PAPER • OPEN ACCESS

Data Mining Applied in School Dropout Prediction

To cite this article: Amelec Viloria *et al* 2020 *J. Phys.: Conf. Ser.* **1432** 012092

Data Mining Applied in School Dropout Prediction

Amelec Viloria[1], Jesús García Guliany[2], William Niebles Núñez[3], Hugo Hernández Palma[4] Leonardo Niebles Núñez[5]

[1]Universidad de la Costa, Barranquilla, Colombia.
[2]Universidad Simón Bolívar, Barranquilla, Atlántico, Colombia
[3]Universidad del Sucre, Sincelejo, Sucre, Colombia.
[4, 5] Universidad del Atlántico, Puerto Colombia, Atlántico, Colombia.

[1]**Email:** aviloria7@cuc.edu.co

Abstract. In recent years, many studies have emerged about regarding the topic of school failure, showing a growing interest in determining the multiple factors that may influence it [1]. Most of the researches that attempt to solve this issue [2] are focused on determining the factors that most affect the performance of students (dropout and failure) at the different educational levels (basic, middle and higher education) through the use of the large amount of information that current computer equipment allows to store in databases. All these data constitute a real gold mine of valuable information about students. But, identifying and finding useful and hidden information in large databases is a difficult task [3]. A very promising solution to achieve this goal is the use of knowledge mining techniques or data mining in education, which has resulted in so-called Educational Data Mining (EDM) [4]. This new area of research is concerned with the development of methods for exploring data in education, as well as the use of these methods to better understand students and the contexts where they learn [5].

1. Introduction

EDM techniques are successfully applied for creating models that predict student´s performance [6], providing promising results that demonstrate how certain sociological, economic, and educational characteristics of students can affect academic performance [7]. It is important to note that most of the research on data mining applied to dropout at the higher education level [8] and, to a greater extent, in the distance education modality [9]. On the contrary, very little information has been found on the application in basic or secondary education, where just simple analyses of the information based on statistical methods have been carried out [10]. There are some differences and/or advantages in applying data mining over just using statistical models [11]:

- In statistical techniques (data analysis), depending on the model, the likelihood of the data is usually used as a quality criterion. But data mining uses a more straightforward criterion, as using the percentage of well classified data.
- In statistics the search is usually carried out by modelling based on a hill-climbing algorithm in combination with a hypothesis test based on a likelihood ratio. In data mining, a meta-heuristic search is often used.
- Data mining is geared to working with billions of data. On the other hand, statistics does not usually work in such a large size databases and high dimensionality.

This paper proposes the use of data mining techniques to detect the factors that most influence the failure or dropout of middle or high school students. In addition, it is proposed to use different data mining techniques because it is a complex problem, where data tend to be highly dimensional (there are many factors that can influence this) and tend to be very unbalanced (most students tend to approve and just a minority tend to fail). The final objective is to detect, as soon as possible, the students who present these factors in order to offer some attention or help to avoid and/or diminish the school failure.

2. Methods

The method proposed is based on data mining techniques and consists of the typical steps of a knowledge extraction process. - Data collection. At this stage all available information is collected from the students. To do this, first select the set of factors that may affect it and then collect them from the different available data sources [12]. Finally, all the information must be integrated into a single dataset. - Pre-processing. At this stage, the data are prepared in order to be able to apply the data mining techniques. Typical pre-processing tasks such as data cleansing, variable transformation and data partitioning are performed first.

In addition, other techniques such as attribute selection and data re-balancing are applied to solve the problems of high dimensionality and unbalance that this type of data set normally presents. - Data mining. At this stage, data mining algorithms are applied to predict school failure as if it were a classification problem. To this end, it is proposed to use classification algorithms based on rules and decision trees because they are "white box" techniques that generate highly interpretable models that allow their direct use in decision-making processes. In addition to the traditional classification, it is also proposed to use classification based on costs or penalties to try to correct the problem of data imbalance [13].

Finally, the different algorithms must be evaluated and compared to determine which of them obtain the best classification results. - Interpretation of results. In this last stage, the models that obtain the best results are analyzed in order to use them in the detection of school failure. For this purpose, the factors that appear in the rules and/or decision trees are analyzed, as well as the values they present and how they are related to other factors. A case study is described below using data from real students to show the usefulness of the proposed method.

3. Data Collection

Dropout School is known as "the problem of a thousand causes" [14] due to the large number of possible causes or factors like personal, academic, physical, economic, family, social, institutional, pedagogical orders, etc., that can affect student´s failure or abandonment. In this specific case study, the data used are from students of the Preparatory Program of three schools in Colombia. All students who participated in this study were new entrants in the 2018-2019 academic year in the upper-middle level of Colombian education.

All the information was collected from three different sources: a) A survey applied to all students in the middle of the year, with the purpose of obtaining information to detect important factors that may have an impact on their school performance. b) The Colombian Institute for the Evaluation of Education (IFCES). When students register for the examination, they are also given a socioeconomic study, from which, part of the information is extracted. c) The School Services Department of each school, where all grades obtained by the students are collected.

Table 1 shows all the variables grouped in the three data sources.

4. Data Pre-Processing

Before applying a data mining algorithm, it is generally necessary to perform some preprocessing tasks, which allow transforming the original data to a more suitable form to be used by the particular algorithm. In this case study the tasks consisted of the data integration, cleaning, transformation and discretization [15]. Data integration consists of grouping all available information about each student from the three data sources into a single electronic data file.

At the cleaning stage, those students who did not have 100% complete information were extracted from the data set. That is, if during the socioeconomic study conducted by the IFECS or during the survey to detect factors affecting academic performance, a student omitted one or more responses, then the student was excluded from the data set.

Table 1. Used variables and source of origin

Source	**Variable**
Survey	Semester and Group, Shift, Motivation Level, Administrative Sanction, No. of Friends, Additional Study Time, Study Form, Place of Study, When Studying, Doubts, Marital Status, Children, Religion, Chosen College Career, Influence on College Decision, Personality, Physical Disability, Serious Illness, Alcoholic Beverages, Smoking, Economic Status, Study Resources, Scholarship, Work, Who Lives with, Mother's Education Level, Father's Education Level, No. of siblings, Order of birth, Space to study, Stimulation of parents, Community inhabitants, Years living in community, Type of transportation, Distance to school, Attendance to classes, Bored in class, Considers useful knowledge, Difficult subject, Taking notes, Excess of homework, No. of students in group, Way of teaching, School infrastructure, Advisor, Interest of the institution.
IFECS	Age, Sex, Department of origin, School of origin regime, Secondary school model, Secondary school average, Mother's job, Father's job, No. of PC in family, Limited for exercise, Frequency of exercise, Time of exercise sessions, Grades in Logical Mathematical Reasoning, Grades in Mathematics, Grades in Verbal Reasoning, Grades in Spanish, Grades in Biology, Grades in Physics, Grades in Chemistry, Grades in History, Grades in Geography, Grades in Civic Formation, Grades in Ethics, Grades in English and Grade of EXANI I.
School Department	Grade in Humanities, Grade in Reading and Writing Workshop, Grade in English, Grade in Computer, Academic Status, Grade in Mathematics, Grade in Physics, Grade in Social Sciences.

In addition, the age of the students was also put in years, because the information provided by the IFECS, contained day, month and year of birth. In the discretization stage, both the grades obtained in the secondary school average and in the subjects attended during the semester changed from numerical format (values from 0 to 10) to nominal or categorical format. Specifically, the tags used and the discretization ranges of the grades were:

Excellent (4.5 to 5); Good (3.0 to 4.4); Deficient (2.0 to 2.9); Very Deficient (less than 2.0) and Not presented. A file was created in Weka's. ARFF format [16], which is the data mining software for testing.

After performing the previous pre-processing tasks, a first data file with 89 attributes/variables is available for 1,268 students. This data file was partitioned (20 partitions) in order to cross-check the classification tests. A partition is the random division of the original data file into two others, one for the training stage and the other for the test stage. Due to the large number of collected attributes (89), an analysis or selection study of attributes was also carried out to determine which of them most influence the output variable or class to be predicted (Academic Status). Several attribute selection methods available in the Weka software were used to select the most relevant variables [17].

When selecting the more frequent attributes, the top 15 attributes, out of 89 attributes, remained. Again, this data file was split into 10 training files and 20 test files. Finally, as mentioned above, the data set is unbalanced. In this case, from 1268 students, 1201 approved and 67 failed or dropped out. Therefore, the data are considered to be unbalanced, that is to say, there is a majority of students who approved as opposed to a minority who failed.

One way to solve the problem is to act in the pre-processing stage of the data, making an over-sampling or balancing of the distribution of classes. For it, there are several re-balancing algorithms and one widely used is the so-called SMOTE (Synthetic Minority Oversampling Technique), available in Weka as a data filter [18].

In general terms, SMOTE synthetically introduces elements of the minority class to balance the data sample, based on the nearest neighbour rule. The created synthetic elements are introduced into the space between the elements of the minority class. Depending on the size of the required over-sampling, the nearest neighbors are randomly selected [19]. In this case the data set with the 15 best attributes and with 1268 instances was partitioned as follows:

Each training file was rebalanced with the SMOTE algorithm so that it had 50% of Approved instances and 50% of Suspended instances, leaving the test files without re-balancing. After performing all data pre-processing tasks, the result was: - 20 training and testing files with all attributes (89 attributes). - 20 training and testing files with just the top 15 attributes. - 20 training and testing files with just the top 15 attributes, where the training files are re-balanced.

5. Results and discussion

In a first experiment, 10 classification algorithms were applied using all available attributes. In a second experiment, just the best attributes or variables were considered. In a third experiment were using the rebalanced data files. In a last experiment, different sorting costs were considered. 10 classification algorithms were selected from those available by the Weka data mining tool. This selection is carried out because the algorithms are all of white box type, that is, it results in an output model understandable to the user, because it obtains either If-Then classification rules or decision trees.

The 5 induction algorithms of classification rules used are: JRip, NNge, OneR, Prism [20] and Ridor.

A decision tree is a set of conditions organized in a hierarchical structure, which contains zero or more internal nodes and one or more sheet nodes. Internal nodes have two or more secondary nodes and contain divisions, which test the value of an expression of attributes. The arcs from one internal node to another secondary (or lower hierarchy) node are labeled with different test outputs from the internal node. Each sheet node has an associated class label. The decision tree is a predictive model in which an instance is classified by following the path of fulfilled conditions from the root to a leaf, which will correspond to a labelled class. A decision tree can easily be converted into a set of classification rules [21].

The 5 decision tree algorithms to be used are J48 [22], SimpleCart [23], ADTree [24], RandomTree and REPTree. In the first experiment, the 10 algorithms were executed using all available information, that is to say, the data files with 89 attributes of the 1,268 students. 20 partitions were cross-validated. In this type of cross validation, training and testing is performed ten times with the different partitions.

The results obtained (the average of 10 executions) with the test files of the application of the classification algorithms are shown in Table 2. In addition to the overall or total accuracy, the percentages for each of the two class values (Approved and Suspended/Abandoned) and a central tendency measure quite used for cases similar to this one of unbalanced data, the geometric mean, were indicated. It can be noted in Table 3 that the accuracy percentages obtained for the total accuracy and for the Approvals are high, but different for those who suspended and the geometric mean. Specifically, the algorithms that obtain the maximum values are: JRip (in the Suspended/Abandoned ratio and geometric mean), and ADTree (in the Suspended/Abandoned ratio and accuracy) [25].

In the second experiment, the files with the best 15 attributes were used, consisting of running the 10 classification algorithms again to be able to check how the selection of attributes affected the prediction. Table 3 shows the results of the cross validation (the mean of the 10 runs) of the classification algorithms using just the 15 best attributes. When comparing Tables 2 and 3, it can be noted that the algorithms improved the accuracy percentage when using just the best attributes. Although there are some algorithms that worsen a little, in general, the trend is an improvement. In fact, better maximum values are obtained than those obtained with all attributes.

Again, the algorithms that obtain these maximum values are the JRip and ADTree. In spite of having obtained better results, a good classification of the minority class Suspended/Abandoned is still not obtained, obtaining a maximum value of just 81.9% of success as opposed to 99.4% of success of the Approved majority class. This may be because the data is very unbalanced. This characteristic of the data is an undesirable fact and can negatively affect the results obtained when applying the classification algorithms, and it is because the algorithms tend to focus on classifying the individuals of the majority class in order to obtain a good percentage of total classification and forget the individuals of the minority class.

Table 2. Cross validation using all 89 available attributes

Algorithm	% Success Approved	% Success Suspended	% Accuracy Total	Geometric Mean
JRip	96.3	77.2	97.1	86.4
NNge	97.4	72.2	97.2	84.1
OneR	97.5	42.6	94.2	63.5
Prism	98.3	26.3	94.3	48.8
Ridor	95.5	64.1	94.6	79.1
ADTree	98.6	75.6	98.5	86.0
J48	96.2	54.4	94.2	71.0
RandomTree	94.6	47.4	92.4	69.2
REPTree	97.1	55.6	95.2	73.4
SimpleCart	96.4	64.2	95.6	78.6

Table 3. Cross validation using the 15 attributes selected as the best

Algorithm	% Success Approved	% Success Suspended	% Accuracy Total	Geometric Mean
JRip	96.0	81.9	94.7	88.0
NNge	97.0	75.7	95.1	87.6
OneR	97.7	42.7	92.7	65.3
Prism	99.4	43.2	93.7	65.4
Ridor	95.5	67.4	92.2	81.9
ADTree	99.4	77.4	97.4	87.2
J48	97.4	54.5	93.7	74.1
RandomTree	98.2	64.3	95.8	77.7
REPTree	96.8	61.2	95.4	77.5
SimpleCart	97.0	65.2	94.2	78.7

In the third experiment, an attempt was made to solve or mitigate this problem of data unbalancing. To do this, the files with the best 15 attributes were used again, but then, the training files were previously re-balanced with the SMOTE algorithm. Table 4 shows the results of this third test. When analyzing this table and comparing it with the previous results of Tables 2 and 3, it is noted that most of the algorithms increased their accuracy in prediction, obtaining new maximum values in almost all measurements, except in the percentage of total accuracy. In this case, the algorithms that obtained the best results were the Prism algorithm, OneR and, again, the ADTree algorithm.

Finally, another way to address the problem of classification of unbalanced data is to conduct a cost-sensitive classification. Traditional classification does not distinguish whether one of the classes to be classified is more important than another, i.e., whether one has a different classification cost. Optimizing the classification rate without considering the cost of errors can often lead to sub-optimal results due to the high cost that the misclassification of a minority instance can cause. In fact, in this particular problem, the focus is the classification of students in the Suspended/Abandoned class (minority class).

The Weka software allows classification considering cost, for which the CostSensitiveClassifier is used and to which both the cost matrix and the classifying algorithm to be used are associated. After making several tests with different costs, it was found that using the matrix [0, 1; 4, 0], the best classification results were obtained, indicating that when making the classification it is taken into account that it is 4 times more important to classify correctly the cases of Suspended/Abandoned than the cases of Approved. Finally, the fourth and last experiment consisted of executing the ten classification algorithms using costs and the files with the best 15 attributes.

Table 4 shows the obtained results. When comparing the results obtained in Table 4 with respect to Table 5, it is noted that, although the percentage of Approved worsened a little, the total accuracy, on the contrary, increased (obtaining the maximum values with respect to all the previous experiments) both the geometric mean and the percentage of successes of Suspended/Dropped out, which is precisely the interest of the study (detecting students at risk). In this case, the algorithms with the best results were Prism, JRip, ADTree and SimpleCart.

Table 4. Cross validation using the best attributes and balancing the training data with the smote

Algorithm	% Success Approved	% Success Suspended	% Accuracy Total	Geometric Mean
JRip	96.7	65.3	94.7	77.7
NNge	97.6	78.5	97.8	86.1
OneR	88.7	87.5	88.4	87.3
Prism	99.7	37.2	95.2	58.0
Ridor	97.8	71.1	94.2	82.3
ADTree	98.3	86.6	96.1	93.2
J48	95.4	75.1	93.7	83.7
RandomTree	95.1	68.2	94.5	78.6
REPTree	95.5	76.0	94.5	82.6
SimpleCart	95.5	77.5	94.5	84.5

Table 5. Cross validation using the best attributes and the cost of classification

Algorithm	% Success Approved	% Success Suspended	% Accuracy Total	Geometric Mean
JRip	97.3	94.2	96.0	93.7
NNge	97.3	72.7	95.8	83.0
OneR	95.1	71.0	93.7	80.5
Prism	98.5	38.7	95.5	54.1
Ridor	95.9	57.4	93.4	73.0
ADTree	97.2	82.7	96.5	89.2
J48	94.7	81.2	95.3	87.1
RandomTree	94.5	69.3	94.0	81.3
REPTree	96.4	66.0	92.6	78.1
SimpleCart	96.3	91.4	97.8	94.5

6. Conclusions

This study, a set of experiments were carried out in order to predict, with a good degree of accuracy, the academic status of students at the end of the first semester through the use of classification algorithms. This objective is not easy to achieve because it is not just an unbalanced data set, but also a multifactorial problem. It is important to comment that a very important task in this work was the collection of information and the preprocessing of the data, since the quality and reliability of the information directly affects the obtained results. It is a hard task which involves investing a lot of time and willingness of whoever is in charge of carrying it out. Specifically, the data from the applied survey had to be captured, in addition to integrating data from three different sources to build the final data set.

With regard to the classification results of the different tests, the main conclusions obtained are: - Classification algorithms can be used successfully to predict student´s academic performance.

- The usefulness of feature selection techniques was shown when many attributes are available, improving the classification of algorithms by using a reduced set of 15 attributes from the 89 initially available.

- Two different ways of addressing the problem of classification of unbalanced data was shown, both by rebalancing the data and by considering different classification costs and applying a cost matrix.

Both ways were successful in improving classification, although the cost matrix obtained the best classification results from the minority class. Regarding the knowledge extracted from the classification models obtained, the main conclusions are:

- The use of classification algorithms of the white box type allows models to be understood by a user who is not an expert in data mining in decision-making processes. In this case, the final objective is to detect students with problems or tendency to Suspend/Abandon in order to prevent it on time.

- With respect to the factors that mostly appeared in the obtained models:

The grades of the subjects in the semester are those that appear in greater measure in the outputs of the classification algorithms, being the most important those that obtained deficient grades or did not present themselves to the subjects of Physics 1, Humanities 1, Mathematics 1 and English 1. In addition, other attributes that appeared in the models were: age (particularly those older than 15), having siblings (particularly 1), the group attended, the (regular) level of motivation to study, not presenting to the Reading and Writing Workshop, living in a large city (particularly in a community of more than 20,000 inhabitants) and considering that the most difficult subject is Mathematics.

It is interesting that the poor grade of a subject such as Humanities, which is generally approved by the majority of students, appears in the models obtained as a factor related to student failure. It should also be noted that, in this study, the students' previous grades were used, focusing on not just social, cultural and demographic attributes for two reasons. First, the ranking results obtained if the attributes from previous grades were removed, got a lot worse. Second, the grades previously obtained by students to predict failure is a widely used resource in similar studies [26], [27].

From the rule models and decision trees generated by the data mining algorithms, a system can be implemented to alert the teacher and parents about students potentially at risk of suspension or abandonment. As an example of possible actions that can help students at risk, it is proposed that once a student at risk is detected, they should be assigned a teacher-tutor to provide both academic support and motivational and guidance to avoid failure of the student. Finally, as a next step in this research more tests will be performed using other data than that captured and pre-processed in this research.

References

[1] L. A. Alvares Aldaco, “Comportamiento de la Deserción y Reprobación en el Colegio de Bachilleres del Estado de Baja California: Caso Plantel Ensenada”, X Congreso Nacional de Investigación Educativa. México, 2009.

[2] F. Araque, C. Roldán, A. Salguero, “Factors Influencing University Drop Out Rates”, Computers & Education, vol. 53, pp. 563–574, 2009.

[3] M. N. Quadril and N. V. Kalyankar, “Drop Out Feature of Student Data for Academic Performance Using Decision Tree Techniques”, Global Journal of Computer Science and Technology, vol. 10, pp. 2-5, 2010.

[4] C. Romero and S. Ventura, “Educational data mining: A Survey From 1995 to 2005”, Expert System with Applications, vol. 33, pp. 135-146, 2007.

[5] M. M. Hernández, “Causas del Fracaso Escolar”, XIII Congreso de la Sociedad Española de Medicina del Adolescente, pp.1-5. 2002.

[6] E. Espíndola, A. León, “La Deserción Escolar en América Latina un Tema Prioritario Para la Agenda Regional”, Revista Iberoamericana de Educación, no. 30, pp. 1-17, 2002.

[7] I. H. Witten and F. Eibe, “Data Mining, practical Machine Learning Tools and Techniques”, Second Edition, Morgan Kaufman Publishers, 2005.

[8] M. A. Hall and G. Holmes, “Benchmarking Attribute Selection Techniques for Data Mining”, Technical Report 00/10, University of Waikato, Department of Computer Science, Hamilton, New

Zealand, Julio 2002. Available: http://www.cs.waikato.ac.nz/~ml/publications/2000/00MH-GHBenchmarking.pdf.
[9] N. V. Chawla, K. W. Bowyer, L. O. Hall, W.P. Kegelmeyer, "Synthetic Minority Over-sampling Technique", Journal of Artificial Intelligence Research, 2002, 16:321-357.
[10] J. Cendrowska, "PRISM: An algorithm for inducing modular rules", International Journal of Man-Machine Studies, vol. 27, no. 4, pp. 349-370, 1987.
[11] J. R. Quinlan, "C4.5: Programs for Machine Learning", Morgan Kaufman Publishers, 1993.
[12] L. Breiman, J. H. Friedman, R. A. Olshen, C. J. Stone, "Classification and Regression Trees", Chapman & Hall, New York, 1984.
[13] Y. Freund and L. Mason, "The Alternating Decision Tree Algorithm", Proceedings of the 16th International Conference on Machine Learning, pp. 124-133, 1999Lantz B. Machine learning with R: learn how to use R to apply powerful machine learning methods and gain an insight into real-world applications. Birmingham: Packt Publ; 2013.
[14] Bucci, N., Luna, M., Viloria, A., García, J. H., Parody, A., Varela, N., & López, L. A. B. (2018, June). Factor analysis of the psychosocial risk assessment instrument. In International Conference on Data Mining and Big Data (pp. 149-158). Springer, Cham.
[15] Gaitán-Angulo, M., Viloria, A., & Abril, J. E. S. (2018, June). Hierarchical Ascending Classification: An Application to Contraband Apprehensions in Colombia (2015–2016). In Data Mining and Big Data: Third International Conference, DMBD 2018, Shanghai, China, June 17–22, 2018, Proceedings (Vol. 10943, p. 168). Springer.
[16] Viloria, A., & Lezama, O. B. P. (2019). An intelligent approach for the design and development of a personalized system of knowledge representation. Procedia Computer Science, 151, 1225-1230.
[17] Viloria A., Lis-Gutiérrez JP., Gaitán-Angulo M., Godoy A.R.M., Moreno G.C., Kamatkar S.J. (2018) Methodology for the Design of a Student Pattern Recognition Tool to Facilitate the Teaching - Learning Process Through Knowledge Data Discovery (Big Data). In: Tan Y., Shi Y., Tang Q. (eds) Data Mining and Big Data. DMBD 2018. Lecture Notes in Computer Science, vol 10943. Springer, Cham
[18] Viloria, A., Bucci, N., Luna, M., Lis-Gutiérrez, J. P., Parody, A., Bent, D. E. S., & López, L. A. B. (2018, June). Determination of dimensionality of the psychosocial risk assessment of internal, individual, double presence and external factors in work environments. In International Conference on Data Mining and Big Data (pp. 304-313). Springer, Cham.
[19] Hox, J., & Maas, C. (2005). Multilevel analysis. Encyclopedia of Social Measurement, 2, 785–793. doi: 10.1016/B0-12-369398-5/00560-0
[20] Mellado A., Suárez, N., Altimir, C., Martínez, C., Pérez J. C., Krause, M., & Horvath, A. (2017) Disentangling the change-alliance relationship: Observational assessment of the therapeutic alliance during change and stuck episodes. Psychotherapy Research, 27(5), 595-607. doi: 10.1080/10503307.2016.1147657
[21] Ogles, B. M. (2013). Measuring change in psychotherapy research. En M. J. Lambert (Ed.), Bergin and Garfields's Handbook of Psychotherapy and Behavior Change (pp.134– 166). New Jersey: Wiley.
[22] Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods (2nd Ed.). Thousand Oaks, California: Sage.
[23] Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R. T., & du Toit, M. (2011). HLM7:
[24] Hierarchical Linear and Nonlinear Modeling. Chicago, IL: Scientific Software International.
[25] Skrondal, A., & Rabe-Hesketh, S. (2004). Generalized latent variable modeling. Boca Raton: Chapman & Hall/CRC
[26] AlShammari, I., Aldhafiri, M., & Al-Shammari, Z. (2013).A Meta-Analysis of Educational Data Mining on Improvements in Learning Outcomes. College Student Journal, 47(2), 326-333.
[27] Baker, R. S. 1. (2011). Data mining for education. In International encyclopedia of education. 3rd ed. Oxford, UK: Elsevier.