Some Intelligent Computing Methods for Classification of Students in ITS

[1]Kiran Mishra, [2]Bhavtosh Mishra, [3]Nikita Kanwar
[1]Dept of CSE, NIT Bhopal, Madhya Pradesh, India
[2]Dept. of Information Technology, Rajkiya Engineering College Banda, UP, India
[2]Centre for Converging Technologies, Jaipur, Rajasthan, India

Abstract
Categorization or classification of students based on their performance is an essential aspect of an intelligent tutoring system (ITS). Different intelligent computing systems as ANN, GA and data mining methods have been deployed in such problem. In this paper, different computing methods such as: Bayesian network, Naïve Bayes, Ada boost. Logistic regression and SMO have been utilized. The classification is based on students' performance on a test in a subject with three types of content A, R, D, AR, AD and ARD. We obtain highest accuracy for Bayesian network and lowest for AdaboostM1. The other methods have accuracy in between them.

Keywords
Data mining, Intelligent Computing, Intelligent Tutoring System, Classification, and Students

I. Introduction

In Intelligent Tutoring System (ITS), categorization of student is an important issue, in the sense that reading material is fetched to them depending on their capability and performance in a test for the subject.

A subject has different difficulty levels such as: very difficult, difficult, and easy. In correspondence to these levels a student may be categorized into excellent good and average. It is advisable and worthwhile that there should be one to one correspondence between the difficulty level of a subject and the capability level of a student. This paper refers to the categorization of students based on their performance in a subject which has seven levels of its content types. The seven types of content are Analytical (A), Reasoning (R), Descriptive (D), AR, AD, AR and ARD. Some of the work done in ITS are given below which reflect student performance measure and its implementation in ITS.

Han et al. (Han et al., 2005) proposed an intelligent tutoring system using case-based student modeling. The proposed system can effectively infer the state of the student's knowledge. The knowledge state is diagnosed through the cases that are generated when the student solves a problem. They have chosen a procedural learning in the physics and designed this domain.

Rongmei et al. (Rongmei et al., 2009) designed model of Distributed Intelligent Internet Tutoring System based on MAS and CBR .They discussed the key technologies such as the establishment of knowledge model based on semantic network, the establishment of cognitive ability of student model, the establishment of individual instruction strategies model.

Rishi et al. (Rishi et al., 2007) conceptualized a Case Based Distributed Student Modeling (agent based) ITS architecture to support student-centered, self-paced, and highly interactive learning. In the system the first step in building an effective learning environment is building a Case Base where the system maintains a rich set of cases (scenario) of student's learning pattern, and employs an efficient and flexible case retrieval system.

Schiaffino et al. (Schiaffino et al., 2008) present eTeacher, an intelligent agent that provides personalized assistance to e-learning students. eTeacher observes a student's behavior while he/she is taking online courses and automatically builds the student's profile. This profile comprises the student's learning style and information about the student's performance, such as exercises done, topics studied, and exam results. In their approach, a student's learning style is automatically detected from the student's actions in an e-learning system using Bayesian networks. Then, eTeacher uses the information contained in the student profile to proactively assist the student by suggesting him/her personalized courses of action that will help him/her during the learning process.

Kerly et al. (Kerly et al., 2008), describes a system which incorporates natural language technologies, database manipulation and educational theories in order to offer learners a Negotiated Learner Model, for integration into an Intelligent Tutoring System. The system presents the learner with their learner model, offering them the opportunity to compare their own beliefs regarding their capabilities with those inferred by the system. A conversational agent, or "chatbot" has been developed to allow the learner to negotiate over the representations held about them using natural language.

Baylari et al. (Baylari et al., 2009) proposed a personalized multi-agent e-learning system based on item response theory (IRT) and artificial neural network (ANN) which presents adaptive tests (based on IRT) and personalized recommendations (based on ANN). These agents add adaptivity and interactivity to the learning environment and act as a human instructor which guides the learners in a friendly and personalized teaching environment.

Lee et al. (Lee et al., 2009) proposed to analyze learners' preferences with a data mining technique. A decision tree was created to illustrate the learning behavior of each cognitive style group. The frequency of using the backward/forward button is the most important feature for classifying the cognitive styles and the second most important features are the frequencies of using overviews and those of having repeated visiting. Findings in their study show that Field Independent learners frequently use backward/forward buttons and spent less time for navigation. On the other hand, Field Dependent learners often use main menu and have more repeated visiting.

Romero et al. (Romero et al., 2005) described the use of data mining methods in e-learning system for providing feedback to courseware authors. The discovered information is presented in the form of prediction rules since these are highly comprehensible and they show important relationships among the presented data. The rules will be used to improve courseware, especially Adaptive Systems for Web-based Education (ASWE). They proposed to use evolutionary algorithms as rule discovery methods, concretely Grammar-Based Genetic Programming (GBGP) with multi objective optimization techniques. They have developed a specific tool named EP Rules (Education Prediction Rules) to facilitate and simplify the knowledge discovery process for usage data in web-based education systems.

Romero et al. (Romero et al., 2007) described a personalized recommender system that uses web mining techniques for recommending a student which (next) links to visit within an adaptable educational hypermedia system. They presented a specific mining tool and a recommender engine that they have integrated in the AHA! system in order to help the teacher to carry out the whole web mining process. They reported on several experiments with real data in order to show the suitability of using both clustering and sequential pattern mining algorithms together for discovering personalized recommendation links.

Merceron et al. (Merceron et al, 2005) used data mining algorithms for discovering pedagogically relevant knowledge contained in databases obtained from Web-based educational systems. These findings can be used both to help teachers with managing their class, understand their students' learning and reflect on their teaching and to support learner reflection and provide proactive feedback to learners.

Kristofic et al. (Kristofic et al., 2005) presented techniques for data mining, which can be used to discover knowledge about students' behavior during learning, as well as techniques, which take advantage of such knowledge to recommend students lessons they should study next. They also described a process of recommendation based on knowledge discovery and present architecture of a web-based system, which uses proposed approach to improve adaptation. Proposed architecture is independent of actual adaptive hypermedia system used.

Hamalainen et al., 2006 designed and implemented a Data Mining System to analyze the study records of two programming courses in a distance curriculum of Computer Science. Various data mining schemes, including the linear regression and probabilistic models, were applied to describe and predict student performance. The results indicate that a DMS can help a distance education teacher, even in courses with relatively few students, to intervene in a learning process at several levels: improving exercises, scheduling the course, and identifying potential dropouts at an early phase.

Mishra & Mishra (2010; 2011) developed a data mining based evaluation method of students performance in learning and characteristics of students and tutor subjects modules in ITS.

II. Classifiers

Different computing methods for classifications are given below. The number in the bracket refers to the website for the classifier method.

A. Naive Bayes [11]

The naive Bayes classifier applies to learning tasks where each instance x is described by a conjunction of attribute values and where the target function f (x) can take on any value from some finite set V. A set of training examples of the target function is provided, and a new instance is presented, described by the tuple of attribute values ($a_1, a_2 \dots a_n$). The learner is asked to predict the target value, or classification, for this new instance. The naive Bayes classifier is based on the simplifying assumption that the attribute values are conditionally independent given the target value. In other words, the assumption is that given the target value of the instance, the probability of observing the conjunction $a_1, a_2 \dots a_n$, is just the product of the probabilities for the individual attributes:

$$P(a_1, a_2 \dots a_n | v_j) = \prod_i P(a_i | v_j)$$

$$v_{NB} = \underset{v_j \in V}{argmax} P(v_j) \prod_i P(a_i | v_j)$$

Where, v_{NB} denotes the target value output by the naive Bayes classifier. In a naive Bayes classifier the number of distinct $P(a_i | v_j)$ terms that must be estimated from the training data is just the number of distinct attribute values times the number of distinct target values-a much smaller number than if we were to estimate the $P(a_1, a_2 \dots a_n | v_j)$ terms as first contemplated.

B. Sequential Minimal Optimization (SMO)[12]

SMO solves the SVM-QP (support vector machine quadratic programming) problem by decomposing it into SVM-QP Sub-problems and solving the smallest possible optimization problem, involving the two Lagrange multipliers, at each step. A Quadratic Problem is maximizing or minimizing a quadratic objective function subject to a set of linear constraints.

SMO is an iterative algorithm for solving the optimization problem. SMO breaks this problem into a series of smallest possible sub-problems, which are then solved analytically. Because of the linear equality constraint involving the Lagrange multipliers a_i the smallest possible problem involves two such multipliers. Then, for any two multipliers a_1 and a_2, the constraints are reduced to:

$$0 \le a_1, a_2 \le C,$$

$$y_1 a_1 + y_2 a_2 = k,$$

Also, this reduced problem can be solved analytically: one needs to find a minimum of a one-dimensional quadratic function. k is the negative of the sum over the rest of terms in the equality constraint, which is fixed in each iteration.

C. Ada-Boost Classifier [14]

Boosting refers to a general and provably effective method of producing a very accurate prediction rule by combining rough and moderately inaccurate rules of thumb. Boosting has its roots in a theoretical framework for studying machine learning called the ―PAC‖ learning model.

The Ada Boost algorithm, solved many of the practical difficulties of the earlier boosting algorithms (which came up with the first provable polynomial-time AdaBoost.M1 Classifier boosting algorithm takes as input a training set of N examples S= $((x_1, y_1)$ ……., $(x_m, y_m))$ where x_i is an instance drawn from some space X and represented in some manner (typically, a vector of attribute values), and $y_i \in Y$ is the class label associated with x_i . The set of possible labels Y is of finite cardinality k. Also, the boosting algorithm has access to another unspecified learning algorithm, called the weak learning algorithm, which is denoted generically as WeakLearn. The boosting algorithm calls WeakLearn repeatedly in a series of rounds. On round t, the booster provides WeakLearn with a distribution D_t over the training set S. In response; WeakLearn computes a classifier or hypothesis $h_t: X \rightarrow Y$ which should correctly classify a fraction of the training set that has large probability on D_t.

D. Bayesian Network [13]

A Bayesian network is a probabilistic directed acyclic graph composed of a set of nodes and a set of edges between nodes. Nodes represent the random data and edge between two nodes represents conditional dependency among nodes. Learning the process of Bayesian network is a two-step process: learning network, learning relationship among data.

- Take domain variable and partitioned into:
Set of attribute, $X = \{X_1, X_2, \dots \dots, X_n\}$ and Class variable, C
- Find the value of attribute X, given C:
Take an example x, compute predictive distribution $P(C, x|S)$ by marginalizing joint distribution

$$P(C,x|S) = \frac{P(C,x|S)}{P(x|S)} \alpha P(C, x|S)$$

- Apply Bayesian Network classification rule:

$$C^* = h_{BNC}(x) = argmax_{j=1\dots m} P(x, c_j|S, \theta_s)$$

Computing $P(x, c_j)$ for each class c_j by joint probability distribution:

$$P(x, c_j, |S, \theta_s) = \prod_{i=1}^{n} P(X_i|pa(X_i)).P(C = c_j|pa(C))$$

E. Logistic [15]

Logistic classification applies to the data set where dependent and independent attribute are dichotomous. The data set used for depression prediction has binary outcome so it cannot be modeled using linear regression. For such data logistic regression is required. The logistic function is used in this model to predict the output of an experiment.

$$f(x) = \frac{1}{1 + e^{-x}}$$

Each attribute or variable or feature has some contribution in the prediction of expected outcome in a data set. The predictive capability of each attribute is measured using maximum likelihood estimation statistics. The logistic model calculates the probability of prediction of a binary outcome using input data set. This model uses likelihood ratio and Wald test to test statistical significance.

III. Data Set Preparation

We have prepared the data set obtained from some test experiments for leaning of C++ by a group of students Mishra & Mishra 2010. The data set consist of 120 instances of different students' performance in seven types of contents as given in the seven columns of the Table 1. The 8th column contain the averaged score and the 9th column contains the category of the student as excellent, very good, good, average and below average respectively in Table 1.

Table 1: Data Set

A	R	D	AR	AD	RD	ARD	score	Result
0.4	0.5	0.6	0.5	0.3	0.5	0.4	0.6	Good
0.5	0.8	0.9	1	0.4	0.6	0.7	0.8	Very Good
0.3	0.4	0.3	0.4	0.2	0.3	0.1	0.3	Below Average
0.4	0.5	0.6	0.4	0.7	0.3	0.6	0.7	Good
0.8	0.9	0.5	1	0.3	0.6	0.9	0.8	Very Good
0.5	0.6	0.9	0.6	0.3	0.6	0.4	0.6	Good
0.4	0.2	0.3	0.4	0.1	0.3	0.4	0.5	Average
0.6	0.5	0.4	0.7	0.6	0.3	0.4	0.6	Good
0.9	0.6	0.5	1	0.6	0.9	0.7	1	Excellent
0.3	0.5	0.4	0.6	0.3	0.7	0.6	0.5	Average
0.3	0.4	0.3	0.4	0.1	0.3	0.2	0.3	Below Average
0.3	0.5	0.6	0.4	0.7	0.4	0.6	0.5	Average
0.9	1	0.7	1	0.3	0.6	0.8	0.8	Very Good
0.6	0.7	0.9	0.5	0.3	0.6	0.3	0.7	Good
0.4	0.2	0.3	0.5	0.3	0.4	0.6	0.5	Average
0.9	0.6	0.5	0.3	0.5	0.6	0.7	0.8	Very Good
0.6	0.9	0.6	0.6	0.4	0.6	0.7	0.7	Good
0.4	0.5	0.9	0.5	0.6	0.4	0.5	0.6	Good
0.6	0.6	1	0.6	0.3	0.5	0.3	0.5	Average
0.3	0.1	0.4	0.4	0.2	0.4	0.3	0.3	Below Average
0.5	0.6	0.3	0.7	0.5	0.6	0.4	0.6	Good
1	0.5	0.6	0.9	0.6	0.8	0.5	1	Excellent
0.4	0.7	0.4	0.6	0.3	0.5	0.7	0.7	Good
0.7	0.6	0.9	0.5	0.4	0.6	0.4	0.8	Very Good
0.3	0.1	0.3	0.4	0.1	0.3	0.4	0.4	Below Average
0.6	0.6	0.3	0.7	0.7	0.4	0.4	0.7	Good

 ISSN : 0976-8491 (Online) | ISSN : 2229-4333 (Print)

0.9	0.6	0.7	1	0.6	0.8	0.6	0.9	Excellent
0.3	1	0.6	0.5	0.9	0.7	0.6	0.8	Very Good
0.8	1	0.9	0.8	0.9	0.9	0.8	1	Excellent
1	0.8	1	0.9	0.8	0.8	0.9	1	Excellent
0.8	0.7	0.3	0.6	0.4	0.6	0.9	0.6	Good
0.8	0.6	0.4	0.9	0.5	0.6	0.7	0.6	Good
0.3	0.4	0.3	0.4	0.2	0.3	0.1	0.3	Below Average
0.5	0.9	0.8	0.6	0.9	0.4	1	0.8	Very Good
0.3	0.4	0.3	0.1	0.9	0.6	0.4	0.5	Average
0.4	1	0.7	0.8	0.9	1	0.6	0.8	Very Good
0.7	0.7	0.6	0.9	0.5	0.3	0.5	0.6	Good
0.9	0.3	0.4	0.3	0.5	0.3	0.3	0.4	Average
0.9	0.5	0.6	0.8	0.7	0.9	0.5	0.7	Good
0.7	0.2	0.3	0.2	0.8	0.9	0.5	0.5	Average
0.3	0.4	0.3	0.4	0.1	0.3	0.1	0.2	Below Average
0.9	0.9	0.5	0.7	0.6	0.8	0.5	0.6	Good
0.5	0.7	0.3	0.4	0.9	0.7	0.8	0.7	Good
0.3	0.3	0.4	0.4	0.2	0.9	1	0.5	Average
0.7	0.5	0.5	0.9	0.3	0.4	0.4	0.5	Average
0.9	0.7	0.5	0.3	0.5	0.7	0.6	0.6	Good
1	0.9	0.9	0.8	0.5	0.6	0.8	0.8	Very Good
0.6	0.8	0.3	0.4	0.4	0.3	0.3	0.4	Average
0.6	0.7	0.9	0.5	0.3	0.5	0.3	0.5	Average
0.3	0.2	0.3	0.4	0.1	0.4	0.3	0.2	Below Average
0.7	0.5	0.9	0.5	0.6	0.8	0.9	0.7	Good
0.3	0.4	0.5	0.9	0.4	0.6	0.5	0.5	Average
0.7	0.7	0.5	0.6	0.9	0.9	0.9	0.7	Good
0.7	0.8	0.7	0.7	0.8	0.7	0.8	0.7	Good
0.3	0.1	0.1	0.3	0.4	0.5	0.6	0.3	Below Average
0.5	0.5	0.8	0.6	0.6	0.3	0.4	0.8	Very Good
0.9	0.5	0.6	0.8	0.7	0.9	0.7	0.7	Good
0.4	0.8	0.5	0.6	0.9	0.6	0.7	0.6	Good
0.9	1	1	1	0.8	0.9	0.8	1	Excellent
1	0.9	1	1	0.9	0.8	1	1	Excellent
0.6	0.5	0.6	0.5	0.8	0.7	0.4	0.6	Good
0.5	0.8	0.9	1	0.4	0.6	0.7	0.8	Very Good
0.3	0.4	0.3	0.4	0.2	0.3	0.1	0.3	Below Average
0.4	0.5	0.6	0.4	0.7	0.3	0.6	0.7	Good
0.8	0.9	0.5	1	0.3	0.6	0.9	0.8	Very Good
0.5	0.6	0.9	0.6	0.3	0.6	0.4	0.6	Good
0.4	0.2	0.3	0.4	0.1	0.3	0.4	0.5	Average
0.6	0.5	0.4	0.7	0.6	0.3	0.4	0.6	Good
0.9	0.8	0.9	1	1	0.9	0.9	1	Excellent
0.3	0.5	0.4	0.6	0.3	0.7	0.6	0.5	Average
0.3	0.2	0.3	0.1	0.1	0.3	0.2	0.3	Below Average
0.3	0.5	0.6	0.4	0.7	0.4	0.6	0.5	Average
0.9	1	0.9	1	0.8	0.6	0.8	0.8	Very Good

0.6	0.7	0.9	0.7	0.7	0.6	0.7	0.7	Good
0.4	0.2	0.3	0.5	0.5	0.4	0.6	0.5	Average
0.9	0.6	0.8	0.3	0.8	0.7	0.8	0.8	Very Good
0.6	0.9	0.7	0.8	0.7	0.9	0.7	0.7	Very Good
0.7	0.6	0.9	0.6	0.7	0.6	0.7	0.6	Good
0.6	0.6	1	0.6	0.7	0.5	0.6	0.5	Average
0.3	0.1	0.4	0.4	0.2	0.4	0.3	0.3	Below Average
0.5	0.6	0.3	0.7	0.5	0.6	0.4	0.7	Very Good
1	0.5	0.6	0.9	0.6	0.8	0.5	1	Excellent
0.4	0.7	0.4	0.6	0.3	0.5	0.7	0.8	Very Good
0.7	0.6	0.9	0.5	0.4	0.6	0.4	0.8	Very Good
0.3	0.1	0.3	0.4	0.1	0.3	0.4	0.4	Below Average
0.7	0.6	0.8	0.7	0.7	0.6	0.7	0.7	Good
0.9	1	0.8	1	1	0.8	0.9	0.9	Excellent
0.9	1	0.8	0.9	0.9	0.7	0.8	0.8	Very Good
0.8	1	0.9	0.8	0.9	0.9	0.8	1	Excellent
1	1	1	0.9	1	1	0.9	1	Excellent
0.8	0.7	0.7	0.6	0.4	0.6	0.9	0.6	Good
0.8	0.6	0.4	0.9	0.5	0.6	0.7	0.6	Good
0.3	0.1	0.3	0.3	0.2	0.2	0.1	0.3	Below Average
0.5	0.9	0.8	0.6	0.9	0.8	1	0.8	Very Good
0.3	0.4	0.3	0.1	0.9	0.6	0.4	0.5	Average
0.7	1	0.7	0.8	0.9	1	0.7	0.8	Very Good
0.7	0.7	0.6	0.9	0.7	0.6	0.7	0.6	Good
0.7	0.3	0.4	0.3	0.5	0.6	0.3	0.4	Average
0.9	0.5	0.6	0.8	0.7	0.9	0.5	0.7	Very Good
0.7	0.6	0.3	0.7	0.8	0.9	0.5	0.5	Average
0.3	0.4	0.3	0.4	0.1	0.3	0.1	0.2	Good
0.9	0.9	0.5	0.7	0.6	0.8	0.5	0.6	Good
0.5	0.7	0.6	0.9	0.9	0.7	0.8	0.7	Very Good
0.3	0.3	0.4	0.4	0.2	0.9	1	0.5	Good
0.7	0.5	0.5	0.9	0.3	0.4	0.4	0.5	Good
0.9	0.7	0.5	0.3	0.5	0.7	0.6	0.6	Good
1	0.9	0.9	0.8	0.5	0.6	0.8	0.8	Very Good
0.6	0.8	0.3	0.4	0.4	0.3	0.3	0.4	Average
0.6	0.7	0.9	0.5	0.3	0.5	0.3	0.5	Average
0.3	0.2	0.3	0.4	0.1	0.4	0.3	0.2	Below Average
0.7	0.5	0.9	0.5	0.6	0.8	0.9	0.7	Good
0.3	0.4	0.5	0.9	0.4	0.6	0.5	0.5	Average
0.7	0.7	0.5	0.6	0.9	0.9	0.9	0.7	Good
0.7	0.8	0.7	0.7	0.8	0.9	0.8	0.8	Very Good
0.3	0.1	0.1	0.3	0.4	0.5	0.6	0.3	Below Average
0.5	0.5	0.8	0.6	0.6	0.3	0.4	0.8	Very Good
0.9	0.5	0.6	0.8	0.7	0.9	0.7	0.7	Very Good
0.9	0.8	1	0.6	0.9	1	0.9	1	Excellent
0.9	1	1	1	0.8	0.9	0.8	1	Excellent
1	0.9	1	1	0.9	0.8	1	1	Excellent

 ISSN : 0976-8491 (Online) | ISSN : 2229-4333 (Print)

IV. Result

The results consisting of True Positive (TP) rate, false Positive Rate (FP), precision, recall, F-measure, MCC, ROC area, PRC area are given below of each classification method as shown from Table 2 obtained by using WEKA tools.

Table 2: Bayes Network: Accuracy is 85%

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.744	0.049	0.879	0.744	0.806	0.728	0.924	0.860	**Good**
0.815	0.065	0.786	0.815	0.800	0.741	0.959	0.917	**Very Good**
1.000	0.038	0.789	1.000	0.882	0.871	0.990	0.853	**Below Average**
0.875	0.021	0.913	0.875	0.894	0.868	0.994	0.975	**Average**
1.000	0.019	0.882	1.000	0.938	0.930	1.000	1.000	**Excellent**

Table 3: Naive Bayes: Accuracy is 84.16%

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.846	0.160	0.717	0.846	0.776	0.661	0.918	0.843	**Good**
0.667	0.032	0.857	0.667	0.750	0.697	0.969	0.914	**Very Good**
1.000	0.010	0.938	1.000	0.968	0.964	0.990	0.852	**Below Average**
0.833	0.021	0.909	0.883	0.870	0.840	0.987	0.943	**Average**
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	**Excellent**

Table 4: Logistic Classifier: Accuracy is 83.33%

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.795	0.074	0.838	0.795	0.816	0.731	0.931	0.912	**Good**
0.852	0.054	0.821	0.852	0.836	0.788	0.980	0.950	**Very Good**
0.933	0.010	0.933	0.933	0.933	0.924	0.993	0.928	**Below Average**
0.958	0.021	0.920	0.958	0.939	0.923	0.980	0.811	**Average**
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	**Excellent**

Table 5: AdaBoostM1 Classifier: Accuracy is 50%

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.923	0.519	0.462	0.923	0.615	0.397	0.664	0.428	**Good**
0.000	0.000	0.000	0.0000	0.000	0.000	0.691	0.320	**Very Good**
0.000	0.000	0.000	0.000	0.000	0.000	0.834	0.303	**Below Average**
1.000	0.188	0.571	1.000	0.727	0.681	0.881	0.510	**Average**
0.000	0.000	0.000	0.000	0.000	0.000	0.634	0.165	**Excellent**

Table 6: SMO Classifier: Accuracy is 80.83%

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.897	0.210	0.673	0.897	0.769	0.650	0.848	0.645	**Good**
0.630	0.022	0.895	0.630	0.739	0.696	0.892	0.701	**Very Good**
1.000	0.029	0.833	1.000	0.909	0.900	0.986	0.833	**Below Average**
0.625	0..10	0.932	0.625	0.750	0.723	0.924	0.745	**Average**
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	**Excellent**

Table 7: Over all Accuracy

Classifiers	Accuracy (%)
Bayes Network	85.00
Naive Bayes	84.16
Logistic	83.33
AdaBoostM1	50.00
SMO	80.83

V. Conclusion

We have computed the classification of students in five types using five methods Bayesian network, Naïve Bayes, Ada boost, Logistic regression and SMO. The basis of classification is based on their performance on a test performance in a subject with three types of content A, R, D, AR, AD and ARD. Highest accuracy we obtain for Bayesian network whereas lowest for Ada boostM1, other methods have accuracy in between.

References

Journals

[1] Hsieh, T., Wang, T.,"A mining-based approach on discovering courses pattern for constructing suitable learning path", Expert Systems with Applications, Vol. 37, pp. 4156–4167, 2010.

[2] Lee, M.W., Chen, S. Y., Chrysostomou, K., Liu, X. "Mining students behavior in web-based learning programs", Expert Systems with Applications, vol.36,pp. 3459–3464,2009

[3] Kerly, A., Ellis, R., Bull, S.,"CALMsystem: A Conversational Agent for Learner Modeling", Knowledge-Based Systems, Vol. 21, pp. 238-246, 2008.

[4] Rishi, O. P., Govil, R., Sinha, M.,"Distributed Case Based Reasoning for Intelligent Tutoring System: An Agent Based Student Modeling Paradigm World Academy of Science", Engineering and Technology 29, 2007.

[5] Schiaffino, S., Garcia, P., Amandi, A.,"eTeacher: Providing personalized assistance to e-learning students", Computers and Education, 2008.

[6] Baylari, A., Montazer, G.A.,"Design a personalized e-learning system based on item response theory and artificial neural network approach", Expert systems with applications, Vol. 36, pp. 8013–8021. 2009

[7] Kiran Mishra, R.B. Mishra,"Characteristic factors and components of different modules in expert system for intelligent tutoring system", Vol. 3, No. 3, pp. 14-42, 2011. Int. Jr of. Advanced Intelligence paradigms, Inder Science Publication.

[8] Kiran Mishra, R.B. Mishra,"Heuristic method for performance evaluation of learning in intelligent tutoring system", Vol. 9, No. 2, pp. 1-7, June, 2010, INFOCOMP Journal of Computer Science.

Conferences

[9] Merceron, A., Yacef, K.,"Educational Data Mining: A Case Study", Proceeding of the 2005 conference on Artificial Intelligence in Education: Supporting learning through Intelligent and Socially Informed Technology, 2005.

[10] Kristofic, A., Bielikova, M.,"Improving adaptation in web-based educational hypermedia by means of knowledge discovery", In 6th ACM conference on Hypertext and hypermedia, Salzburg, Austria, pp. 184–192, 2005.

Books

[11] Romero, C., Ventura, S., Hervás, C., González, P.,"Rule Discovery in Web-Based Educational Systems Using Grammar Based Genetic Programming", In Data Mining, Text Mining and Their Business Applications, pp. 113-126. WIT Press, Essex (UK), 2005.

[12] Hämäläinen, W., Laine, T.H., Sutinen, E.,"Data mining in personalizing distance education courses", In Romero, C. and Ventura, S. (Eds.) Data Mining in E-Learning, WIT Press, Southhampton, Boston, pp. 157-171. 2006.

Website

[13] [Online] Available: http://personal.disco.unimib.it/Vanneschi/McGrawHill_-_Machine_Learning_ Tom_Mitchell.pdf

[14] [Online] Available: https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/tr-98-14.pdf

[15] [Online] Available: http://www.cis.upenn.edu/~mkearns/teaching/COLT/boostingexperiments.pdf

[16] [Online] Available: http://personal.disco.unimib.it/Vanneschi/McGrawHill_-_Machine_Learning_- Tom_Mitchell.pdf

[17] [Online] Available: http://www.sciencedirect.com/science/article/pii/S1532046404000772

Kiran Mishra has done his PhD from Indian Institute of Technology (BHU) Varanasi, India. He published over 25 research paper in Journals and conferences. He has 12yrs of experience in the teaching field, and at present, he is Guest faculty at NIT Bhopal in Department of Computer Science. His area of interest is Data Mining, ITS.

Bhavtosh Mishra has done his M.Tech from UPTU Lucknow, India. He published over 5 research paper. He has 5-year experience in teaching field. At present, he is Guest Faculty at Rajikiya Engineering College, Banda, Varanasi, U.P, India, in the department of information technology. His research area includes Data Mining, ITS.

Nikita Kanwar is pursuing her M. Tech. in Information and Communication Technology at Centre for Converging Technologies, University of Rajasthan, Jaipur, India. She is currently doing her dissertation work on “Intelligent Computing Methods for prediction and classification of Parkinson’s disease” from Indian Institute of Technology (BHU), Varanasi. Her area of interest lies in Machine Learning, Computational Neuroscience, Data Mining.