ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/319955477

Analysis on students performance using naÏve Bayes classifier

Article *in* Journal of Theoretical and Applied Information Technology · August 2017

CITATIONS
43

READS
4,362

3 authors:

Mokhairi Makhtar
Sultan Zainal Abidin University
77 PUBLICATIONS 770 CITATIONS
SEE PROFILE

Hasnah Nawang
Sultan Zainal Abidin University
7 PUBLICATIONS 88 CITATIONS
SEE PROFILE

Syadiah Nor WAN Shamsuddin
Sultan Zainal Abidin University
50 PUBLICATIONS 283 CITATIONS
SEE PROFILE

All content following this page was uploaded by Hasnah Nawang on 12 July 2021.

The user has requested enhancement of the downloaded file.

ANALYSIS ON STUDENTS PERFORMANCE USING NAÏVE BAYES CLASSIFIER

[1] MOKHAIRI MAKHTAR, [2] HASNAH NAWANG, [3] SYADIAH NOR WAN SHAMSUDDIN

[1,3] Lecturer, Department of Informatics and Computing, University of Sultan Zainal Abidin,, Terengganu

[2]*University of Sultan Zainal Abidin, Kuala Terengganu, Terengganu*

E-mail: [1]mokhairi@unisza.edu.my, [2]hasnah@mrsmkber.edu.my [3]syadiah@unisza.edu.my

ABSTRACT

Classification of students' academic performance for Sijil Pelajaran Malaysia (SPM) at early stage of their previous study will able to help in identify the students' achievement to will assist the educators and school management taking the necessary actions. In this research, data mining techniques are used to classify students' of Maktab Rendah Sains MARA (MRSM) Kuala Berang performance based on their performance in certain subjects. The aim of this study is to examine the Naive Bayes algorithm which is one of the classification methods in data mining, to identify the hidden information between subjects that affected the performance of students in Sijil Pelajaran Malaysia (SPM). Data was collected from the second semester obtained from year 2011 until 2014 with the total of 488 students' data were used to train the algorithm. It has been shown that with 10 cross fold-validation that Naive Bayes algorithm can be used for classification of students' performance in early stages of second semester with an accuracy of 73.4%.

Keywords: *Classification, Data Mining, Feature Selection, Naïve Bayesian.*

1. INTRODUCTION

Educational Data Mining is an emerging interdisciplinary research area that deals with the development of methods to explore data originating in the educational context. In recent years, there has been a lot of studies in Educational Data mining that concerned the study of students' behavior and performance. In order to understand the pattern of students' learning or behavior, many researchers have study the outcomes of students especially in higher education. Researchers in the higher educational institution mainly focus the application of data mining to emphasis on the improvement of the students' results[1][2][3], to help educators reduce the drop out ratio in college or university [4] [5], to help students in choosing the right course based on previous achievement [6][7] and many more. The data mining application has been widely used on the educational sector, however it is mostly applied to the higher institutional.

This research focuses on the application of machine learning to the educational data of secondary school. Educators in Maktab Rendah Sains MARA (MRSM) are always facing problem in ensuring the increasing of students' performance starting from semester 1 until the students sit for Sijil Pelajaran Malaysia (SPM). In order to help the educators in MRSM especially MRSM Kuala Berang, a research has been done on the examination data that has been discovered from the Student Information System (SIS). The data will be used in helping educators and school management to analyze and discover useful information in order to help their students. The information gained from the data mining techniques will help teachers classify their students' Student Performance Analysis (SPA) for Sijil Pelajaran Malaysia.

The main task of data mining is to analyze large quantities of data in order to extract previously unknown information and patterns [7]. Some of the data mining methods are classification, regression and association. Methods that used to build the predictive model are classification, cauterization and regression. Mostly researchers used classification in order to predict students' performance. One of the classification models is

Naïve Bayes Classifier that is the probabilistic classifier based on the Bayes Theorem. Naïve Bayes classifier assumes that the effect of the attributes value on a given class is independent on the value of other features [15]. Naïve Bayes is also well known used on text categorization. This article introduces a Naïve Bayes classifier for classifying students' performance and the existence of the relationship between subjects that contribute to the certain classes.

2. RELATED WORK

The literature review revealed that data mining in educational field has been the interest of various researchers during the last few years. The development of data mining models for predicting student performance at various levels and comparison of those models were discussed in a number of research papers. Alaael-Halees [8] has stated that data mining has become an emerging method in educational field to enhance the understanding of the learning process. The application of data mining has been widely used in higher educational system.

There are two main factors in predicting students' performance that are attributes and prediction methods [9]. Their research shown that ten of thirty papers have used GPA as the main attributes to predicts students' performance.

Borkar and Rajeshwari [10] had run a research using association method on 60 datasets of students from the Pimpri Chinchwad College of Engineering, Pune University. Their research used the students' attendance, unit test, graduation percentage and the assignment as the attributes to predict students' performance in university.

Baradwaj & Pal [11] have used Decision Tree algorithm to classify students' dataset obtained from VBS Purvanchal University, Jaunpur (Uttar Pradesh) to predict students' division on the basis. Not only previous semester result has been choose as the attributes but the lab work and the seminar performance also have contribute to the findings. Their research also able to identify those students which needed special attention in order to reduce fail ratio.

In the year 2012. Kumar & Chadha [12] performed a case study of a university to improve the quality of education and discover the factors that affect the academic results using association rules discovery techniques. They discovered that students who have scored badly in their Graduation have done relatively well in their Post-Graduation in the subjects, which are common in both Graduate and Post Graduate courses.

The researchers have used many classifiers in order to study the pattern of students' behavior and performance in academic field such as Decision Tree, Nearest Neighbor, Naïve Bayes and etc. Naive Bayesian is a simple probabilistic classifier based on Bayesian theorem with the (naive) independence assumption [16] (Krishnaiah et al., 2013). While Wang [17] stated that Naive Bayes classifier is recognized as a simple probabilistic classifier based on the application of the Bayesian theorem with strong independence assumptions. In other words, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. Naive Bayesian is a simple probabilistic classifier based on Bayesian theorem with the independence assumption [18].

Recently, many researchers have shown interest in Naïve Bayes algorithm and it is proven has been executed well in many complicated real-world difficulty. For example, [13] concluded that Bayesian classification theories are useful analyses form to predict future data trends and make intelligent decision in distance education system. While [14] in their researchers have found that Naïve Bayes Multi-label Classifier algorithm together with K-means Clustering can help in understanding social media data issues such as heavy study load, hectic schedule and lack of sleep for education enhancement.

Karthika & Sairam, [15] have done a research paper on method to categorize the educational qualification using the Naïve Bayes Classification algorithm. It is found that Naïve Bayes Classification algorithm performs well when the attributes are non-numerical.

Brijesh and Saurabh, [11] have used Bayesian classification method on students' database from Awadh University, Faizabad, UP, India that involved 300 students. Their research is to predict students' division in order to help students and lecturers to improve the division of their students. Their research also helps in identifying those students that need special attention in order to reduce failing ratio.

There are already a large number of research papers discussing various problems within the higher education sector and providing examples for successful solutions reached by using data mining. Therefore, the rationale described in this paper is based on the great potential that is seen in using data mining methods and techniques in university level to be applied in the upper secondary level in MRSM.

3. METHODOLOGY

The methodology adopted in this research starts with the data collection from the MRSM Kuala Berang Information System involved 491 students' data. The method then followed by data selection process in order to identify the relevant attributes need in this research. The next phase is pre-processing stage that covers data integration of three databases and files from different formats using key-based approach. Followed by data cleaning where instances were checked to ensure the data is complete and data transformation where the normalization process covered and finally the data were reduced to 488 out of 491. Before the data were mined, the feature selection is applied in order to increase the accuracy of mining results.

3.1 Feature Selection

Feature selection is a pre-processing stage used to reduce dimensionality and delete irrelevant data to increase learning accuracy and improve result comprehensibility [9]. Irrelevant attributes may add noise to the data and also will affect the accuracy of the model. Furthermore, noise will increase the time in model building. During feature selection process, some subjects have been identified in affected the data mining results. The subjects are English, History, Islamic Education, Mathematics, Additional Mathematics, Chemistry and Physics.

Table 1 shows the results accuracy for each classifier used in this research before the feature selection has been applied. There are five classifiers that have been tested using the dataset that are Naïve Bayes, Random Tree, Multi Class Classifier, Conjuctive Rule and Lazy IB1. Naïve Bayes holds the highest accuracy results as shown in the table.

	Before Feature Selection			
Classifier/ Semester	1	2	3	4
Random Tree	63.40	67.42	63.52	72.54
Naïve Bayes	69.12	72.95	75.82	80.94
Nearest Neighborhood (IB1)	64.82	65.78	68.24	78.28
Multi Class Classifier	61.35	60.66	64.14	67.21
Conjunctive Rule	66.67	66.80	67.83	73.57

Table 1 : Accuracy for Variants Classifier

3.2 Classification Technique Using Naïve Bayes Classifier

The Naïve Bayes classification technique will be applied seem the algorithm had shown the highest accuracy result. The Naïve Bayes classification helps to classify the students' performance in Sijil Pelajaran Malaysia learners based on their grades in certain subjects for semester examination. Naïve Bayes works well with categorical attributes (i.e. A, B, C, etc.) not with numerical attributes. To apply the Naïve Bayes algorithm, the study will also use the attributes as discussed below.

The attributes selected after the feature selection are the grades for subjects that are given below:

- English -BI (Core subject).
- Islamic Education subject - PI (Core subject).
- History – SJ (Core subject).
- Mathematics - MM (Core subject).
- Additional Mathematic -MT (Elective Subject)
- Physics –FZ (Elective Subject)
- Chemistry –KM (Elective Subject)

The situations considered here deals with the details of the students' performance in second semester and Sijil Pelajaran Malaysia. Based on these details, the students are classified into four different classes. Here, four classes are taken into consideration and the attributes above are used to compute the probability that the students' performance will experience event of classes that are:

- Excellent.
- Good.

- Average.
- Poor

Bayes' Theorem provides a way of calculating the posterior probability, P $(x|c)$, from P(c), P $(x|c)$, and P(x). Bayes' Theorem is:

$$P(c|x) = \frac{P(x|c)\,P(c)}{P(x)}$$

Where:
P(c) is the prior probability of class that reflects background knowledge due to the chance of c to be correct.
P(x) is the probability of x to be observed.
P(x|c) is the probability of observing x given a world in c holds.
P(c|x) is the posterior probability of class (target) given predictor (attribute).

Naïve Bayes classifier work as follow:

Let D be a training set of tuples and their associated class labels. As usual, each tuple is represented by an n-dimensional attribute vector, X = $(x_1,x_2,...,x_n)$, depicting n measurements made on the tuple from n attributes, respectively, $A_1,A_2,...,A_n$.

Suppose that there are m classes, $C_1,C_2,...,C_m$. Given a tuple, X, the classifier will predict that X belongs to the class having the highest posterior probability, conditioned on X. That is, the naïve Bayesian classifier predicts that tuple X belongs to the class C_i if and only if $P(C_i|X)> P(C_j|X)$ for 1≤ j≤m, j6=i.

Thus the P(Ci|X) need to be maximize.The class Ci for which P(Ci|X) is maximized is called the maximum posteriori hypothesis.

3.3 Classification Using Naïve Bayes

This dataset samples is based on training data collected from MRSM Kuala Berang starting 2011 until 2014 for semester 2. The dataset consist of academic information of the students in their second semester to be classify with SPM result. The sample data in Table 2 clearly shows 7 attributes as discussed above in order to be used for the classification that involved four category of students' performances in this study.

Table 2: Sample of Dataset

No	Bi	Pi	Sj	Mm	Mt	Fz	Km	Category
1	A	A	A	C+	C	B	D	Good
2	C+	A	A	A	A-	B	B	Good
3	A	A	A	A	D	B	C	Excellent
4	A-	A	A-	A-	C	C+	G	Good
5	B	A	A	A-	C	B	B	Excellent
6	A-	A	A	A	A-	B	B	Excellent
7	C	A	A	A-	C	E	G	Good
8	B	A	A	C+	G	C+	G	Average
9	C+	A	A	D	G	G	G	Average
10	B	A	A	A+	C+	B+	B	Excellent
11	A-	A	A	B+	C	G	G	Good
12	B	B	A-	A	C+	D	G	Average
13	D	B	A	C+	G	G	G	Average
14	B+	B	C	E	G	E	G	Poor
15	A-	B	A	B+	G	C+	C+	Good
16	C+	B	A-	A	D	D	D	Good
17	A-	B	A	A	B+	E	C	Excellent
18	C+	B	A	B+	C	C+	B+	Excellent
19	A-	B	A	A-	E	D	C	Excellent
20	B+	B	A	B+	C	E	D	Average

The Naïve Bayes classification algorithm will calculate the probability of the categories for students' performance in SPM based on grades for subjects as illustrated in Table 2. In order to test students' performance category when a sample of *X* is stated below:

X = PI=A, SJ=A, MM=B+, MT=C, FZ=D

Given the sample *X* as above, the classifier will predict that *X* belongs to the class having the highest a posteriori probability, conditioned on *X* that *X* is belong to class C_i if and only if;

$$P(Ci \mid X) > P(Cj \mid X) \text{ for } 1 \leq j \leq m,\ j \neq i.$$

Thus, we need to maximize P(*X* |Ci)P(Ci), for i = Excellent, Good, Average and Poor. P(Ci), the priori probability of each class, can be estimated based on the training dataset samples:

Class:
P(Category = Excellent) = 7/20 = 0.35
P(Category =Good) = 7/20 = 0.35
P(Category=Average) = 5/20 = 0.25

P(Category=Poor) = 1/20 = 0.05

To compute P(*X*|Ci), for i = Excellent, Good, Average and Poor, we compute the following conditional probabilities:

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

= P(X_1 |C_i) × (X_2 |C_i) × …… × (X_n | C_i)

Where:
k = number of classes in dataset,
n = number of attributes that will be used for mining

P (Excellent | X)
= P (Excellent) P(PI | A) P(SJ | A) P(MM | B+) P(MT | C) P(FZ | D)
= (4/7) × (7/7) × (1/7) × (2/7) × (1/7) × (2/7)
=0.6 × 1 × 0.1 × 0.3 × 0.1
=0.0018

P (Good | X)
= P (Good) P(PI | A) P(SJ | A) P(MM | B+) P(MT | C) P(FZ | D)
= (5/7) × (5/7) × (2/7) × (4/7) × (1/7)
=0.7 × 0.7 × 0.3 × 0.6 × 0.1
=0.0088

P (Average | X)
= P (Average) P(PI | A) P(SJ | A) P(MM | B+) P(MT | C) P(FZ | D)
= (2/5) × (4/5) × (1/5) × (1/5) × (1/5)
=0.4 × 0.8 × 0.2 × 0.2 × 0.2
=0.0026

To find the class that maximizes P(X|Ci)P(Ci), then we need to compute :

P(X | Category=Excellent) P(Category=Excellent)
=0.0018 × 0.35
=0.00063

P(X | Category=Good) P(Category=Good)
=0.0088 × 0.35
=0.0031

P(X | Category=Average) P(Category=Average)
=0.0026 × 0.25
=0.00065

Based on the formula above, P(Good | X) > P(Average | X) > P(Excellent | X). The Naïve Bayesian classifier predicts Category = Good for sample = X.

4. RESULTS AND DISCUSSIONS

This paper shows the comparison of accuracies for the five classifiers (NBC, RT, MCC, IB1 and CR) based on a 10-fold cross validation as a test method. In order to select the most significant features for classification, with the aid of WEKA tool, the feature selection has been done on the dataset for every semester. The parameters that have been selected using the BestFirst have shown different parameters that effected the accuracy for every semester. By applying feature selection in this research, the highest data mining results conducted by Naïve Bayes algorithm. The table below illustrates the results increase in Naïve Bayes classifier for Semester 1 with the 1.43%, Semester 2 with 1.44%. Semester 3 with 2.66% and semester 4 with 1.44%. Method has been used for feature selection used in this research is Best First algorithm.

Table 3: Data Mining Accuracy after Feature Selection Applied to the Dataset

	After Feature Selection			
Classifier/ Semester	1	2	3	4
Random Tree	63.40	71.31	70.90	75.61
Naïve Bayes	70.55	74.39	78.48	82.38
Nearest Neighborhood (IB1)	63.40	69.68	75.95	75.20
Multi Class Classifier	67.28	74.59	77.25	80.12
Conjunctive Rule	66.67	66.81	67.83	73.57

The further step in classification then to test either each subject or the combination of two subjects that will affected each category. The output showing the number of test samples getting classified into four different classes that are tabulated in Table 4.

Table 4: Data Mining Accuracy Based on Subjects

Subject Name	Result (%)
English	68.03
Islamic Education	63.73
History	67.83
Mathematics	67.62
Additional Mathematics	69.47
Physics	68.44
Chemistry	73.36

From the Table 4, the result has shown that Chemistry subject contain the high result percentage among the seven subjects with the 73.36% followed by Additional Mathematics with 69.47%. While the Islamic Education has shown the lowest result percentage with the 63.73%.

The research then deepen with combining certain subjects. To find the probability of dependency between the combination subjects, some core subjects have been combined with the elective subjects and some elective subjects have been combined with the elective subjects also. Table 5 shows the results from the combination of the subjects.

Table 5: Data Mining Accuracy with the Combination of Certain Subjects

Subject 1	Subject 2	Results
Add Mathematics	Chemistry	72.75%
English	Add Mathematics	67.83%
English	Mathematics	68.03%
History	Islamic Edu	63.73%
Physics	Chemistry	73.36%
Chemistry	English	73.16%
Add Mathematics	Physics	70.70%
Add Mathematics	Chemistry	72.75%

Table 5 has shown the highest result after the subject has been combined is the combination of subjects Chemistry and Physics with the percentage 73.36%. The result is the same with the Chemistry subject without the combination of subject followed by the combinations of English and Chemistry subjects with the 73.16%.

Meanwhile the lowest result of the combination subjects is Islamic Education and History. As seen in Table 3 and Table 4, the Elective subject itself with the combination of the elective subjects has shown the higher data mining result compare to the combination with the core subjects. Hence the scenario seen, the educators need to work harder on the elective subjects in order to achieve the excellent class while to not ignoring in emphasizing the core subjects.

After analyzing the Naïve Bayesian model that applied to the data of the second semester, it is observed that Chemistry subject do affected the Excellent Class. The confusion matrix below shows the performance of accuracy on the Chemistry subject. A confusion matrix is a table that is used to describe the performance of a classification model on a set of test data.

4.1 Confusion Matrix:

Excellent Good Average Poor

197 58 2 0
= 197/257 × 100 = 77% : Excellent

19 104 21 0
= 104/144 × 100 = 72% : Good

3 26 57 0
= 57/86 × 100 = 66% : Average

0 0 1 0
= 0/1 × 100 = 0% : Poor

From the results, the knowledge about the prediction model is presented. As it visualizes, the model has able to predict Excellent class with 77%, 72% for the *Good* class and 66% for the *Average* class. However this model unable to predict the poor class.

4.2 Analysis on Chemistry Subject

As the results shown that Chemistry Subject has the highest accuracy result itself or with the combination of other subjects, the research then goes deeper on the chemistry subject's performance to identify the number of students by grades.

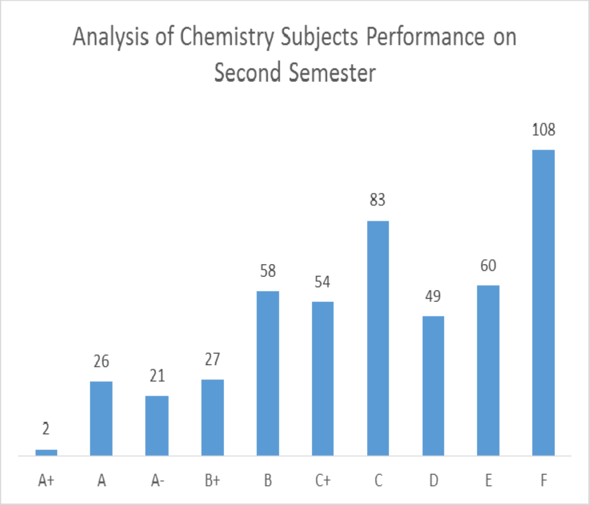

Figure 1: Chemistry Subject Performance on Second Semester

As we can see, 59 students has scored A+, A and A- in Chemistry Subjects while 85 have scored B+ and B. The students who have scored A+, A and A- for this subjects have been categorized as Excellent for their SPM examination and 79/85 students that have scored B+ and B have also categorized in Excellent category. From the Figure above, it is observed that there are students whose obtain E and F grades with 168/488 that is equal to 34% compare to the A+,A,A-,B+ and B grades =30%.

5. CONCLUSIONS

The paper analyzed the potential use of one of the data mining technique called Naïve Bayesian algorithm in enhancing the quality of students' performances at Sijil Pelajaran Malaysia level. The aim of this study is to examine the Naive Bayes algorithm which is one of the classification methods in data mining, to identify the hidden information between subjects that affected the performance of students in Sijil Pelajaran Malaysia (SPM). As discussed in Section 3, it is proven that core elective subject such as Chemistry do affected the *Excellent* category. Students that have scored with A or B in elective subjects are potentially inclined to categorize in *Excellent* class. Thus, this situation can be the matter of concern for the educators and the school management in order to come out with the better approach. This paper should be further enhance in future with different parameters and methods in order to classify the *Poor* category students. However it can be concluded that the students' mastery of Language Subject such as English is as essential as having the mastery in Science and Mathematics Subjects as the results shows the combination of both English and Chemistry is 73.16%. Having full grasp in both the core subject and the elective subject will somehow influence the students' overall academic performance. Proving that both of the subjects are of equally important in determining the students' excellent performance.

For future work we may redefine our techniques with more attributes and data in order to get more accurate outputs that can be useful for teachers in order to improve the students' learning outcomes for semester examination and Sijil Pelajaran Malaysia.

ACKNOWLEDGEMENTS

This work is partially supported by UniSZA.

REFERENCES

[1] A. Badr El Din Ahmed and I. Sayed Elaraby, “Data Mining: A prediction for Student's Performance Using Classification Method,” *World J. Comput. Appl. Technol.*, vol. 2, no. 2, pp. 43–47, 2014.

[2] C. El Moucary, M. Khair, and W. Zakhem, “Improving student's performance using data clustering and neural networks in foreign-language based higher education,” *Res. Bull. Jordan ACM - ISWSA*, vol. 2, no. Iii, pp. 27–34, 2011.

[3] Y. Zhang, S. Oussena, T. Clark, and K. Hyensook, “Using data mining to improve student retention in HE: A case study,” *Proc. 12th Int. Conf. Enterp. Inf. Syst.*, vol. 1, pp. 190–197, 2010.

[4] M. H. I. Shovon and M. Haque, “An Approach of Improving Student ' s Academic Performance by using K-means clustering algorithm and Decision tree,” vol. 3, no. 8, pp. 146–149, 2012.

[5] J. F. Superby, J.-P. Vandamme, and N. Meskens, “Determination of factors influencing the achievement of the first-year university students using data mining methods,” *Proc. Int. Conf. Intell. Tutoring Syst. Work. Educ. Data Min.*, 2006.

[6] N. J. Namdeo, Jyoti, “Predicting Students Performance Using Data Mining Technique with Rough Set Theory Concepts,” *Int. J. Adv. Res. Comput. Sci. Manag. Stud.*, vol. 2, no. 2, pp. 367–373, 2014.

[7] R. V. Monika Goyal, "Applications of Data Mining in Higher Education," *IJCSI Int. J. Comput. Sci. Issues*, vol. 9, no. 2, pp. 113–120, 2012.

[8] A. El-Halees, "Mining Students Data To Analyze Learning Behavior : a Case Study Educational Systems," *Work*, no. February, 2008.

[9] A. Mohamed, W. Husain, and A. Rashid, "The Third Information Systems International Conference A Review on Predicting Student ' s Performance using Data Mining Techniques," *Procedia - Procedia Comput. Sci.*, vol. 72, pp. 414–422, 2015.

[10] S. Borkar and K. Rajeswari, "Predicting students academic performance using education data mining," *Int. J. Comput. Sci. Mob. Comput.*, vol. 2, no. 7, pp. 273–279, 2013.

[11] B. K. Baradwaj and S. Pal, "Mining educational data to analyze students' performance," *Int. J. Adv. Comput. Sci. Appl.*, vol. 2, no. 6, pp. 63–69, 2011.

[12] V. Kumar and A. Chadha, "Mining association rules in student's assessment data," *Int. J. Comput. Sci. Issues*, vol. 9, no. 5, pp. 211–216, 2012.

[13] M. Da, W. Wei, H. Hai-guang, and G. Jian-he, "The Application of Bayesian Classification Theories in Distance Education System," no. July, pp. 9–16, 2011.

[14] M. N. Borde, M. N. Dubey, and M. Aakankshi, "Knowledge Discovery from Social Media Data for Education Enhancement," vol. 5, no. 4, pp. 875–877, 2016.

[15] S. Karthika and N. Sairam, "A Naïve Bayesian Classifier for Educational Qualification," vol. 8, no. July, 2015.

[16] Krishnaiah, V., Narsimha, G., & Chandra, N. S. (2013). Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques, International Journal of Computer Science and Information Technologies, Vol. 4 (1) , 2013, 39 - 45 Diagnosis 4(1), 39–45.

[17] Wang, S. (2015). Detection of Pathological Brain in MRI Scanning Based on WaveletEntropy and Naive Bayes Classifier Detection of Pathological Brain in MRI Scanning Based on Wavelet-Entropy and Naive Bayes Classifier, (April).http://doi.org/10.1007/978-3-319-16483-0.

[18] Dai, Y., & Sun, H. (2014). The naive Bayes text classification algorithm based on rough set in the cloud platform, Journal of Chemical and Pharmaceutical Research, 2014, 6(7): 1636-1643.

[19] Nawang, H., & Makhtar, M. (2017) Key-Based Approach in Data Integration for Students Information System in Secondary School, World Applied Science Journal 35 (2017) 77-81.

[20] Abdel-moneim, M. S., & El-bastawissy, A. H. (2015). Data Quality Based Data Integration Approach, World of Computer Science and Information Technology Journal (WCSIT) 5(10), 155–164.

[21] Baker, R. S. J. D. (2010). Data mining for education. International Encyclopedia of Education, 7, 112–118. http://doi.org/10.4018/978-1-59140-557-3, 3rd Ed. Elsevier, Oxford, UK.

[22] E.Khedr, A., & I. El Seddawy, A. (2015). A Proposed Data Mining Framework for Higher Education System. International Journal of Computer Applications, 113(7), 24–31. http://doi.org/10.5120/19839-1693.

[23] Elgamal, A. F., Mosa, N. A., & Amasha, N. A. (2014). Application of Framework for Data Cleaning to Handle Noisy Data in Data Warehouse, International Journal of Soft Computing and Engineering (IJSCE), ISSN: 2231-2307, Volume-3, Issue-6, January 05, 2014. Application (6), 226–231.

[24] Garcı, E. (2008). Data mining in course management systems : Moodle case study and tutorial, Computers & Education 51, 368–384 51, 368–384. http://doi.org/10.1016/j.compedu.2007.05.016.

[25] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. ACM SIGKDD explorations newsletter, 11(1), 10-18.

View publication stats