Mining Educational Data to Analyze Students' Behavior and Performance

Moustafa M. Kurdi
American University of Technology
AUT
Tyre, Lebanon
mostafa.alkirdi@aut.edu

Hatim Al-Khafagi
American University of Technology
AUT
Tyre, Lebanon
mase.school.hatim@gmail.com

Imad Elzein
International University in Beirut
BIU
Beirut, Lebanon
imad.zein@liu.edu.lb

***Abstract*—Predicting and supporting students' behavior turns out to be all the more difficult because of the vast volume of information in educational databases. Currently in Lebanon, the absence of existing system to dissect and screen or the student performance and progress is not used. We analyze educational data to figure out the reasons behind students' behaviors and make a decision for solutions and treatment paths. A systematical system on anticipating student performance is proposed by utilizing mining data procedures to enhance students' accomplishments. We could actually improve students' achievement, success, and bring the benefits and impacts to students, educators and academic institutions. This paper centers on how the expectation calculation can be utilized to distinguish the most imperative qualities in students' information (behavior). We could really enhance students' accomplishment, achievement, and convey the benefits and effects to students, instructors and scholarly establishments.**

Keywords—Educational Data Mining; EDM; Student Behavior; Student Academic Performance.

I. INTRODUCTION

The appearance of data technology in different fields has lead the vast volumes of data storage in different types like records, documents, archives, pictures, sound, movies, scientific information and numerous new data types. The information gathered from various applications require legitimate strategy for extracting knowledge from large database for better uses.

There are expanding research interests in utilizing information mining in education [1]. This new rising field, called Educational Data Mining (EDM), worries with creating techniques that find learning from information originate from educational conditions [2].

EDM centers around investigating information created in an educational setup by the different intra-associated or divergent systems to create demonstrate for enhancing learning background and institutional viability as appeared in Fig.1. Table I condense the EDM users' at various dimensions. Data mining also called knowledge discovery in databases (KDD) is a famous field of life sciences and commerce [3].

TABLE I. EDM USERS AT VARIOUS LEVELS

	Level of Analysis	*User Benefitted*
Educational Data Mining	**Course Level:** Design and development of Intelligent Curriculum, Conceptual Enhancement and development, Predictive modeling of Success/Failure across the departments	Students, School
	Institutional: Academic Performance and overall Institution Growth and Reputation	Administrators, Funders
	Regional: Comparisons between various Systems	Administrators, Funders
	National and International	Educational Authorities, Governments

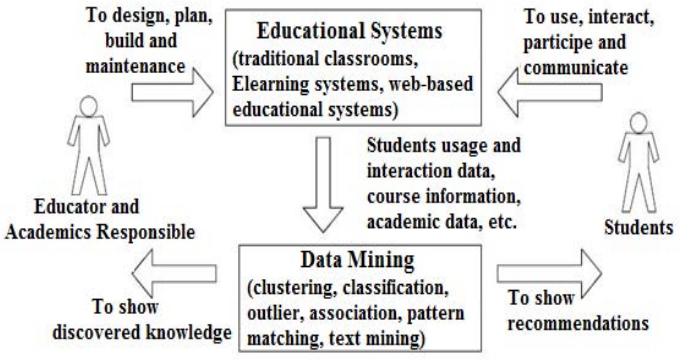

Fig. 1. Educational Data Mining (EDM) flowchart

978-1-5386-6880-1/18/$31.00 ©2018 IEEE

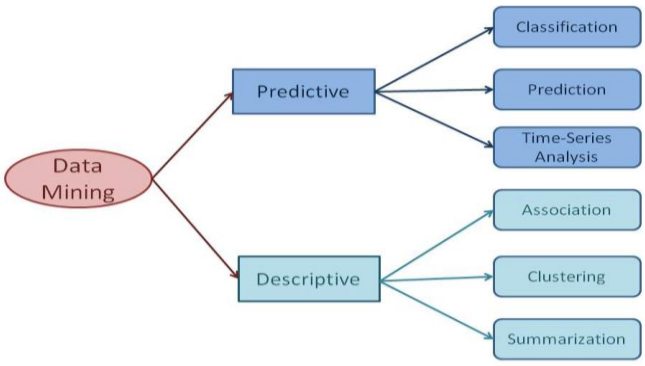

Fig. 2. Data Mining Techniques

Data Mining utilized in educational field to upgrade our comprehension of learning procedure to concentrate on distinguishing, removing and assessing factors identified with the learning procedure, execution, and students' behavior.

The fundamental elements of data mining are re applying different techniques and calculations so as to find and concentrate examples of put away information.

Pattern mining focuses on recognizing decides that depict explicit examples inside the information. Pattern mining comprises of utilizing/creating data mining algorithms to find fascinating, unforeseen and helpful patterns in databases. Pattern mining algorithms can be connected on different kinds of data such as streams, sequence databases, spatial data, transaction databases, strings, and diagrams.

Data mining and knowledge discovery applications have a rich concentration because of its centrality in basic leadership and it has turned into a basic part in different associations (Fig.2). Data mining techniques have been brought into new fields of Measurements, Pattern Redesign, Machine Learning, Databases, Artificial Intelligence and Computation capabilities etc.

Students performance is a basic part for better learning results. This is because one of the criteria for a high quality school is based on its brilliant record of scholarly accomplishments [4]. Currently, there are many techniques being proposed to evaluate students' performance. In order to improve students' achievement, success and behavior, we have to answer these questions.

1. How can we use these data to support the right behavior?
2. How can we correct and fix the indiscipline behavior without harming the student?
3. What are the optimum methods suitable for each student having a unique character?
4. What are the reasons behind this change and what are the right reactions or procedures that should be taken if indiscipline behavior occurred.

Using the real data (not fake; data obtained from Modern Academy for School Education MASE), efficient algorithms, we should be able to extract and analyze the data collected from teachers about their students to figure out:

1. Reasons behind students' behaviors.
2. To predict some specific behaviors and show hidden patterns explaining the reasons of students' actions and behaviors, which are the reflection of these reasons.
3. To draw a treatment path and solutions for students' behavior.

Subsequently, it would help the instructors and teachers in giving a powerful educating methodology. Teachers and instructors could likewise screen their understudies' accomplishments. Students could enhance their learning exercises, enabling the organization to enhance the systems performance [5, 6].

II. DATA MINING TECHNIQUES

There are several strategies and algorithms to process the data mining and to find undetectable patterns and connections to help in decision-making and later on discover solutions.

Some examples about these methods: Classification, Clustering, Association rule, Decision Tree, Genetic algorithm, Nearest Neighbor method, artificial intelligence and neural networks. The main goal of data mining is to extract knowledge, is also known to KDD (Knowledge Data Discovery). To understand the steps of extracting data and the concept of KDD see Fig.3.

A. Classification

It is a classic data mining method based on machine learning. Classification is utilized to characterize everything in a lot of data into one of predefined set of classes or gatherings [7].

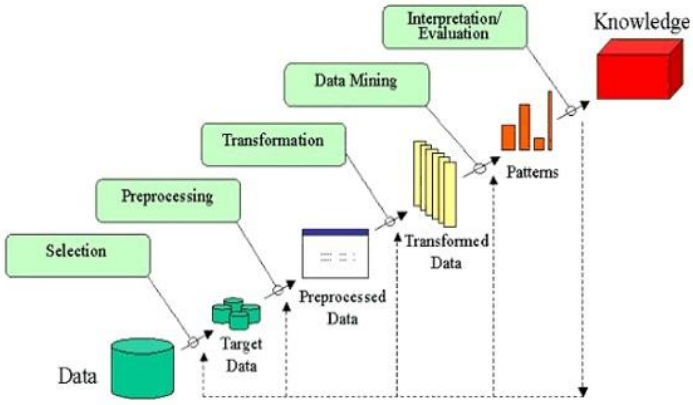

Fig. 3. The stages of extracting knowledge from data

We classified the educational data into two groups: behavioral and educational. The behavioral data contain the student's behavior at classroom, schoolyard, and outdoor activities. The other group is the educational behavior and student's performance in classroom. The two groups are affected by each other but at the same time, they are completely different.

B. Clustering

It is the way toward sorting out items into groups that members are comparable [8]. The cluster is therefore a collection of items that are " comparative" between them and are

"uncomparative" to the items belonging to different clusters or bunches (Fig.4).

It is obvious that we have two groups behavioral and educational. The behavioral group contains similar notes. The educational group is set to record the student's actions and performance inside the classroom.

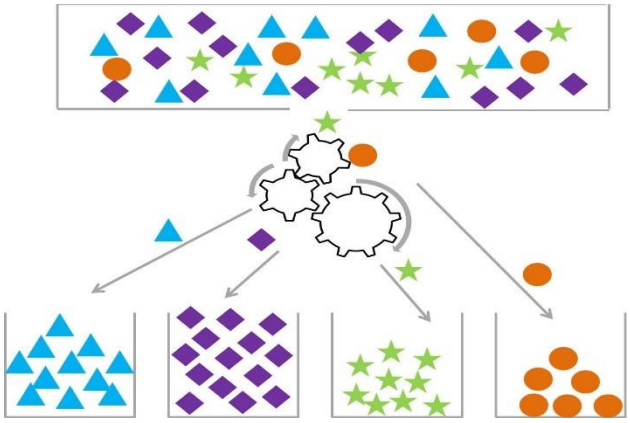

Fig. 4. Clustering Process Techniques

C. Association

It is one of the best-known data mining techniques. In association, a pattern is found dependent on a relationship of a specific thing on different things in a similar exchange [9].

We observed that there is a relationship between behavioral actions and educational performance, if improper behaviors increase then the students' academic performance decrease.

D. Decision tree

A structure incorporates a root node, branches, and leaf nodes. Each internal node indicates a test on an attribute, each branch signifies the result of a test, and each leaf node holds a class name. The highest node in the tree is the root node [10].

At specific level of processing after submitting the student to some kinds of tests, the system should figure out if the student has genetic problem that causes the physical disorder or it is just a reflection of improper environment. According to the result, the system should suggest that a student should contact a physician, consultant, psychologist or psychiatrist (Fig.5).

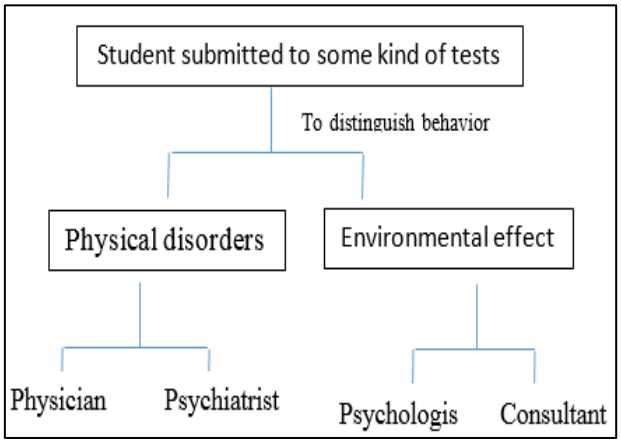

Fig. 5. Student behavior after using specific tests

III. Implementation and Result

A. Data preparations

To get better-input data for data mining techniques, preprocessing for the collected data was done. We gathered information from teachers that were complaining about some students' indiscipline or unacceptable behaviors [11, 12].

The data set used in this study was obtained from Modern Academy for School Education (MASE) School, Abbassieh (South Lebanon) on the sampling method of students' data of year 2017-2018. Initially size of the data is 100. In this step, data stored in various tables was participated in a solitary table after joining process mistakes were evacuated.

Over a period of nine months (one scholar year), we received a huge amount of information describing students' behavior. We have worked closely with teachers and students having different backgrounds, perspectives and problems. Data preparation:

- Collect a lot of information and more specific details about students' behaviors.
- Paraphrase the notes that we received from the teachers.
- Code the input-data to submit correctly in the system.
- Order from our teachers to use these unified coded phrases.

The fieldwork helped us to acquire considerable knowledge to cluster, classify and analyze data.

Figure 6 shows the behavior of a student over a period of seven months, it is a real example (specimen) of one student assigned by "X" as a nickname.

B. Data Selection and transformation

We analyzed the students' data and classified them into two groups (clustering). A few derived variables were selected form MASE database. While some variables of information were extracted from the database [13, 14, 15]. .

We distinguish from graph chart between behavioral and educational notes. Figure 7 shows the behavior of a student "X" over a period of seven months.

It shows two curves: one sums up the notes of unacceptable behavior and the other sums up the educational drawbacks. In further stages, using Decision Tree Techniques (DTT), the system figure out whether the student should be submitted to certain tests or not. This is done according to certain criteria already set, for example:

1. What is the number of notes considered acceptable compared to certain consensus average?
2. What are the types and levels of tests that should be performed?

After reaching diagnoses that depends on the results of the applied test, a decision is made to decide a treatment path.

The viability of the treatment path is monitored through a chart that tracks the student's progress, this chart is the

continuity of the one used in the first place (since it uses the same source of notes provided by teachers).

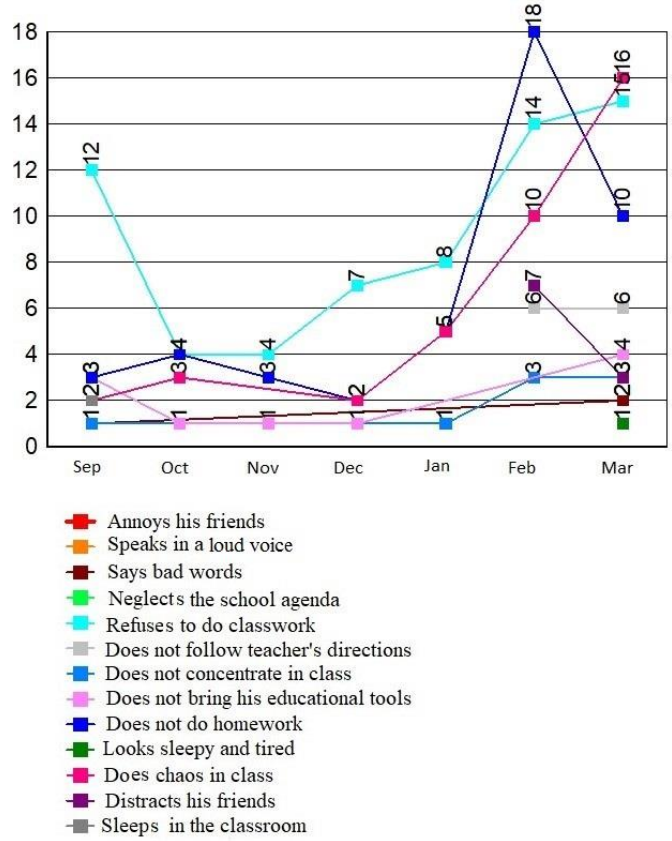

Fig. 6. Students' behavior over a period

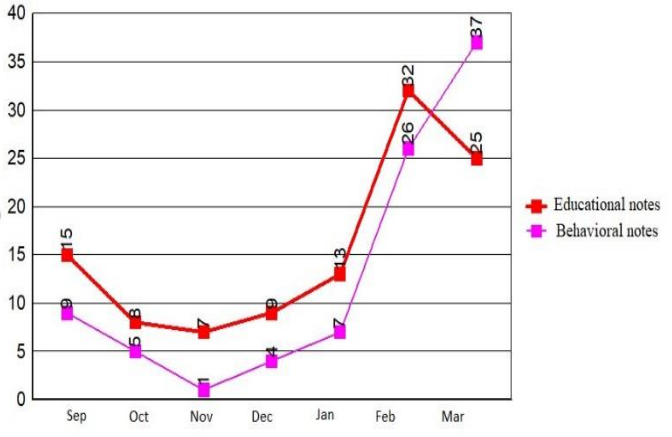

Fig. 7. Educational and behavioral data over a period

IV. Conclusion

The study was conducted on the student behavior related to academic performance by selecting a sample of 100 students (73 males, 27 females) from a group of academic levels (elementary, intermediate, high school) affiliated to Modern Academy for School Education (MASE).

We describes data mining software that allow the users to analyze data from different dimensions, categorize it and summarize the relationships (behavior vs. academic performance) which are identified during the mining process.

Our system (see Fig.6 and Fig.7) should be able to:

1. Analyze the result to conclude whether or not the student should submit to some tests.
2. Provide the tests that allow distinguishing between behaviors that result from physical disorders and behaviors that result from improper environment (society, home, school …).
3. Use the results of the tests to conclude solutions and possible treatment paths.
4. Adapt teaching strategies that attract student to special communication, directing student to a physician, consultant, psychologist, and psychiatrist.
5. Track the student's progress resulting from the treatment by using the same chart and the same source of notes (provided by the teachers).
6. Consider new kinds of behavioral problems and index them under the set classification rules.

In this article, numerous data mining tasks were utilized to make subjective prescient models, which were productively and viably ready to anticipate the students' execution and conduct from a gathered preparing database. First, a survey utilizes attributes from MASE student tables such as multiple personal, social, and academic data related to them. Second, the collected data explores and preprocesses to become suitable for predicting data mining tasks. Third, the implementation of data was presented on the dataset to produce classification models and testing them. Finally, interesting outcomes were drawn from the classification and comparison models [16, 17].

There is a real need to invest the huge amount of notes (students' data) coming from educators to figure out the real causes behind each behavior, and to set a plan that puts the student on the right treatment track.

It is very hard for a psychiatrist to elicit a treatment depending on several hours of close contact with a student in one place (clinic), but for a school, observing a students' behavior is much more realistic. It occurs over long periods and done by several educators and in different locations like classroom, schoolyard, school-theater, school-cafeteria, and outdoor school activities.

This research will help to the students and the educators to enhance the division of the student. This research will also work to distinguish those students, which required uncommon regard for lessen fall flat apportion and making proper move later on, additionally to recognize the issues in beginning times, draw treatment ways and track the suitability and viability of treatment.

References

[1] S. Slater, S. Joksimović, V. Kovanovic, R. S. Baker, and D. Gasevic, "Tools for educational data mining: A review", Journal of Educational and Behavioral Statistics, 2017, Vol. 42, No. 1, pp. 85–106..

[2] U. bin Mat, N. Buniyamin, P. M. Arsad, R. Kassim, An overview of using academic analytics to predict and improve students' achievement: A

proposed proactive intelligent intervention, in: Engineering Education (ICEED), 2013 IEEE 5th Conference on, IEEE, 2013, pp. 126–130.

[3] M. Hussain, M. Al-Mourad, S. Mathew, amd A.Hussein, "Mining Educational Data for Academic Accreditation: Aligning Assessment with Outcomes", Global Journal of Flexible Systems Management, March 2017, Volume 18, Issue 1, pp 51–60.

[4] R. Asif, A. Merceron, S. A. Ali, and N. G. Haider, "Analyzing undergraduate students' performance using educational data mining", Computers & Education, No. 113, pp.177-194, 2017 - Elsevier.

[5] M. Mayilvaganan, D. Kalpanadevi, Comparison of classification techniques for predicting the performance of students academic environment, in: Communication and Network Technologies (ICCNT), 2014 International Conference on, IEEE, 2014, pp. 113–118.

[6] K. F. Li, D. Rusk, F. Song, Predicting student academic performance, in: Complex, Intelligent, and Software Intensive Systems (CISIS), 2013 Seventh International Conference on, IEEE, 2013, pp. 27–33.

[7] K. Bunkar, U. K. Singh, B. Pandya, R. Bunkar, Data mining: Prediction for performance improvement of graduate students using classification, in: Wireless and Optical Communications Networks (WOCN), 2012 Ninth International Conference on, IEEE, 2012, pp. 1–5.

[8] A. Bogarın, C. Romero, R. Cerezo, M. Sanchez-Santillan, Clustering for improving educational process mining, in: Proceedings of the Fourth International Conference on Learning Analytics And Knowledge, ACM, 2014, pp. 11–15.

[9] D. M. D. Angeline, Association rule generation for student performance analysis using apriori algorithm, The SIJ Transactions on Computer Science Engineering & its Applications (CSEA), 2013, pp.12–16.

[10] C. F. Lin, Y. C. Yeh, Y. H. Hung, and R. I. Chang, "Data mining for providing a personalized learning path in creativity: An application of decision trees", Computers & Education, vol. 68, pp. 199-210, 10, 2013.

[11] S. Sembiring, M. Zarlis, D. Hartama, S. Ramliana, E. Wani, Prediction of student academic performance by an application of data mining techniques, in: International Conference on Management and Artificial Intelligence IPEDR, Vol. 6, 2011, pp. 110–114.

[12] G. Elakia, N. J. Aarthi, Application of data mining in educational database for predicting behavioural patterns of the students, Elakia et al,/(IJCSIT) International Journal of Computer Science and Information Technologies 5 (3) (2014) 4649–4652.

[13] S. Natek, M. Zwilling, "Student data mining solution–knowledge management system related to higher education institutions", Expert systems with applications, Vol. 41, No.14, 2014, pp. 6400–6407.

[14] S. Parack, Z. Zahid, F. Merchant, Application of data mining in educational databases for predicting academic trends and patterns, in: Technology Enhanced Education (ICTEE), 2012 IEEE International Conference on, IEEE, 2012, pp. 1–4.

[15] A. Dutt, M. A. Ismail, and T. Herawan, "A systematic review on educational data mining", DOI 10.1109/ACCESS.2017.2654247, IEEE Access.

[16] A. M. Shahiri, W. Husain, N. Abdul Rashid. "A Review on Predicting Student's Performance Using Data Mining Techniques", The Third Information Systems International Conference (ISICO2015), School of Computer Sciences Universiti Sains Malayisa 11800 USM, Penang - Malaysia, Procedia Computer Science, 2015

[17] A. Abu Saa, "Educational Data Mining & Students' Performance Prediction", International Journal of Advanced Computer Science and Applications (IJACSA), DOI: 10.14569/IJACSA.2016.070531, Vol. 7, No. 5, 2016.