Students' Success Prediction based on Bayes Algorithms

Alaa Khalaf Hamoud
Department of Computer
Information Systems
University of Basrah
Basrah, Iraq

Aqeel Majeed
Humadi
Department of Electricity
Misan University
Misan, Iraq

Wid Akeel Awadh
Department of Computer
Information Systems
University of Basrah
Basrah, Iraq

Ali Salah Hashim
Department of Computer
Science
University of Basrah
Basrah, Iraq

ABSTRACT

Prediction in data mining is a sophisticated task that is conducted in various disciplines. Given that the overall success of educational institutions can be measured by their students' success, many studies are dedicated to predicting it. This paper provides a model of student's success prediction based on Bayes algorithms and suggests the best algorithm based on performance details. Two built Bayes Algorithms (naïve Bayes and Bayes network) were used in this model with students' questionnaire answers. The questionnaire consists of 62 questions that cover the fields affecting students' performance the most. The questions refer to health, social activity, relationships and academic performance. The questionnaire is constructed based on a Google form and open-source applications (LimeSurvey); the total number of student answers is 161. To build this model, the tool Weka 3.8 is used. The overall model design process can be divided into two stages. The first stage is finding the most correlated questions to the final class, and the second is applying algorithms and finding the optimal algorithm. A comparison is made between these two Bayes algorithms based on performance details. Finally, the naïve Bayes algorithm is selected as an optimal choice for students' success prediction.

Keywords
Prediction, students' success, naïve Bayes, Bayes network, Weka

1. INTRODUCTION
The ability to predict a student's performance is highly important in educational environments. Students' academic performance is based on diverse factors such as personal, social, psychological and other environmental variables. A promising tool to attain this objective is data mining. Data mining techniques are used to operate on large amounts of data to find hidden patterns and relationships that are helpful in decision making [1]. Data mining is a process of extracting useful knowledge and information from data that are stored in databases and data warehouses. Data mining is an integral part of knowledge discovery in database (KDD) [2].

Data mining is applied in various human endeavours including education. Educational data mining (EDM) is a discipline that concerns developing methods to explore the unique types and patterns of data that originate from educational institutions and environments as well as using those methods for improved understanding. The settings in which students learn are also included in the discipline. Using data mining in education has elicited research attention. Data mining methods or approaches include classifications, clustering, naïve Bayesian, decision trees, neural networks and fuzzy rules [3].

Data mining became highly popular among researchers because many accessible standalone or desktop data mining tools are available in the market, including Microsoft Excel, SPSS, Weka, Protégé (a knowledge acquisition system) and Rapid Miner. A number of these tools (e.g., MS Excel) are freely available to professors who can benefit from their existing knowledge of Excel [4].

Learning Bayesian networks from data is a rapidly growing field of research that has seen a great deal of activity in recent years. This approach is a form of unsupervised learning in the sense that the learner does not distinguish the class variable from the attribute variables in the data. The objective is to induce a network (or a set of networks) that best describes the probability distribution over the training data. This optimization process is implemented in practice by using heuristic search techniques to find the best candidate over the space of possible networks. The search process relies on a scoring function that assesses the merits of each candidate network [5].

Naïve Bayes is the simplest form of Bayesian network, in which all attributes are independently given the value of the class variable. This feature is called conditional independence. Obviously, the conditional independence assumption is rarely true in most real-world applications. A straightforward approach to overcome the limitation of naïve Bayes is to extend its structure to represent explicitly the dependencies among attributes. An augmented naïve Bayesian network, or simply augmented naïve Bayes, is an extended naïve Bayes in which the class node directly points to all attribute nodes, and links among attribute nodes exist [6].

In knowledge discovery and data mining, multiple different prediction methods and techniques are used. Every method or technique has its advantages and disadvantages. Thus, this study uses multiple prediction methods to confirm and verify the results with multiple algorithms. The best result could be selected in terms of accuracy and precision. The objective of this study is to analyse collective student information via a questionnaire and to classify the collected data to predict and categorize student performance. We also seek to elucidate the various factors that affect student success and failure rates in relation to other variables in the data set of students by applying Bayes algorithms. In this paper, we focus on the performance details of two algorithms (naïve Bayes and Bayes network) based on TP rate, FP rate, precision and recall generated by the algorithms when they are applied on the data set.

The rest of the paper is structured into five sections. In Section 2, a review of the related work is presented. Section 3 describes the classification methods implemented in this

study, which include naïve Bayes and Bayes network. Section 4 provides the model implemented in this study, which includes a representation of the data processing, an attribute selection, a result evaluation and applying the algorithm. Section 5 presents the conclusion of this study.

2. RELATED WORKS

Shaziya, Zaheer, and Kavitha [7] presented an approach to predict the performance of students in per-semester exams. This approach is based on a naïve Bayes classifier. The objective is to know the grades that students may obtain in their end-of-semester results. This process helps the educational institution, teachers, and students, i.e., all the stakeholders involved in an educational system. They can benefit from the prediction of students' results in a multitude of ways. Students and teachers can take necessary actions to improve the results of those students whose result prediction is not satisfactory. A training dataset of students is used to build the naïve Bayes model. Then, the model is applied to the test data to predict the end-semester results of students. In this approach, a number of attributes is considered to predict the grade of a student.

Kaur and Singh [8] conducted a study to maintain the education quality of institute by minimizing the diverse effects of psychological, personal and environmental factors on student performance. Student performance is predicted by applying naïve Bayes and the Weka tool of the J48 decision-tree classification technique. By applying data mining techniques to student data, we can obtain knowledge that describes student performance. This knowledge will help improve education quality and student performance while decreasing failure rate. All of these effects will improve institutional quality.

Jayaprakash and Balamurugan [9] proposed a model in which the naïve Bayes algorithm is applied to predict students' academic performance in end-of-semester examinations by analysing student feedback and their performance in mid-semester exams. This study helps educational institutions to identify the weaker students in advance and arrange necessary training before they take their final exams.

Osmanbegović and Suljić [10] developed a model that can derive conclusions on students' academic success. Different methods and techniques of data mining (naïve Bayes, J48 and multilayer perceptron) were compared during the prediction of student success. The researchers applied the data collected from the surveys conducted during the summer semester at the University of Tuzla, Faculty of Economics, academic year 2010–2011, among first-year students. Data were obtained during enrolment. Success was evaluated on the basis of passing grades in the exam. The effect of students' socio-demographic variables, results from high school and from the entrance exam, in addition to attitudes towards studying that can affect success, were all investigated. In future investigations, when identifying and evaluating variables associated with the process of studying, and with the sample increase, producing a model that can stand as a foundation for the development of the decision support system in higher education is possible.

3. PREDICTION ALGORITHMS AND EDUCATIONAL DATA MINING

This section lists a brief explanation of Bayes algorithms that are used for prediction and a brief explanation of EDM.

3.1 Naïve Bayes

Student performance is predicted using a data mining technique called classification rules. The naïve Bayes classification algorithm is used by the administrator to predict student performance based on performance detail. The algorithm is a simple probabilistic classifier that calculates a set of probabilities by counting the frequency and combinations of values in a given dataset. The algorithm uses the Bayes theorem and assumes that all attributes are independent given the value of the class variable. This conditional independence assumption rarely holds true in real-world applications, hence the characterization as naïve. However, the algorithm tends to perform well and learn rapidly in various supervised classification problems [11].

The naïve Bayesian classifier works as follows:

1. We let T be a training set of samples, each with their class labels. k classes, $C_1, C_2, . . . , C_k$ exist. Each sample is represented by an n-dimensional vector, $X = \{x_1, x_2, . . . , x_n\}$ that depicts n measured values of the n attributes, $A_1, A_2, . . . , A_n$, respectively.

2. Given a sample X, the classifier will predict that X belongs to the class having the highest a posteriori probability, conditioned on X. X is predicted to belong to the class C_i if and only if

$P(C_i|X) > P(C_j|X)$ for $1 <= j <= m, j \neq i$.

Thus, we find the class that maximizes $P(C_i|X)$. The class C_i for which $P(C_i|X)$ is maximized is called the maximum posteriori hypothesis. By Bayes' theorem,

$P(C_i|X) = P(X|C_i) P(C_i) P(X)$.

3. Given that P(X) is the same for all classes, only $P(X|C_i)P(C_i)$ needs to be maximized. If the class a priori probabilities, $P(C_i)$, are not known, then we assume that the classes are equally likely, that is, $P(C_1) = P(C_2) = . . . = P(C_k)$, and we would therefore maximize $P(X|C_i)$. Otherwise, we maximize $P(X|C_i)P(C_i)$. The class a priori probabilities may be estimated by $P(C_i) = freq(C_i, T)/|T|$.

4. Given data sets with many attributes, computing $P(X|C_i)$ would be computationally expensive. To reduce computation in evaluating $P(X|C_i)P(C_i)$, the naïve assumption of class conditional independence is made. This assumption presumes that the values of the attributes are conditionally independent of one another, given the class label of the sample. Mathematically,

$P(X|C_i) \approx \prod_{k=1}^{n} P(x_k|C_i)$

The probabilities $P(x_1|C_i)$, $P(x_2|C_i)$, . . . , $P(x_n|C_i)$ can easily be estimated from the training set. Here, x_k refers to the value of attribute A_k for sample X.

(a) If A_k is categorical, then $P(x_k|C_i)$ is the number of samples of class C_i in T that have the value x_k for attribute A_k, divided by $freq(C_i, T)$, the number of sample of class C_i in T.

(b) If A_k is continuous-valued, then we typically assume that the values have a Gaussian distribution with a mean μ and standard deviation ∂ defined by

$g(x, \mu, \partial) = \frac{1}{\sqrt{2\pi\partial}} exp - \frac{(x-\mu)^2}{2\partial^2}$

so that

$p(x_k|C_i) = g(x_k, \mu C_i, \partial C_i)$.

We compute μCi and ∂Ci, which are the mean and standard deviation of values of attribute Ak for training samples of class Ci.

5. To predict the class label of X, P(X|Ci)P(Ci) is evaluated for each class Ci. The classifier predicts that the class label of X is Ci if and only if it is the class that maximizes P(X|Ci)P(Ci) [13].

3.2 Bayesian Network

A Bayesian network (BN) is a graphical model for probability relationships among a set of variables (features). The Bayesian network structure S is a directed acyclic graph (DAG) and the nodes in S have one-to-one correspondence with the features X. The arcs represent casual influences among the features, whereas the lack of possible arcs in S encodes conditional independencies. Moreover, a feature (node) is conditionally independent of its non-descendants given its parents [13].

To construct a learning algorithm for Bayesian networks, two components are defined: a function for evaluating a given network based on the data and a method for searching through the space of possible networks. The quality of a given network is measured by the probability of the data given the network [14].

A Bayesian network begins with a graphical depiction of the relationships among the most important variables in the system of interest. In this depiction, the variables are represented by round nodes and dependence between one variable and another is represented by an arrow. The conditional independence implied by the absence of a connecting arrow between any two nodes simplifies the modelling process by allowing separate sub-models to be developed for each relationship, as indicated by the presence of an arrow. These sub-models may be derived from any combination of knowledge processes, statistical correlations, or expert judgment, depending on the information available about that particular relationship [15][16].

In a Bayesian network, each dependence indicated by an arrow represents a conditional probability distribution that describes the relative likelihood of each value of the down-arrow node, conditional on every possible combination of values of the parent nodes. A node that has no incoming arrows is said to have no parents, and such a variable can be described probabilistically by a marginal (or unconditional) probability distribution. Therefore, the graphical network constitutes a description of the probabilistic relationships among the system variables that amounts to a factorization of the joint distribution of all variables into a series of marginal and conditional distributions [17].

3.3 Educational Data Mining

Educational data mining is an emerging discipline that is concerned with developing methods to explore the unique types of data that come from educational settings and with using those methods to better understand students and the settings in which they learn [18]. The structure of most EDM systems can be broken down into three parts: collection, archiving and analysis. Collection refers to the tools and tutoring systems used to record the relevant information, whether student scores, answers to online quizzes or events from an intelligent tutoring system (ITS). Archiving is the process of storing and browsing the collected data.

For score data, this is a relatively minor issue, but for the vast quantities of data generated by some ITSs, this can be a

significant task. Analysis brings to bear the tools of machine learning and data mining on the collected data to gain a deeper understanding of student learning, find the relationships among questions, and possibly develop a deeper quantitative understanding of cognitive processes in general [19].

Data mining, which is also referred to as KDD, is a process of non-trivial extraction of implicit, previously unknown and potentially useful information (such as knowledge rules, constraints and regularities) from data in the database [18]. The data mining algorithms can be divided into three major categories based on the nature of their information extraction: predictive modelling (also called classification or supervised learning), clustering (also called segmentation or unsupervised learning) and frequent pattern extraction [20].

Table 1. Table captions should be placed above the table

Graphics	Top	In-between	Bottom
Tables	End	Last	First
Figures	Good	Similar	Very well

4. MODEL

As shown in Figure (1), the model construction process passes through four steps starting with data preprocessing and ending with result evaluation. First, the process of building a questionnaire is considered a part of the data preprocessing step. This step includes all processes of preparing a dataset for evaluation, cleaning data, converting data ranges and creating a derived column (Failed) based on the column (Number of Failed Courses). The column (Failed) is created based on a simple condition:

If (Number of Failed Courses>0) then Failed='F'

Else Failed='P';

where F is an abbreviation of Failed, P is an abbreviation of Passed;

The column "Failed" is considered the goal class of the model.
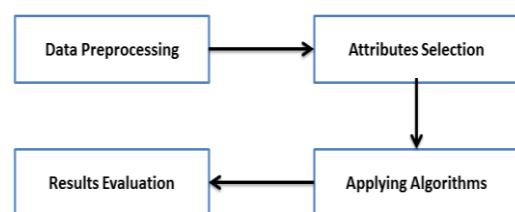

Figure 1 Model Construction Diagram

4.1 Data Preprocessing

Data preprocessing involves the following steps:

4.1.1 Data Collection

The questionnaires were built based on a Google form and an open-source application (LimeSurvey) to collect student answers at College of Computer Science and Information Technology (CSIT), University of Basrah. The first questionnaire (based on LimeSurvey) is constructed to collect answers locally from the college building of CSIT, whereas the Google form is used to collect the answers online. The total number of students' answers is 161 after combining the Comma Separated Values (CSV) files from the Google form and LimeSurvey questionnaire. The research sample (161

answers) represents an acceptable sample of the population of CSIT with 10% as the percentage of error for the result of this study [21].

Table (1) below shows both question description and questions and answers range of the questionnaire. The description of all questions in this table is shortened so it can be seen in tree nodes and can be easily understood. Question ranges were also shortened and converted from nominal to numeric type for ease of use and understanding.

Table 1 Questionnaire's Questions Descriptions

Question	Description	Range	Question	Description	Range
Q1	Dep	IS,CS	Q32	IHvEnrgyEnjoy	1,2,3,4,5
Q2	Age	1,2,3,4	Q33	PractRegular	1,2,3,4,5
Q3	Stage	1,2,3,4	Q34	MyHealthHelp	1,2,3,4,5
Q4	Gender	F,M	Q35	FreshFood	1,2,3,4,5
Q5	Address	IN,OUT	Q36	CnUseLaptpToAchvSucc	1,2,3,4,5
Q6	Status	M,S	Q37	PlanForWeek	1,2,3,4,5
Q7	Work	YES,NO	Q38	PlanDaily	1,2,3,4,5
Q8	LiveWithParent	YES,NO	Q39	PlanToNotReadAgain	1,2,3,4,5
Q9	ParentAlive	0,1,2,3	Q40	PlanToDoFunThing	1,2,3,4,5
Q10	FatherWork	0,1,2	Q41	ContrlMyBudget	1,2,3,4,5
Q11	MotherWork	1,2	Q42	ClrIdeaAbtMyBudget	1,2,3,4,5
Q12	FCourses	0,1,2,3	Q43	CnWork	1,2,3,4,5
Q13	AbsenceDays	0,1,2	Q44	MyEduSuppoMyGoal	1,2,3,4,5
Q14	Credits	0,1,2	Q45	IHvSavPlan	1,2,3,4,5
Q15	GPA	1,2,3,4	Q46	EduIsLiveJob	1,2,3,4,5
Q16	ComCredits	1,2,3,4	Q47	ClrAbotMyLiveGoal	1,2,3,4,5
Q17	YearsOfStudy	1,2,3,4	Q48	ResponAbtMyEdu	1,2,3,4,5
Q18	ListImporPoints	1,2,3,4,5	Q49	RespoAbtMyMyLiveQuality	1,2,3,4,5
Q19	WriteNotes	1,2,3,4,5	Q50	RedyToFacChallng	1,2,3,4,5
Q20	PrepStudySchedule	1,2,3,4,5	Q51	ClrIdeaABoutPlans	1,2,3,4,5
Q21	CalmDurExam	1,2,3,4,5	Q52	Worked Recently	4,5
Q22	LDegNotMakeMeFail	1,2,3,4,5	Q53	KnowledWtBossExpectFrmMe	1,2,3,4,5
Q23	EasCanChosColgStudy	1,2,3,4,5	Q54	EduChoicesToAchivGoal	1,2,3,4,5
Q24	OptimToAchvGoals	1,2,3,4,5	Q55	IHvEnoughMoney	1,2,3,4,5
Q25	CnStudyEvUImpoBothMe	1,2,3,4,5	Q56	RelationWithOthers	1,2,3,4,5
Q26	ExiToMater	1,2,3,4,5	Q57	EnoughBudget	1,2,3,4,5
Q27	ClrIdeaAbtBenifit	1,2,3,4,5	Q58	RequestHelpFromOthers	1,2,3,4,5
Q28	DevRelationWithOthers	1,2,3,4,5	Q59	TryToEnhancMySelf	1,2,3,4,5
Q29	ContrlMyAnger	1,2,3,4,5	Q60	IHvSkillToSchvAcadmWork	1,2,3,4,5
Q30	MakeFriendship	1,2,3,4,5	Q61	IHvSkillsForSelfFeel	1,2,3,4,5
Q31	OpenWithOthers	1,2,3,4,5			

Given that the questionnaire includes more than 60 questions, shortening them is necessary. Question descriptions were shortened to be used by Bayes algorithms. Table (2) lists some question descriptions, which were shortened in the previous table.

Table 2 Question Abbreviations and Descriptions

Question	Abbreviation	Description
Q1	Dep	Your department
Q2	Age	Your age
Q3	Stage	Your stage
Q4	Gender	Your gender
Q5	Address	Where do you live?
Q6	Status	Your status
Q7	Work	Are you working now?
Q8	LiveWithParent	Do you live with your parents?
Q9	ParentAlive	Are your parents alive?
Q10	FatherWork	What is your father's work scope?
Q11	MotherWork	What is your mother's work scope?
Q12	FCourses	Number of courses you fail in per semester
Q13	AbsenceDays	Number of days of absence per semester

Q14	Credits	Number or registered credits per semester
Q15	GPA	Grade point average
Q16	ComCredits	Number of completed credits
Q17	YearsOfStudy	Number of academic years until the present
Q18	ListImporPoints	I can write down the important points while reading the material.
Q19	WriteNotes	During lectures, I can write notes and use them for exam preparation.
Q20	PrepStudySchedule	I prepare a time schedule for studying.
Q21	CalmDurExam	During exams, I stay calm and coherent.
Q22	LDegNotMakeMeFail	Getting low grades does not make me feel like a failure.

The first step in data preprocessing is preparing data for processing by removing rows with empty values and converting data for evaluation and processing. The number of rows with empty values in one or more columns is 11. After these rows are removed, the total number of answers become 150. The second step is converting the row values to process them in Weka 3.8 with its built-in classifiers.

4.1.2 Reliability

Reliability is used to describe the overall consistency of a measure. A measure has a high reliability rate if it produces similar results under consistent conditions. For example, measurements of human height and weight are often extremely reliable [22]. In statistics, the coefficient alpha is the most frequently used method for calculating internal consistency that is used as a measure of reliability for the dependent variable of the study. Cronbach's alpha 0.7 indicates satisfactory internal consistency in reliability [23]. The coefficient alpha is 0.85 for the scaled variables that contain 62 items and 161 respondents, as shown in Table (3).

Table 3 Questionnaire Reliability

Cronbach's alpha	No. of items	No. of respondents	% of respondents
0.85	62	161	100%

4.2 Attribute Selection

Finding the most correlated attributes (questions) to the final class (Failed) is important, as is determining how much they affect the final class. This stage is important because it will show the average of attribute correlation to the final class. This average will help us to find the questions with low correlation average and remove them to improve the accuracy of the results. The questions with high correlation averages can be considered recommendation points for both students and academic staff.

In this step, the filter CorrelationAttributeEval is used to evaluate the correlation between the class and other attributes. This step is important because we want to find the most closely related attributes that affect the class and to ignore the less-related attributes from the model. The attribute evaluator algorithm (CorrelationAttributeEval) evaluates the worth of an attribute by measuring the correlation (Pearson's) between it and the class.

Table (4) shows the average of correlation between the attributes (questions) and the final class with evaluation mode (10-fold cross validation) to ensure highly accurate results. The questions are arranged in ascending order based on correlation average rate to the final class. The questions with high average rates are the most correlated to the final class. To increase result accuracy, the final 20 questions are removed.

Table 4 Correlation Average of Questions

Seq	Question	Average	Seq	Question	Average
1	Q14	0.436	31	Q52	0.101
2	Q25	0.312	32	Q49	0.09
3	Q60	0.234	33	Q19	0.094
4	Q15	0.224	34	Q11	0.086
5	Q46	0.227	35	Q26	0.088
6	Q17	0.225	36	Q42	0.089
7	Q51	0.21	37	Q48	0.077
8	Q47	0.209	38	Q24	0.074
9	Q10	0.174	39	Q20	0.074
10	Q23	0.168	40	Q41	0.065
11	Q27	0.168	41	Q22	0.068
12	Q45	0.173	42	Q55	0.064
13	Q29	0.162	43	Q36	0.065
14	Q57	0.158	44	Q4	0.062
15	Q13	0.159	45	Q5	0.061
16	Q18	0.152	46	Q31	0.054
17	Q35	0.147	47	Q6	0.052
18	Q40	0.145	48	Q3	0.042
19	Q58	0.142	49	Q56	0.04
20	Q43	0.134	50	Q44	0.039
21	Q16	0.135	51	Q32	0.039
22	Q38	0.131	52	Q1	0.043
23	Q8	0.133	53	Q2	0.028
24	Q34	0.131	54	Q54	0.025
25	Q37	0.123	55	Q12	0.023
26	Q59	0.12	56	Q39	0.026
27	Q53	0.113	57	Q33	0.023
28	Q9	0.116	58	Q28	0.023
29	Q30	0.116	59	Q7	0.022
30	Q50	0.106	60	Q21	0.018

4.3 Applying Algorithms

The Weka tool provides built-in algorithms that help us to apply different classifiers and obtain results in an easy and flexible process. Two algorithms will be used in this stage (Bayes net and naïve Bayes). Both of them are applied after removing the questions with less correlation. Removing attributes is helpful to determine the effectiveness of these attributes on the performance and how they can increase or decrease accuracy.

Applying algorithms is conducted based on two steps. The first step is preparing a train data set and applying algorithms, and the second step is applying a test data set on the result of the first step. Train data is prepared by adding 10% of the total number of questionnaire answers (150 answers). A total of 15 arbitrary answers are added to prepare the train data set. The performance details of algorithms are observed after applying the test data set (i.e., the data set consisting of the questionnaire answers).

4.4 Result Evaluation

Table (5) shows the performance details after removing the less correlated attributes to Failed attribute. The table shows details of the two algorithms (naïve Bayes and Bayes net) based on Truly Positive Rate (TP Rate), False Positive Rate (FP Rate), Precision and Recall. Recall is also referred to as the TP Rate or sensitivity, and precision is also referred to as a positive predictive value (PPV). Other related measures used in classification include true negative rate and accuracy.

Table 5 Bayes Algorithm Performance

Classifier	TP Rate	FP Rate	Precision	Recall
Bayes Net	0.655	0.432	0.643	0.655
Naïve Bayes	0.667	0.297	0.706	0.667

The TP Rate of the naïve Bayes algorithm is higher (0.667) compared with Bayes net (0.655) and FP Rate of naïve Bayes is lower (0.297) compared with Bayes Net (0.432). Furthermore, the precision and recall values of naïve Bayes are better (0.706, 0.667) than precision and recall values of Bayes net, as illustrated in Figure (2). Based on these performance details, the naïve Bayes algorithm can be selected as the best algorithm to predict student performance (Failed column). Table (6) shows the predicted column (Failed) of 30 students based on their answers.

Table 6 Predicted Column of Naïve Bayes Algorithms

Answer Number	Failed	Predicted Failed	Answer Number	Failed	Predicted Failed
1	F	F	16	F	F
2	P	P	17	P	P
3	P	P	18	P	P
4	F	P	19	F	F
5	P	F	20	P	P
6	F	F	21	P	P
7	F	F	22	P	P
8	F	F	23	P	P
9	P	P	24	P	P
10	P	P	25	P	P

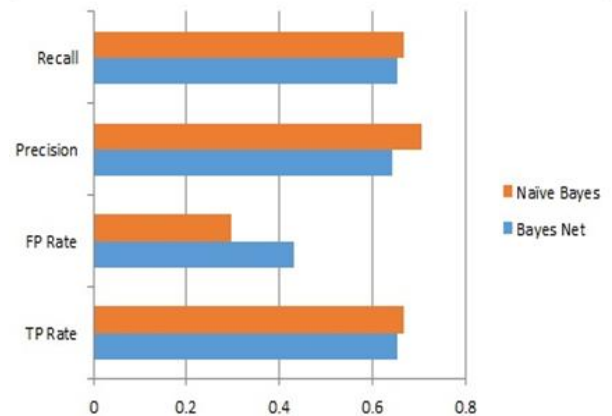| 11	P	P	26	F	F
12	F	F	27	F	F
13	P	F	28	F	P
14	P	F	29	F	F
15	P	F	30	F	F

Figure 2 Performance Chart of Bayes Algorithms

Figure (2) shows the performance chart of the naïve Bayes and Bayes net algorithms based on TP Rate, FP Rate, Recall and Precision. With the naïve Bayes algorithm, all of (Recall, Precision, TP Rate) obtain high values compared with Bayes net, which has the high value in FP Rate. FP Rate in naïve Bayes has the low value. The Bayes net algorithm obtains low values in TP Rate, Precision and Recall. Thus, based on this chart, naïve Bayes is the better choice for prediction process.

5. CONCLUSION

The objective of this study is to explore the possibility of predicting student performance based on applying Bayes algorithms on student questionnaire answers. The questionnaire contains many unimportant questions that can be found by data mining algorithms, thereby helping us to find the less correlated questions. Many factors affect result accuracy. In addition to large data sets, the number of attributes and the clean dataset affect the accuracy. To obtain a highly accurate result, many steps are followed such as cleaning data sets, removing incomplete answers and removing the less correlated questions to the final class. Bayes classifier algorithms are used in this model and the accuracy of prediction is compared to find the optimal one. Finally, the naïve Bayes algorithm is selected as the best algorithm for prediction based on performance detail. The model can be depended on by both students and academic staff to decide the questions/answers that will enhance academic performance and improve institutional success.

6. REFERENCES

[1] Shruthi P and Chaitra B P, "Student Performance Prediction in Education Sector Using Data Mining", International Journal of Advanced Research in Computer Science and Software Engineering. Volume 6, Issue 3, March 2016.

[2] Agrawal Bhawana D, Gurav Bharti B," Review on Data Mining Techniques Used For Educational System". International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com (ISSN

2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 11, November 2014).

[3] Kolo David Kolo, Solomon A. Adepoju, John Kolo Alhassan, "A Decision Tree Approach for Predicting Students Academic Performance", I. J. Education and Management Engineering, 2015, 5, 12-19.

[4] Romero, Cristóbal, Sebastián Ventura, and Enrique García. "Data mining in course management systems: Moodle case study and tutorial." Computers & Education 51.1 (2008): pp. 368-384.

[5] Friedman, Nir, Dan Geiger, and Moises Goldszmidt. "Bayesian network classifiers." Machine Learning 29.2-3 (1997): 131-163.

[6] Zhang, Harry. "The optimality of naïve Bayes." AA 1.2 (2004): 3. Israel, Glenn D., Determining Sample Size. Large databases. Proc. of ACM SIGMOD, pp. 207–216, 1993.

[7] R. Z. G. Humera Shaziya, "Prediction of Students Performance in Semester Exams using a Naïve Bayes Classifier," International Journal of Innovative Research in Science, Engineering and Technology, vol. 4, no. 10, pp. 9823-9829, October 2015.

[8] Kaur, Gurmeet, and Williamjit Singh. "Prediction of Student Performance Using Weka Tool." (2016).

[9] Jayaprakash, Sujith, E. Balamurugan, and Vibin Chandar. "Predicting Students' Academic Performance Using Naïve Bayes Algorithm." 8th Annual International Applied Research Conference ,2015.

[10] Osmanbegović, Edin, and Mirza Suljić. "Data mining approach for predicting student performance." Economic Review 10.1 (2012): 3-12.

[11] George Dimitoglou, James A. Adams, and Carol M. Jim," Comparison of the C4.5 and a Naïve Bayes Classifier for the Prediction of Lung Cancer Survivability", Journal of Computing, Volume 4, Issue 8, 2012.

[12] Pandey, U. K. and Pal, S., "A Data Mining View on Class Room Teaching Language", (IJCSI) International Journal of Computer Science Issue, Vol. 8, Issue 2, March -2011, 277-282, ISSN: 1694-0814.

[13] Leung, K. Ming. "Naïve Bayesian classifier." Polytechnic University Department of Computer Science/Finance and Risk Engineering (2007).

[14] Witten, Ian H., et al. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, 2016.

[15] Varis, Olli. "Belief networks for modelling and assessment of environmental change." Environmetrics 6.5 (1995): 439-444.

[16] Peters, Robert Henry. A critique for ecology. Cambridge University Press, 1991.

[17] Borsuk, Mark E., Craig A. Stow, and Kenneth H. Reckhow. "A Bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis." Ecological Modelling 173.2 (2004): 219-239.

[18] www.educationaldatamining.org.

[19] Titus deLaFayette Winters, Educational Data Mining: Collection and Analysis of Score Matrices for Outcomes-based Assessment, University of California, June 2006.

[20] Agrawal, Rakesh, Tomasz Imieliński, and Arun Swami. "Mining association rules between sets of items in large databases." Acm sigmod record. Vol. 22. No. 2. ACM, 1993.

[21] Israel, Glenn D. "Determining Sample Size. University of Florida IFAS extension." (2009).

[22] Carson B. "The transformative power of action learning." Retrieved 2017 from http://wial.org/executive-board/bea-carson-executive-board/.

[23] Sekaran, Uma, and Roger Bougie. Research methods for business: A skill building approach. John Wiley & Sons, 2016.