Analysis of Students' Misconducts in Higher Education Institutions using Decision Tree and ANNs

Anas H. Blasi
Department of Computer Information Systems
Mutah University
Karak, Jordan
ablasi1@mutah.edu.jo

Mohammed A. Alsuwaiket
Department of Computer Science and Engineering Technology
Hafar Batin University
Hafar Batin, Saudi Arabia
malsuwaiket@uhb.edu.sa

Abstract—**A major problem that the Higher Education Institutions (HEIs) face is the misconduct of students' behavior. The objective of this study is to decrease these misconducts by identifying the factors which cause them on college campuses. CRISP-DM Methodology has been applied to manage the process of data mining and two data mining techniques: J48 Decision Tree (DT) and Artificial Neural Networks (ANNs) have been used to build classification models and to generate rules to classify and predict students' behavior and the location of misconduct in college campuses. They take into consideration seven factors: Student Major, Student Level, Gender, GPA Cumulative, Local Address, Ethnicity, and time of misconduct by month. Both techniques were evaluated and compared. The accuracy results were high for both classification models, whereas the J48 Decision Tree gave higher accuracy.**

Keywords-J48 decision tree; artificial neural networks; machine learning; student misconduct; student behavior

I. INTRODUCTION

Higher Education Institutions (HEIs) constitute an environment in which students, families, educators and community members have opportunities to learn, teach, and grow. Educators do their best to provide students with a stable and positive learning environment. Often, HEIs face problems which impede the educational process, problems related with students' prohibited behavior which results in dangerous misconducts in the short and the long term. In fact, the challenges associated with students with severe behavior problems are increasing [1]. In addition, these behaviors are reflecting on other students, either in or out of the class. Effective HEIs are safe institutions. Therefore, HEIs appreciate the need to prevent misconducts that threaten safety, and they need highly effective ways to respond effectively to such misconducts. To achieve HEIs' mission to educate students and maintain campus safety, they must be able to recognize and prevent these misconducts before they occur. Recently, many HEIs' departments and local school districts have taken steps to decrease these behaviors and support all the developed ways and enhance the educational process to keep the educational environment secure and safe for all students. The subject of student behavior misconducts is one of the most significant and major subjects in HEIs. However, many researches concentrate on student's behaviors and their inter-relations to decrease the

misconducts that occur in college campuses. Unfortunately, the literature is very limited in propositions of models to predict these behaviors to reduce the number of college student's misconducts. There many studies describing the danger of problematic behaviors and different solutions have been suggested to decrease the number of misconducts that affect negatively the environment of HEIs' and its safety, but very limited studies have used Data Mining (DM) methodologies and techniques in this area.

Applying DM in higher education is a recent research field which is also known as Educational Data Mining (EDM). EDM is concerned with developing different ways for exploring the unique types of data that come from the education environment. The process of applying DM to educational systems can be interpreted from different points of view [2]. From an educational viewpoint, it can be shown as an iterative cycle of hypothesis formation. In this process, the objective is not only turning data into useful information, but also to filter the mined knowledge for decision-making about how to enhance the environment of education to improve students' behavior and learning. From a DM viewpoint, it can be very similar to the general Knowledge Discovery and Data mining (KDD) process. There is increasing interest in using DM in higher education. Authors in [3, 4] focused on attendance rates and they found that a lower attendance rate negatively affects a student's GPA. The study in [5] focused especially on students' alcohol consumption. It was found that there is a negative affect increment on student GPAs and alcohol usage. Authors in [6] used the DM prediction technique to find the effective factors that determine a student's test grade, and then these factors were adjusted to improve the students' performance. Authors in [7-9] applied different DM techniques to improve the education process in the higher education system by improving different assessments methods. Authors in [10] used DM techniques to develop new curricula and help engineering students choose an appropriate major. Regarding alcohol issues, the research on the spread of alcohol-related problems at traditional four-year institutions is vast, even though little is known about alcohol consumption and related issues among college students.[11-13].

Data mining and criminology analysis can be divided into two main areas: misconduct control which tends to use

Corresponding author: Anas H. Blasi

www.etasr.com   *Blasi & Alsuwaiket: Analysis of Students' Misconducts in Higher Education Institutions using Decision …*

knowledge from the analyzed data to control and prevent the occurrence of misconduct, and criminal suppression which endeavors to catch a perpetrator in the act by using the person's history recorded in DM [14]. Studies of DM in terms of misconducts and intelligence analysis for national security are in their ongoing growth stage. The following techniques have been discussed in misconduct DM: entity extraction, clustering, and classification.

The current study focuses on nine problematic behaviors or behavior issues: alcohol use, drug issues, theft of property or services, endangering, threatening or causing physical harm, stalking or communicating in a manner likely to cause injury, distress, or emotional or physical discomfort and which serves no legitimate purpose, harassment, including sexual harassment, sexual assault, rape, hazing, unauthorized entry into or use of University premises, and violations of published University regulations or policies.

II. RESEARCH METHODOLGY

CRISP-DM (CRoss-Industry Standard Process for Data Mining) is a DM model that deals with DM from different perspectives. The purpose of this model is to increase knowledge of the data [15]. CRISP-DM life cycle consists of six stages, which are described below.

A. Business Understating Phase

The first phase determines the objectives and requirements of the study and the used techniques. Decision Tree (DT) and Artificial Neural Networks (ANNs) will be used in this study. Different tools (WEKA tool, Orange Canvas, Minitab, and MATLAB) will be used for the implementation.

B. Data Understanding Phase

The data of this study were collected from the office of student conduct at Binghamton University. The dataset consists of the data of 1374 students (1017 male, 357 female). The considered attributes of the data set are shown in Table I. All students in this sample have been charged with one of the referred attributes/behaviors, whereas the whole student population of the University is 13000 students.

C. Data Preparation Phase

The data preparation phase consists of five tasks: data selection, data cleaning, data construction, data integration, and data formatting [16]. Data were cleaned and prepared to be suitable for the modeling and implementation phase.

D. Modeling Phase

This section describes in detail the classification techniques which will be used in this study.

1) J48 Decision Tree

The DT technique is one of the most widely used methods in DM [17, 18]. DT is used to split data into a set of branches. The construction of the tree is conducted from top to bottom in a recursive divide-and-conquer manner. The main objective of DT learning is to select the attribute that is most useful for classifying samples [19]. A suitable measure of the significance of an attribute is the information gain that it measures. This measure is used to pick an attribute from the candidate

attributes at each step while building the DT [20]. DTs are used to represent decision making (visually and explicitly). The result of the classification can be an input for the decision making step and the tree can be a descriptive means for calculating conditional probabilities [1].

TABLE I. ATTRIBUTE DESCRIPTIONS

Attribute	Description	Possible values	Attribute type
Academic_Major	Student academic major	Art, Undeclared, Languages, Science, Management	Nominal
Std_Level	Student level	Freshman, Junior, Sophomore, Senior	Nominal
Student_Gender	Student gender	Male, Female	Nominal
GPA_cume	Student accumulative GPA	Between 2 and 4	Numeric
Local_city	Student's address	On-campus, Off-campus	Nominal
Misconduct_Location	The misconduct location	On-campus, Off-campus	Nominal
Misconduct_Date	Misconduct date	Between January and December	Date
Charge	Charge "Student Behavior"	Alcohol, Drug, Theft, Published University Policy, Endangering, Harassment, Disobedience, Unauthorized entry	Nominal
Ethnicity	Student's ethnicity	American, Black_non_Hispani, White_non_Hispanic, Hispanic, Asian, White-not specific	Nominal

J48 is a version of an earlier algorithm developed by J. Ross Quinlan. The very popular C4.5 offers powerful ways to express structures in data. C4.5 implementation in WEKA toolkit is known as the J48. J48 employs two pruning methods [21]:

- Sub-tree replacement: nodes in a DT are replaced with a leaf reducing the number of tests along a certain path.

- Sub-tree raising: a node moved upwards towards the root of the tree, replacing other nodes along the way. Sub-tree raising often has a negligible effect on decision tree models.

J48 DT technique will be applied to the data when the output is the "charge (behavior)", while the second application will take place when the output is the "misconduct location". Each time the data will be trained and validated. Figure 1 shows the results of the training set using J48 DT when the output is "Charge", which has an accuracy of is 97.8894% and a mean square error of 0.0688. Figure 2 shows the results of a 50 fold cross validation run. The correctly classified instances (accuracy) in this case are 98.1077% and the root mean squared error is 0.0697. If we compare the results of the 50 fold cross validation with the training set, we conclude that the accuracies of the two tests are close. The test was repeated many times in different folds and the results were close in every run.

```
=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances      1345           97.8894 %
Incorrectly Classified Instances      29            2.1106 %
Kappa statistic                    0.9725
Mean absolute error                0.0095
Root mean squared error            0.0688
Relative absolute error            4.9418 %
Root relative squared error       22.2392 %
Total Number of Instances          1374
```

Fig. 1.   Training test evaluation when the output is "Charge".

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      1348           98.1077 %
Incorrectly Classified Instances      26            1.8923 %
Kappa statistic                    0.9752
Mean absolute error                0.009
Root mean squared error            0.0697
Relative absolute error            4.6806 %
Root relative squared error       22.5389 %
Total Number of Instances          1374
```

Fig. 2.   Cross-validation test evaluation when the output is "Charge".

Table 2 demonstrates the generated rules of the training set when the output is "Charge".

TABLE II.   GENERATED RULES OF THE TRAINING SET WHEN THE OUTPUT IS "CHARGE"

No	Generated Rules
1	If gpa_cume <= 2.84 and gpa_cume <= 2.79: alcohol (498.0)
2	If gpa_cume <= 2.84 and gpa_cume > 2.79: disobedience (40.0/1.0)
3	If gpa_cume > 2.84 and gpa_cume <= 3.45 and gpa_cume <= 3.3: drugs (311.0/7.0)
4	If gpa_cume > 2.84 and gpa_cume <= 3.45 and gpa_cume > 3.3 and gpa_cume <= 3.42: endangering (79.0)
5	If gpa_cume > 2.84 and gpa_cume > 3.42: harassment (22.0/1.0)
6	If gpa_cume > 3.45and gpa_cume <= 3.6: publish_university_policy (118.0/19.0)
7	If gpa_cume > 3.45 and gpa_cume > 3.6 and gpa_cume <= 3.93: theft (266.0)
8	If gpa_cume > 3.45 and gpa_cume > 3.6 and gpa_cume > 3.93: unauthorized_entry (40.0/1.0)

The results of applying J48 decision tree on all the inputs individually with the output "Charge" show that the student GPA has the better rank, then misconduct date, student level, gender, ethnicity, academic major, and finally student local address. The input with higher accuracy can help to predict the charge better than the lower accuracy. Figure 3 shows the results of the training test using J48 DT when the output is "Misconduct Location". The correctly classified instances (accuracy) are 99.7817% and the root mean squared error is 0.0465. Figure 4 shows the results of a 50-fold cross validation run. The accuracy in this case is 99.7817% and the root mean squared error is 0.0467. If we compare the results of the 50-fold cross-validation and the training set, we can see that the accuracies of the two tests are close to each other. The test was repeated many times in different folds and the results were always very close.

Table III shows a sample of the generated rules (37 rules totally) of the training set when the output is "Misconduct location".

```
=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances      1371           99.7817 %
Incorrectly Classified Instances       3            0.2183 %
Kappa statistic                    0.9941
Mean absolute error                0.0043
Root mean squared error            0.0465
Relative absolute error            1.1662 %
Root relative squared error       10.8018 %
Total Number of Instances          1374
```

Fig. 3.   Training test evaluation when the output is "Misconduct Location".

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      1371           99.7817 %
Incorrectly Classified Instances       3            0.2183 %
Kappa statistic                    0.9941
Mean absolute error                0.0043
Root mean squared error            0.0467
Relative absolute error            1.1674 %
Root relative squared error       10.832  %
Total Number of Instances          1374
```

Fig. 4.   Cross-validation test evaluation when the output is "Misconduct Location".

TABLE III.   GENERATED RULES OF THE TRAINING SET WHEN THE OUTPUT IS "MISCONDUCT LOCATION"

No	Generated Rules
1	If local_city = on_campus and gender = male and misconduct_date = january: on_campus (15.0)
2	If local_city = on_campus and gender = male and misconduct_date = february: on_campus (32.0)
3	If local_city = on_campus and gender = male and misconduct_date = march: on_campus (209.0)
4	If local_city = on_campus and gender = male and misconduct_date = april and academic_major = art: on_campus (21.0)academic_major = management: off_campus (5.0)
5	If local_city = on_campus and gender = male and misconduct_date = april and academic_major = undeclared: on_campus (18.0)

2) Artificial Neural Network

An ANN will be used to classify inputs into a set of target categories. The Neural Network Pattern Recognition Tool by MATLAB will assist data selection, network creating and training, and performance evaluation using mean square error and confusion matrices. A two-layer feed-forward network, with a sigmoid hidden layer and output neurons can classify vectors arbitrarily well, given enough neurons in its hidden layer. The network will be trained with scaled conjugate back-propagation gradient (trainscg). The ANN classifier will be applied to the data twice, at first when the output is the "Charge (Behavior)" and the second time when the output is the "Misconduct Location". Each time the data will be trained and validated. The data in the ANN have been divided into three groups: the first group is the training set which is presented to the network during training, and the network is adjusted according to its error, the second group is the validation which is used to measure network generalization, and to halt training when the generalization stops improving, and the third is the testing which has no effect on training and provides an independent measure of network performance during and after training. In this study, the training sample consisted of 962 samples (70%), the validation sample of 206 samples (15%),

and the testing sample consisted of 206 (15%). The number of the hidden neurons is 50. Figure 5 shows the ANN diagram when the output (target) is the "Charge" (the number of charges is 8). The inputs are 7 (Student Major, Student level, Student Ethnicity, Student Local Address, Student's GPA, and Misconduct time).
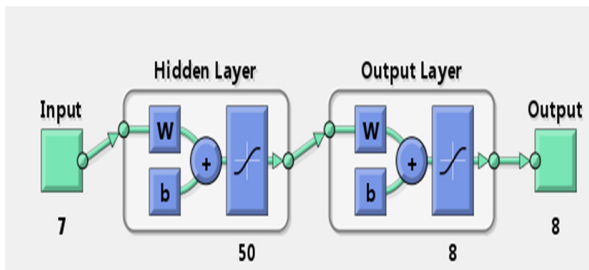

Fig. 5.   ANN diagram when the output (target) is "Charge".

Figure 6 shows the training, validation, and testing results. However, these results are represented by the MSE (Mean Squared Error), which is the average squared difference between outputs and targets. For MSE, lower values are better, since zero means no error) and %E (percent error: indicates the fraction of samples which are misclassified). A value of %E=0 means no misclassifications whereas 100 indicates maximum misclassifications.

Fig. 6.   Training, validation, and testing results when the target is "Charge".

Figure 7 shows the neural network diagram when the output (target) is the "Misconduct Location. The inputs are the same as before. Figure 8 shows the results of the training, validation, and testing. These results are again represented by MSE.
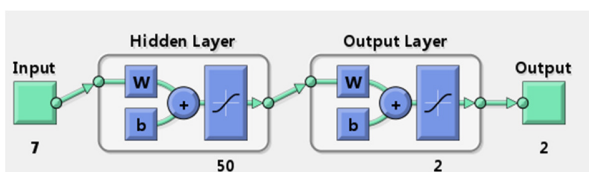

Fig. 7.   ANN diagram when the output (Target) is "Misconduct Location".

Fig. 8.   Training, validation, and testing results when the target is "Misconduct Location".

III.   RESULTS AND DISCUSSION

From the summarized results in Table IV, it can be concluded that the DT can be considered a better classifier than the ANN in terms of accuracy and RMSE when the target is "Charge". However, when the target is "Misconduct Location", both classifiers have the same Accuracy and little difference in RMSE.

TABLE IV.   DM RESULTS

DM Technique	Target			
	"Charge"		"Misconduct Location"	
DT - J48	Accuracy	RMSE	Accuracy	RMSE
	98.1	0.0697	99.8	0.0467
ANN	Accuracy	RMSE	Accuracy	RMSE
	91	0.115	99.8	0.014

According to the results of this study, the predictor that resulted from the J48 decision tree, can be used to predict and thus decrease the students' behavior misconducts by considering the generated rules shown in Tables II and III. Also, and by knowing the factors that affect students' behavior that cause misconducts, we can decrease the future number of misconducts in college campuses. For instance, the misconduct time is highly correlated (high accuracy) with both outputs (Charge, Misconduct Location). This factor can be used to predict the charge and the misconduct location, so from this point of view, some plans can be suggested in accordance to the predicted charge and location in order to decrease the students' misconducts in college campuses.

IV.   CONCLUSION AND FUTURE WORK

Two different DM techniques, ANNs and J48 DT have been applied in data referring to the misconducts of students in HEIs. CRISP DM methodology has been applied in this study to mine student's data and identify the factors which cause misconducts in college campus. The phases of CRISP DM methodology have been used to manage all the processes in this study. The implementation part of the study has been presented and discussed. The WEKA tool and MATLAB have been used to build the classification models and the results were compared. As a result, the J48 DT can be considered a more accurate classifier and predictor. Regarding future work, the use of different methods such as fuzzy logic [22], optimization algorithms [23], and other DM techniques such as k-NN and decision tree [24] will be considered.

REFERENCES

[1]   V. Berikov and A. Litvinenko, "Methods for statistical data analysis with decision trees," *Novosibirsk, Sobolev Institute of Mathematics*, 2003.

[2]   C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to 2005," *Expert Systems with Applications*, vol. 33, no. 1, pp. 135–146, Jul. 2007, https://doi.org/10.1016/j.eswa.2006.04.005.

[3]   G. Durden and L. Ellis, "The Effects of Attendance on Student Learning in Principles of Economics," *American Economic Review*, vol. 85, no. 2, pp. 343–46, 1995.

[4]   D. Romer, "Do Students Go to Class? Should They?," *Journal of Economic Perspectives*, vol. 7, no. 3, pp. 167–174, Sep. 1993, https://doi.org/10.1257/jep.7.3.167.

[5]   A. M. Wolaver, "Effects Of Heavy Drinking In College On Study Effort, Grade Point Average, And Major Choice," *Contemporary Economic Policy*, vol. 20, no. 4, pp. 415–428, 2002.

[6]   M. Didonet Del Fabro and P. Valduriez, "Towards the efficient development of model transformations using model weaving and matching transformations," *Software & Systems Modeling*, vol. 8, no. 3, pp. 305–324, Jul. 2009, https://doi.org/10.1007/s10270-008-0094-z.

[7]   M. Alsuwaiket, A. H. Blasi, and K. Altarawneh, "Refining Student Marks based on Enrolled Modules' Assessment Methods using Data Mining Techniques," *Engineering, Technology & Applied Science Research*, vol. 10, no. 1, pp. 5205–5210, Feb. 2020, https://doi.org/10.48084/etasr.3284.

[8]   M. Alsuwaiket, A. H. Blasi, and R. A. Al-Msie'deen, "Formulating Module Assessment for Improved Academic Performance Predictability in Higher Education," *Engineering, Technology & Applied Science Research*, vol. 9, no. 3, pp. 4287–4291, Jun. 2019, https://doi.org/10.48084/etasr.2794.

[9]   N. Delavari, M. R. Beikzadeh, and S. Phon-Amnuaisuk, "Application of enhanced analysis model for data mining processes in higher educational system," in *2005 6th International Conference on Information Technology Based Higher Education and Training*, Jul. 2005, pp. F4B/1-F4B/6, https://doi.org/10.1109/ITHET.2005.1560303.

[10] K. Prasada Rao, M. V. P. Chandra Sekhara, and B. Ramesh, "Predicting Learning Behavior of Students using Classification Techniques," *International Journal of Computer Applications*, vol. 139, no. 7, pp. 15–19, Apr. 2016, doi: 10.5120/ijca2016909188.

[11] K. M. Coll, "An Assessment of Drinking Patterns and Drinking Problems among Community College Students: Implications for Programming," *Journal of College Student Development*, vol. 40, no. 1, pp. 98–100, 1999.

[12] C. A. Presley, P. W. Meilman, J. R. Cashin, and J. S. Leichliter, "Alcohol and Drugs on American College Campuses: Issues of Violence and Harassment," NCJRS, 187865, 1997.

[13] F. D. Sheffield, J. Darkes, F. K. Del Boca, and M. S. Goldman, "Binge drinking and alcohol-related problems among community college students: implications for prevention policy," *Journal of American college health: J of ACH*, vol. 54, no. 3, pp. 137–141, Dec. 2005, https://doi.org/10.3200/JACH.54.3.137-142.

[14] A. Malathi and S. S. Baboo, "An Enhanced Algorithm to Predict a Future Crime using Data Mining," *International Journal of Computer Applications*, vol. 21, no. 1, pp. 1–6, May 2011, https://doi.org/10.5120/2478-3335.

[15] P. Chapman *et al.*, *CRISP-DM 1.0: Step-by-step data mining guide*. SPSS, 2000.

[16] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Haryana, India; Burlington, MA, USA: Morgan Kaufmann, 2011.

[17] M. Kantardzic, *Data Mining: Concepts, Models, Methods, and Algorithms*, 2nd ed. Hoboken, NJ, USA: Wiley-IEEE Press, 2011.

[18] T. Mitchell, "Decision Tree Learning," in *Machine Learning*, The McGraw-Hill Companies, Inc., 1997, pp. 52–78.

[19] A. El-Halees, "Mining Students Data to Analyze Learning Behavior: A Case Study," in *The 2008 international Arab Conference of Information Technology (ACIT2008)*, Sfax, Tunisia, Dec. 2008.

[20] C. H. Yu, S. A. DiGangi, A. Jannasch-Pennell, W. Lo, and C. Kaprolet, "A Data-Mining Approach to Differentiate Predictors of Retention," presented at the EDUCAUSE Southwest Conference, Austin, TX, USA, Feb. 2007.

[21] I. Witten, E. Frank, M. Hall, and C. Pal., *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed. Morgan Kaufmann, 2016.

[22] A. Blasi, "Scheduling food industry system using fuzzy logic," *Journal of Theoretical and Applied Information Technology*, vol. 96, no. 19, pp. 6463–6473, Oct. 2018.

[23] A. Blasi, "Performance increment of high school students using ANN model and SA algorithm," *Journal of Theoretical and Applied Information Technology*, vol. 95, no. 11, pp. 2417–2425, Jun. 2020.

[24] M. A. A. Lababede, A. H. Blasi, and M. A. Alsuwaiket, "Mosques Smart Domes System using Machine Learning Algorithms," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 11, no. 3, 40/30 2020, https://doi.org/10.14569/IJACSA.2020.0110347.