

Predicting Student Performance in Higher Education Institutions Using Decision Tree Analysis

Alaa Khalaf Hamoud, Ali Salah Hashim, Wid Aqeel Awadh *

University of Basra, Basra (Iraq)

Received 3 November 2017 | Accepted 22 January 2018 | Published 9 February 2018



ABSTRACT

The overall success of educational institutions can be measured by the success of its students. Providing factors that increase success rate and reduce the failure of students is profoundly helpful to educational organizations. Data mining is the best solution to finding hidden patterns and giving suggestions that enhance the performance of students. This paper presents a model based on decision tree algorithms and suggests the best algorithm based on performance. Three built classifiers (J48, Random Tree and REPTree) were used in this model with the questionnaires filled in by students. The survey consists of 60 questions that cover the fields, such as health, social activity, relationships, and academic performance, most related to and affect the performance of students. A total of 161 questionnaires were collected. The Weka 3.8 tool was used to construct this model. Finally, the J48 algorithm was considered as the best algorithm based on its performance compared with the Random Tree and RepTree algorithms.

KEYWORDS

Prediction, Students' Success, Decision Tree, J48, Random Tree, REPTree, Weka.

DOI: 10.9781/ijimai.2018.02.004

I. INTRODUCTION

RECENTLY, the increasing volume of data and using them to increase students' performance is one of the major challenges of higher education institutions (HEIs). Higher education institutions are generally interested in the success of students during their study [1, 2]. Higher education institutions, in a lifetime of teaching, have a large data set of student information stored in their databases. However, storage is not a problem. Handling data, extracting relevant patterns, and discovering knowledge stored in the massive database are tremendously difficult. Accordingly, data mining can be considered as a very promising tool to attain these objectives [3].

Data mining is used to detect and extract relevant as well as worthy information from a very huge volume of data. Currently, this process has acquired a great deal of attention as well as concerns from the information industry and society. This technique is also receiving significant attention in data analysis as well as it been recognized as a newly emerging tool for analysis [4]. Predicting student performance is a very beneficial in providing students as well as lecturers with the necessary assistance in the learning process. Predicting the possible outcomes of the learning process on the basis of the results of prediction can help an organizations change the outcome of new students and performance by adjusting the factors that contributed to the past performance [5].

Predicting of student performance also aids educational planners and decision makers and administrators to adequately plan for changes

in student population in any direction (i.e., find factors which increase performance or decrease it). However, coming up with a manual set of rules required to predict student performance is commonly difficult. For these reasons, there is an extreme need to achieve these goals [6]. Particularly, data mining tools became very popular among researchers and users because of their ease of use and availability such as the following tools (most of them are available online) Microsoft Excel, SPSS, Weka, Protege (a knowledge acquisition system) and Rapid Miner. A number of these tools (e.g., MS Excel-based) are freely available to HEI professors with which they can benefit with their existing knowledge of the Excel application. The accessibility, availability, ease of use and understandability are the most reasons for the inclusion these tools in the research. Weka is chosen as a data mining analysis tool for supporting conclusion because of its highly readable and understandable results [7].

We aim to analyze collective student information through a questionnaire (based on LimeSurvey and Google Form) as well as classify the collected data to predict and categorize student performance. We also seek to elucidate the different factors that affect student performance (success and failure rates) in relation to other variables in the data set of students by applying decision tree algorithms. The work seeking to find the possibility of depending decision tree algorithms' results in support academic decisions related to improve academic performance and give a proper road map for students and lecturers. Therefore, this paper uses multiple classification and prediction methods to confirm and verify the results with multiple Weka classifiers. We select the best result in terms of accuracy and precision based on performance results. This newly discovered knowledge can help students as well as instructors in carrying out enhanced educational quality. Identifying possible underperformers at the beginning of the semester/year and increasing the attention allotted for them will aid their educational process as well as improve students' grades. In fact, underperformers, as well as good performers, can benefit from this study by exerting extra effort

* Corresponding author.

E-mail addresses: alaa7alaf@gmail.com (Alaa Khalaf Hamoud), alishashim2009@gmail.com (Ali Salah Hashim), umzainali@gmail.com (Wid Aqeel Awadh).

in conducting quality projects and research through obtaining help and attention from their instructors.

The paper organized as follow: in section 2, the recent related works to academic performance are presented; section 3 explains some proposed decision tree algorithms that used in the model (J48, REPTree and Random Tree). Section 4 lists and explains the proposed model while the final section explains the conclusion points that concluded from the model.

II. RELATED WORK

Natek Srečko, and Moti Zwilling [8] focused on data mining for small data sets of student and answered the research questions by comparing two data mining tools. The best model was chosen based on results evaluation. They compared their prediction of student performance for the academic year 2012/2013 using the attribute: “Final grade”, with the actual Final grade of the same students. After that, the authors chose the best parameters for filling the data set from sample data mining techniques. The conclusions of the work were very promising and encouraging for HEIs in terms of integrating data mining tools as very important part of their knowledge management systems in higher education.

Alaa Khalaf Hamoud [9] constructed a model on the basis of an experimental data set on Portuguese students from two courses (Mathematics; 395 instances) and Portuguese (Portuguese language course; 659 instances). Paulo Cortez and Alice Silva, from the University of Minho, Portugal, collected and analyzed the data. Three decision tree algorithms (J48, RepTree, and Hoeffding Tree (VFDT)) were applied to this work. The results confirmed that the J48 algorithm was most suitable for classifying and predicting the willingness of students to complete higher education and success in their courses.

Mishra Tripti, Dharminder Kumar, and Sangeeta Gupta [10] used different classification algorithms to construct a prediction model of performance on the basis of the academic and social integration as well as various emotional students’ skills which other studies, thus far, have not considered. The J48 (implementation of C4.5) and Random Tree algorithms have been applied to students’ records of colleges affiliated with Guru Gobind Singh Indraprastha University to predict third semester performance. Clearly, Random Tree was more accurate compared with J48 in performance prediction.

Muluken Alemu Yehuala [11] built models and tested them using a sample data set of 11,873 regular undergraduate students. WEKA application software was used for the building model and analyzing it. The results offered supportive and constructive recommendations to academic decision makers in universities to enhance supporting and making decisions. Furthermore, the results will aid in rebuilding and modification process of curriculum structure to improve student academic performance. Based on previous research results and findings, students are able to decide about their preferred field of study before they register and enroll. These findings verify that the Ethiopian Higher Education Entrance Certificate Examination (EHEECE) results, number of courses per semester, gender and number of students in the class, as well as the majority of study are the considerable factors that affect student performance. Accordingly, based on the results, the level of student success can be controlled. Therefore, preventing educational institutions from incurring serious financial strains is possible.

III. DECISION TREE

A decision tree is one of data mining classification technique which used to build a top down tree like model based on the attributes of a given data set. Moreover, a decision tree is a predictive modeling

technique that used to predicate, classify, or categorize given data objects on the basis of a previously generated model using a training data set with the same features (attributes) [12].

The structure of the generated tree includes a root node as well as internal and leaf (terminal) nodes. The first or root node is the first top node which has no incoming nodes and one or more outgoing edges. An internal or middle node has one incoming edge and one or more outgoing edges, where each internal node denotes a test on an attribute and each edge represents a result of the test. Finally, the leaf node represents the final suggested (predicted attribute (label) of a data object [13].

The decision tree classification technique is performed in two stages [14]: tree building and pruning. Tree building stage follows top to down manner. During this stage, the tree is recursively partitioned until the data items belong to the same class label. This stage is very tedious and consumes a lot of computation processes since the training data set is repeatedly reprocessed. Tree pruning stage is done in bottom up manner. This practice is performed to improve the prediction and classification accuracy results of the algorithm by minimizing over fitting (noise or considerable detail in the training data set). Over fitting may result misclassification error in the decision tree algorithm. Decision tree offers many advantages to data mining, some of which are the following:

- Decision tree can be clearly understood by the analyst and any end user.
- Decision tree can handle different kinds of input data, namely, nominal, numeric, and textual.
- Decision tree can process erroneous data sets or missing or uncompleted values.
- Decision tree has a high level of performance with a minimal amount of effort and time.
- Decision tree can working in data mining applications over a variety of platforms [15].

In this study, three decision tree algorithms were used on collective student data, namely, J48, Random Tree, and RepTree.

A. J48

J48 is used for both of classification and prediction operations. For classification, J48 was chosen (on the basis of the C4.5 algorithm from machine learning) given that this algorithm is one of the most used in Weka tools that offers stability among the precision, speed, and interpretability of results. In addition this algorithm classifies data in the form of a decision tree with which we can easily identify weak students. Classification learning as a part of Educational Data Mining (EDM) was also implemented to predict the performance of students [16].

B. REPTree

The reduced-error pruning (REP) tree as a decision tree-learning algorithm can be considered a fast classifier based on the principles of computing information gain with entropy and minimizing the error arising from variance [17]. REPTree generates many of trees and applies regression tree logic in changing the iterations. Subsequently, the algorithm selects the best one from all spawned trees. Based on variance and result information, the algorithm builds a regression decision tree. Further, this algorithm prunes the trees based on using back-fitting method and reduced-error pruning. As in C4.5, this algorithm can also works with missing or uncompleting values by splitting the identical instances into pieces [18].

C. Random Tree

The random tree algorithm selects a test on the basis of a specific number of random features at each node without pruning. Commonly,

Random trees refer to those randomly built and have nothing to do with machine learning. The merit of building a random tree is the efficiency of training and minimal memory requirements. To create a random decision tree, Random Tree algorithm uses only one pass over the data [17, 19].

IV. MODEL

Fig. 1 shows the construction process of the model, which passes through four steps. First, building a questionnaire is considered a part of the data preprocessing step. This step includes preparing the data set for evaluation, cleaning data, converting data range, and creating a derived column (Failed) based on another column (Number of Failed Courses). The derived column (Failed) is created on the basis of a simple equation: If (Number of Failed Courses > 0), then Failed = 'F'. Else Failed = 'P'; F = abbreviation of Failed, P = abbreviation of Passed. The column Failed is considered the goal class of the model.

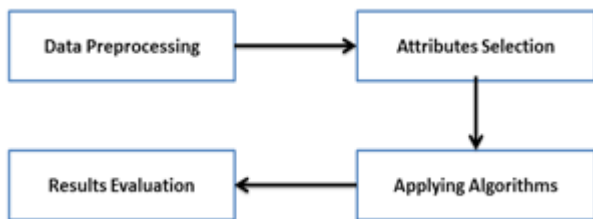


Fig. 1. Model Construction Diagram.

A. Data Preprocessing

1) Data Collection

The questionnaires that were built on Google Forms and an open source application (LimeSurvey) were used to conduct a survey on students from the College of Computer Science and Information Technology (CSIT), University of Basrah. The first questionnaire (built Lime Survey) was used to locally collect data from CSIT, whereas Google Forms was used to collect data over the Internet. A total of 161 questionnaires were completed after combining the CSV files from Google Forms and the LimeSurvey questionnaire. The research sample (161 answers) represents an acceptable sample CSIT population with a 10% margin of error for the result [20].

Table I shows the description of all answers to the questionnaire. Question possible value was shortened and converted from a nominal to a numeric type for ease of use and understanding. Response values of questions (Q18-Q60) are of the form {1; 2; 3; 4; 5} where 1; 2; 3; 4; 5 represents the answers “Strongly Disagree”, “Disagree”, “Neutral”, “Agree”, “Strongly Agree” respectively.

The initial step in data preprocessing is preparing data by removing rows with empty values and converting data for evaluation as well as processing. A total of 8 rows with empty values are under one or more columns. After removing these rows, we obtain a total 151 answers. Subsequently, row values are converted for data processing in the Weka 3.8 tool with its built-in classifiers.

TABLE I. QUESTIONNAIRE DESCRIPTION

Class	Variable	Description	Possible Value
Demographic Data	Q1	Department	IS, CS
	Q2	Age	18,19,20, >20
	Q3	Stage,	1,2,3,4
	Q4	Gender	Female Male
	Q5	Address	In Basra Out of Basra
Social Information	Q6	You're Status?	Married Single
	Q7	Are you working?	YES, NO
	Q8	Are you live with your parents?	YES,NO
	Q9	Are you parent a live?	YES, NO
	Q10	Are your father working?	YES, NO
	Q11	Are your mother working?	YES,NO
Academic Information	Q12	No. of Fail Courses	0,1,2,>2
	Q13	No. of Absence Days	0,1-5, 5-10, >10
	Q14	No. of Credits	<12 , 13-17, >17
	Q15	GPA	<60,61-70,71-80, > 80
	Q16	No. of Complete Credits ,	<36 , 36-71, 72-107, >107
	Q17	No. of Years of Study	1,2,3,> 3
Study Skill	Q18	I can pick out the important point in the material that I read easily.	1,2,3,4,5
	Q19	I write notes and later I use them for preparing for tests.	1,2,3,4,5
	Q20	I schedule my time carefully for study.	1,2,3,4,5
	Q21	I am cool, very calm, and collected during the exam process.	1,2,3,4,5
Motivation	Q22	Getting a low test score does not make me feel that I am a failure.	1,2,3,4,5
	Q23	I made my own decision to go to get a college education	
	Q24	I study even when less important things distract me.	1,2,3,4,5
	Q25	I am excited about the courses I take.	1,2,3,4,5
	Q26	I have a very clear idea of the benefits which I expect to get from my education.	1,2,3,4,5
Personal Relationship	Q27	I develop and maintain supportive relationships with others.	1,2,3,4,5
	Q28	I control my upsets or anger without blaming others.	1,2,3,4,5
	Q29	I have ability for making friendships and create worthy relationships in any new place.	1,2,3,4,5
	Q30	I am open with people whom I don't especially like in order to learn from them.	1,2,3,4,5

Health	Q31	I have enough energy for studying and still fully enjoying other areas of my life.	1,2,3,4,5
	Q32	I regularly exercise.	1,2,3,4,5
	Q33	My emotional health supports my ability to learn.	1,2,3,4,5
	Q34	I control the alcohol and drugs I put into my body.	1,2,3,4,5
Time Management	Q35	I can use a computer effectively to enhance my success in school.	1,2,3,4,5
	Q36	I plan for each single day of the week.	1,2,3,4,5
	Q37	I plan my day by deciding what is important first.	1,2,3,4,5
	Q38	I don't study hard on the night before exam because of good planning.	1,2,3,4,5
	Q39	I plan times regularly for doing fun things.	1,2,3,4,5
Money Management	Q40	I plan for spending or saving my money, and I control of my personal finances.	1,2,3,4,5
	Q41	I have a clear view of the financial resources available to me to support my education.	1,2,3,4,5
	Q42	I can make money.	1,2,3,4,5
	Q43	My education supports my long range financial goals.	1,2,3,4,5
	Q44	I have an emergency savings plan which I follow as needed.	1,2,3,4,5
Personal Purpose	Q45	I see education as something I will be do through all of my life.	1,2,3,4,5
	Q46	I know that I am responsible for my own education.	1,2,3,4,5
	Q47	I responsible of the quality of my life.	1,2,3,4,5
	Q48	Ready to face challenges even where I'm not sure how to meet them.	1,2,3,4,5
	Q49	I have a clear idea about the best career plans for me	1,2,3,4,5
Career Planning	Q50	I know a lot about myself in relation to what kind of profession would be best for me.	1,2,3,4,5
	Q51	I have worked in the profession areas that interest me, or at least seen for myself what they require.	1,2,3,4,5
	Q52	I know the potential qualities that my employers will expect of me.	1,2,3,4,5
	Q53	I know education choices and schedules that I must follow in order to reach my career goal.	1,2,3,4,5
Resource Needs	Q54	I have enough money to complete my education and still meet all of my other financial commitments.	1,2,3,4,5
	Q55	My relationships with friends, family, and others support my educational goals.	1,2,3,4,5
	Q56	The available resources to me are enough to meet my needs.	1,2,3,4,5
Self_Esteem	Q57	I do not hesitate to ask for help when needed from family members, friends, or other appropriate individuals.	1,2,3,4,5
	Q58	I actively try to improve myself for my own benefit and for others as well.	1,2,3,4,5
	Q59	I am able to manage time and other life demands effectively.	1,2,3,4,5
	Q60	I am sure that I have the ability and skills to do college-level work successfully.	1,2,3,4,5

2) Reliability

Reliability is a feature of data set used for characterizing the overall uniformity of a measure. A measure is said to have a high level of reliability if it gives similar results under consistent conditions. For

example, the measurements of people's height and weight are generally highly reliable [21]. In statistics, the coefficient alpha method is the most frequently used for the measuring of internal uniformity that is used as an indicator of reliability for the dependent variable of the study. Based on [22], it can be said that the value 0.7 indicates satisfying internal consistency in reliability. Table II shows that the coefficient alpha is 0.85 for the scaled variables that contain 60 items and 161 respondents.

TABLE II. QUESTIONNAIRE RELIABILITY

Cronbach's alpha	No.of items	No. of respondents	% of respondent
0.85	60	161	100%

B. Attribute Selection

Finding the most correlated attributes (questions) to the final class (Failed) and how much they affect the final class is important. Significantly, this stage shows the average correlation of the attributes to the final class. In turn, this average will help us find the questions with low correlation and remove them to improve the accuracy of the results. Questions with high correlation can be considered as recommendation points for students and academic staff.

In this step, the filter CorrelationAttributeEval is used to evaluate the correlation between the class and other attributes. This step is crucial because we want to find the most closely related attributes that affect the class and ignore the less-related attributes from the model. The attribute evaluator algorithm (CorrelationAttributeEval) evaluates the worth of an attribute by measuring the correlation (Pearson's) between the attribute and the class.

TABLE III. CORRELATION AVERAGE OF QUESTIONS

Seq.	Question	Average	Seq.	Question	Average
1	Q14	0.436	31	Q52	0.101
2	Q25	0.312	32	Q49	0.09
3	Q60	0.234	33	Q19	0.094
4	Q15	0.224	34	Q11	0.086
5	Q46	0.227	35	Q26	0.088
6	Q17	0.225	36	Q42	0.089
7	Q51	0.21	37	Q48	0.077
8	Q47	0.209	38	Q24	0.074
9	Q10	0.174	39	Q20	0.074
10	Q23	0.168	40	Q41	0.065
11	Q27	0.168	41	Q22	0.068
12	Q45	0.173	42	Q55	0.064
13	Q29	0.162	43	Q36	0.065
14	Q57	0.158	44	Q4	0.062
15	Q13	0.159	45	Q5	0.061
16	Q18	0.152	46	Q31	0.054
17	Q35	0.147	47	Q6	0.052
18	Q40	0.145	48	Q3	0.042
19	Q58	0.142	49	Q56	0.04
20	Q43	0.134	50	Q44	0.039
21	Q16	0.135	51	Q32	0.039
22	Q38	0.131	52	Q1	0.043
23	Q8	0.133	53	Q2	0.028
24	Q34	0.131	54	Q54	0.025
25	Q37	0.123	55	Q12	0.023
26	Q59	0.12	56	Q39	0.026
27	Q53	0.113	57	Q33	0.023
28	Q9	0.116	58	Q28	0.023
29	Q30	0.116	59	Q7	0.022
30	Q50	0.106	60	Q21	0.018

Table III shows the average correlation between questions/attributes and the final class with the mode of evaluation (10-fold cross validation) to ensure a high level of accuracy. The questions were arranged in an ascending order on the basis of the average rate of correlation to the final class. Questions with high average rates are most correlated to the final class. Later, the last twenty questions will be removed to increase the accuracy of the result.

C. Applying Algorithms

The Weka tool provides built-in algorithms that help in the application of different classifiers and obtain results in an easy and flexible process. Three algorithms will be used in this stage (J48, Random Tree, and RepTree) before (less correlated questions) and after the removal of attributes. The attribute removing process is profoundly helpful in discovering the effect of these attributes on performance and how they can increase or decrease accuracy.

Table IV shows details on performance before the removal of attributes from the data and further shows the three algorithms (J48, Random Tree and RepTree) according to their attributes, namely, True Positive (TP), False Positive (FP) rates, Precision, and Recall.

TABLE IV. ALGORITHM PERFORMANCE BEFORE THE REMOVAL OF ATTRIBUTES

Classifier		TP Rate	FP Rate	Precision	Recall
Without Attribute Filter	J48	0.529	0.485	0.539	0.529
	Random Tree	0.608	0.442	0.601	0.608
	Rep Tree	0.621	0.448	0.609	0.621
With Attribute Filter	J48	0.516	0.499	0.526	0.516
	RandomTree	0.641	0.373	0.646	0.641
	Rep Tree	0.634	0.440	0.623	0.634

The first row reveals the performance of the algorithms without an attribute filter (CorrelationAttributeEval). The second presents the performance after applying the attribute filter combined with the decision tree algorithm. A meta classifier (Attribute Selected Classifier) allows us to combine the classifier algorithm with the attribute evaluator and the search method to get highly accurate results.

The effectiveness of the attribute filter on the algorithms (Random Tree and RepTree) is evident. TP Rate, Precision, and Recall in both algorithms are increased, while FP Rate decreased after applying the attribute filter with algorithms (Random Tree and RepTree) while it decreased with the J48 algorithm.

TABLE V. PERFORMANCE DETAILS AFTER THE REMOVAL OF ATTRIBUTES

Classifier		TP Rate	FP Rate	Precision	Recall
Without Attribute Filter	J48	0.634	0.409	0.629	0.634
	Random Tree	0.595	0.430	0.597	0.595
	Rep Tree	0.634	0.440	0.623	0.634
With Attribute Filter	J48	0.621	0.423	0.616	0.621
	Random Tree	0.614	0.412	0.615	0.614
	Rep Tree	0.601	0.488	0.583	0.601

D. Result Evaluation

Result evaluation is the final stage in the model construction process. Based on Tables IV and V, the accuracy of the J48 classifier after the removal of the less correlated attributes is apparently higher compared with that of the RepTree and Random tree classifiers. The TP rate attribute takes a value of 0.634, which is the highest value for this attribute, whereas Precision (refers to a positive predictive value) also gets the highest value at 0.629. Recall (refers to a TP rate) with a value of 0.634 and an FP rate with a value of 0.409 makes the RepTree classifier as the nominated algorithm for the model (see Fig. 2).

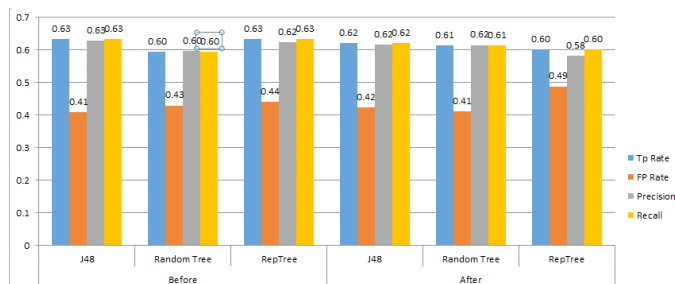


Fig. 2. Classifier Performance Chart.

Fig. 3 shows the result tree of the J48 algorithm, which is the best approach to show the most correlated factors/questions to the final class (failed). Each node in the tree is a question, and its branches are drawn on the basis of the answers. Each node can be considered a decision. The tree can be used also for predicting failure or success by answering questions in the nodes of the tree.

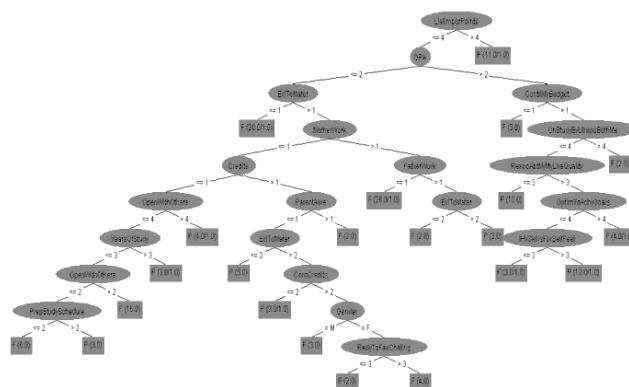


Fig. 3. J48 Classifier Tree.

V. CONCLUSION

This study aims to explore and test the process of applying the decision tree algorithms with questionnaire of students to seek the factors that affect student success/failure. Data mining algorithms and especially decision tree algorithms can be the best solution for predicting students' performance because they provide high accurate results and give road map for both academic staff and students. Based on the results of the model, it can be said that a number of factors (attributes) can affect the accuracy of the result tree and overall student academic performance. Attributes such as Age, Work, Gender, Stage, and Status had less effect on student success, whereas GPA, Credits, List Important Notes, Father Work, and Fresh Food had the most significant effect on the final class. The attribute evaluator algorithms can be used to find closely related questions that negatively or positively affect the success of students. The questionnaire contains many unimportant questions that can be discovered by the data-mining algorithms. Large data set and number of attributes in the data set affects the accuracy of the decision tree. The model can be utilized by students and academic staff to decide which questions/answers will enhance academic performance and improve the success of institutions.

REFERENCES

- [1] A. M. Abaidullah, N. Ahmed, and E. Ali. "Identifying Hidden Patterns in Students' Feedback through Cluster Analysis." International Journal of Computer Theory and Engineering 7.1 (2015): pp. 16-20.
- [2] M. Goyal and R. Vohra. "Applications of Data Mining in Higher Education." International journal of computer science 9.2 (2012): pp. 113-120.

- [3] M. A. Yehuala. "Application of Data Mining Techniques for Student Success and Failure Prediction (The Case of Debre_Markos University)." *International Journal of Scientific & Technology Research* 4.4 (2015): pp. 91-94.
- [4] K. M. Osei Bryson. "Towards Supporting Expert Evaluation of Clustering Results Using a Data Mining Process Model." *Information Sciences* 180.3 (2010): pp. 414-431.
- [5] R. Bekele and M. Wolfgang. "A Bayesian Approach To Predict Performance of a Student (bapps): A Case With Ethiopian Students." *Algorithms* 22.23 (2005): 24.
- [6] K. D. Kolo, S. A. Adepoju, and J. K. Alhassan. "A Decision Tree Approach for Predicting Students Academic Performance." *International Journal of Education and Management Engineering (IJEME)* 5.5 (2015): pp. 12-19.
- [7] C. Romero, V. Sbastián, and G. Enrique. "Data mining in Course Management Systems: Moodle Case Study and Tutorial." *Computers & Education* 51.1 (2008): pp. 368-384.
- [8] S. Natek and M. Zwilling. "Student Data Mining Solution Knowledge Management System Related To Higher Education Institutions." *Expert systems with applications* 41.14 (2014): 6400-6407.
- [9] A. Kh. Hamoud. "Selection of Best Decision Tree Algorithm for Prediction and Classification of Students' Action". *American International Journal of Research in Science, Technology, Engineering & Mathematics*, 16(1), September- November, 2016, pp. 26-32. Available at SSRN: <https://ssrn.com/abstract=3027086>.
- [10] T. Mishra, K. Dharminder and G. Sangeeta. "Mining Students' Data For Performance Prediction." *Fourth International Conference on Advanced Computing and Communication Technologies*. 2014.
- [11] M. A. Yehuala. "Application of Data Mining Techniques for Student Success and Failure Prediction (The Case of Debre_Markos University)." *International Journal of Scientific & Technology Research* 4.4 (2015): pp. 91-94.
- [12] Bhargava, G. Sharma, R. Bhargava and M. Mathuria, "Decision Tree Analysis on J48 Algorithm for Data Mining", *International Journal of Advanced Research in Computer Science and Software Engineering*, Volume 3, Issue 6, June 2013.
- [13] Amjad Abu Saa, "Educational Data Mining & Students' Performance Prediction." *International Journal of Advanced Computer Science and Applications* 7.5 (2016): pp.212-220.
- [14] M. N. Anyanwu, and S. G. Shiva. "Comparative Analysis of Serial Decision Tree Classification Algorithms." *International Journal of Computer Science and Security* 3.3 (2009): pp. 230-240.
- [15] L. Rokach and O. Maimon. "Data mining with decision trees: theory and applications." *World scientific*, 2014.
- [16] K. N.Shah, et al. "Study of Data Mining in Higher Education-A Review." (*IJCSIT*) *International Journal of Computer Science and Information Technologies* 6.1 (2015): pp. 455-458.
- [17] I. H. Witten, E. Frank, and M.A. Hall. "Data Mining: Practical Machine Learning Tools and Techniques." *Morgan Kaufmann*, 2016.
- [18] S. K. Jayanthi and S. Sasikala. "REPTree Classifier For Identifying Link Spam in Web Search Engines." *IJSC* 3.2 (2013): pp. 498-505.
- [19] G. Jagannathan, K. Pillaipakkamnatt, and R. N. Wright. "A practical Differentially Private Random Decision Tree Classifier." *Data Mining Workshops, 2009. ICDMW'09. IEEE International Conference on. IEEE*, 2009.
- [20] G. D. Israel. "Determining Sample Size. "University of Florida IFAS extension." (2009).
- [21] B. Carson. "The Transformative Power of Action Learning." *Chief Learning Officer*. Retrieved 2017, from: <http://wial.org/executive-board/bea-carson-executive-board/>.
- [22] U. Sekaran, and R. Bougie. "Research Methods For Business: A Skill Building Approach." *John Wiley & Sons*, 2016.



Alaa Khalaf Hamoud

Alaa Khalaf Hamoud is a lecturer in Computer Information Systems, University of Basrah, Iraq. He received BSc degree from Computer Science Department, University of Basrah in 2008 with first ranking college student. He also received his MSc degree from the same department with first ranking department student. He participated in (seven months) IT administration course in TU Berlin-Germany.

His scientific interests are data mining, data warehousing.



Ali Salah

Ali Salah received M.sc in Information Technology from Tenaga University, (Malaysia) and B.sc in Computer Science from Basra University (Iraq). His current research interests includes cloud computing and data mining.



Wid Akeel

Wid Akeel received M.sc in Information Security (2012) from Basra University, (Iraq) and B.sc in Computer Science (2006) from Basra University (Iraq). Her current research interests includes information security, Steganography, image processing and data mining.