

An Educational Data Mining Approach to Explore The Effect of Using Interactive Supporting Features in an LMS for Overall Performance Within an Online Learning Environment

Ashwaq Al-Musharraf

Al Imam Mohammad Ibn Saud Islamic University (IMSIU)
Riyadh, Saudi Arabia

Mona Alkhatabi

Al Imam Mohammad Ibn Saud Islamic University (IMSIU)
Riyadh, Saudi Arabia

Summary

The abstract Today, Saudi Arabia has many universities that provide distance education programs. This study focuses on e-learning and distance education deanships at Al-Imam Muhammad Ibn Saud Islamic University. The university has been making significant efforts to improve the learning process and to provide high quality e-learning. As part of their efforts in supporting the educational process for students, they offer optional live virtual lectures and online quizzes on a weekly basis. This study aimed to investigate the effect of attending live virtual lectures (AVL) and taking online quizzes (TOQ) on distance learners' overall performance. The study applied the Naïve Bayes Classifier on learners' records, and found a significant positive correlation between attending live virtual lectures and academic performance. There was also a significant positive correlation between taking online quizzes and academic performance.

Key words:

Distance education, educational data mining, classification technique, student performance.

1. Introduction

Distance education (DE) is defined as 'any learning process, in which students are geographically distant'. DE offers a way for students that are restricted from learning because of socio-economic (or other) circumstances, to continue their education. Moreover, it is an option for people with different learning capabilities and disabilities (Harasim 2000, Abdulaziz 2012).

Today, DE is delivered through information and communication technology (ICT), which has caused major changes in DE and the way it is administered. DE has experienced a rapid growth in popularity because it reduces environmental costs and increases flexibility/convenience to students. A learning management system (LMS) is an e-learning, web-based application used to plan, implement, manage and assess a

learning process, via the World Wide Web. Today, LMSs have become very popular in higher education institutions (HEI) as a medium through which to deliver learning materials and for interaction and collaboration between learners and instructors (Lepouras et al. 2014, Bednar et al. 2013).

An LMS can provide vast amounts of data about students based on their interaction and behaviors, such as: frequency of watching videos, number of quizzes taken, quiz results, virtual lecture attendance times, test grades, etc. (Lepouras, Katifori, Vassilakis, Antoniou and Platis 2014). When faced with a large amount of data in an educational system, educational data mining is a key process which makes sense of the complex dataset and provides insight results (Siemens and Gasevic 2012)

This study aims to discover the effects of attending the optional live virtual lectures and taking online quizzes in distance learners' overall performance. The data set consists of 22,983 (female and male) distance learners. The data was collected from the course 'Introduction to computer science', where learners were registered at Al-Imam Muhammad Ibn Saud Islamic University in the second semester of 2014/2015.

2. Distance education

Distance education (DE) has been developing for decades. It has been defined as 'education given in a non-traditional manner, and offered off-campus' (Ferdiana 2013). DE has changed drastically and these changes have affected the way universities deliver knowledge (Larreamendy-Joerns and Leinhardt 2006). This new mode of education is delivered to students through the Internet, so students can access the education system in any place and at any time.

Distance education has many advantages for learners. First, the flexibility of DE facilitates learning to those who are not able to attend traditional classes. Learners can attend their classes from a convenient location with an Internet connection. DE allows learners to complete their work

without worrying about a lesson schedule. Second, there is no commuting as DE provides courses through an LMS enabling learners to take their courses online at home. Thus, DE saves time and costs of transportation. Third, is the option to learn while working; DE gives students an opportunity to complete their learning according to their own schedule. For learners who work, it is much easier than trying to complete one's education through a traditional educational program. DE fits learning around work and home life. This gives students (especially women), the opportunity to complete their education while simultaneously managing their home/work life (Abdulaziz 2012).

2.1 Distance Education in Saudi Arabia

During the last decade, many universities in Saudi Arabia have adapted distance-learning programs and have been using technologies such as Blackboard and Tadarus as an LMS to deliver and manage courses. These universities are aiming to develop the education systems of the Kingdom of Saudi Arabia, and include King Abdulaziz University (KAU), King Faisal University (KFU), King Saud University (KSU), Saudi Electronic University, and Al-Imam Muhammad Ibn Saud Islamic University (Abdulaziz 2012).

King Abdulaziz University (Jeddah) established the Faculty of Distance Learning in 2007/2008. This university offers distance learning in the western region of Saudi Arabia. King Faisal University (Al-Hasa) introduced the Deanship of E-Learning in 2010. King Saud University (Riyadh) established their Deanship of Distance Education in 2010 (Abdulaziz 2012). Saudi Electronic University established distance learning in 2011. Programs at the Saudi Electronic University are not offered wholly online. Students normally attend their lectures physically, or by using virtual classes through the Internet (Alshathri 2015). Al-Imam Muhammad Ibn Saud Islamic University (Riyadh) (Abdulaziz 2012), is the university under focus in this study.

The Al-Imam Muhammed Ibn Saud University offers bachelor degrees to students via Tadarus as the main learning management system.

The Deanship of E-Learning and Distance Education at the university provides extra optional live virtual lectures (transmitted via the Internet) on a daily basis. The lectures are recorded and uploaded onto the Tadarus system. Moreover, the university provides online quizzes and live virtual lectures, available anywhere and anytime.

This study will coordinate with the Deanship of E-learning and Distance Education, at Al-Imam Muhammad Ibn Saud Islamic University, to extract the dataset from the LMS. The next section will shed some light on educational data mining and its process to improve the understanding of

learners' behaviors and interactions in learning environments.

3. Educational Data Mining

Learning management systems (LMSs) accumulate large amounts of information, in terms of learners' interactions, and these need a data mining technique to discover real knowledge about these interactions and to interpret them. Data mining is a powerful tool that uses algorithms to discover knowledge from a great amount of data (Wang et al. 2012).

Nowadays, there is increasing interest in analyzing valuable information in educational databases to learn more about student behavior and the factors that have an effect on learners' performance. The use of data mining for educational data is known as educational data mining (EDM) (Pathan et al. 2014).

Educational data mining has been considered by many researchers and is an emerging research area in the education field (Al-Razgan et al. 2013). It is the process of converting data collected from educational system or LMSs into useful information or patterns (Anupama and Vijayalakshmi 2012). EDM is a branch of knowledge discovery from educational data mining algorithms to understand the students' learning behaviors (Romero and Ventura 2010; Mohamad and Tasir 2013).

Recently, higher education institutions (HEI) have sought to improve and manage educational processes to enhance students' performance. To achieve this objective, EDM is being used as a suitable technique to improve understanding of learning processes, and student behavior and performance through identifying, extracting and evaluating the variables related to the learning process (Shravya and Ch 2013; Agrewal et al. 2012) (Khan et al. 2014).

Educational data mining could enable schools to identify students receiving low grades and examine their weaknesses, recommend improvements, and assist HEIs to make better decisions about their educational activities (Kumar and Ramaswami 2010).

Many researchers have utilized EDM to answer research questions (Hani et al. 2013). EDM can answer questions such as which learners are at risk? Which learners are more likely to withdraw from the course? How good was the quality of learners' participation? (Kabra and Bishkar 2011). This study will use EDM to answer the following questions: How does attending live virtual lectures affect the overall performance of distance learners? How does taking online quizzes affect the overall performance of distance learners?

There are various algorithms and techniques in data mining that can operate on large amounts of data in an educational database to discover patterns and relationships

(Pratiwi 2013), for example: classification, clustering, regression or association rules. Through these techniques, new knowledge can be gained.

In fact, the most useful data mining technique in an educational context is classification (Bharadwaj and Pal 2012), and this technique will be used in this study. A classification technique is used to classify each item in the dataset into one of a set of predefined classes (Sukumaran 2013). This technique deals with the behaviors and characteristics of groups of learners already created. In this study, one of the datasets will classify learners based on their behavior in attending virtual lectures, and another dataset will classify learners based on their behavior of taking online quizzes.

The classification technique involves training and testing. In training, the data is analyzed through classification algorithms. In testing data, the accuracy will be estimated based on the classification rules (Padmanaban 2014). There are many different classification techniques, for example, Naïve Bayes classifier, neural networks, decision trees, and K-nearest neighbor (KNN) as reviewed by Kiang (2003). This study will use the Naïve Bayes classifier (NBC) algorithm.

There are a number of free and open source data mining tools, which can be used for research, for example, RapidMiner, R, WEKA, and KNIME. Some researchers have studied the comparison between these data mining tools, for example, Al-Odan and Al-Daraiseh (2015) or Jovic et al. (2014). This study will use RapidMiner, as presented in Gamulin et al. (2014). RapidMiner is an open source, free data mining tool implemented in JAVA and provides a wide range of different data mining algorithms. It has flexible operators for data input and output in different file formats and has more than 100 learning schemes for regression, classification and clustering tasks (Graczyk et al. 2009).

This study will apply educational data mining to discover knowledge and patterns within the educational data. It will use RapidMiner to apply the Naïve Bayes algorithm.

3.1 Edm for Evaluating Students' Academic Performance (Sap)

Researchers are interested in extracting knowledge from education databases for many reasons, but specifically, to evaluate learners' academic performance at universities. Many researchers use final grades to evaluate learners' academic performance. The main representative attribute, or measure, is learners' performance (Kazanidis et al. 2012, Ahmed and Elaraby 2014).

One study by Ghassemi, Taghiyareh and Orooji, use 'grade' as the measure for a student's performance. They use 'final results' to represent the progress of a student in the course's collaborative activities. These authors also focus on common collaborative activities in LMSs: forum

discussion, collaborative wiki and online communication, which they believe reflect learners' skills (Ghassemi et al. 2013). The present study will use final grade as a measure of learners' academic performance, and will group the final grade value into five categories: A, B, C, D and F (Ahmed and Elaraby 2014).

A number of researchers have published work that studied various EDM techniques used to better understand and enhance student performance. Aziz et al. (2015), Borkar and Rajeswari (2013), and Ahmed et al. (2014) applied association rules data mining techniques on learner data to evaluate knowledge about learners' academic performance. Other researchers used clustering techniques to study students' performance (Tiwari et al. 2014 and Ktona et al. 2014).

However, many of the previous studies interested in evaluating and improving learners' academic performance applied a classification technique, where most of the studies used the Naïve Bayes Classifier in the educational field, as summarized in Table 1.

Table 1: Classification techniques studies

<i>Author</i>	<i>DM technique</i>	<i>Summary and purpose</i>	<i>Year</i>
Bharadwaj and Pal	Decision tree, classification technique	This study aimed to study attendance, class test, seminar and assignment marks to extract and evaluate students' performance, then help to identify the dropouts and students needing special attention	2011
Kumar, Bharadwaj and Pal	C4.5, ID3 and CART decision tree, classification technique	This study aimed to predict the performance in final exams at the end of the semester. It identified the weaker students who needed special attention and students who were likely to pass and be promoted to the next level.	2012
Bharadwaj and Pal	Bayes classification	This study aimed to find the factors that have a significant influence on students' performance and identify the students needing special attention to reduce failure and take appropriate action at the right time	2012

Marinal	Decision tree, Naïve Bayes, Naïve Bayes Tree, K-Nearest Neighbor and Bayes network algorithm	This study aimed to compare five classification algorithms for predicting students' grades, particularly for engineering students. It concluded that excellent results can be seen through IBK, decision tree and Bayes network algorithm.	2014
Marvin and Mewati	Naïve Bayes tree classification technique	This study aimed to predict students' performance. It concluded that some attributes had significant influence over students' performance, such as gender and test score attributes.	2014
Pathan et al.	Decision tree, classification technique	This study aimed to develop students' C program skills	2014
Aziz et al.	Naïve Bayes classifier	This study aimed to predict students' performance and concluded that students' family income, hometown, and gender parameters contribute towards students' academic performance	2015
Mashael and Muna	J48 decision tree, classification technique	This study aimed to predict students' final GPA and identify the most important course based on their grade in previous courses	2015

The present study will focus on 'the effect of attending live virtual lectures and taking online quizzes on learners' academic performance' through the Tadarus system, which focuses on the probabilities of each final grade (learners' academic performance) in each class (the number of live virtual lectures attended, and the number of online quizzes taken) to attain this objective. The Naïve Bayes classification technique will be employed in this study to classify the dataset based on these criteria, and will then find the probabilities.

4. Work Methodology Framework

This research is descriptive and quantitative, as this study will collect quantitative data that describes learners' behaviors. The research methodology in this study consists of three phases:

1. Data collection.
2. Data preprocessing.
3. Data mining.

Figure 1 illustrates the research methodology:

- Data will be collected from the Tadarus database.
- This data will be exported into a Microsoft Excel sheet.
- The data will be processed in preparation for mining techniques. This process consists of 3 steps:
 1. Data selection
 2. Data cleaning
 3. Data normalization

A data mining technique will be applied to discover the effect of attending live virtual lectures (AVL) and taking online quizzes (TOQ) on distance learners' overall performance.

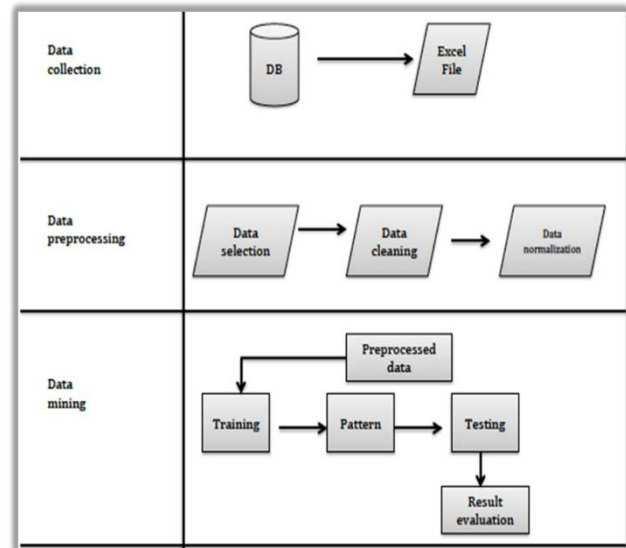


Figure 1 Work Methodology.

4.1 Data Collection

Data will be collected from learners registered on the preparatory year at Al-Imam Muhammad Ibn Saud Islamic University. The dataset specifically relates to distance learners in the course 'Introduction to Computer Science', which was offered in the second semester of the academic year 2014/2015.

- The data will be exported into a Microsoft Excel sheet, containing the following information:
- Student gender
- Number of taken online quizzes (TOQ)
- Number of attendances at live virtual lecture (AVL)
- Final exam mark (final grade).

4.2 Data Processing

The study will apply preprocessing to the dataset, to prepare it for data mining techniques. This stage will improve input data quality for the mining and consists of three steps: data selection, data cleaning and data normalization.

In the data selection process, four parameters will be selected for the mining process: student gender, number of online quizzes, number of live virtual lectures and final exam mark (final grade). These parameters will be extracted from the Tadarus system database. All selected parameters will be saved in an Excel file.

Next, the data cleaning process is as follows:

- All student records without a final exam mark (i.e. final grade) will be removed.
- The number 0 will replace all blank cells (missing values) in the 'attending virtual lectures' and 'taking online quizzes' attributes. The missing value means that the student never attended virtual lectures, and never took online quizzes, respectively.

The next step is the data normalization process, where the numerical attribute in 'final grade', will be transformed to a categorical attribute, and will be classified into one of five symbols: A, B, C, D and F.

Now, each record will have four attributes:

- Attendance at virtual lectures - AVL (nominal attribute)
- Taken online quizzes - TOQ (nominal attribute)
- Student gender (categorical attribute)
- Exam mark - final grade (categorical attribute)

Table 2 demonstrates the dataset attributes:

Table 2: Data Attributes

<i>Parameters type</i>	<i>Parameters</i>	<i>Description</i>	<i>Value</i>
Independent parameters	Gender	Student gender	(Male, Female)
	AVL	Attending virtual lectures in Tadarus system	1 = one time, 2 = two times, 3 = three times, 4 = four times, 5 = five times, ... Etc.
	TOQ	Taking online quizzes in Tadarus system	1 = one time, 2 = two times, 3 = three times, 4 = four times, 5 = five times, ... Etc.
Dependent parameters	Final grade	Student final grade in final exam	A = excellent, B = very good, C = good, D = acceptable, F = failed

4.3 Implementation of a Data Mining Model

As mentioned earlier, one of the research objectives is to investigate the effect of attending live virtual lectures and taking online quizzes on distance learners' overall performance. Thus to achieve this objective, a data mining technique will be applied.

4.3.1 Classification Technique and Naïve Bayes Classifier

The Naïve Bayes classification technique will be applied because classification is often used in educational data mining (Aziz et al. 2015). The Naïve Bayes classification helps to classify learners based on their attendance of virtual lectures (i.e. number of times) and frequency in taking online quizzes (i.e. number of times).

Naïve Bayes works with categorical attributes (i.e. A, B, C, etc.) not with numerical attributes. Therefore, we must convert numerical variables to categorical variables before applying the Naïve Bayes classifier. For that reason, the final grades were converted from numerical to categorical variables using the categorical values 'A', 'B', 'C', 'D', and 'F'. To apply the Naïve Bayes algorithm, the study will also use the following nominal attributes:

- Number of attendances at virtual lectures (AVL)
- Number of online quizzes taken (TOQ)

The Naïve Bayes classification algorithm will calculate the probability of the final exam mark (final grade = A or B or C or D, etc.) in each class (i.e. from the number of times which the student attended an online lecture, or took an online quiz).

For instance: $X = A, B, C, D, \text{ or } F$ and $C = 1, 2, 3, 4, 5, 6$

X is a set of events C is a set of classes

To compute the probability that a student record will experience event $X = A$ (student gets an A), IF s/he attends a lecture just once (i.e. $C = 1$), we use the Bayes theorem. Consider the illustration in the following formula, as mentioned in Islam et al. (2007).

$$P(C|X) = P(C)P(X|C)/P(X)$$

A conditional probability of event X , given class C , means the possibility of event X occurring when class C occurs.

4.4 Data Mining Process In Rapidminer Tool

As mentioned before, this study will use RapidMiner to apply the Naïve Bayes classification algorithm because it has an easy graphical user interface, which allows users to drag and drop operators in an easy way to create data analysis workflow (Kitcharoen et al. 2013).

4.4.1. Classification Technique and Naïve Bayes Classifier

The data mining process in RapidMiner conducted in this study has many steps, as shown in Figure 2.

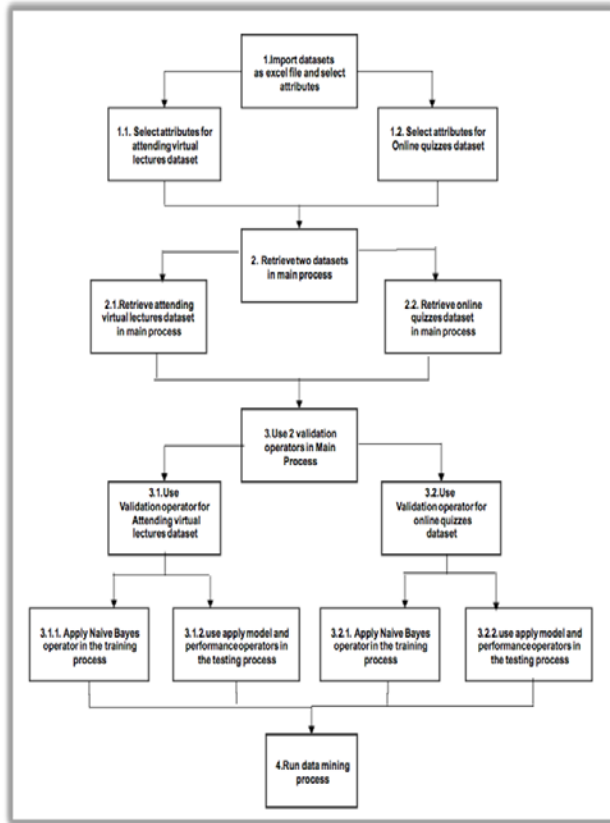


Figure 2: Data mining process

1. The dataset will be imported as an Excel file showing attendance at virtual lectures and online quizzes taken.
2. Retrieve operators will be used to:
 - a. Load the AVL dataset.
 - b. Load the TOQ dataset.
3. Two validation operators will be used for all datasets in the main process. These are mainly used to estimate model accuracy. Validation is a nested operation, which has two sub processes:
 - a. Training sub-process
 - b. Testing sub-process

Each validation operator has an input port (tra);

- a. The (tra) port accepts the dataset (training dataset) for the training model.
- b. The same dataset will be used in the testing sub-process for the testing model.

Both validation operators have three output ports:

- a. (mod) To deliver the Naïve Bayes model

- b. (tra) To pass the exampleset without change
- c. (ave) To return the performance vector

The Naïve Bayes operator will be used in the training model, Naïve Bayes operator has (tra) input port, which accepts the dataset and is used for input into the Naïve Bayes operator.

It has two output ports:

- a. (mod), from where the Naïve Bayes model will be delivered
- b. (exa), where the exampleset will be passed without changes. This port is usually reuses the same exampleset in further operators.

The training model will be applied in the testing sub-process. The performance of the model will be measured in the testing sub-process.

5. Results and Discussion

This study collected a dataset from 23,863 distance learners (female and male). Then, it applied preprocessing on the data set. The data set initially had 23,863 learner records, which became 22,983 records after preprocessing. This new figure (22983) was used for data mining; 880 student records were disqualified (removed) due to missing values in the final grades.

This study used the RapidMiner tool and applied the Naïve Bayes classifier algorithm for data mining on the large amount of educational data, as mentioned above. Table 3 shows the learners' results after data mining: 9% got an A grade, 18% got a B grade, 14% got a C grade, 33% got a D grade and 25% got an F grade.

Table 3 : Learners grade results

<i>Index</i>	<i>Nominal value</i>	<i>Absolute count</i>	<i>Fraction</i>
1	D	7674	0.334
2	F	5813	0.253
3	B	4118	0.179
4	C	3276	0.143
5	A	2102	0.091

5.1 Attending Virtual Lectures: Results

As introduced earlier, the first research question was: How does attending live virtual lectures affect the overall performance of distance learners? This study applied data mining (the Naïve Bayes classification technique) to answer this question. The dataset is classified into 12 classes; Starting from 'Class 0' to 'Class 11', as shown in Figure 3.

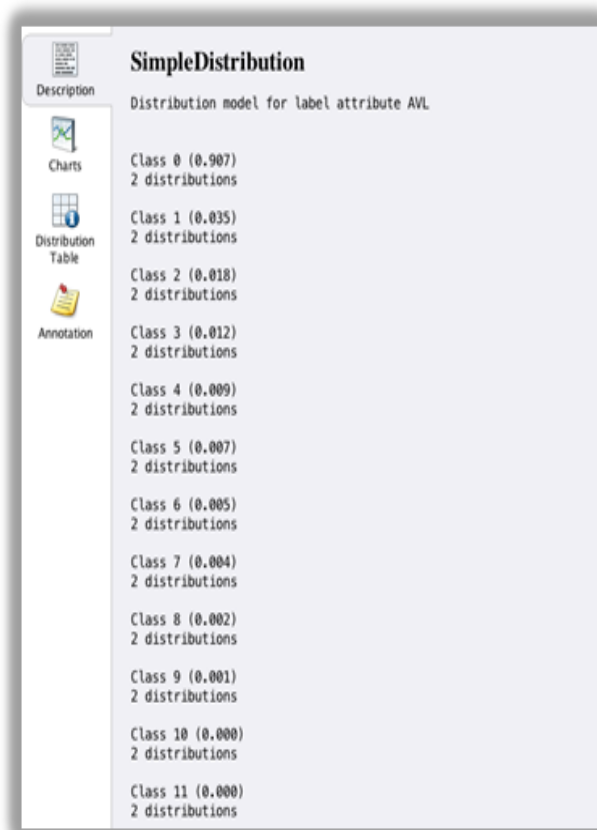


Figure 3: Distribution Model for Label Attribute AVL

This study found that:

- The minimum AVL is 0 (distance learners never attended a virtual lecture).
- The maximum AVL is 11 (distance learner attended 11 virtual lectures).

Table 4 illustrates that 91% of learners NEVER attended any lectures, 3% of learners attended virtual lectures once, 2% of learners attended the lectures twice, 1% attended lectures three times, 0.9% attended four times, 0.7% of learners attended five times, 0.5% of learners attended lectures six times, 0.4% attended the lectures seven times, 0.2% of learners attended eight times, 0.1% attended lectures nine times, 0.0% attended lectures ten times and 0.0% of learners attended lectures 11 times.

Table 4 Learners grade results.

#Of attending lectures times	#Of Distance learners	% Distance learners	Gender
0 (never)	20,846 students	91%	72% male, 28%

			female
1 time	805 students	3%	17% male, 83% female
2 times	416 students	2%	15% male, 85% female
3 times	286 students	1%	11% male, 89% female
4 times	196 students	0.9 %	6% male, 94% female
5 times	156 students	0.7 %	3% male, 97% female
6 times	119 students	0.5 %	4% male, 96% female
7 times	89 students	0.4 %	0% male, 100% female
8 times	42 students	0.2%	5% male, 95% female
9 times	19 students	0.1%	0% male, 100% female
10 times	8 students	0.0%	0% male, 100% female
11 times	1 student	0.0%	0% male, 100% female

5.1.1 The Impact of Attending Virtual Lectures on Learners' Performance

As shown in Table 5, the results show the probabilities of the final grades in each class (according to the number of times of attending virtual lectures). The table also illustrates the gender probability of learners attending virtual lectures. The accuracy was 90.69%, and the classification error was 9.31%.

In more detail, we can say that:

DISTANCE LEARNERS WHO NEVER ATTENDED ANY VIRTUAL LECTURES. Ninety-one percent of distance learners did not attend any virtual lectures (totalling 20,846 learners), 72% of these were male and 28% female. Of these learners, 8% got grade A, 17% got grade B, 14% got grade C, 34% got grade D and 27% got grade F, as shown in Table 6.

Table 5 The Impact of AVL on Learners' Performance.

Attribute	Parameter	0	1	2	3	4	5	6	7	8	9	10	11
GPA	Value=F	0.270	0.098	0.089	0.094	0.082	0.064	0.042	0.090	0.000	0.000	0.000	0.000
GPA	Value=D	0.339	0.313	0.296	0.311	0.218	0.218	0.218	0.202	0.119	0.158	0.250	0.000
GPA	Value=C	0.139	0.181	0.173	0.210	0.143	0.212	0.176	0.124	0.024	0.053	0.250	0.000
GPA	Value=B	0.171	0.261	0.272	0.224	0.306	0.244	0.303	0.225	0.405	0.158	0.125	0.000
GPA	Value=A	0.081	0.147	0.171	0.161	0.189	0.263	0.261	0.360	0.452	0.632	0.375	1.000
GPA	Value=unknown	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Gender	Value=Male	0.717	0.163	0.147	0.108	0.056	0.026	0.042	0.000	0.048	0.000	0.000	0.000
Gender	Value=Female	0.283	0.833	0.853	0.892	0.944	0.974	0.958	1.000	0.952	1.000	1.000	1.000
Gender	Value=unknown	0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 6 Distance Learners: never AVL.

GPA	A	B	C	D	F
%	8%	17%	14%	34%	27%

Distance learners who attended lectures only once. There were 805 distance learners (3%) who attended a virtual lecture on one occasion: 17% were male and 83% female. 15% of learners got grade A, 26% got grade B, 18% got grade C, 31% got grade D and 10% got grade F, as shown in Table 7.

Table 7 Distance Learners: AVL once.

GPA	A	B	C	D	F
%	17%	27%	17%	30%	9%

Distance learners who attended lectures twice. Two percent of distance learners attended virtual lectures twice (416 learners), 15% were male and 85% female. 17% of learners got grade A, 27% got grade B, 17% got grade C, 30% got grade D and 9% got grade F, as shown in Table 8.

Table 8 Distance Learners: AVL two times.

GPA	A	B	C	D	F
%	17%	27%	17%	30%	9%

Distance learners who attended lectures three times. One percent of distance learners attended virtual lectures on three occasions (286 learners), 11% were male and 89% female; 16% of learners got grade A, 22% got grade B,

21% got grade C, 31% got grade D and 9% got grade F, as shown in Table 9.

Table 9 Distance Learners: AVL three times.

GPA	A	B	C	D	F
%	16%	22%	21%	31%	9%

Distance learners who attended lectures four times. There were 196 distance learners (0.9%) that attended virtual lectures four times, 6% were male and 94% female; 19% of learners got grade A, 31% got grade B, 14% got grade C, 28% got grade D and 8% got grade F, as shown in Table 10.

Table 10 Distance Learners: AVL four times.

GPA	A	B	C	D	F
%	19%	31%	14%	28%	8%

Distance learners who attended lectures five times. There were 156 distance learners (0.7%) who attended virtual lectures five times, 3% were male and 97% female; 26% of learners got grade A, 24% got grade B, 21% got grade C, 22% got grade D and 6% got grade F, as shown in Table 11.

Table 11 Distance Learners: AVL five times.

GPA	A	B	C	D	F
%	26%	24%	21%	22%	6%

Distance Learners who Attended Lectures six Times. Of the 0.5% of distance learners who attended virtual lectures six times (119 learners), 4% were male and 96% were female; 26% of learners got grade A, 30% got grade B, 18% got grade C, 22% got grade D and 4% got grade F, as shown in Table 12.

Table 12 Distance Learners: AVL six times.

GPA	A	B	C	D	F
%	26%	30%	18%	22%	4%

Distance learners who attended lectures seven times. Of the 0.4% of distance learners that attended virtual lectures seven times (89 learners), 0% were male and 100% were female; 36% of learners got grade A, 23% got grade B, 12% got grade C, 20% got grade D and 9% got grade F, as shown in Table 13.

Table 13 Distance Learners: AVL seven times.

GPA	A	B	C	D	F
%	36%	23%	12%	20%	9%

Distance learners who attended lectures eight times. Of the 0.2% of distance learners that attended virtual lectures eight times (42 learners), 5% were male and 95% were female; 45% of learners got grade A, 41% got grade B, 2% got grade C, 12% got grade D and 0% got grade F, as shown in Table 14.

Table 14 Distance Learners: AVL eight times.

GPA	A	B	C	D	F
%	45%	41%	2%	12%	0%

Distance learners who attended lectures nine times. Only 19 distance learners (0.1%) attended virtual lectures nine times, 0% were male and 100% were female; 63% of learners got grade A, 16% got grade B, 5% got grade C, 16% got grade D and 0% got grade F, as shown in Table 15.

Table 15 Distance Learners: AVL nine times.

GPA	A	B	C	D	F
%	63%	16%	5%	16%	0%

Distance learners who attended lectures ten times. Eight distance learners (0.0%) attended virtual lectures ten times. Of these, 0% were male and 100% female; 37% of learners got grade A, 13% got grade B, 25% got grade C, 25% got grade D and 0% got grade F, as shown in Table 16.

Table 16 Distance Learners: AVL ten times.

GPA	A	B	C	D	F
%	37%	13%	25%	25%	0%

Distance learners who attended lectures eleven times. Only one distance learner (0.0%) attended virtual lectures eleven times, 0% were male and 100% were female; 100% of learners got grade A, 0% got grade B, 0% got grade C, 0% got grade D and 0% got grade F, as shown in Table 17.

Table 17 Distance Learners: AVL eleven times.

GPA	A	B	C	D	F
%	100%	0%	0%	0%	0%

The study results show that a large percentage (91%) of students did NOT attend virtual lectures, thus the results demonstrate that there is a problem with distance learner attendance of virtual lectures. This result answered the first question: whereby, the results proved that as the level of attendance of virtual lectures increased, the failure rate decreased. Thus, there is a significant positive correlation between attending virtual lectures (AVL) and academic performance (final grade). The relationship is directly proportional.

5.2 TAKING ONLINE QUIZZES: RESULTS

The second question considered the following: 'How does taking online quizzes affect distance learners' overall performance?' This study applied data mining (Naïve Bayes classification technique) again, on the second dataset of 'taking online quizzes' to answer the second question. The dataset is classified into five classes: From Class 0 to Class 4. Where '0', '1', '2', '3', and '4' relate to the number of times that a learner takes an online quiz, as shown in Figure 4.

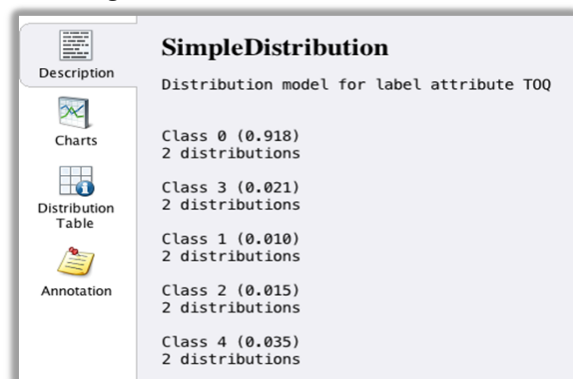


Figure 4: Distribution model for label attribute TOQ

The results show that the minimum number of times for taking an online quiz was 0. This means that there were learners who never took an online quiz. The maximum number of times taking an online quiz is four. This means that there were learners who took all four quizzes.

Table 18 describes the results: 92% of distance learners never took online quizzes, 1% of distance learners took online quizzes once, 2% of distance learners took online quizzes twice, 2% of distance learners took online quizzes three times and 3% of distance learners took online quizzes four times.

Table 18 TOQ Times Results.

#Of Taking online quizzes times	#Of Distance learners	% Distance learners	Gender
0(never)	21102 students	92%	71% male, 29% female
1 time	236 students	1%	23% male, 77% female
2 times	345 students	2%	23% male, 77% female
3 times	490 students	2%	5% male, 95% female
4 times	810 students	3%	0% male, 100 % female

5.2.1 The Impact of Taking Online Quizzes on Students' Performance

Table 19 illustrates the results of data mining, with regard to the impact of taking online quizzes on students' final grades. It demonstrates:

- The probabilities of the final grades, in each class (number of times that the student took an online quiz).
- The gender probabilities of students who took online quizzes.

The accuracy was 91.82%, and the classification error was 8.18%.

Table 19 The Impact of TOQ on Learners' Performance.

Attribute	Parameter	0	3	1	2	4
GP A	Value=F	0.269	0.076	0.076	0.078	0.072
GP A	Value=D	0.339	0.288	0.314	0.287	0.247
GP A	Value=C	0.140	0.182	0.195	0.200	0.152
GP A	Value=B	0.171	0.282	0.275	0.270	0.274

GP A	Value=A	0.082	0.173	0.140	0.165	0.256
GP A	Value=unknown	0	0.000	0.000	0.000	0.000
Gender	Value=Male	0.712	0.051	0.229	0.229	0.000
Gender	Value=Female	0.288	0.949	0.771	0.771	1.000
Gender	Value=unknown	0	0.000	0.000	0.000	0.000

In more detail:

Distance learners who never answered an online quiz (Class 0). Of the 92% of distance learners that DID NOT TAKE ANY online quizzes (21,102 learners), 71% were male and 29% were female; 8% of learners got grade A, 17% got grade B, 14% got grade C, 34% got grade D and 27% got grade F, as shown in Table 20.

Table 20 Distance Learners: never TOQ.

GPA	A	B	C	D	F
%	8%	17%	14%	34%	27%

Distance learners who Answered One Online Quiz (Class 1). There were 236 distance learners (1%) that took online quizzes on one instance; 23% of these were male and 77% were female, 14% of learners got grade A, 27% got grade B, 19% got grade C, 31% got grade D and 8% got grade F, as shown in Table 21.

Table 21 Distance Learners: TOQ once.

GPA	A	B	C	D	F
%	14%	27%	19%	31%	8%

Distance learners who Answered two Online Quiz (Class 2). Of the 345 distance learners (2%) that took online quizzes twice, 23% were male and 77% were female; 16% of learners got grade A, 27% got grade B, 20% got grade C, 29% got grade D and 8% got grade F, as shown in Table 22.

Table 22 Distance Learners: TOQ twice.

GPA	A	B	C	D	F
%	16%	27%	20%	29%	8%

Distance learners who Answered three Online Quiz (Class 3). There were 2% of distance learners that took online quizzes three times (490 learners), 5% were male and 95% were female; 17% of learners got grade A, 28% got grade B, 18% got grade C, 29% got grade D and 8% got grade F, as shown in Table 23.

Table 23 Distance Learners: TOQ three times.

GPA	A	B	C	D	F
%	17%	28%	18%	29%	8%

Distance learners who Answered four Online Quiz (Class 4). Of the 3% of distance learners that took online quizzes four times (810 learners), 0% were male and 100% were female; 26% of learners got grade A, 27% got grade B, 15% got grade C, 25% got grade D and 7% got grade F, as shown in Table 24.

Table 24 Distance Learners: TOQ four times.

GPA	A	B	C	D	F
%	26%	27%	15%	25%	7%

The study results show that a large percentage (92%) of distance learners did NOT take any online quizzes. The results also answer the second research question: where the results proved that as the number of online quizzes taken increased, the failure rate decreased. The study concludes that there is a significant positive correlation between taking online quizzes and academic performance (final grade). The relationship is directly proportional.

5. Conclusion

Al-Imam Muhammad Ibn Saud Islamic University aims to provide high-quality e-learning experiences, especially to its distance learners. To improve the educational process, and enhance learners' academic performance, the university offers virtual lectures and online quizzes through its Tadarus learning management system.

This study has shown that there are a high percentage of distance learners who do not attend live virtual lectures and don't take online quizzes. This study focused on how attending live virtual lectures and taking online quizzes affects distance learners' overall performance. By applying data mining techniques on learners' records, the study discovered the effect of attending virtual lectures and taking online quizzes on learners' final results. Through visualization tools and questionnaires, it may be possible to change learner attitudes towards attending live virtual lectures and taking online quizzes.

Data mining results revealed a significant positive correlation between attendance at live virtual lectures and learners' academic performance (final grade). This means that when instances of attendance at live virtual lectures increased, failure rates decreased. Data mining results also revealed a significant positive correlation between the frequency of taking online quizzes and learners' academic performance (final grade). This means that when the frequency of taking online quizzes increased, the failure rates decreased.

For the student, the study's results gave an insight into the importance of attending live virtual lectures and taking online quizzes. Moreover, the results could be used to convince learners that attending virtual lectures and taking online quizzes improves their final mark.

Future researchers could use the chosen methodology for future research in various ways. Here are some suggestions:

- Researchers could use the data mining technique at different universities in different countries, to discover the factors that have an effect on learners' performance, and then comparisons could be drawn between them.
- Researchers could also apply the data mining technique on other attributes in the Tadarus system, and could examine gender as the available attribute in this study and discover how gender will affect learners' performance.
- Researchers could apply different data mining techniques to answer the same research questions.
- Researchers could use the visualization tool in this study for different purposes, as a persuasive technique.

References

- [1] ABDULAZIZ, A. 2012. An Exploration of Distance Learning in Saudi Arabian Universities: Current Practices and Future Possibilities. *International Journal of Instructional Technology and Distance Learning* 9, pp. 21-28.
- [2] AGREWAL, S., PANDEY, G. and TIWARI, M. 2012. Data Mining in Education: Data Classification and Decision Tree Approach. *International Journal of e-Education, e-Business, e-Management and e-Learning*, 2, pp. 140-144.
- [3] AHMED, A.B.E.D and ELARABY, I.S. 2014. Data Mining: A prediction for Student's Performance Using Classification Method *World Journal of Computer Application and Technology*, 2, pp. 43-47.
- [4] AHMED, S., PAUL, R. and MD LATIFUL HOQUE, A.S. 2014. Knowledge Discovery from Academic Data using Association Rule Mining In *Proceedings of the Computer and Information Technology (ICCIT)*, 2014 17th International Conference, Dhaka 2014 IEEE, pp. 314 - 319.
- [5] AL-ODAN, H.A. and AL-DARASEH, A.A. 2015. Open Source Data Mining Tools. In *Proceedings of the 1st International Conference on Electrical and Information Technologies ICEIT*, Marrakech 2015 IEEE, pp. 369 - 374.
- [6] AL-RAZGAN, M., AL-KHALIFA, A.S. and AL-KHALIFA, H. 2013. Educational Data Mining: A Systematic Review of the Published Literature 2006-2013. In *Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng 2013)*, 285, pp. 711-719..
- [7] ALSHATHRI, S.M. and MALE, T. 2015. Students and Instructors Perceptions of Blended Learning in the First Electronic University in the Arab World (Saudi Electronic University) In *Proceedings of the British Education Studies Association* 2015, pp. 1-23.

- [8] ANUPAMA K. S. and VIJAYALAKSHMI, M.N. 2012. Mining of student academic evaluation records in higher education In Proceedings of the 2012 International Conference on Recent Advances in Computing and Software Systems Chennai 2012 IEEE, pp. 67–70.
- [9] AZIZ, A. A., ISMAIL, N.H., AHMAD, F. and HASAN, H. 2015. A Framework for Students' Academic Performance Analysis using Naïve Bayes Classifier In Proceedings of the ICIDM 2015.
- [10] BEDNAR, S., HUSAR, J., HRICOVA, R., LIPTAKOVA, A. and MARTON, D. 2013. Adoption of ILIAS Web Learning System for distance education. In Proceedings of the Emerging eLearning Technologies and Applications (ICETA), 2013 IEEE 11th International Conference, Stara Lesna 2013 IEEE, pp. 139-144.
- [11] BHARADWAJ, B. and PAL, S. 2012. Data Mining: A prediction for performance improvement using classification. IJCSIS 9, pp. 136-140.
- [12] BORKAR, S. and RAJESWARI K., R. 2013. Predicting Students Academic Performance Using Education Data Mining. IJCSMC 2, pp. 273 - 279.
- [13] FERDIANA R. 2013. Distance Learning Lifecycle Management An Agile Approach to Manage the Lifecycle of a Distance Learning. In Proceedings of the ICITEE, Yogyakarta 2013 IEEE, pp. 106 - 112.
- [14] GAMULIN, J., GAMULIN, O. and KERMEK, D. 2014. Comparing classification models in the final exam performance prediction In Proceedings of the 7th International Convention on Information and Communication Technology, Electronic and Microelectronics (MIPRO), 2014, pp. 663-668.
- [15] GHASSEMI, S., TAGHIYAREH, F. and OROOJI, F. 2013. Developing A Learner-Centered Criteria to Assessing Group Activities in Blended Learning Environments In Proceedings of the 2013 5th Conference on Information and Knowledge Technology (IKT), Shiraz 2013 IEEE, pp.150 - 154.
- [16] GRACZYK, M., LASOTA, T. and TRAWIŃSKI, B. 2009. Comparative Analysis of Premises Valuation Models Using KEEL, RapidMiner, and WEKA. In Computational Collective Intelligence. Semantic Web, Social Networks and Multiagent Systems Springer Berlin Heidelberg, pp 800-812.
- [17] HANI, H., HOOSMAND., S. and MIRAFZAL, H. 2013. Identifying the Factors Affecting the Success and Failure of E-learning Students Using Cluster Analysis In Proceedings of the 7th international conference on e-commerce in developing countries: with a focus on e-security (ECDC), 2013, Kish Island 2013 IEEE, pp. 1 - 12.
- [18] HARASIM, L. 2000. Online education as a new paradigm in learning. Elsevier 3, pp. 41-61.
- [19] ISLAM, M.J., WU, Q.M.J., AHMADI, M. and SID-AHMED, M.A. 2007. Investigating the Performance of Naïve-Bayes Classifiers and K- Nearest Neighbor Classifiers In Proceedings of the International Conference on Convergence Information Technology, Gyeongju, 2007 IEEE, pp. 1541 - 1546.
- [20] JOVIC, A., BRKIC, K. and BOGUNOVIC, N. 2014. An overview of free software tools for general data mining. In Proceedings of the 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija 2014 IEEE, pp. 1112 - 1117.
- [21] KABRA R.R. and BISHKAR R.S. 2011. Performance Prediction of Engineering Students using Decision Trees International Journal of Computer Applications, 36, 8-12.
- [22] KAZANIDIS, I., VALSAMIDIS, S. KONTOGIANNIS, S. and KARAKOS, A. 2012. Measuring and mining LMS data. In Proceedings of the 16th Panhellenic Conference on Informatics, Piraeus, 2012 IEEE, pp. 296 - 301.
- [23] KHAN, A.M., GHARIBI, W. and PRADHAN K.S. 2014. Data Mining Techniques for Business Intelligence in Educational System: A Case Mining. In Proceedings of World Congress on Computer Applications and Information Systems, WCCAIS 2014, pp. 94-99.
- [24] KIANG, M.Y. 2003. A comparative assessment of classification methods. Elsevier 35, pp. 441–454.
- [25] KITCHAROEN, N., KAMOLSANTISUK, S., ANGSOOMBOON, R., and ACHALAKUL, T. 2013. RapidMiner Framework for Manufacturing Data Analysis on the Cloud. In Proceedings of the 10th International Joint Conference on Computer Science and Software Engineering (JCSSE), IEEE, pp. 149 - 154.
- [26] KTONA, A., XHAJA, D. and NINKA, I. 2014. Extracting Relationships Between Students' Academic Performance and Their Area of Interest Using Data Mining Techniques. In Proceedings of the IEEE 2014.
- [27] KUMAR, S. P. and RAMASWAMI, K.S. 2010. Fuzzy K-Means Cluster Validation for Institutional Quality Assessment. In Proceedings of the Communication and Computational Intelligence (INCOCCI), 2010 International Conference, Dec 2010, pp. 628 - 635.
- [28] LARREAMENDY-JOERNS, J. and LEINHARDT, G. 2006. Going the distance with online education Review of Educational Research 76, pp. 567-605.
- [29] LEPOURAS, G., KATIFORI, A., VASSILAKIS, C., ANTONIOU, A. and PLATIS, N. 2014. Towards a Learning Analytics Platform for Supporting the Educational Process In Proceedings of the The 5th International Conference on Information, Intelligence, Systems and Applications, Chania 2014 IEEE, pp. 246 - 251.
- [30] MOHAMAD, K. and TASIR, Z. 2013. Educational data mining: A review. In Proceedings of the The 9th International Conference on Cognitive Science 2013 Elsevier, pp. 320-324.
- [31] PADMANABAN, H. 2014. Comparative Analysis of Naive Bayes and Tree Augmented Naive Bayes Models In Department of Computer Science San José State University, 65.
- [32] PATHAN A. A, HASAN, M., AHMED, M.F. and FARID, M.D. 2014. Educational Data Mining: A Mining Model for Developing Students' Programming Skills. In Proceedings of the Software, Knowledge, Information Management and Applications (SKIMA), 2014 8th International Conference, Dhaka 2014 IEEE, pp. 1 - 5.
- [33] PRATIWI, N. 2013. Predicting Student Placement Class using Data Mining. In Proceedings of the Teaching, Assessment and Learning for Engineering (TALE), 2013 IEEE International Conference, Bali 2013 IEEE, pp. 618 - 621.
- [34] ROMERO, C. and VENTURA, S. 2010. Educational Data Mining: A Review of the State-of-the-Art. In Proceedings of

- the Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions 2010 IEEE, pp. 601 - 618.
- [35] SHRAVYA, S.G. AND CH, R. 2013. Data Mining: Performance Improvement In Education Sector Using Classification And Clustering Algorithm. IJIRD 2, pp. 101-106.
 - [36] SIEMENS, G. AND GASEVIC, D. 2012. Guest Editorial - Learning and Knowledge Analytics. Journal of Educational Technology & Society, 15, pp. 1-2.
 - [37] SUKUMARAN, S. 2013. A study on classification techniques in data mining. In Proceedings of the 4th International Conference on Computing, Communications and Networking Technologies, Tiruchengode 2013 IEEE, pp. 1 - 7.
 - [38] TIWARI, M., SINGH, R. and VIMAL, N. 2014. An Empirical Study of Applications of Data Mining Techniques for Predicting Student Performance in Higher Education. International Journal of Computer Science and Mobile Computing, 2, pp. 53 – 57.
 - [39] WANG, J., Lu, Z., WU, W. and LI, Y. 2012. The Application of Data Mining Technology Based on Teaching Information. In Proceedings of the The 7th International Conference on Computer Science & Education, Melbourne, Australia 2012 IEEE, 652-657.

Ashwaq Almusharraf hold a Bachelors degree in Information Systems and a Masters degree in the same field graduated from Al Imam Mohammad Ibn Saud Islamic University (IMSIU). Currently she is working as a technical support engineer at Naseej Company.

Mona Alkhatabi - graduated the PhD studies in the School of Computing, Informatics and Media of the University of Bradford in 2010. She joined the Department of Information System, College of Computer and Information Sciences as Assistant Professor in the same year. Dr. Mona is currently the Vice-Dean of E-learning and Distance Education Deanship at Al-Imam Mohammad Ibn Saud Islamic University (IMSIU), Saudi Arabia. She has over 5 years educational experience in the field of Mathematics, with a bachelor degree from Education College in Riyadh, Saudi Arabia. After she was awarded an MSc in Computing with Distinction from the University of Bradford in 2006, she started her research in the quality of e-learning content. She has number of published research in her specialization field. Also, she works as a reviewer for a number of scientific publishing houses. For more details: www.malkhatabi.com