example of stream data. Other 4 attributes represent student_ID, repeat, attendance and difficulty_level respectively.

Last but not the least, class attribute is class_name , (class label) which have assigned to students based on rules generated by their performance.

Name of each attribute and its brief description has been covered in section below:

1. Student_ID: An integer which uniquely identifies student such as enrollment number. It takes value from {1, 2…}
2. Repeat: A number which represents frequency (number of times) the student is taking the course. It takes value from {0, 1, 2, 3…}
3. Attendance: A number which shows level of attendance student has presented during the course. It takes value from {0, 1, 2, 3…}
4. Difficulty: A number which shows level of difficulty. It takes value from{0,1,2,3,4,5}

Students have been evaluated out of 5 points {1, 2, 3, 4, 5} for each question described below.

Q1: Student's marks in subject1 of given course
Q2: Student's marks in subject2 of given course
Q3: Student's marks in subject3 of given course
Q4: Student's marks in subject4 of given course
Q5: Timely submission of assignments given for homework
Q6: Quality of contents of assignments given for homework
Q7: Laboratory performance level
Q8: Field work performance level
Q9: Quiz performance level
Q10: Project performance level
Q11: Interaction with teaching faculty during classroom sessions (raising doubts)
Q12: Arrived on time for classroom sessions
Q13: Level of Communication skills
Q14: Active participation in classroom discussion among classmates and teaching faculty
Q15: Participation and performance level in extracurricular activities
Q16: Student have shown positive approach towards the course
Q17: Regularity of study during course
Q18: Students active participation in technical events organized during the course
Q19: Students performance in midterm 1
Q20: Students performance in midterm 2
Q21: Students ability of leadership
Q22: Students ability to work in a team
Q23: Family background of a student (education level of family members)
Q24: Students locality background (rural, urban, semi urban; family members)
Q25: Students dedication towards the course
Q26: Student learned from the course and helped himself/herself to look at lite and world with a new perspective
Q27: Student followed the syllabus during studies and also did out of box learning to gain knowledge of something new and extra related to academics beyond contents of assigned syllabus
Q28: Students attitude towards learning (positive approach)
Q29: Utilization of assigned resources for study purpose properly (i.e. library (books, journals, magazines, research articles, video lectures, digital library etc.)
Q30: Disciplined behavior of student during the course
Q31: Student's nature (was it helpful and cooperative among other students taking same course? What was the level?)

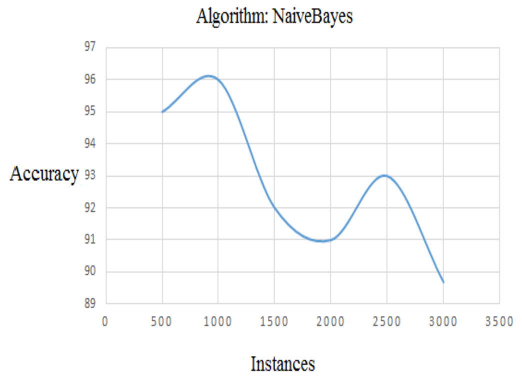
Class Attribute:
This is the final field of dataset consisting of 3 classes viz. class1, class2, class3

5. RESULTS AND COMPARISONS

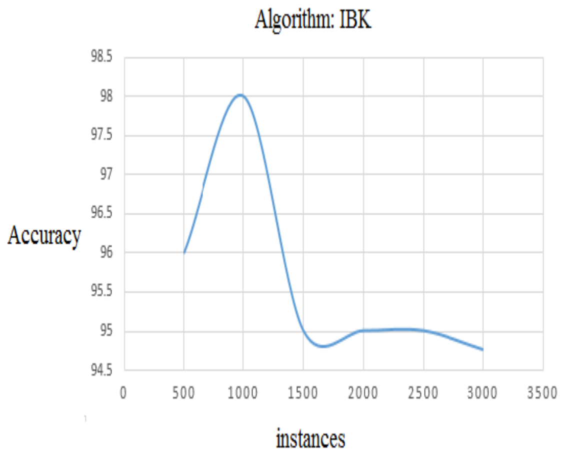
This section consists of the graphs which shows performance of each algorithm on student dataset. The results are obtained on each instance since this is an application of incremental algorithms and data. Weka's Knowledge Flow tool is used to build the model and test the system of algorithms on student evaluation application. This is done for simulation purpose.

Number of instances are plotted on x-axis and Accuracy is plotted on y-axis. All the graphs show the performance in terms of accuracies. Name of the algorithm is shown above the graph

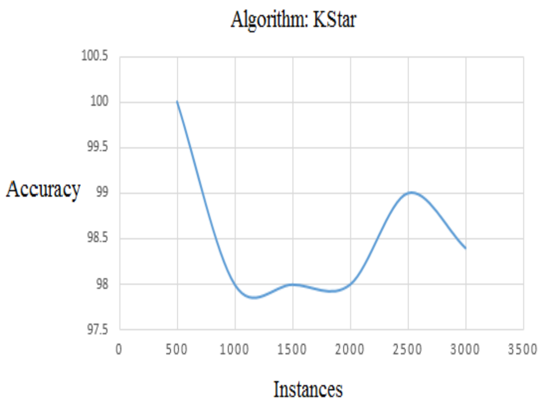
1) NaiveBayes



2) IBK



3) KStar



4) NNGe

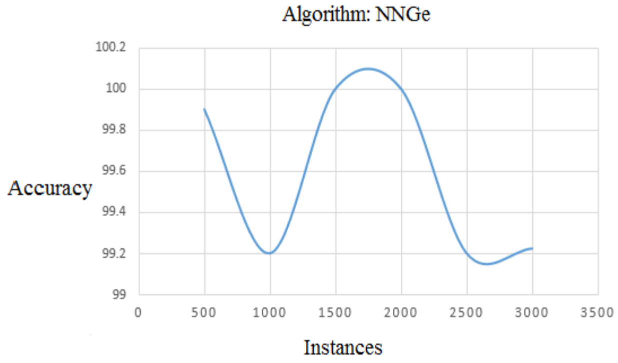


Table 1. Summary of performance of Incremental algorithms on Student dataset

	Naïve Bayes	IBK	KStar	Nearest Neighbor
Accuracy (in %)	89.6959	94.7635	98.0405	99.223
Time taken to evaluate model (in sec)	3	3	260	2

Table 2. Summary of performance of Incremental algorithms on Student dataset

	Naïve Bayes	IBK	KStar	Nearest Neighbor
Incorrectly classified instances (Error) (in %)	10.3041	5.2365	1.9595	0.777
Kappa statistic	0.8309	0.9144	0.9675	0.9871

By observing the accuracy (table 1), it is clear that NNGe algorithm performs best and comparatively Naïve Bayes is in lowest rank in terms of performance. Naïve Bayes can work well when dataset is small. Here, for experiments the algorithms are tested on datasets containing 3000 instances. It is also seen that this conclusion which we have drawn are true for larger datasets. They extremely work well on stream data.

Table 3. Summary of performance of Incremental algorithms on Student dataset (errors)

	Naïve Bayes	IBK	KStar	Nearest Neighbor
Mean absolute error	0.0693	0.038	0.0152	0.0052
Root mean Squared error	0.2482	0.1874	0.1111	0.072

Table 2 and table 3 show performance of algorithms in terms of different kind of errors like incorrectly classified instances, mean absolute error, and root mean squared error. Table 2 shows kappa statistic. These two tables support our statement about the performance of NNGe algorithm, and proves that NNGe is best in such an environment in student data prediction application.

The incremental algorithm chosen here are the one which gives good results and are standard one. While taking out readings of the above graph, it can be seen very clearly how the data arrives,i.,e.in incremental way. Batch of 500 instances is stored and one by one each batch is taken. There can also be the case where we can take data single instance wise. But these two are different approaches and can be applied according to nature of the data. In other words, these are two ways of how incremental data is processed.

These algorithms are tested in this experiment on student dataset, and prediction is done. In the same way it is desirable to test them in another stream data application.

6. CONCLUSION AND FUTURE SCOPE

The development in the area of education is a new inspiration for incremental learning algorithms. Real world applications can use such a dynamic strategies to design solutions to the problems and utilize available resources. Such dynamic systems continuously receiving recent samples of data which in turn will be stored. There is area known as online learning whereby each training sample is examined only once. So in incremental applications online learning is necessary instead of batch learning. Here, in this paper, a student dataset has been built up to test their performance which is an application of incremental environment. Four incremental algorithms i.e. NaiveBayes, IBK, KStar and NNGE are applied one by one (in instance mode of course) on the student evaluation dataset. Algorithm's performance is recorded in the form of accuracy, time taken to build and execute. It is observed that KStar classifies well, but requires more time as compared to NaiveBayes and IBK .But among all of them, NNGE algorithm worked nicely in terms of accuracy and also took less time. This paper suggests to use NNGe algorithm for evaluating student's performance like in datasets which has been used here. In future, ensemble technique can be incorporated with incremental learning for achieving better result.

7. REFERENCES

[1] Fong, Simon, Zhicong Luo, and Bee Wah Yap. "Incremental Learning Algorithms for Fast Classification in Data Stream." In Computational and Business Intelligence (ISCBI), 2013 International Symposium on, pp. 186-190. IEEE, 2013

[2] Smith, Michael R., and Tony Martinez. "Improving classification accuracy by identifying and removing instances that should be misclassified." In Neural Networks (IJCNN), The 2011 International Joint Conference on, pp. 2690-2697. IEEE, 2011.

[3] Bunkar, Kamal, U. K. Singh, B. Pandya, and Rajesh Bunkar. "Data mining: Prediction for performance improvement of graduate students using classification." In Wireless and Optical Communications Networks (WOCN), 2012 Ninth International Conference on, pp. 1-5. IEEE, 201.

[4] Bunkar, Kamal, U. K. Singh, B. Pandya, and Rajesh Bunkar. "Data mining: Prediction for performance improvement of graduate students using classification." In Wireless and Optical Communications Networks (WOCN), 2012 Ninth International Conference on, pp. 1-5. IEEE, 2012M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[5] Bhardwaj, Brijesh Kumar, and Saurabh Pal. "Data Mining: A prediction for performance improvement using classification." arXiv preprint arXiv:1201.3418 (2012).

[6] Nguyen N., Paul J., and Peter H., A Comparative Analysis of Techniques for Predicting Academic Performance. In Proceedings of the 37th ASEE/IEEE Frontiers in Education Conference. pp. 7-12, 2007.

[7] J. Han and M. Kamber, "Data Mining: Concepts and Techniques," Morgan Kaufmann, 2000

[8] Delavari N. & Beikzadeh M. R & Shirazi M. R. A., "A New Model for Using Data Mining in Higher Educational System", in Proceedings of 5th International Conference on Information Technology Based Higher Education and Training (ITHET), Istanbul, Turkey, May 31 to June 2, 2004

[9] John G. Cleary, Leonard E. Trigg: K*: An Instance-based Learner Using an Entropic Distance Measure. In: 12th International Conference on Machine Learning, 108-114, 1995.

[10] Cover, Thomas, and Peter Hart. "Nearest neighbor pattern classification." Information Theory, IEEE Transactions on 13, no. 1 (1967): 21-27

[11] Martin, Brent. "Instance-based learning: nearest neighbour with generalisation." PhD diss., University of Waikato, 1995.

[12] Roy, Sylvain. "Nearest neighbor with generalization." Christchurch, New Zealand (2002).

[13] Kaushik H. Raviya, Biren Gajjar, "Performance Evaluation of Different Data Mining Classification Algorithm Using WEKA".

[14] Ms S. Vijayarani1, Ms M. Muthulakshmi, Comparative Analysis of Bayes and Lazy Classification Algorithms, International Journal of Advanced Research in Computer and Communication Engineering,Vol. 2, Issue 8, August 2013

[15] George H. John, Pat Langley: Estimating Continuous Distributions in Bayesian Classifiers. In: Eleventh Conference on Uncertainty in Artificial Intelligence, San Mateo, 338-345, 1995

[16] Kotsiantis, Sotiris B. "An incremental ensemble of classifiers." Artificial Intelligence Review 36, no. 4 (2011): 249-266.

[17] Garg, Bandana. Design and Development of Naive Bayes Classifier. Diss. North Dakota State University, 2013

[18] Karim, Masud, and Rashedur M. Rahman. "Decision Tree and Naïve Bayes Algorithm for Classification and Generation of Actionable Knowledge for Direct Marketing." Journal of Software Engineering & Applications 6, no. 4 (2013).

[19] D. Aha, D. Kibler (1991). Instance-based learning algorithms. Springer, Machine Learning. M:January 1991, Volume 6, Issue 1, pp 37-66

[20] Kotsiantis, S., Christos Pierrakeas, and P. Pintelas. "Predicting Student;s 'Performance In Distance Learning Using Machine Learning Techniques." Applied Artificial Intelligence 18, no. 5 (2004): 411-426.

[21] E. H. Wang and A. Kuh, "A smart algorithm for incremental learning," in Proc. Int. Joint Conf. Neural Netw., vol. 3, 1992, pp. 121–126

[22] B. Zhang, "An incremental learning algorithm that optimizes networsize and sample size in one trial," in Proc. IEEE Int. Conf. Neural Netw.,1994, pp. 215–220.

[23] F. S. Osorio and B. Amy, "INSS: A hybrid system for constructive machine learning," Neurocomput., vol. 28, pp. 191–205, 1999. R. Schapire, "Strength of weak learning," Machine Learn., vol. 5, pp. 197–227, 1990

[24] Y. Freund and R. Schapire, "A decision theoretic generalization of on-line learning and an application to boosting," Comput. Syst. Sci., vol. 57, no. 1, pp. 119–139, 1997

[25] R. Schapire,Y. Freund, P. Bartlett, andW. S. Lee, "Boosting the margins: A new explanation for the effectiveness of voting methods," Ann. Stat., vol. 26, no. 5, pp. 1651–1686, 1998

[26] Learn++: An Incremental Learning Algorithm for Supervised Neural Networks, Robi Polikar, Member, IEEE, Lalita Udpa, Senior Member, IEEE, Satish S. Udpa, Senior Member, IEEE, and Vasant Honavar, IEEE transactions on systems, man, and cybernetics—part c: applications and reviews, vol. 31, no. 4, november 2001

[27] N. Littlestone and M.Warmuth, "Weighted majority algorithm," Inform. Comput., vol. 108, pp. 212–261, 1994.

[28] Romero, Cristóbal, and Sebastian Ventura. "Educational data mining: A survey from 1995 to 2005." Expert Systems with Applications 33, no. 1 (2007): 135-146

[29] Arruabarrena, R., Pe´rez, T. A., Lo´pez-Cuadrado, J., & Vadillo, J. G. J.(2002). On evaluating adaptive systems for education. In Adaptive hypermedia (pp. 363–367).

[30] Ingram, A. (1999). Using web server logs in evaluating instructional web sites. Journal of Educational Technology Systems, 28(2), 137–57.

[31] T. Kidera, S. Ozawa, S. Abe, An incremental learning algorithm of ensemble classifier systems, in: Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN'06), 2006, pp. 3421–3427.

[32] J. Macek, Incremental learning of ensemble classifiers on ECG data, in: Proceedings of the IEEE 18th Symposium on Computer-Based Medical Systems (CBMS'05), Dublin, 2005 pp. 315–320

[33] M.D. Muhlbaier, R. Polikar, An ensemble approach for incremental learning in nonstationary environments, in: 7th International Workshop on Multiple Classifier Systems, Prague, Lecture Notes in Computer Science, vol. 4472, 2007,pp. 490–500

[34] R. Polikar, S. Krause, L. Burd, Ensemble of classifiers based incremental learning with dynamic voting weight update, in: Proceedings of the IEEE International Join Conference on Neural Networks (IJCNN'03), 2003, pp.2770–2775

[35] Luo, Jianhui, et al. "Model-based prognostic techniques [maintenance applications]." AUTOTESTCON 2003. IEEE Systems Readiness Technology Conference. Proceedings. IEEE, 2003

[36] Martin, Brent. "Instance-based learning: nearest neighbour with generalisation." PhD diss., University of Waikato, 1995.

[37] Sylvain Roy (2002). Nearest Neighbor With Generalization. Christchurch, New Zealand.

[38] Salzberg, Steven. "A nearest hyperrectangle learning method." *Machine learning* 6, no. 3 (1991): 251-276.

[39] B.K. Bharadwaj and S. Pal. "Data Mining: A prediction for performance improvement using classification", International Journal of Computer Science and Information Security (IJCSIS), Vol. 9, No. 4, pp. 136-140, 2011

[40] Ramaswami M., and Bhaskaran R., CHAID Based Performance Prediction Model in Educational Data Mining, IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 1, No. 1, 2010.

[41] Shannaq, B., Rafael, Y. and Alexandro, V. (2010) 'Student Relationship in Higher Education Using Data Mining Techniques', Global Journal of Computer Science and Technology, vol. 10, no. 11, pp. 54-59.

[42] Martens, David, Bart Baesens, and Tom Fawcett. "Editorial survey: swarm intelligence for data mining." *Machine Learning* 82.1 (2011): 1-42.

[43] Frank, Eibe, et al. "Weka." *Data Mining and Knowledge Discovery Handbook*. Springer US, 2005. 1305-1314.

[44] Rajeswari, P., and G. Reena. "Analysis of liver disorder using data mining algorithm." *Global Journal of Computer Science and Technology* 10, no. 14 (2010).

[45] Norén, G. Niklas, Johan Hopstadius, Andrew Bate, Kristina Star, and I. Ralph Edwards. "Temporal pattern discovery in longitudinal electronic patient records." *Data Mining and Knowledge Discovery* 20, no. 3 (2010): 361-387.