

# Mining Educational Data to Predict Students' Academic Performance

Mona Al-Saleem<sup>1,3(✉)</sup>, Norah Al-Kathiry<sup>1</sup>, Sara Al-Osimi<sup>1</sup>,  
and Ghada Badr<sup>1,2</sup>

<sup>1</sup> King Saud University, Riyadh, Saudi Arabia  
{Zamil.mona, norahabdulaziz.cs}@gmail.com,  
{srmo2008, badrghada}@hotmail.com

<sup>2</sup> IRI-the City of Scientific Research and Technological Applications,  
Alexandria, Egypt

badrghada@hotmail.com

<sup>3</sup> Qassim University, Qassim, Buraydah, Saudi Arabia

**Abstract.** Data mining is the process of extracting useful information from a huge amount of data. One of the most common applications of data mining is the use of different algorithms and tools to estimate future events based on previous experiences. In this context, many researchers have been using data mining techniques to support and solve challenges in higher education. There are many challenges facing this level of education, one of which is helping students to choose the right course to improve their success rate. An early prediction of students' grades may help to solve this problem and improve students' performance, selection of courses, success rate and retention. In this paper we use different classification techniques in order to build a performance prediction model, which is based on previous students' academic records. The model can be easily integrated into a recommender system that can help students in their course selection, based on their and other graduated students' grades. Our model uses two of the most recognised decision tree classification algorithms: ID3 and J48. The advantages of such a system have been presented along with a comparison in performance between the two algorithms.

## 1 Introduction

According to Wook (2009): [1] “data mining is the process of automatically extracting useful information and relationships from immense quantities of data”. In this sense data mining is not about acquiring specific information; rather, it is about answering a question or approving theory. Another definition states: “Data mining algorithms can predict future trends and behaviours and enable businesses to make proactive, knowledge-driven decisions” [2]. These definitions argue that data mining could be very beneficial in every aspect of our lives.

One of the most famous applications of data mining is its use to support and improve higher education. In this context, data mining techniques can have a direct impact on the workforce being provided to industry and, as a result, can directly affect the economy. However, academic failure among university students is one of the

biggest problems affecting higher education. One of the solutions involves predicting a student's performance in order to help the student to choose the right course and increase their performance. Using data mining we can study and analyse information from previous students (which is the main subject of this paper). The various data mining techniques, which include classification, clustering and relationship mining, can be applied in educational settings to predict the performance of a student.

In this paper we present a data mining technique that analyses previous students' data in order to predict future students' performance on specific courses. To achieve this and support the development of higher education in our country, we use ID3 and J48 decision tree classification algorithms. The data set used in this paper is student data from the Computer Science Department, Computer and Information Sciences College, King Saud University of Saudi Arabia (KSU).

The rest of this paper is organised as follows. In Sect. 2, we present a brief background about the paper. Research related to our objectives is reviewed in Sect. 3. In Sect. 4, the methodology and experimental results are discussed. Conclusions are drawn in Sect. 5.

## 2 Decision Tree Classification Algorithms

Classification generally refers to the mapping of data items into predefined groups and classes [4]. It is also referred to as supervised learning and involves learning and classification. In the learning phase, training data are analysed by classification algorithm; during the classification phase, test data are used to estimate the accuracy of the classification rules [5]. Data mining involves various classification techniques such as decision tree algorithms, Bayesian Classification, and classification by Back Propagation, Support Vector Machine and K Nearest Neighbour.

Decision tree algorithms are used for gaining information for the purpose of decision-making and are one of the most widely used algorithms in data mining. The decision tree starts with a root node from which users take actions. From this first node, users split further nodes recursively according to the decision tree learning algorithm [6]. The final result is a decision tree in which each branch represents a possible scenario of decision and its outcome. The two widely used decision tree learning algorithms discussed in this paper are: ID3 and J48 (which is an implementation of C4.5).

### 2.1 ID3 (Iterative Dichotomiser 3) [6]

Quinlan Ross introduced ID3 in 1986. It is based on Hunt's algorithm. The tree is constructed in two phases: tree building and pruning. ID3 uses information gain measure to choose the splitting attribute. It only accepts categorical attributes in building a tree model.

To build the decision tree, information gain is calculated for each and every attribute, with the attribute containing the highest information gain designated as a root

node. The attribute is labelled as a root node and the possible values of the attribute are represented as arcs. Then all possible outcome instances are tested to check whether they fall under the same class or not. If all the instances fall under the same class, the node is represented with a single class name; if not, the splitting attribute is chosen to classify the instances.

## 2.2 J48 (Implementation of C4.5)

This algorithm is a successor to ID3 developed by Quinlan Ross. Unlike ID3 it handles both categorical and continuous attributes. It splits the attribute values into two partitions based on the selected threshold so that all the values above the threshold sit as one child and the remaining sit as another child. It uses Gain Ratio as an attribute selection measure to build a decision tree.

## 3 Data Mining in Higher Education

Universities process a vast amount of data each year, including information on the enrolment of students and academic performance. Therefore, these institutions require a suitable data mining tool to process past data and come up with solutions to resolve current situations. When data mining was introduced, the application of data mining techniques boosted many industries such as business, telecommunications, banking and education. In the education sector, data mining was defined as “the process of converting raw data from educational systems to useful information that can be used to inform design decisions and answer research questions” [3]. The prediction and prevention of academic failure among students has long been debated within each higher education institution. It has become one of the most important reasons for using data mining in higher education.

Performance prediction is a prominent field of research in educational data mining. There are many researchers who have addressed this problem, drawing similar conclusions to ours but using different techniques and tools. Vialardi et al. [7] developed a recommender system to support new students to choose better academic itineraries based on the acquisition of knowledge from past students' academic performance. Their system was based on using the C4.5 algorithms to provide the rules used in the recommender system to decide if the enrolment of the student on a specific course has a good probability of success or not. Experiments on the performance of the system give a success rate for 80 % of cases in terms of predicting the results for each student.

Another piece of research [8] developed a model to predict students' academic performance by taking into account socio-demographic and academic variables. The model, which is based on using the Naïve Bayes classifier and the Rapid miner software, reached accuracy of 60 % correct classification. Bharadwaj and Pal [9] conducted a study on student performance by selecting 300 students from five different colleges. By means of Bayesian classification method, which examines 17 attributes, it was found that factors including students' living location and medium of teaching were highly interlinked with student academic performance.

From the range of data mining techniques that exist, classification, and in particular decision tree algorithms, are the most effective modelling functions since they can be used to find the relationship between a specific variable, target variable and other variables. Al-Radaideh et al. [10] proposed a prediction approach that used data mining classification techniques to evaluate data that may affect student performance. They used three different classification methods: ID3, C4.5 and the Naïve Bayes. The results indicated that the decision tree model had better prediction accuracy than the other two models.

Another research study used decision tree algorithms to predict the results of students in their final semester based on the marks they obtained in previous semesters. They used two algorithms, J48 (implementation of C4.5) and Random Tree, to predict the result of the fifth semester based on the marks obtained by the students in the previous four semesters. They applied the two algorithms to the students' records and found that Random Tree is more accurate in predicting performance than the J48 algorithm.

## **4 Proposed Workflow**

Data mining educational information is a vibrant and important field of research. Generally, the majority of research for the prediction of student academic performance shares the same methodology: data collection and preprocessing followed by the implementation of the classification algorithm and the analysis of results.

### **4.1 Data Collection and Preprocessing**

Data preparation is the most complex step in the data mining process as it involves the collection of data and preprocessing. The collection of data includes the identification of data resources and the gathering of relevant data from these resources. Following this initial step, the collected data should be preprocessed to make it suitable for data mining algorithm implementation, which will also provide a greater understanding of the data collected. Preprocessing includes four important stages: data cleaning, which includes studying the data to remove noise (which are mostly the errors and biased values that can affect the accuracy of results) and dealing with unknown or missing data values; removal of duplicate values; smoothing noisy data and resolving inconsistencies.

Preprocessing also includes the integration of data from multiple resources, data values transformation from one domain to another and normalisation. Another preprocessing task is data reduction, which might be in the form of a feature selection to reduce the data dimensionality (which eliminates features that are not much use in model construction and produce better performance by reducing the number of used objects). Preprocessing also involves data discretisation, which reduces the number of values for a given continuous attribute by dividing its range into intervals. This helps to replace the numerous values of a continuous attribute by a small number of interval labels and supports concept hierarchy generation, which is used for discretisation.

Finally, the resulted data should be transferred into a suitable file format for the desired application, for example the WEKA file format (.arff).

## 4.2 Algorithm Implementation

After preparing the data, the dataset is divided into training and test datasets. The training dataset is used to construct the classification model and the test dataset is used either to test the performance of the constructed classification model or to compare predictions to the known target values [12]. Different algorithms are used for the classification and prediction of student performance; however, the decision tree classification is the most frequently used algorithm for this type of application.

## 4.3 Analysis of Results

After applying the algorithm, the resultant data patterns should undergo further processing. The post-processing stage includes pattern evaluation to measure different parameters such as accuracy. Post-processing also includes pattern selection because not all of the resultant patterns are of interest so a further selection may be beneficial. Finally, it is important that the result is interpreted correctly so that it is effectively viewed and understood.

# 5 Case Study

Our proposed system is modelled on predicting students' grades using former students' grades as a basis. Our methodology consists of four steps:

- a. Collecting and preparing educational data.
- b. Building the models of ID3 and J48 classification.
- c. Evaluating the models using one of the evaluation methods.
- d. The best model will be chosen as the model for online prediction.

We used NetBeans IDE as the programming environment and JAVA programming language in implementation. WEKA classes were integrated as a jar library.

## 5.1 Data Preparation

The dataset used in this work was obtained from King Saud University of Saudi Arabia, Computer Science Department, Computer and Information Sciences College. It was a sample of CS Master student records from session 2003 to 2011. Microsoft Excel was used to save the collected data for preprocessing; it was then stored as an .arff file for data mining tasks.

Preprocessing the data was a time consuming process; the original data were stored in an un-normalised database table with a row each for student ID, term, course and grade for each course. As part of the preprocessing stage we:

1. Filtered all the unneeded columns (like registration term and registration place), as those have no influence on the required results.
2. Removed the old courses that are no longer available in the department course plan.
3. Filled the empty cells where values were missing with the average of the grades taken in the same course by other students.
4. Transformed data to a transactional format where each row was indexed by student ID and each column indexed by course. At the intersection of each row and column, a value of the grade taken for each course is recorded.
5. Finally, our data form included 112 records and 37 attributes (one student ID, four core courses and 32 elective courses). We split our data into 82 records for training data and 30 records for test data. These training data are used to construct the model while the test data is used in evaluating. Table 1 shows an example of the normalised data.

**Table 1.** Example of the used dataset

	CS 541	CS 524	CS 512	CS 581	CS 595	CS 597
424221334	B+	B	A+	B+	B+	B+
426221091	A	B	A+	B	B+	A+
426221091	B	A	C	C+	A	A+
426221091	B+	A	B	C+	B+	A+
426221091	A	A	A+	B+	A	A+
426221091	A	A	C+	C+	A+	A
426221091	A+	A+	A	A	B+	A+

### 5.2 Offline Classification

We selected two classification models: ID3 and J48. The performance of the two models was compared to decide on the best approach for online prediction. Each model was built for each elective course as a class attribute based on the grades in the core courses. The average accuracy in prediction grades for all elective courses was computed to decide the best model.

Table 2 shows the analysis of the classification algorithms and the accuracy percentages for each elective course.

Table 2 shows that the performance of both algorithms is satisfactory; however, higher overall accuracy (83.75 %) was attained through J48 implementation compared to ID3, which had 69.27 % accuracy.

**Table 2.** Comparison of classification algorithms

Elective Courses	Classification Algorithms	
	J48	ID3
CS 595	50 %	50 %
CS 597	76.66 %	63 %
CS 598	96.66 %	73.33 %
CS 519	60 %	40 %
CS 551	70 %	70 %
CS 562	63.33 %	13.33 %
CS 572	36.66 %	26.66 %
CS 588	80 %	56.66 %
CS 520	90 %	73.33 %
CS 522	100 %	60 %
CS 523	70 %	46.66 %
CS 526	96.66 %	90 %
CS 527	86.66 %	63.33 %
CS 528	76.66 %	73.33 %
CS 529	100 %	100 %
CS 530	100 %	100 %
CS 546	73.33 %	60 %
CS 547	100 %	100 %
CS 548	100 %	100 %
CS 549	100 %	100 %
CS 558	80 %	53.33 %
CS 561	100 %	86.66 %
CS 563	56.66 %	6.66 %
CS 566	96.66 %	83.33 %
CS 579	100 %	100 %
CS 573	100 %	76.66 %
CS 574	96.66 %	96.66 %
CS 576	66.66 %	46.66 %
CS 578	100 %	80 %
CS 586	83.33 %	83.33 %
CS 589	96.66 %	83.33 %
CS 590	76.66 %	60 %
Accuracy Average	83.75 %	69.27 %

### 5.3 Online Classification

Based on the results shown in Table 2, the J48 algorithm was used for online prediction of students' grades. Our small system allows students to ask for predicted grades for one or more courses. The classification models are built using training data and specific attributes, using students' records to get a prediction. Attributes used to construct the model include the elective courses passed by a student together with core courses.

The system also recommends the best courses for students based on two criteria: maximum accuracy value and maximum grade value for all predicted courses. The goal is to help the students to choose a suitable course that enhances their academic attainment level.

The remainder of this section presents some examples of how the system interacts with students, how the results are displayed and an explanation of how the system works.

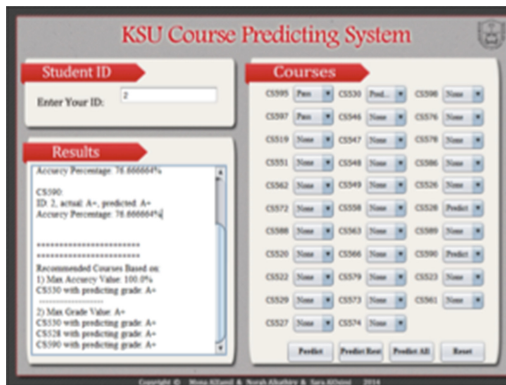


Fig. 1. Example 1

In Fig. 1, the student selected CS595 and CS597 as pass courses and CS530, CS528 and CS590 as predicted courses. Here, three classification models were built by using the training student's dataset after he or she pressed the predict button.

After building the models, the system predicted the student's grade in CS530, CS528 and CS590. In addition, it gave an accuracy percentage for this prediction by comparing predicted grades with actual grades:



CS530:  
 Predicted Grade: A+  
 Accuracy Percentage: 100.0%

CS528:  
 Predicted Grade: A+  
 Accuracy Percentage: 76.666664%

CS590:  
 Predicted Grade: A+  
 Accuracy Percentage: 76.666664%

The system also recommended the best courses based on maximum accuracy value and maximum grade value:

Recommended Courses Based on:  
 1) Max Accuracy Value: 100.0%  
 CS530 with predicting grade: A+

2) Max Grade Value: A+  
 CS530 with predicting grade: A+  
 CS528 with predicting grade: A+  
 CS590 with predicting grade: A+

Here, the system has 100 % accuracy in predicting course CS530, so it is recommended for the student. Moreover, all three courses gave an A+ prediction, which means that all three courses are recommended.

**KSU Course Predicting System**

**Student ID**  
 Enter Your ID: 3

**Results**

Predicted Grade: A+  
 Accuracy Percentage: 70.0%

CS561:  
 Predicted Grade: B+  
 Accuracy Percentage: 100.0%

.....

Recommended Courses Based on:  
 1) Max Accuracy Value: 100.0%  
 CS522 with predicting grade: A  
 2) Max Grade Value: A+  
 CS529 with predicting grade: A+  
 CS579 with predicting grade: A+  
 CS573 with predicting grade: B+  
 CS561 with predicting grade: B+

**Courses**

CS595	None	CS530	None	CS598	None
CS597	None	CS546	None	CS576	None
CS519	None	CS547	None	CS578	None
CS551	None	CS548	None	CS586	None
CS562	None	CS549	None	CS526	None
CS572	None	CS558	None	CS528	None
CS588	None	CS563	None	CS589	None
CS520	None	CS566	None	CS590	None
CS522	None	CS579	None	CS523	None
CS529	None	CS573	None	CS561	None
CS527	None	CS574	None		

Predict Predict Rest Predict All Reset

Copyright © Mona AlZamil & Norah Alkathary & Sara AlOstini 2014

Fig. 2. Example 2

In Fig. 2, a separate student selected some pass courses and obtained a prediction for the rest of the courses. The attributes used to construct these models were the pass courses together with four core courses and each of the rest of the courses, which were used as class attributes for each model. The system gave a predicted grade for each course along with an accuracy percentage:

CS588:
Predicted Grade: A
Accuracy Percentage: 80.0%
CS522:
Predicted Grade: A
Accuracy Percentage: 100.0%
CS520:
Predicted Grade: A
Accuracy Percentage: 90.0%
CS529:
Predicted Grade: A+
Accuracy Percentage: 100.0%
CS527:
Predicted Grade: B+
Accuracy Percentage: 86.666664%
CS579:
Predicted Grade: A+
Accuracy Percentage: 100.0%
CS573:
Predicted Grade: B+
Accuracy Percentage: 100.0%
CS574:
Predicted Grade: B
Accuracy Percentage: 96.666664%
CS589:
Predicted Grade: B+
Accuracy Percentage: 96.666664%
CS590:
Predicted Grade: A+
Accuracy Percentage: 76.666664%
CS563:
Predicted Grade: B+
Accuracy Percentage: 56.666668%
CS566:
Predicted Grade: A+
Accuracy Percentage: 96.666664%
CS523:
Predicted Grade: A+
Accuracy Percentage: 70.0%

The system also recommended the best courses for students based on maximum accuracy value and maximum grade value:

<p><b>Recommended Courses Based on:</b></p> <p>1) <b>Max Accuracy Value: 100.0%</b> CS522 with predicting grade: A CS529 with predicting grade: A+ CS579 with predicting grade: A+ CS573 with predicting grade: B+ CS561 with predicting grade: B+</p> <p>-----</p> <p>2) <b>Max Grade Value: A+</b> CS529 with predicting grade: A+ CS579 with predicting grade: A+ CS590 with predicting grade: A+ CS566 with predicting grade: A+ CS523 with predicting grade: A+</p>
--

## 6 Conclusion

To support the aims of higher education institutions, we proposed a model that would predict student performance in future courses based on the grade of previous students in core and elective courses. The workflow for our system was presented. A case study was also presented, using student records obtained from the department of Computer Science in KSU, where programmes consist of core and elective courses. The prediction system is based on the decision tree classification methods, specifically ID3 and J48. First, a preprocessing step was applied to the data and then the model was built using training data for each elective course as a class attribute based on the grades in the core courses. By calculating the average accuracy for each algorithm we found that J48 achieved a better accuracy performance of 83.75 % when compared to an accuracy of 69.27 % for the ID3 algorithm. On that basis we used the J48 algorithm for online prediction to enable the students to get a prediction for one or more courses.

## References

1. Wook, M., Yahaya, Y.H., Wahab, N., Isa, M.R.M., Awang, N.F., Hoo Yann, S.: Predicting NDUM student's academic performance using data mining techniques. In: Second International Conference on Computer and Electrical Engineering, ICCEE, pp. 357–361 (2009)
2. Hoe, A.K., Ahmad, M.S., Tan Chin, H., Shanmugam, M., Gunasekaran, S.S., Cob, Z.C., Ramasamy, A.: Analyzing students records to identify patterns of students' performance. In International Conference on Research and Innovation in Information Systems (ICRIIS), pp. 544–547 (2013)
3. Heiner, C., Baker, R., Yacef, K.: Preface. In: Workshop on Educational Data Mining at the 8th International Conference on Intelligent Tutoring Systems (ITS), Jhongli, Taiwan (2006)

4. Dunham, M.H.: *Data Mining: Introductory and Advanced Topics*. Pearson Education India, Delhi (2006)
5. Taruna, S., Pandey, M.: An empirical analysis of classification techniques for predicting academic performance. In: *IEEE International Advance Computing Conference (IACC)*, pp. 523–528 (2014)
6. Bunkar, K., Singh, U.K., Pandya, B., Bunkar, R.: Data mining: prediction for performance improvement of graduate students using classification. In: *Ninth International Conference on Wireless and Optical Communications Networks (WOCN)*, pp. 1–5 (2012)
7. Vialardi, C., Bravo, J., Shafti, L., Ortigosa, A.: Recommendation in higher education using data mining techniques. In: *International Working Group on Educational Data Mining (2009)*
8. Garcia, E.P.I., Mora, P.M.: Model prediction of academic performance for first year students. In: *10th Mexican International Conference on Artificial Intelligence (MICAI)*, pp. 169–174 (2011)
9. Bhardwaj, B.K., Pal, S.: Data mining: a prediction for performance improvement using classification. *Int. J. Comput. Sci. Inf. Secur. (IJCSIS)* **9**, 136–140 (2012)
10. Al-Radaideh, Q.A., Al-Shawakfa, E.M., Al-Najjar, M.I.: Mining student data using decision trees. In: *International Arab Conference on Information Technology (ACIT)*, Yarmouk University, Jordan (2006)
11. Anupama Kumar, S., Vijayalakshmi, M.N.: Mining of student academic evaluation records in higher education. In: *International Conference on Recent Advances in Computing and Software Systems (RACSS)*, pp. 67–70 (2012)
12. Krishna Kishore, K.V., Venkatramaphanikumar, S., Alekhya, S.: Prediction of student academic progression: a case study on Vignan University. In: *International Conference on Computer Communication and Informatics (ICCCI)*, pp. 1–6 (2014)