Al-Baha University Journal of Basic and Applied Sciences 6(2) (2022) 25–30

Article available at Al-Baha University Journal of Basic and Applied Sciences

Journal of Basic and Applied Sciences

Journal homepage: https://portal.bu.edu.sa/web/jbas/

An Intelligent Approach for Improving Students' Performance

Sami A. Al Ghamdi*

Faculty of Computer Science and Information Technology, Al-Baha University, Al-Aqiq, Saudi Arabia

ARTICLE INFO

Article history:
Received 18 January 2022;
Received in revised form 15 May, 2022;
Accepted 19 May, 2022

Keywords:
Educational data mining
Decision trees
Association rules
J48 algorithm
Apriori algorithm

ABSTRACT

The employment of data mining approaches to resolve educational tasks have increases tremendously in the past decade. This paper investigates the efficiency of employing a combination of data mining techniques to extract insightful knowledge from students' information. Particularly, association rules algorithm known as Apriori algorithm is used to find rules between courses. In addition, a decision tree classifier known as J48 is applied on the data to predict students' performance. The data went through a features selection process where features with high influence were selected. Experiments are performed on WEKA, a data mining and machine learning open-source software, on a real-world dataset collected from the Information Systems department at Al-Baha university. The results show that J48 classifier outperformed two other classifiers (IBk and Naïve Bayes) in terms of accuracy.

1. Introduction

The focus of data mining (DM) is on detecting unexpected relationships between observational data sets and presenting a summarization of useful and understandable results to the data owner [1]. Recently, there has been an increased interest in research linking DM with educational systems, forming a new area of research known as Educational Data Mining (EDM) [2]. Over the last few decades, universities have collected a tremendous amount of educational data related to students' information and their educational behavior. These data repositories provide a valuable educational data that could be analyzed and used to understand learning and learners. EDM aims to explore data that was generated on and educational setup using DM approaches such as prediction methods, clustering, and relationship mining, in order to discover hidden patterns in a distinctive type of data with an educational nature to enhance the learning experience and institutional effectiveness [2,3].

Different DM approaches have been employed to resolve many tasks and applications in educational environment. The following EDM subjects/tasks were introduced by Castro et al. [4] and Romero and Ventura [2]: applications that deal with the assessment of the learning performance of students, applications providing learning recommendations and course registration approaches that study teachers and students' feedback in e-learning courses, detecting of unusual learning behaviors among students, grouping students based on their customized features, personal characteristics, etc., the process of constructing courseware and learning contents automatically, and planning and scheduling future coursework for students.

This paper employs a combination of DM approaches in order to predict the students' performance. A dataset of students' information is obtained from Al-Baha University. The famous association rule algorithm known as Apriori algorithm was applied on the data to generate rules with related courses. Next, the decision tree classifier known as J48 was employed to classify the item sets generated by Apriori algorithm. The performance of J48 was compared against two classifiers: IBk (Instance-Based k) and Naive Bayes, where J48 had the best accuracy among them. The features of the student's data are considered and attributes of higher influence are selected. The experiments are performed on WEKA [5], the open-source software that offers a wide range of machine learning algorithms for DM tasks.

The rest of the paper is organized as follows: the related work is discussed first. Then, the research methodology is explained in detail. Next, the data pre-processing and characteristics are discussed. After that, the experimental work is explained and the analysis of results is shown. Finally, we discuss the conclusion and future work.

2. Related Work

Various educational data extraction techniques were used in many studies to classify and predict student performance [6-11]. The goal of such type of studies is to improve the quality of education in educational institutions

There is a large body of work that studies the use of data mining techniques in educational environment. The studies focused on different topics such as the assessment of the learning performance of students, provide course registration recommendations based on the learning behaviour of students, and detect unusual learning behaviours among students.

* Corresponding author: Faculty of Computer Science and Information Technology, Al-Baha University, Al-Aqiq, Saudi Arabia.

E-mail address: samialghamdi@bu.edu.sa.

The following overviewed studies investigate the usage of a combination of data mining approaches (classification, clustering, association rules, …etc.). The main objective of these studies is to enhance the educational environment.

The work in [12] used a decision tree algorithm known as C4.5 to classify students based on their performance to *normal* (student does not need advising), near to risk (student needs advising), in risk (student needs extensive advising). The main attributes the advisor would rely on are the *difference between the registered* and the *gained credit hours* by a student. Kappa coefficient and ROC area measure tools were used to evaluate the algorithm's results.

The work in [6] and [13] studies the association between courses and students' behavior and suggest course recommendation technique. In [6], the authors developed a personalized course sequence recommendation method that could improve the student's GPA and contributes into shortening their graduation time. The work in [13] builds a model to support students' courses selection using a combination of data mining algorithms. Apriori algorithm is used to discover the association rules amongst studied courses. After that, J48 algorithm (a decision tree classification algorithm) is used to predict the most suitable future courses for students to take based on their historical and current data. K-Means clustering algorithm was applied on the data to cluster data based on similarities on major, clustering by learning results, course's score, etc.

In [8], the coursers to be registered next semester is generated using association rules. This *smart academic advising system* used real-data obtained from the University of Jordan and applied Apriori algorithm on the data to get the list of suggested courses sorted by their confidence.

Alhakami et al. [10] applied J48 decision tree algorithm on students' data set to analyze students' performance using various data mining techniques such as classification and clustering. The J48 classification algorithm was applied to the data set. The results showed that the students' performance can be affected by many factors such as: GPA, standardized tests, high school major, sex, age, nationality and city.

Aher and Lobo [7] presented a Course Recommendation System using a combination of clustering and association rules algorithms. First, Simple K-means clustering algorithm grouped similar data into the same clusters. Then, Apriori association rule algorithm was applied on each cluster to find associated courses. Positive association rules are considered as recommended courses for students.

The study in [11] is concerned predicting the students' graduation grades using the decision tree algorithm known as Iterative Dichotomiser 3 (ID3) on various features such as high school grade, entrance examination score.

The work in [9] applied a combination of DM techniques on students' data to evaluate their academic performance. First, (attributes) with the higher influence were chosen (12 out of 24 attributes were chosen). Then, four classification algorithms (J48, PART, BayesNet and Random Forest) were applied. Apriori algorithm was also used to find the best association rules amongst the data. Experimental results showed that the Random Forest algorithm is best choice for the chosen input data set.

The work in [14] proposed an enhanced version of the decision-tree algorithm C4.5 called *improved C4.5* by adding a grid search function in order to get more accuracy in classifying and predicting the students' performance. The algorithm was applied on students' data and its performance was compared against traditional C4.5 algorithm and three other algorithms (Random Forest, Support Vector Machine (SVM) and Naive Bayes). The results showed that the improved C4.5 had better classification accuracy than the other four algorithms (including traditional C4.5).

The authors in [15] investigated the most influential attributes in predicting the students' performance. Two famous classification algorithms, Bayesian Network and Decision Tree, were used to classify the student's data. The results showed that Bayesian Network has better accuracy than Decision Tree, and the student's attendance and GPA in the first semester were are the influential attributes on the students' performance.

In this work, we choose the attributes that have more influence on the data and find the association between studied courses in order to build a course prediction model using a decision tree algorithm.

3. Methodology

This work applies several DM techniques to build a course prediction model for college students. First, the well-known association rules approach known as Apriori algorithm is applied to find association rules between courses. Then, the decision-tree algorithm known as J48 is applied on these rules to build a prediction model for the courses to be taken by students. The following sections give a brief introduction to Apriori and J48 algorithms.

Apriori Algorithm: Apriori is an algorithm designed to discover association rules in item sets. The basic idea of the frequent itemset mining is to find frequent itemset whose support (Frequency of occurrence of an itemset) is greater than or equal to a minimum support (minsup) threshold. The Naïve approach to this problem is the Brute-force algorithm where we consider each itemset in the lattice, and count the support of each candidate by scanning the data. This approach has a time complexity of ~ O(NMw), and a space complexity O(M), where N is the number of transactions, M is the number of possible item sets, and w is the maximum width of a transaction.
The Apriori principle enhances the Naïve approach by considering the following:

- If an itemset is frequent, then all of its subsets must be frequent.
- If an itemset is not frequent, then all of its supersets cannot be frequent.
- The support of an itemset never exceeds the support of its subsets

Pseudocode:

1- Let C_k = candidate itemsets of size k.
2- Let L_k = frequent itemsets of size k.
3- Let $k = 1$, C_1 = all the items in the transaction set.
4- While C_k != Ø.
5- Find frequent itemsets in C_k and put them in L_k.
6- Generate a collection of candidate itemsets C_{k+1} of size $k+1$ using L_k.
7- $k=k+1$.

J48 Decision Tree: The J48 (also referred to as C4.5) decision tree algorithm is the implementation of algorithm ID3 [16] (Iterative Dichotomiser 3) that is developed by WEKA project team. J48 has additional features such as considering missing values, derivation of rules, continuous attributes value ranges, and pruning of decision tree.

Data Collection

This work uses real-world data that involves students and courses in the department of Information Systems at Al-Baha University. The data contains the attributes: semester, student ID, student name, CRN, course code (course code & course number), course name, grade and degree.

The preprocessing stage involves removing irrelevant attributes and selecting more influential features which also contributed in reducing dimensionality. The selected data set files were transformed into Attribute Relation File Format (ARFF, which is a file format compatible with WEKA).

Algorithm

Step1: *Enter data after Pre-Processing.*

Step2: *Apply Apriori algorithm to the data.*
Step3: *Configure the expected value for each course based on the type of pattern classification.*
Step4: *For each course, calculate the probability that it will be the first course in the student's transaction log.*
Step5: *Update the expected value for each node in the tree, (except for the root node). The expected value of its dependent nodes is calculated using the probability values computed in Step 3.*
Step6: *Generate the rules according to the created tree.*

Tracking the decision tree helps in deciding which course to recommend next. The decision tree is created by taking patterns of transaction records as input and updating the value for each node based on the frequency in the transaction records. Looking at students' current transaction records, future transactions can now be predicted by matching them to the decision tree. One transaction is executed from multiple transactions. Each node has a value indicating the probability that the course has been studied.

We build our methodology on several different layers of stages that improve the quality of data and put it in a framework to make it suitable for predicting a suitable course for the students. Starting from data collection and then data preprocessing, appropriate classification with J48 classifier, formation of association rules, prediction, and similarity calculation with Apriori algorithm. Probabilistic prediction for a group of courses is done by finding matches and similarities between the current position of the student and the probability tree. It is not a good and acceptable implementation of Apriori's algorithm directly to classify student record transactional data across a skewed data set. Also, the use of decision trees is not always ideal due to the existence of outliers, so the use of tree pruning is used to make it smaller and less complex. These association patterns can provide knowledge that helps build a decision tree away from data anomalies, based on association rules we can improve the prediction of typical courses that can be offered as a recommendation for advising students.

In our research, we will use Apriori algorithm to discover the association rules among the students. Then we apply J48 classifier that produces a prediction group model for the best available options based on association rules, which improves the results of our predictions.

Based on the simple rule of association, it is said that the courses that are repeated at least 65%, can be considered as courses that are successfully passed by the students, meaning that they appeared repeatedly in the transaction records, and the additional processing is only posted on the courses that appeared frequently. The remaining courses that occurred less than 65% are deleted. The practicable implication rules are those where the left-hand-side (LHS) itemset has a positive effect on the right-hand-side (RHS) itemset.

Correlation analysis is performed based on Apriori algorithm, in two basic steps: 1) finding frequent itemsets, 2) finding the association rules. The Apriori algorithm is one of the methods for searching databases. The Apriori principal assumes that subsets of frequent itemsets are also frequent. The Aproiri support element is used to mine collections of duplicate items from the database.

In Table 1, each cell has a value of 0 or 1, where 0 means that the student has not studied or has not passed the course yet, and 1 means that the student has passed the course. After Apriori is applied, based on support and confidence, a list of strong association rules is obtained. We check if Course N is on the left side of the generated association rules, and we check if the Course on the right is not passed by the student. After that, we can recommend the course to the student.

Table 1: The transactions matrix (1 means course was passed, 0 means course not passed yet)

Student/ Course	Course 1	Course 2	Course 3	Course ….	Course N
Student 1	1	0	1	……	1
Student 2	0	1	1	……	1
Student …	……	……	……	……	……
Student M	1	0	1	……	1

4. Experimental Work and Results

4.1. Apriori Algorithm Results

Table 2 shows the best rules found after applying Apriori algorithm. It can be noticed that some rules should be eliminated for several reasons such as pre-requests constraints. For example, English Language 2 - 11010211 ==> 1. Software Engineering – 41011322 cannot be studied together because the curse 11010211 is a pre-request to the course 4101322. We apply the Apriori association rule algorithm to each cluster then we choose the best rules found in that group to classify the course that gives the correct result.

Table 3 shows a summary of the results obtained after applying the Apriori algorithm, indicating the frequent itemsets that were purified, the number of rules found, and the average of confidence and support for each rule.

From the results, we have noticed that the greater the support value, the less the number of itemsets and the smaller the number of the resulting item set, the more logical the relations between them. The change in both support and confidence, by increasing or decreasing, leads to the emergence of different rules with different degrees of logic. When a large amount of intuitive and illogical results emerged when identifying all the traits to find the relationship between them, which led us to define specific traits and perform operations on them to discover more logical relationships between the courses then refer to 4 lists, Hence, it can be considered as one parameter under the requirement classification, and the approved rules were as shown in Table 2. Table 4 shows the classification of courses based on the course requirement in the study plan.

4.2. Decision Tree Classifiers

The decision tree is used in decision analysis to help determine the strategy that will lead to achieving the goal. The decision tree is a flowchart-like structure, where each internal node represents a "test" of the property, each branch represents the test output and each final node represents the decision taken after calculating all the characteristics. The path from the root to the final node represents classification rules.

- Decision Contract - These are usually represented by squares.
- Opportunity nodes (probability) - are represented by circles.
- Final knots - are represented by squares.

Table 2: Result of association rule algorithm using WEKA

NO.	Results
1	Best rules found: 1.UR=11020101 ==> CR=41010121 2.CR=41010131 ==> UR=41010132 3.CR=41010141 ==> UR=11020101 4.CR=41010131 ==> UR=11020101 5. CR=41010131 ==> UR=11020106 -- Best rules found: 1.UR=11010113 ==> CR=41021315 2.CR=41010121 ==> UR=11010113 3.CR=41010132 ==> UR=11020204 4.CR=41010121 ==> UR=11010113 5.PRC=41021312 ==> UR=11010112 -- Best rules found: 1.PRE =41021611 ==> CR=41011213 2.CR=41011322 ==> PRC=41021463 3.CR=41021212 ==> UR=11020204 4.CR=41010131 ==> PRC=41021322 5.PRC =41021421 ==> UR=11010101
2	Best rules found: 1.UR=41011213 ==> CR=41011213 2.CR=41010131 ==> UR=41010132 3.CR=41010132 ==> UR=11020107 4.CR=41010131 ==> UR=11010010 5.CR=41010131 ==> UR=11020106 -- Best rules found: 1.UR=11010113 ==> CR=41011322 2.CR=41011212 ==> UR=11010111 3.CR=41010132 ==> UR=11010111 4. PRC=41021463 ==> UR=11020101 5.CR=41011327 ==> UR=11010211 -- Best rules found: 1.PRE=41012213 ==> CR=41011213 2.CR=41010132 ==> PRC=41021234 3.CR=41021212 ==> UR=11020204 4.CR=41010141 ==> PRC=41021321 5.PRC=41021611 ==> UR=11010101

Table 3: Summary of results after applying Apriori algorithm

Confidence	Support	Set n. Rules	Sets large item	N. item
0.9	0.1	10	5	1
0.2	0.01	100	7	1
0.2	0.0010	200	6	2

A decision tree is used to start from observations about a specific element to conclude about the value that that element carries, represented by the leaves of the tree, while the element itself is represented by the branches of the tree. One method of predictive modeling is a decision learning tree.

Clustering techniques are used to improve performance, we use the common approach between association rules and decision trees, and for accurate results built-in single classifier. By comparing three classifiers the performance is evaluated in terms of F-measure.

Table 4: Classification of courses based on requirement level (university (UR), college (CR), program compulsory (PRC), and program electives requirements (PRE))

NO.	Course category, course name, and course code
1	**University Requirements (UR)** 1. English Language 1-11010010 2. English Language 2 - 11010211 3. The Holy Quran - 11020107 4. Islamic Education 1- 11010101 5. Islamic Education 2 - 11010111 6. Essay and Research Writing Skill- 11010112 7. Health Education- 11020204 8. Voluntary Community Service- 11010113
2	**College Requirements (CR)** 1. Introduction to Computing - 41010121 2. Calculus for Computing 1- 41010131 3. Physics for Computing 1- 41010141 4. Probability and Statistics - 41010132 5. Discrete Structures for Computing - 41011212 6. Algorithm and Data Structures - 41021212 7. Computer Networks - 41011213 8. Software Engineering - 41011322 9. Human-Computer Interaction- 41011321 10. Web Development- 41011327
3	**Program Requirements Compulsory (PRC)** 1. Foundations of Information Systems-41021211 2. Accounting for CIS-41021261 3. Data and Information Management- 41021221 4. Marketing for CIS-41021463 5. Mathematics for Business Analysis- 41021234 6. Systems Analysis and Design- 41021315 7. IS Project Management- 41021311 8. Application Development- 41021312 9. Enterprise Systems- 41031313 10. Economics- 41021362 11. IS Strategy, Management and Acquisition- 41021321 12. Business Process Management- 41021322 13. Enterprise Architecture- 41021222 14. Senior Project for CIS 1- 41021411 15. Senior Project for CIS 2- 41021421 16. IT Security and Risk Management- 41031412 17. Big Data- 41021413 18. Decision Models for Management- 41021414 19. Professional Issues- 41021422 20. Multimedia- 41031423
4	**Program Requirements Electives (PRE)** 1. Systems Integration- 41031323 2. Operating Systems- 41011314 3. Software Testing- 41031424 4. IS Innovation and New Technologies- 41021324 5. IT Infrastructure- 41031314 6. IT Audit and Controls- 41021323 7. Digital Logic Design- 41012213 8. Recommender Systems- 41011614 9. Selected topics in CIS- 41021611

The J48 algorithm acts as a primary classifier in our approach to predict the courses of the current study, then the combined classifier is built, and the training data and test data are used to verify the prediction and accuracy of the results. Prediction results can be compared with the training data and test data with the help of association rules. It is also divided into test and training groups. In order to obtain unbiased results, we use the (k-fold) cross-validation procedure. That is, to predict the performance of the model when used while training through the use of a limited sample. Through our belief in the rule that it is better to train with small-sized balanced data rather than large unbalanced data. Three measures were used to perform base tests on data sets (F-measure, Precision, Recall). Recall (also known as sensitivity) quantifies the number of positive class predictions made out of all positive examples in the dataset. Precision (also called positive predictive value) quantifies the number of positive class predictions that belong to the positive class. F-Measure is a measure of a test's accuracy. It is calculated from the precision and recall of the test

Based on the algorithms' performance results in Table 5, we can clearly see that the proposed J48 classifier based on Apriori algorithm outperforms the Naïve Bayes and IBk classifiers, where precision, recall and F-measure all higher than the two other classifiers. Hence, this is a good indication to decide on adopting the approach proposed on this work.

Table 5: Comparison results of precision, recall, F-measure for Naïve Bayes, IBK, and J48 classifiers

Classifier	Precision	Recall	F-measure
Naive Bayes	0.944	0.981	0.962
IBK	0.922	0.904	0.913
J48	0.976	0.998	0.986

4.3. Student Course Features (SCF) Models

Input parameters: the student's degree in the course, student ID, student name, CRN, Requirements, course code, Course name, grade, GPA, semesterID, the number of academic hours successfully passed by the student, the number of academic hours remaining for the student, the student's number, the number of academic courses and the number of academic courses remaining for the completion of the study plan. The order of the input variables is shown as determined through the use of J48 algorithm. It was found that the indicator of CRN, the number of academic hours that the student successfully passed and the number of academic hours remaining for the student have a lower ranking, indicating that these variables are the least influential. To verify the validity of the J48 algorithm for selecting the relevant input variable, three student course features (SCF) models (SCF-1, SCF-2, and SCF-3) were developed to assess the accuracy of SCF prediction.

4.4. Testing the Performance of SCF-3 Model

Three SCF models (SCF-1, SCF-2, SCF-3) are developed using WEKA. The SCF-1 model utilized: course, student ID, student name, course code, Course name, grade, GPA, and semester ID. SCF-2 model used course, student ID, student name, course code, course name, grade, and semesterID as input parameters. The SCF-3 used inputs as semesterID, StudentID, CourseCode, Grade and Requirements. Bayesian network (BayesNet) algorithm is used for training the SCF models. The reason because you are using a model that is robust enough so that it can be applied to similar situations. In this capacity, a model is in effect a "pattern" that may be reused as long as it matches sufficiently to the circumstances of the environment [17], and the number of data samples used in SCF model, and indicate the number of input and output variables.
The local conditional probability distributions test is they are sufficient to represent the joint probability distribution of the domain. More concretely (Pr) Eq. (1).

$$\Pr(p_{ij1}, p_{ij2}, \ldots, p_{ijr_i} \mid \mathcal{G}) = \mathrm{Dir}(\alpha_{ij1}, \alpha_{ij2}, \ldots, \alpha_{ijr_i}) = \Gamma(\alpha_{ij}) \prod_{k=1}^{r_i} \frac{p_{ijk}^{\alpha_{ijk}-1}}{\Gamma(\alpha_{ijk})} \quad (1)$$

Table 6: SCF models prediction accuracy

Models	Pr (%)	Highlights
SCF -1	16.84	Good prediction accuracy
SCF -2	16.32	Good prediction accuracy
SCF -3	8.05	High prediction accuracy

It is clear from Table 6 that SCF-3 model show higher prediction accuracy than SCF-1and SCF-2 identifying SemesterID, StudentID, CourseCode, Grade, and Requirements as the most relevant features for SCF prediction. Therefore, J48 algorithm in WEKA can be used to find relevant input variables for SCF feature.

There are different techniques of selecting an attribute or features but the choice was to use the manual techniques, in which the trait is usually chosen based on human understanding (mental classification) of the data set. In our current case, it was mental classification agreed with the outcome of the three SCF models, which clearly and strongly identified the best parameters that could lead us to the correct prediction, which are represented in the Human understanding (mental classification) agreed with the outcome of the three SCF models.
The decision tree classifier will use J48 genetic algorithm. In order to create a decision tree based on the attribute values available in the training data.

The total number of students is 1820 and the number of courses in the study plan is 47 courses. The classification for courses is shown in Table 4 and is as follows: University Requirement (UR) – College Requirement (CR) – Program Requirement Core (PRC) - Program Requirement Elective (PRE), as shown in Table 1, where the number of attributes is five (Semester, StudentID, CourseCode, Grade, Requirements).
SCF-3 Model: A dataset of 66% was selected from the 1,820 data records, and the results were as follows:

When entering the training data, the correct prediction was made by 98.4% for the number of 1261 transactions, and the incorrectly predicted rate was 1.5% for the number of 20 transactions. the number of Leaves: 292, and the size of the tree: 297.

Detailed Accuracy by Class section of the output demonstrates break down of the classifier's prediction accuracy. The measures (Precision, Recall, F- Measure, ROC area) are useful for comparing classifiers. Table 7 shows detailed accuracy by Class. The area under the ROC curve is determined by the true positive and false positive rate. ROC area=1 indicates a perfect prediction, while ROC area=0.5 indicates a random prediction.

The ROC area values in the detailed accuracy by class section of the algorithm output indicate that class 1 (University) has the best prediction because it has the highest ROC area value. Class 4 (Elective) has the weakest prediction because it has the lowest ROC value.

Table 7: Detailed Accuracy by Class

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
University	0.986	0.001	0.996	0.986	0.991	0.997
College	0.906	0.001	0.993	0.906	0.942	0.991
Program	0.998	0.042	0.976	0.998	0.986	0.990
Elective	0.976	0.000	1.000	0.976	0.988	0.986
Weighted Avg	0.983	0.983	0.027	0.983	0.983	0.967

4.5. Confusion Matrix

Confusion matrix is a popular matrix that is used to describe the performance of a classification model. The matrix shows how many tests data instances have been assigned to each class. Each column in the matrix represents the Predicted class, and each row represents the Actual class. Classifier performance is usually evaluated using the data in the matrix, and the size of the matrix is based on the number of categories.

- **TP (True Positive)**: The number of states that the classifier correctly predicts of class p that actually belongs to class p
- **FN (False Negative)**: The number of cases that the classifier incorrectly predicts of class p that actually belongs to class n
- **TN (True Negative)**: The number of states that the classifier correctly predicts of class n that actually belong to class p
- **FP (False Positive)**: The number of cases that the classifier incorrectly predicts of class n that actually belong to class p
- **Recall or True positive rate (TP Rate)**: The rate of cases that are classified for a specific category = TP / TP + FN
- **False-positive rate (FP Rate)**: Rate of incorrectly classified cases for a given class. Calculate as follows: FN / TP + FN
- **Precision:** Rate the correct cases for a given class, and the classifier correctly predicts them. It is calculated as follows: TP / TP + FP.

As Table 8 shows, we have four classes: University, College, Program, and Elective, which are denoted as a, b, c, and d respectively. Therefore, the size of the Confusion Matrix is 4x4. The number of correctly classified instances is the sum of numbers on a matrix diagonal from top left to bottom right. The model made 1259 correct predictions (272+145+802+40) out of 1281 predictions in total.

Table 8: Confusion matrix results

Course Classification	a	b	c	d
a = University	272	0	4	0
b = College	0	145	15	0
c = Program	1	1	802	0
d = Elective	0	0	1	40

5. Conclusion and Future Work

In the current study, we propose an aggregate form Student Course Features (SCF) based on the data available in the Al-Baha University database. The proposed model is primarily intended to assist students who are part of the CIS Study Plan by predicting students' performance using historical data. We integrated Apriori algorithm with J48 decision tree classifier. three student course features (SCF) models (SCF-1, SCF-2, and SCF-3) were developed to assess the accuracy of SCF prediction. In this study, J48 algorithm was implemented using WEKA to identify input variables relevant to predicting SCF using the three SCF models where the most relevant input variables for predicting SCF are: semesterID, StudentID, CourseCode, Grade, and Requirements. Our future work is to focus more on demographic and social impacts on improving academic student rates by integrating artificial intelligence into a prototype that uses recommendation systems by building a full profile for each student while improving prediction accuracy.

Acknowledgment

This research is a part of a project entitled "Intelligent System for Academic Advising". This project was funded by the Deanship of Scientific Research, Albaha University, KSA (Grant No. 3/1440). The assistance of the deanship is gratefully acknowledged.

References

1. C. C. Aggarwal, *Data Mining: The Textbook*. Springer, 2015.
2. C. Romero and S. Ventura, "Educational data mining: a review of the state of the art," *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 40, no. 6, pp. 601–618, 2010.
3. A. Dutt, M. A. Ismail, and T. Herawan, "A Systematic Review on Educational Data Mining," *IEEE Access*, vol. 5, pp. 15991–16005, 2017, doi: 10.1109/ACCESS.2017.2654247.
4. F. Castro, A. Vellido, A. Nebot, and F. Mugica, "Applying data mining techniques to e-learning problems," in *Evolution of teaching and learning paradigms in intelligent environment*, Springer, 2007, pp. 183–221.
5. "Weka 3 - Data Mining with Open Source Machine Learning Software in Java." https://www.cs.waikato.ac.nz/ml/weka/ (accessed May 05, 2021).
6. [6] J. Xu, T. Xing, and M. Van Der Schaar, "Personalized course sequence recommendations," *IEEE Trans. Signal Process.*, vol. 64, no. 20, pp. 5340–5352, 2016.
7. S. B. Aher and L. M. R. J. Lobo, "Combination of machine learning algorithms for recommendation of courses in E-Learning System based on historical data," *Knowl.-Based Syst.*, vol. 51, pp. 1–14, Oct. 2013, doi: 10.1016/j.knosys.2013.04.015.
8. R. Shatnawi, Q. Althebyan, B. Ghalib, and M. Al-Maolegi, "Building a smart academic advising system using association rule mining," *ArXiv Prepr. ArXiv14071807*, 2014.
9. S. Hussain, N. A. Dahan, F. M. Ba-Alwib, and N. Ribata, "Educational data mining and analysis of students' academic performance using WEKA," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 9, no. 2, pp. 447–459, 2018.
10. H. AlHakami, T. Alsubait, and A. S. Al-Jarallah, "Data mining for student advising," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, 2020.
11. A. O. Ogunde and D. A. Ajibade, "A data mining system for predicting university students' graduation grades using ID3 decision tree algorithm," *J. Comput. Sci. Inf. Technol.*, vol. 2, no. 1, pp. 21–46, 2014.
12. M. Al-Sarem, "Building a Decision Tree Model for Academic Advising Affairs Based on the Algorithm C 4-5," *ArXiv151104026 Cs*, Nov. 2015, Accessed: Feb. 24, 2021. [Online]. Available: http://arxiv.org/abs/1511.04026
13. Y. THI and L. BA, "Educational data mining for supporting students courses selection," *Int. J. Comput. Sci. Netw. Secur.*, vol. 19, no. 7, pp. 106–110, 2019.
14. M. H. Sadiq and N. S. Ahmed, "Classifying and Predicting Students’ Performance using Improved Decision Tree C4.5 in Higher Education Institutes," *J. Comput. Sci.*, vol. 15, no. 9, Art. no. 9, Sep. 2019, doi: 10.3844/jcssp.2019.1291.1306.
15. A. U. Khasanah and Harwati, "A Comparative Study to Predict Student's Performance Using Educational Data Mining Techniques," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 215, p. 012036, Jun. 2017, doi: 10.1088/1757-899X/215/1/012036.
16. J. R. Quinlan, *C4.5: programs for machine learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993.
17. M. Karim and R. M. Rahman, "Decision tree and naive bayes algorithm for classification and generation of actionable knowledge for direct marketing," 2013.

AUTHORS GUIDELINES

Types of paper

Contributions falling into the following categories will be considered for publication: original research paper, review, letter to the editor, and case report.

Author(s) should ensure selecting the appropriate article type from the list of options when making their submission. Authors contributing to special issues should ensure that they select the special issue article type from this list.

Ethics in publishing

The author(s) are entirely responsible for the scientific ethics and the accuracy of statements and data contained in the manuscript.

Human and animal rights

If the work involves the use of animal or human subjects, the author should ensure that the work described has been carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans http://www.wma.net/en/30publications/10policies/b3/index.html; EU Directive 2010/63/EU for animal experiments http://ec.europa.eu/environment/chemicals/lab_animals/legislation_en.htm; Uniform Requirements for manuscripts submitted to Albaha University Journal of Basic and Applied Sciences https://portal.bu.edu.sa/web/jbas.

Authors should include a statement in the manuscript that informed consent was obtained for experimentation with human subjects. The privacy rights of human subjects must always be observed.

Conflict of interest

All authors are requested to disclose any actual or potential conflict of interest including any financial, personal or other relationships with other people or organizations within three years of beginning the submitted work that could inappropriately influence, or be perceived to influence, their work.

Submission declaration and verification

Submission of an article implies that the work described has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis or as an electronic preprint), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder. To verify originality, your article may be checked by a detection service.

Authorship

All authors should have made substantial contributions to all of the following: (1) the conception and design of the study, or acquisition of data, or analysis and interpretation of data, (2) drafting the article or revising it critically for important intellectual content, (3) final approval of the version to be submitted.

Changes to authorship

Authors are expected to consider carefully the list and order of authors before submitting their manuscript and provide the definitive list of authors at the time of the original submission. Any addition, deletion or rearrangement of author names in the authorship list should be made only before the manuscript has been accepted and only if approved by the journal Editor. To request such a change, the Editor must receive the following from the corresponding author: (a) the reason for the change in author list and (b) written confirmation (e-mail, letter) from all authors that they agree with the addition, removal or rearrangement. In the case of addition or removal of authors, this includes confirmation from the author being added or removed.

Only in exceptional circumstances will the Editor consider the addition, deletion or rearrangement of authors after the manuscript has been accepted. While the Editor considers the request, publication of the manuscript will be suspended. If the manuscript has already been published in an online issue, any requests approved by the Editor will result in a corrigendum.

Copyright

Authors are requested to complete the copyright form when submitting a manuscript. When the manuscript accepted, the authors agree that the copyright for their paper is transferred to BUJBAS. All forms of reproduction of text, figures or tables out of this journal are prohibited without permission from the publisher.

Author rights

Author or his employer or institution has the right to reuse the work. Authors transfer copyright to the publisher as part of a journal publishing agreement, but have the right to:

Share their article for personal use, internal institutional use and scholarly sharing purposes.

Retain patent, trademark and other intellectual property rights (including raw research data).

Proper attribution and credit for the published work.

Role of the funding source

Author(s) are requested to identify who provided financial support for the conduct of the research. If the funding source(s) had no such involvement, then this should be stated.

Open access

Every peer-reviewed research article appearing in this journal will be published open access. This means that the article is universally and freely accessible via the internet in perpetuity, in an easily readable format immediately after publication. The author does not have any publication charges for open access.

Language (usage and editing services)

Please write your text in English (American or British is accepted, but not a mixture of these).

Submission

Submission to this journal proceeds totally online. Use the following guidelines to prepare your article. Via the homepage of this journal (https://portal.bu.edu.sa/web/jbas) you will be guided stepwise through the creation and uploading of the various files. The system automatically converts source files to a single Adobe Acrobat PDF version of the article, which is used in the peer-review process. Please note that even though manuscript source files are converted to PDF at submission for the review process, these source files are needed for further processing after acceptance. All correspondence, including notification of the Editor's decision and requests for revision, takes place by e-mail and via the author's homepage, removing the need for a hard-copy paper trail. If you are unable to provide an electronic version, please contact the editorial office prior to submission [bujs@bu.edu.sa].

Referees

On first submitting an article, authors must suggest the names of five potential referees. The suggestions should include names e-mail addresses and details of their affiliations. Authors should only include suggested reviewers with whom they have no previous working relationship. It is at the discretion of the Editors as to whether any of these referees are used to review the manuscript.

PREPARATION

Peer review

This journal operates a single blind review process. All contributions will be initially assessed by the editor for suitability for the journal. Papers deemed suitable are then sent to a minimum of two independent expert reviewers to assess the scientific quality of the paper.

The Editor is responsible for the final decision regarding the acceptance or rejection of articles. The Editor's decision is final.

Use of word processing software

It is important that the file be saved in the native format of the word processor used. The text should be in single-column format. Keep the layout of the text as simple as possible. Most formatting codes will be removed and replaced on processing the article. In particular, do not use the word processor's options to justify text or to hyphenate words. However, do use bold face, italics, subscripts, superscripts etc. When preparing tables, if you are using a table grid, use only one grid for each individual table and not a grid for each row. If no grid is used, use tabs, not spaces, to align columns. The electronic text should be prepared in a way very similar to that of conventional manuscripts.

To avoid unnecessary errors, you are strongly advised to use the 'spell-check' and 'grammar-check' functions of your word processor.

Article structure

Subdivision - numbered sections

Article should be divided into clearly defined and numbered sections. Subsections should be numbered 1.1 (then 1.1.1, 1.1.2,...), 1.2, etc. (the abstract is not included in section numbering). Use this numbering also for internal cross-referencing: do not just refer to 'the text'. Any subsection may be given a brief heading. Each heading should appear on its own separate line.

Essential title page information

- **Title.** Concise and informative. Titles are often used in information-retrieval systems. Avoid abbreviations and formulae where possible.
- **Author names and affiliations.** Please clearly indicate the given name(s) and family name(s) of each author. Present the authors' affiliation addresses (where the actual work was done) below the names. Indicate all affiliations with a lower- case superscript letter immediately after the author's name and in front of the appropriate address. Provide the full postal address of each affiliation, including the country name and, if available, the e-mail address of each author.
- **Corresponding author.** Clearly indicate who will handle correspondence at all stages of refereeing and publication, also post-publication. Ensure that the e-mail address is given and that contact details are kept up to date by the corresponding author.
- **Present/permanent address.** If an author has moved since the work described in the article was done, or was visiting at the time, a 'Present address' (or 'Permanent address') may be indicated as a footnote to that author's name. The address at which the author actually did the work must be retained as the main, affiliation address. Superscript Arabic numerals are used for such footnotes.

Abstract

A concise and factual abstract is required. The abstract should state briefly the purpose of the research, the methodology, the principal results and major conclusions. An abstract is often presented separately from the article, so it must be able to stand alone. For this reason, References should be avoided. Also, non-standard or uncommon abbreviations should be avoided, but if essential they must be defined at their first mention in the abstract itself.

Keywords

Immediately after the abstract, provide a maximum of 6 keywords. Be sparing with abbreviations: only abbreviations firmly established in the field may be eligible. These keywords will be used for indexing purposes.

Introduction

State the objectives of the work and provide an adequate background, avoiding a detailed literature survey or a summary of the results.

Material and methods

Provide sufficient detail to allow the work to be reproduced. Methods already published should be indicated by a reference: only relevant modifications should be described.

Theory/calculation

A theory section should extend, not repeat, the background to the article already dealt with in the Introduction and lay the foundation for further work. In contrast, a calculation section represents a practical development from a theoretical basis.

Results

Results should be clear and concise.

Discussion

Discussion should explore the significance of the results of the work, not repeat them. A combined results and discussion section is often appropriate. Avoid extensive citations and discussion of published literature.

Conclusions

The main conclusions of the study may be presented in a short conclusions section, which may stand alone or form a subsection of a discussion or results and discussion section.

Acknowledgements

Collate acknowledgements in a separate section at the end of the article before the references. Individuals who provided help during the research (e.g., providing language help, writing assistance or proof reading the article, etc.) may be listed.

References

Citation in text

Please ensure that every reference cited in the text is present in the reference list (and vice versa). Any references cited in the abstract must be given in full. Unpublished results and personal communications are mentioned in the text. Citation of a reference as 'in press' implies that the item has been accepted for publication.

Web references

As a minimum, the full URL should be given and the date when the reference was last accessed. Any further information, if known (DOI, author names, dates, reference to a source publication, etc.), should also be given. Web references can be listed separately (e.g., after the reference list) under a different heading if desired, or can be included in the reference list.

References in a special issue

Please ensure that the words 'this issue' are added to any references in the list (and any citations in the text) to other articles in the same special issue.

Reference style

Text: Indicate references by number(s) in square brackets in line with the text. The actual authors can be referred to, but the reference number(s) must always be given.

List: Number the references (numbers in square brackets) in the list in the order in which they appear in the text.

Examples:

Reference to a journal publication:

[1] Van der Geer J, Hanraads JAJ, Lupton RA. The art of writing a scientific article. J Sci Commun. 2000;163:51–9.

Reference to a book:

[2] Strunk Jr W, White EB. The elements of style. 3rd ed. New York: Macmillan; 1979.

Reference to a chapter in an edited book:

[3] Mettam GR, Adams LB. How to prepare an electronic version of your article? In: Jones BS, Smith RZ, editors. Introduction to the electronic age, New York: E-Publishing Inc; 1999, p. 281–304.

Note: Shortened form for last page number. e.g., 51–9 may be used, and that for more than 6 authors the first 6 should be listed followed by “et al.”

Appendices

If there is more than one appendix, they should be identified as A, B, etc. Formulae and equations in appendices should be given separate numbering: Eq. (A.1), Eq. (A.2), etc.; in a subsequent appendix, Eq. (B.1) and so on. Similarly, for tables and figures: Table A.1; Fig. A.1, etc.

Abbreviations

Define abbreviations that are not standard in this field in a footnote to be placed on the first page of the article. Such abbreviations that are unavoidable in the abstract must be defined at their first mention there, as well as in the footnote. Ensure consistency of abbreviations throughout the article.

Units

Follow internationally accepted rules and conventions: use the international system of units (SI). If other units are mentioned, please give their equivalent in SI.

Math formulae

Please submit math equations as editable text and not as images. Present simple formulae in line with normal text where possible and use the solidus (/) instead of a horizontal line for small fractional terms, e.g., X/Y. In principle, variables are to be presented in italics. Powers of (e) are often more conveniently denoted by exp. Number consecutively any equations that have to be displayed separately from the text (if referred to explicitly in the text).

Footnotes

Footnotes should be used sparingly. Number them consecutively throughout the article. Please indicate the position of footnotes in the text and list the footnotes themselves separately at the end of the article.

Artwork

General points

- Author(s) should ensure using uniform lettering and sizing of your original artwork.
- Aim only to use the following fonts in your illustrations: Times New Roman and Symbol.
- Number the illustrations according to their sequence in the text.
- Use a logical naming convention for your artwork files.
- Provide captions to illustrations separately.
- Size the illustrations close to the desired dimensions of the published version.
- If the electronic artwork is created in a Microsoft Office application (Word, PowerPoint, Excel) then please supply 'as is' in the native document format.

Please do not:

- Supply files that are optimized for screen use (e.g., GIF, BMP, PICT, WPG); these typically have a low number of pixels and limited set of colors;
- Supply files that are too low in resolution;
- Submit graphics that are disproportionately large for the content.

Color artwork

Please make sure the artwork files are in the format of (TIFF (or JPEG), EPS (or PDF), or MS Office files) and with the appropriate resolution. If, together with the accepted article, submit a usable color figures then BUJBAS will ensure, at no additional charge, that these figures will appear in color online regardless of whether or not these illustrations are reproduced in color in the printed version.

Figure captions

Ensure that each illustration has a caption. Supply captions separately, not attached to the figure. A caption should comprise a brief title (not on the figure itself) and a description of the illustration. Keep text in the illustrations themselves to a minimum but explain all symbols and abbreviations used.

Tables

Please submit tables as editable text and not as images. Tables can be placed either next to the relevant text in the article, or on separate page(s) at the end. Number tables consecutively in accordance with their appearance in the text and place table notes below the table body. Please, do not duplicate results described elsewhere in the article and avoid using vertical rules.

Submission checklist

The following list could be useful during the final checking of an article prior to submission for review. Please consult this Guide for Authors for further details of any item.

Ensure that the following items are present:

One author has been designated as the corresponding author with contact details:

- E-mail address
- Full postal address.

All necessary files have been uploaded, and contain:

- Keywords.
- All figure captions.
- All tables (including title, description, footnotes).

Further considerations

- Manuscript has been 'spell-checked' and 'grammar-checked'.
- References are in the correct format for this journal.
- All references mentioned in the Reference list are cited in the text, and vice versa.
- Permission has been obtained for use of copyrighted material from other sources (including the Internet).

AFTER ACCEPTANCE

Proofs

One set of page proofs (as PDF files) will be sent by e-mail to the corresponding author or, a link will be provided in the e-mail so that authors can download the files themselves.

Offprints

The corresponding author, at no cost, will be provided with a personalized link providing free access to the final published version of the article on **BUJBAS**. This link can also be used for sharing via email and social networks.

AUTHOR INQUIRIES

You can track your submitted and accepted article at https://portal.bu.edu.sa/web/jbas.