

USING OF DATA MINING TECHNIQUES TO PREDICT OF STUDENT'S PERFORMANCE IN INDUSTRIAL INSTITUTE OF AL-DIWANIYAH, IRAQ

Y.K. Salal, Yasskhudheirsalal@gmail.com,

S.M. Abdullaev, abdullaevsm@susu.ru

South Ural State University, Chelyabinsk, Russian Federation

The aim of paper is to show the benefits of the educational data mining (EDM) techniques, in order to understand about of the factors which lead to technical student's success and failure, and predict their performance and determine the individual learning ability in engineering sciences. For these goals, we use the individual data of 311 student and their grades that were collected in Industrial Institute of Al-Diwaniyah city (Iraq) during 2015–2017 academic years, in order to predict the results of final theoretical exam in industrial drawing by applying EDM techniques, such as association rules mining, classification with decision tree algorithm learning, clustering with Apriori algorithm and anomaly detection implemented as the output model of the clustering. Using Microsoft SQL Server Business Intelligence Development Studio 2012 platform and based on Cross Industry Standard Process for Data Mining, we prepare of 13 nominal and numerical attributes for each student and consequently apply and finally evaluate all 4 EDM techniques. We conclude that: 1) association rules were revealed that the most important factor which contribute to the failure of the student is the "project" attribute; 2) decision tree classification permit to the teacher predict the future students and to correct the student's prediction path, but 3) clustering collects of the students into successful and failure groups and helps to the teacher to guide each group separately, and 4) to detect anomaly by an extension DMX for SQL and correct the education process for students located on the border of the cluster.

Keywords: individual learning, data mining techniques, SQL server business intelligence development studio, clustering, classification, association rules, anomaly detection.

Introduction

The vast amount of data needs special tools to analyze and extract the hidden knowledge. These tools come from several scientific fields such as statistics, machine learning and artificial intelligence, all of which contributed to the birth of a new scientific field, namely "Data Mining" or Knowledge Discovery in Databases DM – KDD [1].

Educational institutions like any other institutions, needs to analyze data in order to increase the number of graduates and improve the educational process as a whole, one of the most promising ways to achieve this goal is to apply data mining techniques on educational field. This sub-specialty in data mining is called Educational Data Mining (EDM) [2].

We need to provide an institutional management for a helpful and deductive recommendation to outdo the problem with students that have low grades, to improve educational performance. For better understand students' academic performance and learning styles, discovery of new patterns of knowledge [3]. This paper investigates of the EDM using mining techniques as clustering, classification, association rules detection, and anomaly detection, a case study of data collected from the industrial institute in Iraq, for possible to draw an individual learning trajectory, therefore, the verification of the individual characteristics. It is considered a problem by itself and needs solutions [4].

EDM is interest with applying, developing, researching, and computerized process to uncover patterns in large aggregate of educational data [5]. We will answer the following questions: how can we preprocess the data, how to apply data mining methods on the dataset, and finally how can we benefit from the discovered knowledge.

1. Methodology and Methods

There are four EDM techniques applied to achieve the purpose for this paper:

1. Association rules detection: to understand the most closely related features that lead to the suc-

Краткие сообщения

cess or failure of students, therefore, the students should avoid these reasons to increase the likelihood of success.

2. Classification: to predict the success or failure of students in the theoretical exam at the end of the semester, to increase the opportunity to avoid failure.

3. Clustering: grouping similar students in terms of academic level within individual groups are successful and failed groups. The teacher can deal with each group separately, based on their academic achievement.

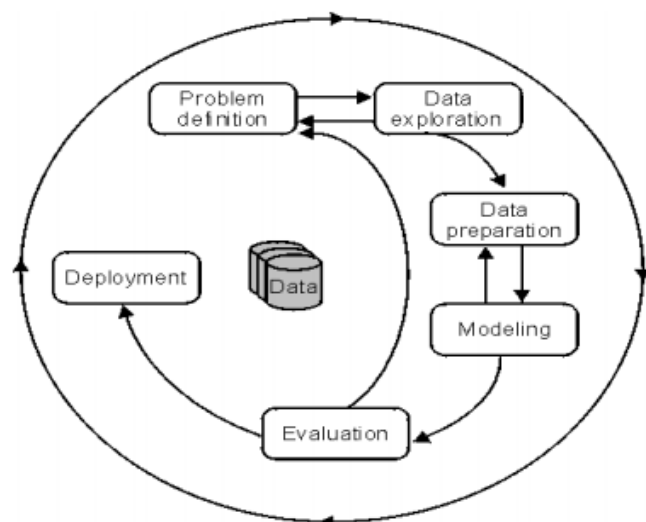


Fig. 1. CRISP-DM process diagram: life cycle of data

methodology for developing data mining technique (Fig. 1), comes up to dissolve the dilemmas that existed in data mining project developments [8].

2. Data collection and preparation

Data preparation is important step and the most critical part of data mining process [9]. Dataset collected for the period 2015–2017 academic years, industrial institute of Al-Diwaniyah city in Iraq, this dataset consists of 301 instances with 13 attribute for two different data types (numerical, nominal), the final result “Pass” indicating to the possibility of student eligible for the final semester exam (theoretical exam). For preparing data, at first the numeric attributes converted to nominal to be compatible with the various algorithms in this paper, and also it will be easy to comprehend to reader. Table 1 shows the selected attributes of the mining process.

Table 1

Database attributes

Attribute	Num.	Nominal data	Description
Stuld		1-311	ID number of the student
First_Name		X	Student's first name (X: to keep student name secret)
Last_Name		Y	Student's last name (Y: to keep student name secret)
Sex		M, F	Male, Female
Family_income	1,2,3	Poor (P), Average (A), Good (G)	Where: P=1, A=2, G=3
Project	1,2,3	P,A,G	Where: P=1, A=2, G=3
Homework	1,2,3	P,A,G	Where: P=1, A=2, G=3
Extra_curricular	1,2,3	P,A,G	Where: P=1, A=2, G=3
Attendance	1,2,3	P,A,G	Where: P=1, A=2, G=3
Sum	0..15	P,A,G	Sum of above numeric data. P=[0..5], A=[6..10], G=[11..15]
Practical_exam	0..15	P,A,G	P=[0..5], A=[6..10], G=[11..15]
Total	0..30	P,A,G, Very Good (VG), Excellent (E)	Total=Sum + Practical_exam. P=[0..10], A=[11..15], G=[16..20], VG=[21..25], E=[26..30]
Pass		No, Yes	Where: “No” when Total=[0..17], “Yes” when Total=[18..30]

To address the empty cells, we assigned the mean value for empty numeric cells, with regard to empty nominal cells, inserted the “missing” word. Thus, the process of data preparation and cleaning has been completed.

3. The application of data mining

Four techniques applied in this part on the dataset; association rules detection, clustering, classification, and anomaly detection. The dataset divided into two samples, the first sample of the dataset consist of 70% for 218 students to represent the modules for training algorithms, as for the second sample consist of 30% for 93 students to testing algorithms. Before applying these algorithms, it is important to identify the selected attributes and how to use them.

There are four types of attributes in CRISP-DM [10]:

Key: indicates that the attribute is a key in the relational spreadsheet.

Input: the attribute is used as input for the algorithm.

Predict: indicates this attribute that required to expect it value, and can used as an input or output for the algorithm.

Predict-only: this attribute that required to expect it value, but can be used only as algorithm output.

Attributes used in CRISP-DM

Table 2

Structure	Classification	Association Rules	Clustering
Attendance	Input	Input	Input
Extra_curricular	Input	Input	Input
Family_Income	Input	Input	Input
Homework	Input	Input	Input
Pass	PredicOnly	PredicOnly	PredicOnly
Practical_exam	Input	Input	Input
Project	Input	Input	Input
StuId	Key	Key	Key

In Table 2 all the attributes are used only nominal values, as well as the three algorithms used these attributes, with regard to the anomaly detection; we applied a query on the clustering result, in section 3.4.

3.1. Association rules detection

The extraction of association rules relies on the Apriori algorithm, to find the most frequent elements, then generate rules as follows:

$A \geq B$ (Support = 2%, Confidence = 70%). Identifying both parameters is very important, because it contributes to the exclusion of non-important rules. So, when Support = 2%, its means that A and B are exists together by 2% of the total number of records, and when the Confidence = 70%, it's means B exists by 70% of the records containing A.

The main purpose of applying the mining technique of extracting the association rules is to reveal the affect factors on the success or failure student in the practical exam (student will be eligible for the theoretical exam or not). Microsoft association rules used to achieve this purpose, by adjusting the most important variables of the algorithm “three variables” (MINIMUM_IMPORTANCE, MINIMUM_PROBABILITY, and MINIMUM_SUPPORT) while the others variables left to take a default values.

Determine the “minimum” of these variables, its means excluding all the rules that less than “minimum”. Therefore, adjusted the value of the variable to an integer greater than “1”, means determining the minimum of variable as (absolute value), but if specify a decimal between “0-1”, here, determine the demand as a percentage.

MINIMUM_SUPPORT: is defined as $\text{Support}(\{A, B\}) = \frac{\text{Number of transactions}(A, B)}{\text{Total number of records}}$, its represents the number of records, that containing both events A, B of the total number of records. Therefore, MINIMUM_SUPPORT means the minimum number of records which contain A and B together to create the rule, and thus excludes all rules that not identical this condition [11].

MINIMUM_PROBABILITY: is the one of characteristics of the association rules, defined as:

$$\text{Probability (B/A)} = \text{Support (A/B)} / \text{Support (A)} . \quad (1)$$

MINIMUM_IMPORTANCE: Is a measurement property of the base and elements group together, also called (interesting score or lift score), and allows the measurement of correlation A and B with each other defined as follows [11]:

$$\text{Importance } (\{A, B\}) = \text{Probability (A/B)} / \text{Probability(A)} \times \text{Probability(B)}. \quad (2)$$

Where:

If Importance = 1 then A, B are independent of each other.

If Importance < 1 then A, B are negative correlation.

If Importance > 1 then A, B are positive correlation.

Definition of rules:

$$\text{Importance (A} \geq \text{B)} = \text{Log [P (B/A)} / \text{P (B/not A)}]. \quad (3)$$

Where:

If Importance = 0 then A, B are independent.

If Importance > 0 then the probability of B increases when A is present or an integer.

If Importance < 0 then the probability of B decreases, if A is present.

Determine values for the previous variables, based on the researcher's desire, so as to give the required results more accurately, when using very small values, many rules will be generated and many of which will be unimportant. In contrast, when large values using will generate very few rules and delete these rules may be useful for the researcher.

3.2. Classification

The Microsoft decision tree based on the ID3 algorithm, a decision tree is a tree structure flowchart, where each node takes one value or a range of values for one attribute. For that, each branch represents a result of the test; the tree leaves offer the distributions of classes [12]. The most current influential attribute is calculated by using the entropy criterion, where choose the attribute that gives less entropy.

MINIMUM_SUPPORT: the number of cases to be present in any node in the tree. Here, the variable value = 7, because the database is relatively small.

SCORE_METHOD: choose among three algorithms, to determine when a node in the decision tree separated into two or more nodes, "Entropy" is selected here, the possibility of all cases:

"1" = Entropy value.

"3" = Bayesian with K2 Prior value.

"4" = Bayesian Dirichlet Equivalent with Uniform Prior value.

SPLIT_METHOD: This variable determines how the node is divided into tree; we choose the "Complete" [11]. The possibility of all cases as follows:

"1" = Value of "Binary": node is divided into two nodes exclusively, so that if our attribute (Practical_Exam) has three values as good, average, poor, becomes (Practical_Exam = good, Practical_Exam = not good).

"2" = Value of "Complete": The node is divided into all the possible values.

So, the attribute which has two values, is divided into two branches, while it has three values, divided into three branches etc.

"3" = Value of "Both": Apply two previous options together and the algorithm will select variable automatically.

3.3. Clustering

Each object is more similar to an object in the same cluster and minimal similar to objects in another clusters [13], so that the distance between the clusters points closer to each other, and away from the points of other clusters.

The Microsoft clustering algorithm in the default case is based on a Scalable Expectation Maximization, to implement the algorithm we need only one variable, (number of clusters). Here, we determined the value of the variable with only two clusters, where we collected the successful students in the practical exam in the cluster and the failure students in another cluster.

3.4. Anomaly detection

The aim of anomaly detection is the process of finding the patterns whose behavior is not normal in a dataset [14]. To find strange results in the clustering process, as the student status, who does not match his/her academic performance during the semester. Since, the SQL-SBIDS does not include a default option to implement this process; therefore, a Data Mining Extensions (DMX) for SQL has been implemented on the output model of the clustering process.

```
SELECT TOP 3 t. [Stuld], Cluster ( ) AS [Cluster], PREDICTCASELIKELIHOOD ( ) AS [Likelihood]
FROM Clustering
NUTURAL PREDICTION JOIN
OPENQUERY ([Database DS], 'SELECT * FROM MARKS') AS t
WHERE PREDICTCASELIKELIHOOD ( ) < 0.5
ORDER BY PREDICTCASELIKELIHOOD ( ) ASC;
```

Fig. 2. DMX query for anomaly detection

DMX language developed by Microsoft in 1999 is designed to create an independent software interface of other companies, and depend on pre-defined concepts for database developers, to create and modify knowledge models resulting from data mining techniques [11]. In other words, DMX in the field of mining, as SQL in the field of databases. DMX query for anomaly detection among students demonstrated on the Fig. 2.

4. Results and Discussion

After the variables have been adjusted, the models are processed in SQL-SBIDS, and then we got three models, a model for each algorithm. In Table 3 shows the resulting of association rules, where observed there are four rules that descending order according to the “Importance” factor, these rules show the most relevant factors to the “Pass” class: Project, Homework, Family_Income and Extra_curricular.

Association rules results

Table 3

Probability	Importance	Rule
0,770	0,682	Homework=Poor, Quiz=Poor→Pass=NO
0,770	0,682	Project=Poor→Pass=No
0,850	0,654	Project=Poor, Family_Income=Poor→Pass=No
0,783	0,652	Extra_Curricular=Poor, Family_Income=Poor → Pass=No

The second rule is more important than the first rule, because we can predict the “Pass” class by using only one attribute is “Project”, while the first rule needs two attributes to achieve the same expectation with the same probability and importance. It is worth noting that the model did not produce any rules which determining when the value of “Pass” is equal to “Yes”. All resultant rules predicts the associated factors with the students failure, so if we want to know the factors that cause success, all we have to do is reflect the results. For example, if the bad project grade leads to a student failure in a probability of 0,77 and 0,682 of confidence, the student should seek high marks in the “project” to raise the probability of success.

Fig. 3 show the decision tree result. To understand the rules of decision tree, must start from the root and move to the nodes until reach the leaves, where each path in the tree represents a rule in the “If-then” form:

```
IF (Practical_exam='poor') Then Pass='No'
IF (Practical_exam='good' and Extra_curricular='average') Then Pass='Yes'
IF (Practical_exam='good' and Extra_curricular='good') Then Pass='Yes'
IF (Practical_exam='good' and Extra_curricular='poor') Then Pass='Yes'
IF (Practical_exam='average' and Attendance='average') Then Pass='No'
IF (Practical_exam='average' and Attendance='good') Then Pass='Yes'
IF (Practical_exam='average' and Attendance='poor') Then Pass='No'
```

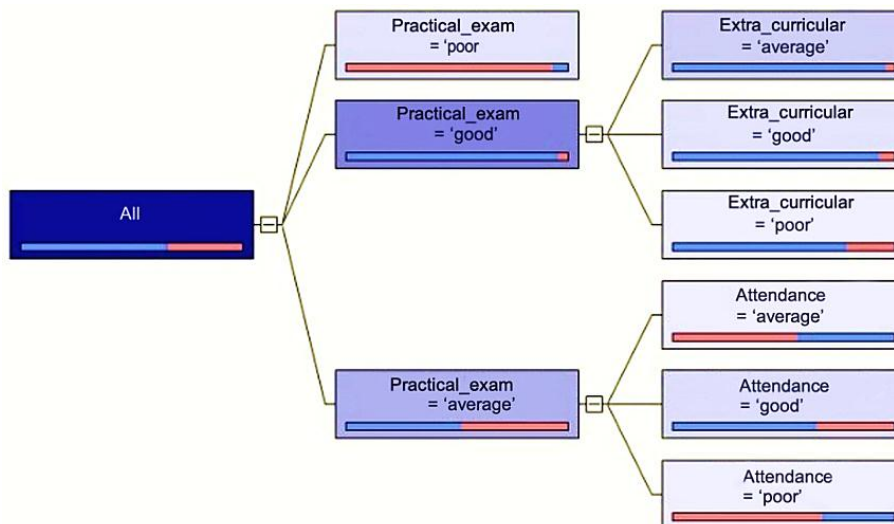


Fig. 3. Microsoft Decision Tree results

From the previous rules we can observe that the Practical_Exam attribute is the most important attribute to predict “Pass” class, because it represents the root of the tree, where it gave less entropy, which is logical because the degree of practical examination is 15 degrees, which constitutes 50% of the grades of final practical exam 30 degrees. Knowing these rules is very useful for teachers and students, because they can pre-determined whether the student is qualified for the theoretical exam or not, therefore they can take appropriate decisions, must the student make a lot of effort on the practical exam or improve his activity and increase the attend of sessions.

Fig. 4a and b shows the two clusters resulting of the applied of the Microsoft Clustering algorithm. Successful students that are belonged to the cluster 1 on Fig. 4a have different characteristics of their peers in the cluster 2 (failure). The length of the line for each attribute or property on Fig. 4b indicates how important it to its cluster, which were arranged according to their importance to the cluster.

The clustering process is a great interest in knowing the characteristics of each group, and thus can the teacher to deal with each group according to their academic level.

- a)
- | | |
|---|--|
| <ul style="list-style-type: none"> ○ Cluster 1 ○ Pass = Yes ○ Practical_exam = good ○ Project = good ○ Extra_curricular = good ○ Family_income = good, average ○ Homework = good | <ul style="list-style-type: none"> ○ Cluster 2 ○ Pass = No ○ Project = poor ○ Extra_curricular = poor ○ Family_income = poor ○ Practical_exam = poor, average ○ Homework = poor |
|---|--|

b)

Cluster 1: Cluster 1 Cluster 2: Cluster 2

Discrimination score for Cluster 1 and Cluster 2			
Variables	Values	Favors Cluster 1	Favors Cluster 2
Pass	Yes	██████████	
Pass	No		██████████
Project	poor		██████████
Extra_curricular	poor		██████████
Practical_exam	good	██████████	
Family_income	poor		██████████
Project	good	██████████	
Extra_curricular	good	██████████	
Family_income	good	██████████	
Practical_exam	poor		██████████
Practical_exam	average		██████████
Homework	poor		██████████
Homework	good	██████████	
Family_income	average		██████████

Fig. 4. Two clusters resulting (a) of the applied of the Microsoft Clustering algorithm and b) importance of attributes

Fig. 5 shows the selected results of a predefined DMX query application demonstrated on Fig. 2. The first record demonstrate that the student no. 94 qualified for the theoretical exam at the end of the semester, but he/she is a member of the second cluster (students group of expected to fail) with a probability of 0.07.

Stuld	Pass	Cluster	Likelihood
94	Yes	Cluster 2	0.0658452552565521
109	No	Cluster 1	0.1524454525451255
85	Yes	Cluster 1	0.1254141125522497

Fig. 5. Results of a DMX query apply of anomalies detection

In the second record, the student no.109 is not eligible for the exam, but it belongs to the cluster1 (students group of expected to success) with a probability of 0.15, while third record, we note that the student is eligible for the exam and belongs to the cluster of expectations of successful students with less score. So what is the problem? In this case gives us a good example to show that, although the student belongs to the right cluster, he/she does not necessarily have to be near the center of the cluster (as a Euclidean distance).

Therefore, the teacher can benefit from the results and look more deeply at the student level during the semester and analyze the situation of students to find the convenient solutions to ameliorate student performance.

Conclusion

This paper highlights the possibilities of applying data mining techniques in the academic field; SQL-SBIDS program was implemented to analyze student's data association rules, classification, clustering, and anomaly detection.

The application of the technique of the association rules was revealed the most factor that caused the failure of the student is the Project.

The application of classification by decision tree algorithm, an easy-to-understand tree was obtained, and the teacher able to predict the future results, through which he could take appropriate action to correct the student's prediction path.

The application of the clustering technique, the students collected into two groups (successful, failure), to understanding what distinguishes each group, which helps the teacher to lead and guide each group separately. An extension DMX for SQL has been implemented on the output model of the clustering process, to find anomaly detection, which is very important for the teachers to correct the path of the education process.

We hope that further research in the field of EDM will help us to resolve the principal problems of computer systems of individual instruction [15].

References

1. Han J., Kamber M., Pei J. *Data Mining Concepts and Techniques*. Morgan Kaufman. Third Edition, USA, 2011. 744 p.
2. Romero C., Ventura S., Pechenizkiy P., Baker M.R. *Hand Book of Educational Data Mining*. CRC Press, USA, 2010. 535 p.
3. Fernandes E. et al. Educational Data Mining: Predictive Analysis of Academic Performance of Public School Students in the Capital of Brazil. *Journal of Business Research, Elsevier Inc.*, 2019, vol. 94, no. 1, pp. 335–343. DOI: 10.1016/j.jbusres.2018.02.012
4. Abdullaev S.M., Lenskaya O.Yu., Salal Y.K. [Computer System of Individual Education: Features of the Student Model]. *University of the XXI Century in the System of Continuous Education. Proceedings of the IV International Scientific Practical Conference*. Chelyabinsk, SUSU Publishing Center, 2018, pp. 7–13. (in Russ.)
5. Matsebula F., Mnrandla E. A Big Data Architecture for Learning Analytics in Higher Education. *AFRICON, IEEE Trans.*, 2017, pp. 951–956. DOI: 10.1109/AFRCON.2017.8095610

6. Zimek S. *Outlier detection*. In: Encyclopedia of Database Systems. Springer, New York, NY, 2017, pp. 1–5.
7. Brown M.S. *What IT Needs To Know About The Data Mining Process*. Published by Forbes. Available at: <https://www.forbes.com/sites/metabrown/2015/07/29/what-it-needs-to-know-about-the-data-mining-process/#9c974cc515>.
8. Mariscal G., Marbán Ó., Fernández C. A Survey of Data mining and Knowledge Discovery Process Models and Methodologies. *The Knowledge Engineering Review*, 2010, vol. 25, no. 2. pp. 137–166. DOI: 10.1017/S0269888910000032
9. Hou Z. Data Mining Method and Empirical Research for Extension Architecture Design. *International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)*. IEEE Trans., 2018, pp. 275–278. DOI: 10.1109/ICITBS.2018.00077
10. *Introducing Business Intelligence Development Studio*. Available at: [https://msdn.microsoft.com/it-it/library/ms173767\(v=sql.105\).aspx](https://msdn.microsoft.com/it-it/library/ms173767(v=sql.105).aspx).
11. MacLennan J., Tang Z., Crivat B. *Data Mining with SQL Server 2008*. Wiley Publ., Indianapolis, Indiana, US. 2008. 672 p.
12. Noughabi Z.E.A. et al. Predicting Students' Behavioral Patterns in University Networks for Efficient Bandwidth Allocation: A Hybrid Data Mining Method (Application Paper). *17th Int. Conf. on Information Reuse and Integration (IRI)*. IEEE Trans., 2016, pp. 102–109. DOI: 10.1109/IRI.2016.21
13. Zhang W., Qin S. A Brief Analysis of the Key Technologies and Applications of Educational Data Mining on Online Learning Platform. *IEEE 3rd International Conference on Big Data Analysis (ICBDA)*. IEEE Trans., 2018, pp. 83–86. DOI: 10.1109/ICBDA.2018.8367655
14. Agrawal S., Agrawal J. Survey on Anomaly Detection Using Data Mining Techniques. *Procedia Computer Science*, 2015, vol. 60, pp. 708–713. DOI: 10.1016/j.procs.2015.08.220
15. Abdullaev S.M., Lenskaya O.Yu., Salal Ya.K. Computer Systems of Individual Instruction: Background and Perspectives. *Bulletin of the South Ural State University. Ser. Education. Educational Sciences*, 2018, vol. 10, no. 4, pp. 64–71. (in Russ.) DOI: 10.14529/ped180408

Received 31 October 2018

УДК 004.8, 004.9

DOI: 10.14529/ctcr190111

ИСПОЛЬЗОВАНИЕ МЕТОДОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ ДЛЯ ПРОГНОЗА УСПЕВАЕМОСТИ СТУДЕНТОВ ИНДУСТРИАЛЬНОГО ИНСТИТУТА ЭЛЬ-ДИВАНИИ, ИРАК

Я.К. Салал, С.М. Абдуллаев

Южно-Уральский государственный университет, г. Челябинск, Россия

Исследуются перспективы применения методов интеллектуального анализа данных (Educational data mining – EDM) в техническом образовании республики Ирак с целью выявления значимых факторов обучения и прогноза успеваемости студента и внедрения индивидуального обучения для студентов инженерных направлений. Для этих целей: 1) нами создана база данных, содержащая индивидуальную информацию и оценки 311 студентов, проходивших обучение в 2015–2017 годах в индустриальном институте города Эль-Дивания; 2) сделаны оценки факторов обучения техническому дизайну и черчению и попытка прогноза результатов выпускного экзамена по этой дисциплине на основе методов EDM, реализованных на платформе Microsoft SQL Server Business Intelligence Development Studio 2012 (поиск ассоциативных правил; классификация с обучением дерева принятия решений; кластеризация с алгоритмом Apriori; обнаружение аномалий обучения). Базируясь на стандарте Cross Industry Standard Process for Data Mining, мы подготовили по 13 номинальных и числовых атрибутов для каждого из студентов, провели обучение методом EDM и затем оценили их преимущест-

ва, заключив, что: 1) ассоциативные правила помогли обнаружить наиболее важный фактор, ведущий к провалу студента на экзамене; 2) дерево решений незаменимо в прогнозе итоговой успеваемости студента, позволяя выбрать траекторию обучения; 3) кластеризация собирает студентов в отдельные по успешности коллективы; 4) обнаружение аномалий данных помогает педагогу находить студентов, находящихся в граничных состояниях. Делается общий вывод о необходимости продолжения работ по апробации EDM и коллективных итогов обучения.

Ключевые слова: индивидуальное обучение, методы интеллектуального анализа данных, SQL server business intelligence development studio, кластеризация, классификация, ассоциативные правила, обнаружения аномалий данных.

Литература

1. Han, J. *Data Mining Concepts and Techniques* / J. Han, M. Kamber, J. Pei. – Morgan Kaufman. Third Edition, USA, 2011. – 744 p.

2. *Hand Book of Educational Data Mining* / C. Romero, S. Ventura, P. Pechenizkiy, M.R. Baker. – CRC Press, USA, 2010. – 535 p.

3. *Educational Data Mining: Predictive Analysis of Academic Performance of Public School Students in the Capital of Brazil* / E. Fernandes et al. // *Journal of Business Research*, Elsevier Inc. – 2019. – Vol. 94, no. 1. – P. 335–343. DOI: 10.1016/j.jbusres.2018.02.012

4. Абдуллаев, С.М. Компьютерные системы индивидуального обучения: особенности модели студента / С.М. Абдуллаев, О.Ю. Ленская, Я.К. Салал // *Университет XXI века в системе непрерывного образования. Материалы IV Международной научно-практической конференции, 11–12 октября 2018, Челябинск.* – Челябинск, 2018. – С. 7–14.

5. Matsebula, F. *A Big Data Architecture for Learning Analytics in Higher Education* / F. Matsebula, E. Mnrandla // *AFRICON*. – *IEEE Trans.*, 2017. – P. 951–956. DOI: 10.1109/AFRCON.2017.8095610

6. Zimek, S. *Outlier detection* / S. Zimek // *Encyclopedia of Database Systems*. – Springer, New York, NY, 2017. – P. 1–5.

7. Brown, M.S. *What IT Needs To Know About The Data Mining Process* / M.S. Brown. – Published by Forbes. – <https://www.forbes.com/sites/metabrown/2015/07/29/what-it-needs-to-know-about-the-data-mining-process/#9c974cc515>.

8. Mariscal, G. *A Survey of Data mining and Knowledge Discovery Process Models and Methodologies* / G. Mariscal, Ó. Marbán, C. Fernández // *The Knowledge Engineering Review*. – 2010. – Vol. 25, no. 2. – P. 137–166. DOI: 10.1017/S0269888910000032

9. Hou, Z. *Data Mining Method and Empirical Research for Extension Architecture Design* / Z. Hou // *International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)*. – *IEEE Trans*, 2018. – P. 275–278. DOI: 10.1109/ICITBS.2018.00077

10. *Introducing Business Intelligence Development Studio*. – [https://msdn.microsoft.com/it-it/library/ms173767\(v=sql.105\).aspx](https://msdn.microsoft.com/it-it/library/ms173767(v=sql.105).aspx).

11. MacLennan, J. *Data Mining with SQL Server 2008* / J. MacLennan, Z. Tang, B. Crivat. – Wiley Publ., Indianapolis, Indiana, US, 2008. – 672 p.

12. *Predicting Students' Behavioral Patterns in University Networks for Efficient Bandwidth Allocation: A Hybrid Data Mining Method (Application Paper)* / Z.E.A. Noughabi et al. // *17th Int. Conf. on Information Reuse and Integration (IRI)*. *IEEE Trans.* – 2016. – P. 102–109. DOI: 10.1109/IRI.2016.21

13. Zhang, W. *A brief analysis of the key technologies and applications of educational data mining on online learning platform* / W. Zhang, S. Qin // *IEEE 3rd International Conference on Big Data Analysis (ICBDA)*. – *IEEE Trans.* – 2018. – P. 83–86. DOI: 10.1109/ICBDA.2018.8367655

14. Agrawal, S. *Survey on Anomaly Detection Using Data Mining Techniques* / S. Agrawal, J. Agrawal // *Procedia Computer Science*. – 2015. – Vol. 60. – P. 708–713. DOI:10.1016/j.procs.2015.08.220

15. Абдуллаев С.М., Ленская О.Ю., Салал Я.К. Компьютерные системы индивидуального обучения: предпосылки и перспективы / С.М. Абдуллаев, О.Ю. Ленская, Я.К. Салал // *Вестник ЮУрГУ. Серия «Образование. Педагогические науки»*. – 2018. – Т. 10, № 4. – С. 64–71. DOI: 10.14529/ped180408

Краткие сообщения

Салал Ясс Кхудейр, аспирант кафедры системного программирования, Южно-Уральский государственный университет, г. Челябинск; Yasskhudheirsalal@gmail.com.

Абдуллаев Санжар Муталович, д-р геогр. наук, профессор кафедры системного программирования, Южно-Уральский государственный университет, г. Челябинск; abdullaevsm@susu.ru.

Поступила в редакцию 31 октября 2018 г.

ОБРАЗЕЦ ЦИТИРОВАНИЯ

Salal, Y.K. Using of Data Mining Techniques to Predict of Student's Performance in Industrial Institute of Al-Diwaniyah, Iraq / Y.K. Salal, S.M. Abdullaev // Вестник ЮУрГУ. Серия «Компьютерные технологии, управление, радиоэлектроника». – 2019. – Т. 19, № 1. – С. 121–130. DOI: 10.14529/ctcr190111

FOR CITATION

Salal Y.K., Abdullaev S.M. Using of Data Mining Techniques to Predict of Student's Performance in Industrial Institute of Al-Diwaniyah, Iraq. *Bulletin of the South Ural State University. Ser. Computer Technologies, Automatic Control, Radio Electronics*, 2019, vol. 19, no. 1, pp. 121–130. DOI: 10.14529/ctcr190111
