International Journal of Engineering Education Vol. 33, No. 5, pp. 1598–1617, 2017
Printed in Great Britain
0949-149X/91 $3.00+0.00

Development of an Interactive Game-Based Learning Environment to Teach Data Mining*

MEHMET CENGIZ[1], KOKTEN ULAS BIRANT[2], PELIN YILDIRIM[3] and DERYA BIRANT[2]
[1] Dokuz Eylul University, The Graduate School of Natural and Applied Sciences, Izmir, Turkey.
[2] Dokuz Eylul University, Department of Computer Engineering, Izmir, Turkey.
[3] Manisa Celal Bayar University, Department of Software Engineering, Manisa, Turkey.
E-mail: {mehmetcengiz, ulas, pelin, derya}@cs.deu.edu.tr

Game-based learning has become a popular topic in all levels of education. A number of computer games have been developed to teach different subjects such as mathematics, English language, medicine, and music. This paper presents the first study that proposes the development of edutainment games to teach data mining techniques with the scope of game-based learning. The aim of this study is to provide an environment that is both fun and enables the achievement of learning goals in data mining training in computer engineering. An escape game called Mine4Escape, which consists of different rooms to teach different data mining techniques (classification and association rule mining), has been developed for individuals at the undergraduate and post-graduate levels. The advantages of the proposed approach are discussed in comparison with traditional data mining training. In addition, this paper describes a dynamic scoring system designed for game-based learning. Finally, an experimental study was carried out to evaluate the performance of our learning environment by analyzing feedback received from a test group consisting of 39 undergraduate and graduate students in computer engineering. The findings from the questionnaire show that it is possible to enhance knowledge acquisition about data mining via the game-based approach. However, the degree of learning interest and information acceptance changes according to students' age, gender, educational level, and game habits.

Keywords: game-based learning; data mining; education; computer engineering

1. Introduction

Recent technological innovations and developments reveal a great demand for data mining to extract valuable knowledge from huge amounts of data. Data mining is increasingly being used in a range of areas, including marketing, health, communication, banking, and sports. Along with this growing interest in data science has come an increasing demand for learning data mining, especially in departments of computer engineering; however, educating trainees who are interested in information technologies in the field of data mining is considered a difficult task. To resolve concerns about this challenge and motivate the trainees, this paper proposes a game-based learning approach as an alternative learning tool.

Game-based learning (GBL) is a kind of educational approach that provides learning through games to make the learning activity more enjoyable and interesting. It plays an important role in the learning process for students at all ages and enhances learners' interest, attitude, and degree of technology acceptance, as well as improving their achievements in problem-solving activities [1]. Game-based learning has a positive outcome on students' motivation, comprehension, and retention of newly taught concepts [2]. Considering this motivation, this paper proposes an interactive game-based learning environment to teach data mining techniques to students in a fun manner.

The main contributions of this study are four-fold. First, a game-based data mining learning (GBDML) approach is proposed and its advantages are listed for the first time. Second, this paper introduces a novel game-based learning environment that has never been implemented before to teach data mining to trainees. Third, it describes a novel scoring mechanism designed for game-based learning, which takes into account several factors such as the degree of task difficulty, learners' background, and knowledge level. Fourth, this is the first study that presents the findings about GBDML by describing the experimental results according to learners' age, gender, educational level, and game habits.

As shown in Figure 1, GBDML is an interdisciplinary area that can be conceptualized as the combination of three main fields: education, games, and data mining. The intersection of these three areas also forms other subareas closely related to GBDML such as game-based learning, game-based data mining, and learning data mining. Among these subareas, the field most related to GBDML is game-based data mining, which is generally looking for new patterns in data through a game strategy.

The remainder of this paper is structured as follows: in the following section, related literature and previous studies are summarized. Section 3 provides relevant background information. Section 4 describes the

* Accepted 20 April 2017.

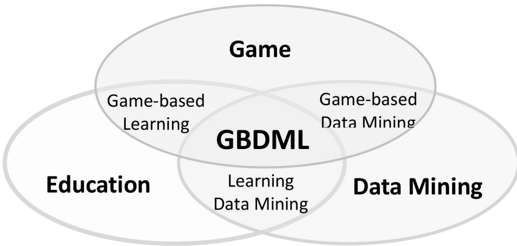


Fig. 1. Main areas related to learning data mining through the game-based approach.

proposed GBDML approach and explains its main benefits. In Section 5, the experimental study is presented and the main findings about players' data mining learning skills are discussed. Section 6 contains some concluding remarks and directions for future study.

2. Related work

Game-based learning is increasingly being used in a range of areas including medicine [3], food [4], electric & electronics [5], civil engineering [6], tourism [7], robotics [8], finance [9], and mechanical engineering [10]. Mortara et al. [11] developed an educational game to help people learn about cultural heritage in an engaging way. Ross et al. [12] have also proposed a game called Space Tug Skirmish to handle complicated sociotechnical systems engineering problems.

Several studies have been conducted on applying the game-based learning approach to the education sector [13]. Browne et al. [14] developed three tablet applications to increase the adult literacy rate and they proved that the applications were effective and successful at increasing learner engagement when tested with participants. Another study by Faghihi et al. [15] proposed a novel system named Math_Dungeon, which aims to help students teach college level mathematical concepts entertainingly by integrating video game elements and artificial intelligence. The researchers discussed the impacts of their proposed system on the students and concluded that the system was more successful than traditional learning activities. Juzeleniene et al. [16] explained the GABALL project that was carried out to improve the language and culture skills of managers at micro, small and medium-sized enterprises using Serious Game.

A number of game-based learning approaches have also been proposed to impart theoretical and applied knowledge in the computer science discipline. Game-based learning studies in computer engineering are listed in Table 1 with their domains and purposes. Existing research has led to many discussions and ideas regarding how best to teach introductory computer programming. Several studies have been designed to help learners to improve computer assembly skills. The study of game-based learning application for the data mining domain, however, is not yet available. This present study is different from the aforementioned studies as it proposes a game developed to teach data mining techniques to computer engineering students in an entertaining way.

Whereas some studies related to the topic aim to demonstrate potential uses, some studies are interested in more systematic assessments of games' potential benefits. Several studies have described the effects of game-based learning on students' learning performances [17, 18]. Cojocariu and Boghian [19] compared traditional and digital game-based learning approaches, considering their advantages and disadvantages. According to comparison results, they suggested game-based learning as an educational type for preschool and elementary school students in particular. Ucus [20] collected elementary school teachers' opinions about game-based learning using semi-structured interviews and obtained views that specified the positive impact of game-based learning on education.

Since the development of the internet and technology, several studies on digital game-based learning [21] have been carried out with the strategy of multiplayer online games despite the fact that [22] multi-player games may generate over-competitiveness, making the users ignore the essence of learning. Thus, the present research aims to develop a single-player game, called Mine4Escape, that targets data mining learning and improves data mining skills.

3. Background

3.1 Data mining

Data mining is the process of finding meaningful and potentially useful knowledge from large-scale data by combining several disciplines such as machine learning, artificial intelligence, database management, high-

Table 1. Game-based learning studies in computer engineering

Authors	Year	Domain	Purpose
Hooshyar et al. [1]	2016	C++ programming language	To improve programming skills of students.
Soflano et al. [23]	2015	Structured Query Language (SQL)	Teaching SQL with a developed game to gain better learning outcomes from 120 higher education students.
Ouahbi et al. [24]	2015	Algorithms and programming	Motivating students to learn algorithms and programming concepts and specifying their programming levels using Scratch environment.
Bourbia et al. [25]	2014	Assembly of computer	Helping learners to improve their computer assembly skills.
Hou and Li [26]	2014	Assembly of computer	Increase learners' knowledge of computer assembly by using educational adventure game.
Rodriguez-Cerezo et al. [27]	2014	Computer language implementation	Helping students design and implement computer languages with language processing exercises.
Corral et al. [28]	2014	Object oriented programming	Teaching object oriented programming with C# programming language through a game-oriented approach.
Seng and Yatim [29]	2014	Object oriented programming	Guiding students to learn object oriented programming through an edutainment game.
Nunohiro et al. [30]	2013	Programming learning support system	Providing programming learning support by programming training with a puzzle-solving interface.
Kazimoglu et al. [31]	2012	Computer programming	Enhancing the problem-solving abilities of students who are learning introductory computer programming.
Schmitz et al. [32]	2011	IT knowledge	Enhancing learners' IT knowledge by developing a game.
Lee and Ko [33]	2011	Computer programming	Enhancing the computer programming skills of 116 self-described novice programmers.
Papastergiou [34]	2009	Computer memory concepts	Teaching computer memory concepts to high school students.
Kuk and Jovanovic [35]	2014	Computer architecture	Supporting efficient learning in computer architecture course.
Minovic et al. [36]	2011	Computer networks	Motivating university students to learn network programming in computer networks course.
Mladenovic et al. [37]	2016	Computer programming	Improving programming skills of university undergraduate novice programmers.

performance computing, and statistics. The main data mining tasks are classification, clustering, and association rule mining (ARM). This paper focuses on the development of an interactive game-based learning environment that enables learners to perform two data mining tasks (classification and ARM) after a data preparation step. In addition, a clustering method was used to cluster feedback received from a test group in the experimental studies.

3.1.1 Data preparation

Data preparation is the step of knowledge discovery in databases (KDD) that makes data suitable for data mining processes. The data preparation stage generally involves four steps:

- *Data Collection and Integration*: Collecting data from disparate sources and combining them into one structure.
- *Data Selection and Reduction*: Making the dataset smaller according to data mining objectives and eliminating redundant attributes and tuples.
- *Data Preprocessing*: Detecting and correcting the inconsistencies and errors, removing noises and outliers, handling missing values, and smoothing data.
- *Data Transformation*: Converting data from one format to a destination format, i.e., data discretization and data normalization.

In the present study, data preparation was designed as a learning objective and the proposed model enables trainees to learn two data preparation steps by practice: data collection by clicking the objects in the room and data preprocessing by filling missing values manually. For example, in the game Mine4Escape, the tuple of the dataset that contains more than half number of attributes with missing values is ignored. Our game model allows users to apply data mining techniques on the dataset obtained after completing data preparation steps.

3.1.2 Classification

Classification is one of the most widely used data mining tasks and constructs a model to categorize objects into pre-defined classes based on their characteristics. Classification algorithms give successful results for many areas such as medical diagnosis, document categorization, marketing, and banking. In the current study, Mine4Escape game is designed to teach Naive Bayes and C4.5 decision tree algorithms for the classification task because these algorithms can be easily applied if the dataset contains a low number of records.

Naive Bayes is a well-known classification algorithm that utilizes Bayes' theorem to evaluate unknown conditional probabilities. As given in Equation (1), the probability model is evaluated by multiplying the probability of each attribute given a specified class and the prior probability of the particular class considering the distribution of the class:

$$l = argmax_{m\in\{1,\dots,M\}} p(C_m)\prod_{i=1}^{n} p(x_i|C_m) \tag{1}$$

where $p(C_m)$ is the prior probability of C_m for each of M classes and $P(x_i|C_m)$ is a conditional probability of x input vector such as $x=(x_1, x_2, \ldots, x_n)$ representing some n features. To specify a class label l of x, maximum likelihood is used.

Another classification algorithm, C4.5, is a supervised learning technique that generates a tree used for categorizing unknown target attribute values. At each node of the tree, the algorithm chooses the attribute that most effectively splits the set of training samples into subsets.

3.1.3 Association rule mining

Association rule mining (ARM) is used to discover interesting relationships among set of items in data. In this study, the Apriori, which is a well-known ARM algorithm, was used both to teach the algorithm depending on the learning objective and to evaluate the performance of our learning environment.

Let $\Gamma = \{i_1, i_2, \ldots, i_m\}$, $\Gamma \neq \emptyset$, be a set of m distinct literals, called items. In our study, an item is an (attribute, value) pair. Let dataset D be a set of n transactions, such that $D = \{t_1, t_2, \ldots, t_n\}$, where each transaction t_i is a set of items such that $t_i \subseteq \Gamma$. A set $I \subseteq \Gamma$ is called an *itemset* and a transaction t_i satisfies I if all the items of I also exist in t_i. In particular, an itemset with k items is called a *k-itemset*. *Support* of an itemset I is denoted by $sup(I)$ and given in Equation (2):

$$sup(I)=|\{t \in D \mid I \subseteq t\}| \,/\, |D| \tag{2}$$

where support of an itemset I is defined as the percentage of transactions t in D containing I. Support is used to measure the strength of an itemset. The itemset I is called *frequent* if its support is greater than some user-defined threshold *minSup*, i.e., if $sup(I) \geq minSup$. An association rule r is a conditional implication of the form r: $X \Rightarrow Y$, where X and Y are itemsets for which $X, Y \subset \Gamma$, $X \neq \emptyset$, $Y \neq \emptyset$ and $X \cap Y = \emptyset$. Support value of the rule r: $X \Rightarrow Y$ is calculated as the number of transactions that contains both X and Y itemsets divided by total number of transactions. Rules are discovered having domain knowledge specified as a minimum support threshold (*minSup*).

3.1.4 Clustering

Clustering is an unsupervised learning method that identifies similar objects and groups them into clusters by using a similarity measure. It helps in better understanding the characteristics of the data because fewer groups are more easily interpreted. In this study, the K-means++ clustering algorithm was used to group the feedback received from users in the experimental studies. The K-means++ algorithm is an improvement version of the standard K-means algorithm with a specific method to choose the initial centroids of the k clusters.

4. Game-Based Data Mining Learning (GBDML)

This is the first study that proposes the development of edutainment games to teach data mining techniques within the scope of game-based learning. To serve this purpose, an interactive GBL environment was selected and developed to teach classification and association rule mining techniques to trainees. This section explains the details about the environment with the advantages and disadvantages of the proposed approach.

4.1 Advantages and disadvantages of GBDML

As a side effect of the technological growth, game-based learning often generates negative connotations because of its close association with video games. Educational games, however, are typically designed with the goal of helping the players to learn something. The GBDML has many advantages over learners and educators. The main advantages of GBDML are improving trainees' learning motivation, attracting trainees to participate in learning, and further promoting their autonomous learning abilities. The following are other advantages of data mining training when delivered through the game-based learning approach.

Skill development
GBDML allows data mining learners to develop cognitive skills (i.e., learn, think, remember, reason, concentrate, solve problems) and physical skills (i.e., hand-eye coordination, muscle skills, reaction agility) simultaneously.

Long term memory
Researchers have found that games have direct, positive effects on learning by enhancing long-term knowledge retention. Repetition strengthens the learner's memory, which helps him or her retain the information for a longer time. Thus, the knowledge and skills acquired through GBDML will be retained longer than information given by traditional learning methods.

Engagement & motivation
GBDML improves learners' data mining skills by providing motivation in a fun way. Entertainment keeps them coming back to learn even more. Prior studies show that the motivation for learning provided by the game-based approach is higher when compared to traditional methods. Because games include rules, definitive objectives, and competition, they deliver an interactive experience that promotes a sense of achievement for the players. Games contain reward systems that entice them to learn more to earn more rewards.

Immediate feedback
GBDML provides instant feedback on data mining learners' performance so they get an idea of their understanding, as well as what else they need to know. Instead of having to wait days or even weeks for an assignment or test grade to be returned, data mining learners get instantaneous results about whether their decisions are correct. Real-time feedback helps learner to know his or her status, so he or she can decide to repeat a level or move on to the next concept. While learners are playing the game to learn data mining, they get their score immediately according to their correct answers in the game. In case of a wrong step, instant feedback notifies learners why they are wrong and what they should have done.

Digital literacy
GBDML provides digital literacy, which is the ability to use information technologies. It requires both cognitive and technical skills and is an important skill for a lifetime of technology use.

Cost effective
The cost of teaching large numbers of people is lower compared to traditional learning methods. If the game is developed for the right target group, GBDML provides an economical advantage for data mining learners.

Learning pace tailored to individuals
Everyone's learning capability is different: some can learn a data mining technique with only one case study and others require through several examples. In the traditional learning approach, most teachers must teach mixed ability groups and may not explain a concept more than once, so students who have difficulty understanding the topic can become frustrated when they cannot keep up with their peers. GBDLM, on the other hand, allows learners to practice newly acquired knowledge repeatedly until they understand it fully.

Although the GBDML approach has many advantages, it has also some disadvantages, as enumerated below.

Waste of time
If the content of an edutainment data mining game is not clearly designed by the educators, learners may waste their time on playing the game. For this reason, educators should prepare the content of the game to be appropriate for the ages and skill levels of specific groups.

Health problems
Using the computer for long period of time can lead to the emergence of certain health problems. Too much gaming can lead to lowering eyesight, back pains and headaches.

4.2 Game design

To demonstrate the applicability of the GBDML approach, an adventure game called Mine4Escape (Mine for Escape) was implemented with the aim of teaching data mining techniques to students. For the implementation of the game, an advanced, free game engine (Unreal Engine) platform was used. Several pedagogical theories and strategies were implemented to design the context and tasks of this game. First, a *problem-solving strategy* was adopted to promote trainees' learning through problem-solving process. Problem-solving has been suggested as a meaningful learning activity [38] that engages learners in a cognitive learning process. Second, a *goal-setting strategy* was applied to design a task that encourages learners to complete it by interacting with objects in the game.

Mine4Escape game has been developed for individuals at the undergraduate and post-graduate levels. It consists of different rooms with objects to teach different data mining techniques (classification and association rule mining). Fig. 2 shows the unified representation of game rooms and their teaching objectives, and Fig. 3 shows screenshots of certain rooms, including the music, colorful, horror and subway room. First, the tutorial room helps trainees to learn detailed information about data mining concepts, techniques, and algorithms. Second, the player will find hidden parts of the dataset by exploring the room and clicking on the objects. To exit one room, the player should first collect data and then solve the data mining riddle. In the problem-solving step, an example is also given to help the player. When the player gathers all dataset pieces and solves the riddle correctly, he or she gets the key to exit the room.

The work presented in this paper contains certain limitations that must be acknowledged. First, the proposed game does not include a competition mechanism that can be one of the design elements of game environments. The creation of competition environment in a serious game can enhance the motivation of the participants. To improve the benefits of the game, a competition mechanism can be developed for the GBDML approach in future research. Another limitation of the game concerns the age of learners, who are expected to be at the undergraduate and post-graduate levels (i.e., aged between 18 and 50). The upper and lower time limits (10 and 100 minutes) used in the scoring mechanism is another limitation of the proposed game environment.

The game presented in this article has been developed both for beginners and students who already have some knowledge about data mining techniques and want to improve their skills in this area. The players are not required to have background knowledge of data mining; beginners' game play is not limited because tutorials, explanations and examples are supported to provide an overview of data mining and its techniques. Trainees with some prior knowledge, however, may enhance their knowledge. In this way, while some students can learn data mining knowledge from scratch, others can benefit from the opportunity to improve their data mining skills. In addition, learners who are interested in information technologies can be adapted to the game better. In this study, we aimed to teach simple data mining algorithms such as Naive Bayes, C4.5 and Apriori through game-based learning; however, Mine4Escape can also be enhanced by including advanced data mining methods such as neural networks and support vector machines.

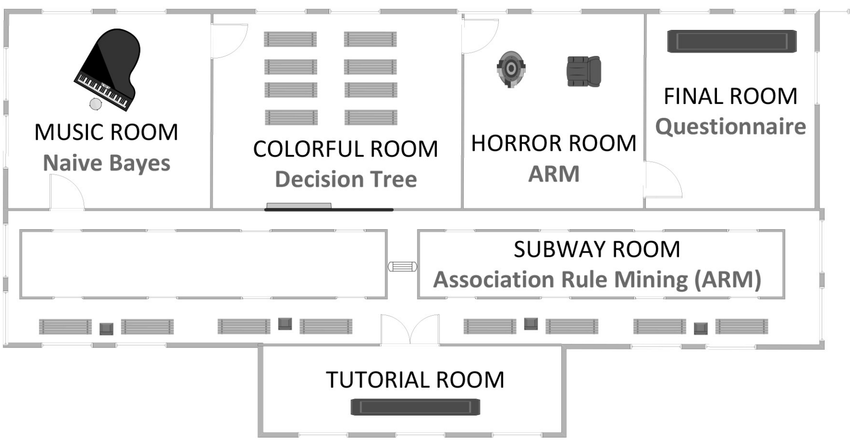
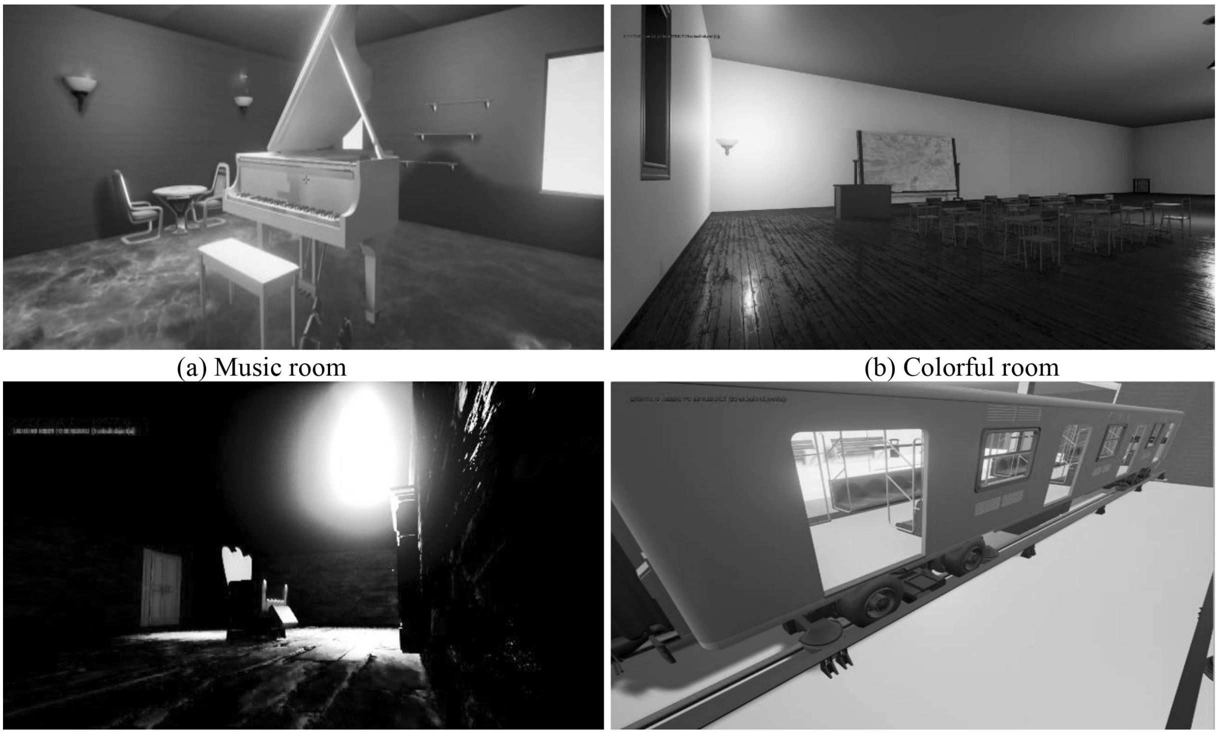


Fig. 2. Unified representation of game rooms and their teaching objectives.



(a) Music room (b) Colorful room

(c) Horror room (d) Subway room

Fig. 3. Screenshots from game rooms.

To increase the game appeal for players who lack prior data mining knowledge, the following game elements can be supported in the game:

- *Tutorial:* A tutorial can be provided to explain basic data mining concepts. The players who have intermediate-level knowledge about data mining can skip the tutorial.
- *Examples:* Several examples can be given for each topic to illustrate how to perform the various data mining tasks.
- *Feedback:* Guided feedback can be delivered to the player while learning is taking place. Learners are immediately informed of how well they are performing and how they might go about making those improvements.
- *Reward:* More points may be given to individuals with no data mining backgrounds than highly educated individuals.
- *Options:* In the problem-solving parts of the game, multiple options can be provided, allowing the player to choose between them. The difficulty level of the game can be changed by representing different datasets ranging from simple to complex according to players' level of knowledge about data mining concepts.
- *Hint:* While a learner solves a problem asked in the game, hints about the tasks can appear on the screen.

4.3 Game design elements

Game design elements are the parts used to build games and are characteristic of smaller games that are found in most games. In this study, the following GBDML elements were utilized to make the learning process more interesting.

- *Learning processes:* Opportunities to learn are provided through a variety of methods: tutorials, examples, problems and tasks. For example, one of the game goals is to answer all questions correctly, which is equivalent to the learning goal.
- *Feedback:* Feedback is given on whether questions are answered correctly.
- *Reward:* Task and scoring mechanisms are used to motivate players to learn more.
- *Rules:* Rules are defined to match the educational objectives. For example; if the player answers the questions in one room correctly, he or she will pass on to the next room.

- *Environment:* The virtual environment in which the game occurs is supported by the rules and solution methods.
- *Choices:* The player is encouraged to choose objects by clicking on them to interact with them and perform certain actions.
- *Challenges:* Finding the right difficulty level is the secret to successful learning. If the game is too difficult, the player can give up; if it is too easy, the player can become bored. The proposed game makes learning just difficult enough that it continues to be fun.
- *Mystery:* Mystery is provided by the gap between available information and unknown information.

4.3.1 Story

One possible method for binding learning with the game is to use educational story as the backbone for the game. Story is the game framework including the background and other information needed in the playing process. In our proposed game, narrative storytelling method was used for the game story to help the students learn data mining techniques more efficiently. The story of the proposed game concerns a clumsy chemist and his chemical experiment. "One day, a chemist was trying to do a chemical experiment when an accident occurred in the laboratory. He was poisoned by gas escaping from the chemical reaction. He needs an antidote to survive. To produce the antidote, the player should gather the components necessary to produce the antidote. The player must visit several rooms, collecting different objects from different rooms by applying data mining techniques. At the end of each data mining application, the player gains a component of the antidote and a key to enter another room. After the player gathers all components of the antidote, the chemist will survive."

4.3.2 Goal

Goals correspond to task descriptors such as objectives, outcomes, and problems to be solved. Goals can be based on the story, or learning goals can be presented as game goals. Combining learning goals with game goals is the easiest method for gamifying education and provides a motivating learning experience in the flow of game play. Table 2 presents the learning goals and game goals defined in Mine4Escape. GBDML can be considered goal-oriented and rule-driven activities.

4.3.3 Reward

Reward is a game element that satisfies users and motivates them to achieve more. When a player executes an action about the requested subject to be taught, he or she gains an achievement. In most games, achievements exist simply to upgrade game mechanics. Games can provide *extrinsic rewards* (i.e., points, badges) and *intrinsic rewards* where tasks are rewarding by their nature. The proposed game in this study includes both mechanisms: *tasks* (see section 4.2.4) as an intrinsic reward and *scoring mechanism* as an extrinsic reward.

Table 2. The types of learning aligned with the specified learning and game goals

Type of Learning	Learning Goal	Game Goal
Problem-based learning	To practice data mining techniques such as Naive Bayesian and Decision Tree .	To escape from a room, the player should answer all questions correctly.
Feedback-based learning	To get higher scores, the learner tries to improve his or her data mining skills and learning abilities. Game points are similar to grades in educational system.	Providing real-time feedback on the accuracy of trainees' answer to help them to know their status and to understand what else they need to know. Thus, the player can decide to repeat a concept or move further to learn next concept.
Action-based learning	To demonstrate data preparation steps such as data collection, selection and preprocessing.	To obtain the dataset, the player should click on the objects
Inquiry-based learning	To inspire students to learn data mining concepts for themselves through their own investigation.	The player should be allowed to take his or her research in a variety of directions.
Example-based learning	To provide sample data mining examples to help students learn.	Before getting a question, the player should examine an example solution.
Lecture-based learning	To help trainees to learn detailed information about data mining concepts.	The player should visit the tutorial room.
Scenario-based learning	To prompt the learner to work through a storyline based around a scenario.	The player tries to help the chemist who was poisoned by gas escaping from a chemical reaction.
Task-based learning	To teach data mining tasks such as association rule mining and classification.	The player should gather all components of the antidote for the chemist.

A *scoring mechanism* can be used for setting goals, giving feedback and representing users' status or reputation. Points are awarded to users after completing certain challenges or reaching achievements. Getting a good score is just one reason that people play games; players engage with games to make changes based on their feedback. In addition, users are always competing against each other; the maximum score that they get is a benchmark for competition. Based on the collected points, players can be listed on the scoreboard. Game points are similar to grades in the educational system. In GBDML, one of the major ideas is that scores or winning positions motivate users to improve their data mining skills and learning abilities.

The new scoring mechanism proposed in GBDML is based on the calculation of five parameters from player to playing habits: age of player, educational level of player, knowledge level of player on data mining, interaction with objects, and time to solve problems.

To calculate the score according to the player's age, a logarithmic based function was developed as given in Equation (3). Increasing age is negative factor in the learning process and the game is therefore designed such that older players gain more points. There are lower and upper age limits denoted by $L(age)$ and $U(age)$, beyond which younger or older learners are not in the target group. Fig. 4a shows the changes of points based on player age with lower and upper limits set at 18 and 50, respectively.

$$Age \rightarrow \alpha = \begin{cases} Age.\log_{U(age)} Age + (Age.\log_{U(age)} Age)/2, & L(age) \leq Age < U(age) \\ U(age) + U(age)/2, & Age \geq U(age) \\ L(age), & Age < L(age) \end{cases} \tag{3}$$

In the proposed scoring mechanism, if the players complete the task assigned to them correctly, they will be given a set number of points as defined in Equation (4). This is easiest method how to implement competition in games. All data mining tasks are timed with lower and upper time limits that are denoted by $L(time)$ and $U(time)$. When learning efficiency is a factor, the time taken to learn data mining is a critical performance measure in GBDML. Therefore, if the player completes a data mining task, the remaining time will be awarded to the player as bonus points, which will encourage the user to continue to play and practice. If the player fails to complete a data mining task, they will receive no points. Fig. 4b shows the changes of points over time with lower and upper limits set as 10 and 100 minutes, respectively. An increase in the duration of activity in game rooms acts as a penalty to the user's score. Time affects score logarithmically according to the lower and upper limits, and the score that will win decreases sharply or vaguely.

$$Time \rightarrow \beta = \begin{cases} 2.L(time).\log_{Time} U(time), & L(time) \leq Time < U(time) \\ 2.L(time), & Time \geq U(time) \\ (U(time)/2 - Time), & Time < L(time) \end{cases} \tag{4}$$

Interaction points can be designed to keep participation active in the learning course. In the proposed scoring mechanism, a player also gains points from his or her interactions with the objects in the game room. In this study, we developed an algorithm to calculate interaction points that denoted by γ (see Appendix). The player

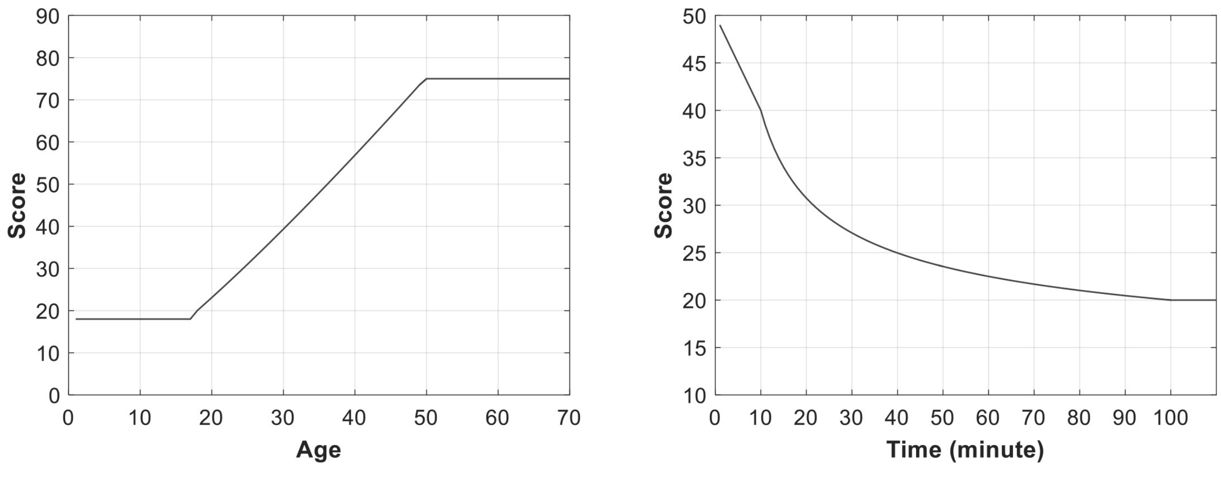


Fig. 4. Scoring mechanism according to the (a) age and (b) time.

collects data mining related objects by clicking on interactive objects in order to handle a dataset. The selection of the same objects again has been prevented using flag mechanisms for both interactive objects and data mining objects.

Points can be awarded according to two different user levels: education level and previous data mining knowledge level. By considering the inverse proportion between these levels and points, Equations (5), (6) and (7) are constructed.

$$Education\ Level \rightarrow \Phi = \begin{cases} 4, & High\ School \\ 3, & Undergraduate \\ 2, & MSc \\ 1, & PhD \end{cases} \tag{5}$$

$$Data\ Mining\ Knowledge\ Level \rightarrow \xi = \begin{cases} 5, & None \\ 4, & Beginner \\ 3, & Intermediate \\ 2, & Upperintermediate \\ 1, & Advanced \end{cases} \tag{6}$$

$$Relationship\ between\ levels \rightarrow \eta = \xi * \Phi \tag{7}$$

In the proposed game, the unlocking of a game room is also termed as a reward. As shown in Equation (8), the weights of game rooms are set according to their difficulty in the interval between 1 and 2. The weight value of the game room is multiplied by the sum of other point aforementioned.

$$Game\ room\ weights \rightarrow \varphi = \begin{cases} 1, & Tutorial\ room \\ 1.2, & Subway\ room \\ 1.4, & Music\ room \\ 1.6, & Colorful\ room \\ 1.8, & Horror\ room \\ 2, & Final\ room \end{cases} \tag{8}$$

Finally, overall score, denoted as δ, is calculated as shown in Equation (9), with the possibility of getting fractional number that is rounded down. The score can also be calculated by the weighted sum of the different point types.

$$Overall\ Score \rightarrow \delta = \delta + \lfloor \varphi * (\alpha + \beta + \gamma + \eta) \rfloor \tag{9}$$

The score can also be calculated by the weighted sum of the different point types to change the importance of the scoring parameters, as shown in Equation (10):

$$Weighted\ Score \rightarrow \delta = \delta + \lfloor \varphi * (w_a * \alpha + w_t * \beta + w_i * \gamma + w_l * \eta) \rfloor \tag{10}$$

where w_a, w_t, w_i, and w_l are user-defined weight values for the different point types: age, time, interaction and level, respectively.

4.3.4 Tasks

A task is a meaning-focused activity that requires learners to integrate and apply multiple skills. Therefore, from the learner's perspective, a task can be considered as an activity with a learning outcome and a set of procedural guidelines to follow to reach this target. Tasks should be achievable with a reasonable amount of effort. The learner will be bored if the task is too easy or stressed if the task is too difficult.

In the game proposed in this paper, the tasks are essentially problem-solving activities related to data mining. The game includes data-focused activities and some exercises that require learners to practice or learn data mining skills. In each room, learners are allowed to solve a different data mining problem. In each room, rewards are categorized into short-time and long-time. A *short-time reward* is a simple key that the player is given when he or she completes a data mining task in order to escape from the room and enter another one. A *long-time reward* is to complete all data mining tasks to gather all the components of the antidote for the chemist. To illustrate the conception of GBDML, typical game tasks are listed and explained below.

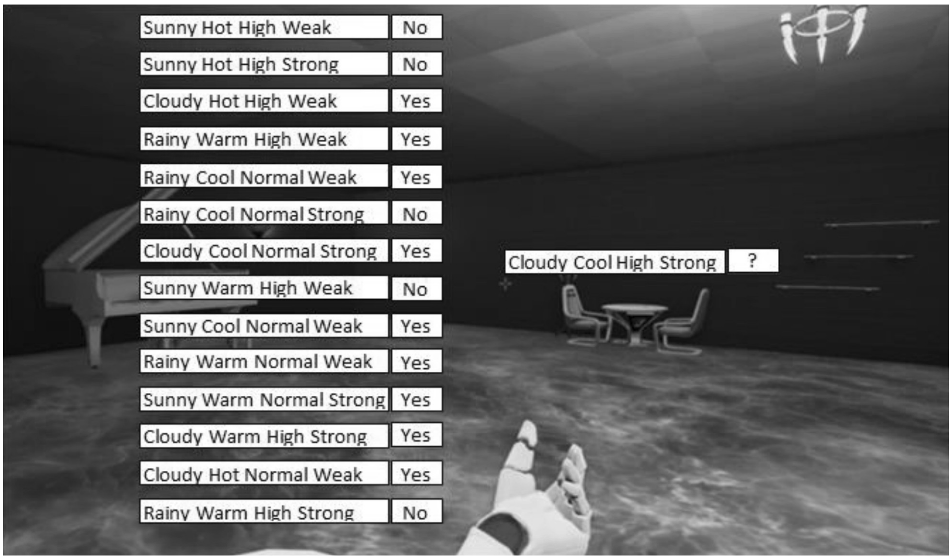


Fig. 5. Data preparation (Task 1) and Bayesian classification (Task 2).

Task 1—Data Preparation
As shown in Fig. 5, the player collects data mining related objects by clicking on interactive objects in order to handle a dataset. After that, some data preparation operations (see section 3.1.1) such as data selection, preprocessing, and transformation (i.e., data discretization and data normalization) can be applied to it as a problem-solving approach.

Task 2—Bayesian Classification
This task is greatly simplified from asking the user to fill the blank textbox using Naive Bayes algorithm (see section 3.1.2), as shown in Fig. 5. Small datasets (14 rows and 5 columns) are preferred to allow users to solve the problem in a short time period.

Task 3—Association Rule Mining
Figure 6 shows an example question related to association rule mining. Two randomly chosen items are given to the player, who is expected to write the correct support and confidence values (see section 3.1.3) in the blank

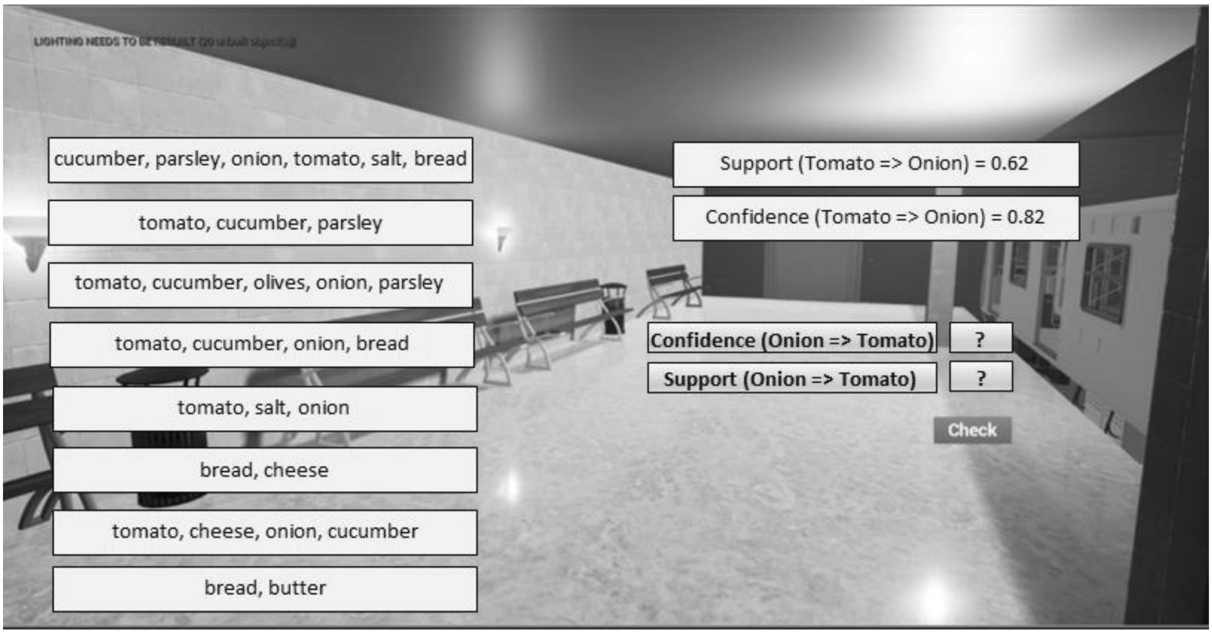


Fig. 6. Association rule mining (Task 3).

textboxes. Although the main challenge is solving the data mining problems, we must calibrate varying difficulties of the task for different type of users or game levels. If challenges are too difficult to solve, the target group can lose interests. To provide the medium difficulty, an example solution can be given as shown in Fig. 6.

Task 4—Decision Tree

Another room with a different environment has been designed to allow players to solve a decision tree problem. The player walks around, interacts with objects to collect data, and applies the C4.5 decision tree algorithm to it.

5. Experimental results

An experimental study was carried out to get feedback from a test group and evaluate the performance of GBDML. The participants were asked to answer a questionnaire requesting their opinions about the game and the effect of the game on their knowledge of data mining concepts.

5.1 Participants

The study was performed with 39 undergraduate and graduate students in computer engineering. The number of men and women in the test group is almost equal. Before the experience, some of the participations did not have any knowledge about data mining, while others were intermediate-level learners of data mining. The prepared survey collected background information from the participants that may affect their performance, learning capability and enjoyment in the game. The survey included questions about participants' age, gender, education level, job, game habits, and their previous data mining knowledge levels.

5.2 Questionnaire

The questions and answer choices in the questionnaire are given in Table 3. The questions ask the respondents about their ability to learn data mining concepts. The questionnaire consists of both closed-ended and open-ended questions. The *closed-ended questions* relate to aspects specific to each of the data mining skills, whereas the *open-ended questions* allow participations to provide input regarding how the game could be improved to further develop their data mining skills. It should be noted that the questionnaire also contains a choose-one-or-specify question, which is a hybrid question that mixes a closed-ended list with an open-ended response. The questionnaire in this study generally consists of Likert scale questions with a five-category scale; however, it also includes one three-point scale question and one seven-point scale question.

Survey items fall into three general content categories: demographic, factual and attitudinal. *Demographic items* ask respondents for information about their backgrounds. The questionnaire prepared in this study includes several demographic questions, i.e., age, gender, job and education level. *Factual items* ask about respondents' experiences. In our questionnaire, there are several factual questions about previous data mining experiences and game habits. *Attitudinal items* ask for respondents' opinions and perceptions of a topic. The participants were required to answer attitudinal questions about data mining tasks they completed. Some questions are also associated with both game enjoyment and preferences for future play.

In this study, we measured five factors in investigating the motivation for computer game play and the effects of game play on learning data mining: competence, intuitive control and autonomy. *Competence* refers to players' feelings of being capable of solving data mining problems within the game. *Intuitive control* refers to the ease with which players learned how to play a game (i.e., easy or hard to play). *Autonomy* refers to players' perceived degrees of data mining concepts in the game.

5.3 Analyzing survey data

After participants finished playing the game and escaped from all the game rooms, they completed a questionnaire in a certain amount of time allotted for them. This section presents survey results, data mining performed on questionnaire answers and interpretation of data mining results.

5.3.1 Application of decision tree technique to analyze survey data

Figure 7 shows the decision tree generated by applying the C4.5 algorithms to survey data to discover significant parameters and rules in GBDML. The accuracy rate of a classification algorithm is generally defined as the closeness of the predicted values to the actual values and is calculated to evaluate the success of the algorithm on the selected dataset. In this study, the overall classification accuracy was calculated as 94.8%.

According to the decision tree, the most significant criteria that effects GBDML is "education level." The PhD students who have intermediate/beginner level knowledge of data mining consider GBDML useful, while

Table 3. The questions and response options in the questionnaire

Attributes	**Values**	**Questions**
Age	Numerical	How old are you?
Gender	– Male – Female	Please select your gender.
Education level	– High school – Undergraduate – MSc – PhD	Please select your education level.
Job	– Student – Computer engineer – Research assistant – Other	Please select your job category.
Data mining education	– Yes – No	Have you ever taken a data-mining course?
Data mining level	– None – Beginner – Intermediate – Upper intermediate – Advanced	What is your knowledge level of data mining?
Game habit	– Never – Rarely – Sometimes – Often – Always	How often do you play computer games?
Learning method(s)	– Theoretical information – Research / Literature review – Solution samples – Project development – Coding existing algorithms – Develop a new algorithm – Game based	What are the effective ways to learn data mining? (You can select more than one answer.)
Game opinion	– Funny / Boring – Educational / Not educational – Absorbing / Not absorbing – Easy / Hard	What is your opinion about this game?
Data preparation (Learning outcome 1)	– No Idea – Never – Few – Middle – High	Did you learn something about the "Data Preparation" topic?
Association rule mining (Learning outcome 2)	– No Idea – Never – Few – Middle – High	Did you learn something about "Association Rule Mining"?
Classification (Learning outcome 3)	– No Idea – Never – Few – Middle – High	Did you learn something about "Classification"?
GBDML	– No idea – No – Yes	Is it possible to learn data mining through game-based learning?
Deficiencies of the game	–	What are the deficiencies of the game in terms of learning data mining?
Errors in the game	–	Did you get any error while playing the game?
Features need to be added	–	What features can be added to the game?

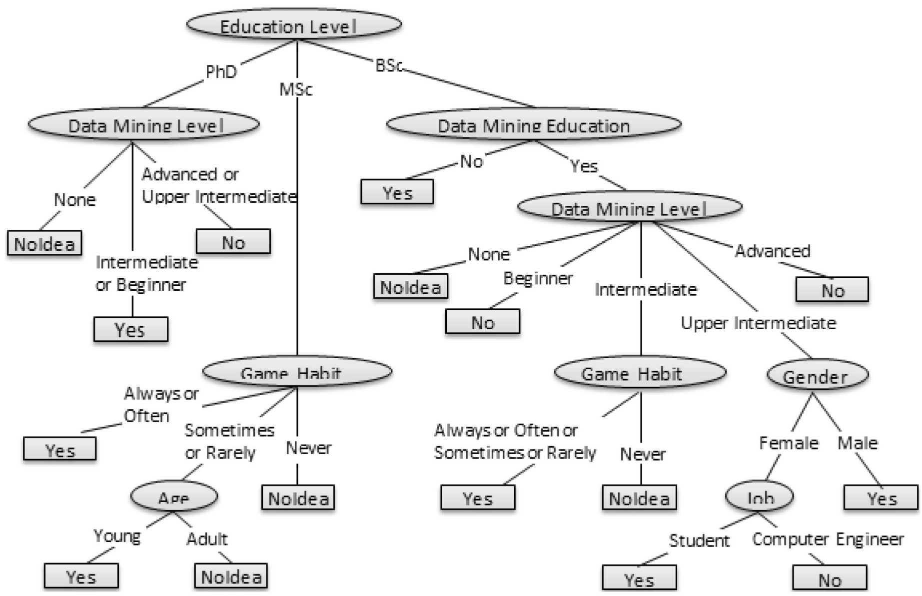


Fig. 7. Decision tree constructed on survey data.

others find it ineffective. On the other hand, opinions about the benefits of GBDML differ based on MSc students' game habits. Participants who play computer games frequently and younger students who can be called "free time gamers" interpret GBDML as successful.

According to the decision tree, the most variety in the answers come from by undergraduate students. All undergraduate students with no knowledge of data mining background find GBDML helpful. If a participant has taken a data mining course before, however, the answer changes according to that person's gaming habits, gender, and job. Generally, positive opinions about GBDML and high gaming habits are directly proportional. In addition, opinions are also varied according to participants' gender. Generally, males have positive opinions, while the answer varies for women according to their jobs.

5.3.2 Cluster analysis of survey data

The K-Means++ algorithm was applied to the survey dataset to cluster participants according to their responses to the questionnaire to identify groups of individuals who reported similar experiences with GBDML. The most important challenge of the K-means++ algorithm is to determine the best k value, in other words, the optimal number of clusters. For this purpose, we used the elbow method, which chooses the k at which the sum of squared error (SSE) produces an "elbow effect" in the graph. Let $C=\{c_1, c_2,...,c_k\}$ be k clusters, each containing some of the x_j's Let μ_i be the mean of cluster c_i such that $\mu_i = 1/|c_i| \, \Sigma_{x_j \in c_i} x_j$. We define the squared error between μ_i and the points in c_i by $\mathrm{SSE}(c_i) = \Sigma_{x_j \in c_i} ||x_j - \mu_i||^2$. The sum of squared error over all k clusters is $\mathrm{SSE}(C) = \Sigma_{i=1}^{k} \Sigma_{x_j \in c_i} ||x_j - \mu_i||^2$. As shown in Fig. 8, SSEs were calculated by executing the algorithm repeatedly with varying k input values from 1 to 8, and then the optimal number of cluster was determined to be 5.

Final cluster centroids are given as follows:

Attributes: age, gender, education level, job, data mining education, data mining level, game habit, GBDML
Cluster 1—young, male, BSc, student, no, none, sometimes, yes (33%)
Cluster 2—adult, male, MSc, computer engineer, yes, intermediate, sometimes, no idea (15%)
Cluster 3—young, female, BSc, student, yes, intermediate, rarely, yes (31%)
Cluster 4—young, male, MSc, computer engineer, yes, upper intermediate, rarely, yes (15%)
Cluster 5—young, female, MSc, computer engineer, yes, upper intermediate, never, no (5%)

According the clustering results, the individuals in three clusters have positive opinion about GBDML, while those in one cluster have no idea and those in the last group have an unfavorable attitude toward it. The first cluster consists of individuals who are generally male undergraduate students under age 26 who do not have any knowledge about data mining. The second group is simply the union of the adult computer engineers who have an MSc degree. This group evaluated GBDML as neutral because they have already taken a data-mining

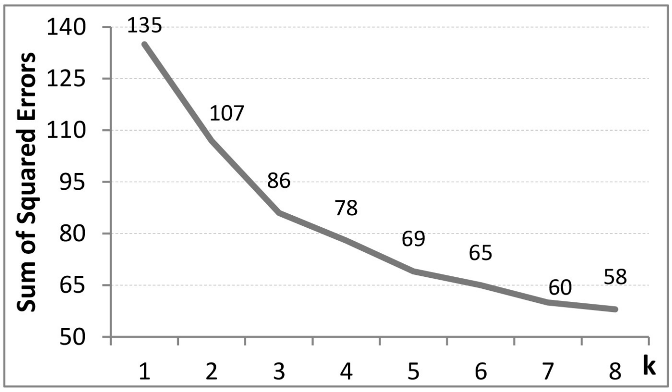


Fig. 8. Sum of Squared Errors for K-Means++.

course. The last user group did not find GBDML particularly useful because participants in this group are generally female and do not play digital games. The results show that the benefits of GBDML changes according to their background knowledge of data mining. The centroids of cluster 1 and cluster 3 show that undergraduate students have generally positive experiences in learning data mining through game-based techniques.

5.3.3 *Association rule mining to analyze survey data*

The Apriori algorithm was executed on questionnaire data with minimum support of 2% to understand the relationships between user profiles and their feedback about GBDML. Table 4 shows some of the rules

Table 4. Some rules discovered by association rule mining

Length	Frequent Itemsets		Support (%)
2-itemsets	Opinion = Yes	Age = Young	56.41
		Education Level = BSc	53.85
		Job = Student	46.15
		Gender = Male	35.90
	Opinion = No Idea	Gender = Female	17.95
		Game Habit = Never	7.69
		Age = Adult	7.69
	Opinion = No	Data Mining Education = Yes	7.69
		Education Level = PhD	5.13
		Job = Other	2.56
		Data Mining Level = Advanced	2.56
3-itemsets	Opinion = Yes	Education Level = BSc, Age = Young	51.28
		Gender = Male, Job = Student	28.21
		Data Mining Level = Intermediate, Job = Computer Engineer	10.26
	Opinion = No Idea	Gender = Female, Game Habit = Never	5.13
		Education Level = MSc, Gaming Time = Never	5.13
	Opinion = No	Education Level = PhD, Age = Adult	5.13
4-itemsets	Opinion = Yes	Job = Student, Data Mining Education = No, Education Level = BSc	28.21
		Age = Young, Gender = Female, Education Level = BSc	25.64
	Opinion = No Idea	Job = Research Assistant, Gender = Female, Game Habit = Never	2.56
	Opinion = No	Data Mining Level = Advanced, Data Mining Education = Yes Education Level = PhD	2.56
5-itemsets	Opinion = Yes	Gender = Male, Education Level = BSc, Job = Student, Age = Young	23.1
	Opinion = No Idea	Game Habit = Never, Data Mining Education = Yes, Job = Student Age = Young	5.13
6-itemsets	Opinion = Yes	Data Mining Level = None, Job = Student, Education Level = BSc, Data Mining Education = No, Age = Young	28.21
	Opinion = No Idea	Game Habit = Never, Gender = Female, Education Level = MSc, Data Mining Education = Yes, Age = Young	2.56
	Opinion = No	Data Mining Level = Advanced, Gender = Male, Education Level = PhD, Data Mining Education = Yes, Age = Adult	2.56

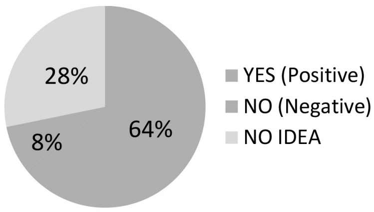


Fig. 9. Overall opinion about GBDML.

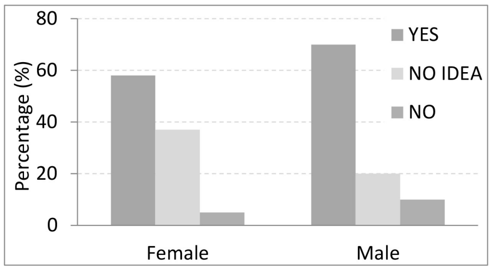


Fig. 10. The percentage of respondents according to their genders and opinions about GBDML.

discovered by association rule mining. The mining results show that the main factors related to satisfaction with GBDML are age, education level, job and gender. In addition, positive opinions about GBDML have a strong relationship with gender = male and age = young. According to the rules obtained, it is also possible to say that if the participants do not have any information about data mining, they find this game useful. The gaming habit of the learner is another important factor that affects his or her ability to learn data mining through game-based technique, especially for females. The obtained results from the questionnaire responses showed that the developed game can be used to teach data mining concepts to BSc students.

5.4 Discussion

Figure 9 presents the overall opinion of GBDML. The results show that the game-based approach is a promising alternative for teaching data mining concepts because more than 60% of participants evaluated it as positive, while only 8% expressed negative opinions.

In this study, we attempted to identify the factors that influence learning data mining through game-based techniques. Analysis of the survey data shows that three characteristics of trainees affect usefulness of GBDML: gender, age and gaming habits.

Figure 10 shows the percentage of respondents according to their gender and opinions about GBDML. Both males and females generally expressed positive feedback about the proposed approach; however, males show a stronger preference for digital games compared to females and males play these games for longer than females. For this reason, it is possible that game-based learning is more acceptable to males than females.

Figure 11 shows the opinions of the participants about GBDML based on their ages. Mining the results shows that interest in computer games decreases with age. Thus, the age of players is inversely proportional to their opinions about the usefulness of GBDML and young learners may therefore benefit more than older ones from the use of computer games for learning. Young trainees have opportunities to achieve learning outcomes more quickly, so they experienced feelings of satisfaction as they improved their data mining skills.

Figure 12 clearly shows that the gaming habits of participants are another factor that influences their opinions of GBDML. If a user is not familiar with computer games, he or she may have trouble adapting to the environment and may therefore have difficulty learning data mining. According to the survey results, participants who often play computer games have positive opinion. In contrast, individuals who rarely play digital games specify that they have no idea or negative opinions about the usefulness of the proposed approach.

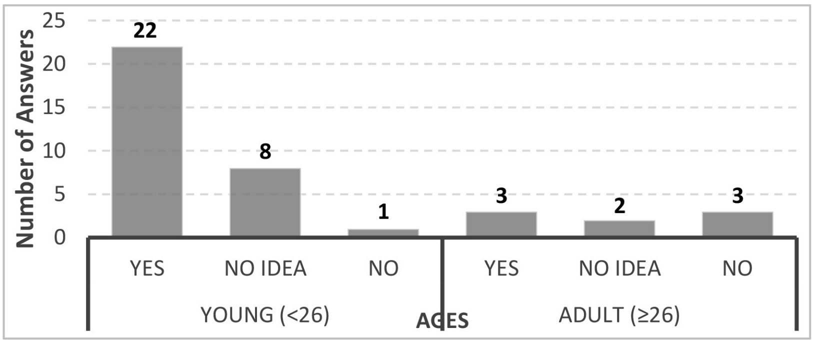


Fig. 11. Opinions of participants about GBDML according to age.

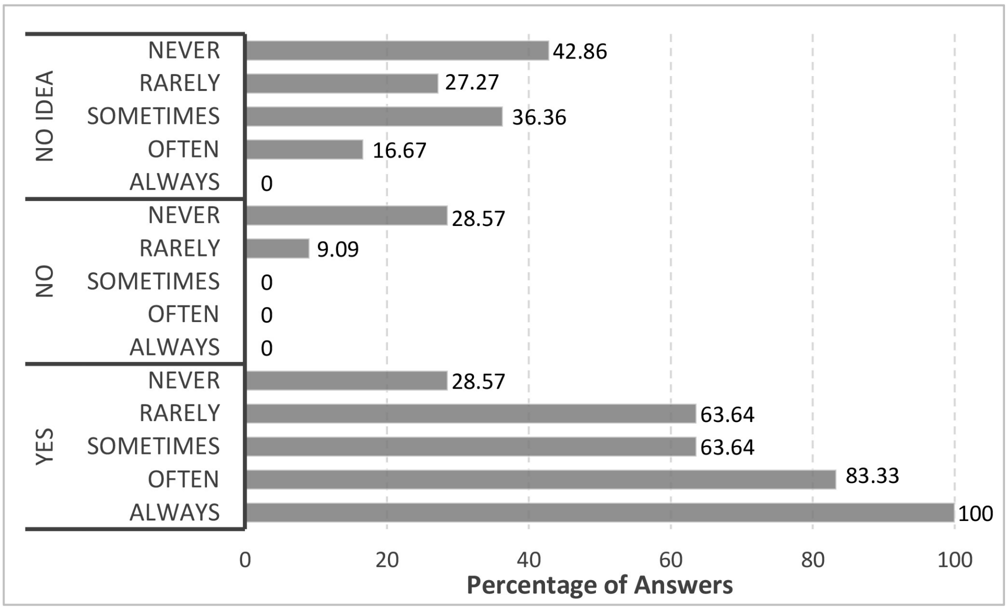


Fig. 12. The percentage of respondents according to their gaming habits and opinions about GBDML.

As a result, the opinions about GBDML change according to learners' age, gender and game habits. In addition, their jobs, education levels and background knowledge about data mining are also secondary factors that influence their opinions. It is possible to say, however, that GBDML provides a positive experience that keeps learners motivated to continue playing and learn data mining.

6. Conclusion

Game-based learning is an enjoyable and interesting learning approach that promotes the improvement of students' skills, learning, and understanding capabilities much more efficiently than classroom learning. Considering this motivation, our study proposes teaching data mining techniques with the scope of game-based learning. The aim of this study is to provide a game environment that enables the achievement of learning goals in data mining training programs. The proposed game, Mine4Escape, was structured as a puzzle game that is based on the story of escaping from the locked rooms. The basic knowledge necessary to start a game is being a BSc student with some education in Computer Engineering. The learning goals of the proposed system are as follows: learning basic data mining concepts, practicing on data mining techniques, demonstrating data preparation steps, illustrating the various data mining tasks and improving data mining skills. The game was designed to be consists of six rooms: music, colorful, horror and subway rooms for teaching Naïve Bayes, decision tree, and ARM techniques, in addition to the final and tutorial rooms for the application of the questionnaire and providing a guide for players, respectively. In addition, this article also proposes a dynamic scoring mechanism that awards different points to the different players by considering several factors such as the degree of task difficulty, learners' background and their knowledge level.

An experimental study was carried out to evaluate the performance of the proposed approach by conducting a survey to receive feedback from a test group. In this survey, the participants stated their opinions about the game and the effect of the game on their knowledge of data mining concepts. According to the results of survey, 64% of participants evaluated it as positive. On the basis of the findings from the questionnaire, it is possible to say that knowledge acquisition about data mining can be enhanced by the game-based approach when comparing with the traditional data mining training methods. According to the results, positive opinions were generally ranked by young participants with high game playing habits. The results also indicate that the degree of learning interest and learners' information acceptance changes according to participants' age, gender, educational level, and gaming habits.

In future research, a competition mechanism can be added to the GBDML approach to investigate its effect on learning data mining. In addition, a "play with multiple users" option can also be included to provide learners with the option of playing as a team to overcome obstacles infused in the game related to data mining. Furthermore, a mobile game version of the proposed study can be implemented to guide students to improve their learning motivation and effectiveness given their growing popularity. In addition to classification and association rule mining in the present study, other data mining tasks such as clustering or sequential pattern mining can also be applied as learning outcomes in future research.

Acknowledgements—We would like to thank the Game-Based Learning Student Group from the Department of Computer Engineering at Dokuz Eylul University that has supported this work in the development and testing stages.

References

1. D. Hooshyar, R. B. Ahmad, M. Yousefi, M. Fathi, S. Horng and H. Lim, Applying an online game-based formative assessment in a flowchart-based intelligent tutoring system for improving problem-solving skills, *Computers and Education*, **94**, 2016, pp. 19–36.
2. D. Yousef, S. Baadel and R. Makad, Exploratory study on the impact of game-based learning on student engagement, *International Conference on Web and Open Access to Learning (ICWOAL),* Dubai, 25–27 November 2014, pp. 1–4.
3. E. Pesare, T. Roselli, N. Corriero and V. Rossano, Game-based learning and gamification to promote engagement and motivation in medical learning contexts, *Smart Learning Environments*, **3**(5), 2016, pp.
4. Y. C. Yang, C. Wang, M. Tsai and J. Wang, Technology-enhanced game-based team learning for improving intake of food groups and nutritional elements, *Computers & Education*, **88**(C), 2015, pp. 143–159.
5. M. Callaghan and M. Cullen, Game based learning for teaching electrical and electronic engineering, *Advances in Computer Science and its Applications*, **279,** 2014, pp. 655–660.
6. S. Al-Jibouri, M. Mawdesley, D. Scott and S. Gribble, The use of a simulation model as a game for teaching management of projects in construction, *International Journal of Engineering Education*, **21**(6), 2005, pp. 1195–1202.
7. R. Pace and A. Dipace, Game-Based learning and lifelong learning for tourist operators, *Cultural Tourism in a Digital Era—Springer Proceedings in Business and Economics*, 2015, pp. 185–199.
8. A. Kinsheel, Constructivist game-based robotics simulator in engineering education, *International Journal of Engineering Education*, **29**(4), 2013, pp. 1024–1036.
9. J. Minovic, M. R. Markovic and B. Draskovic, Financial engineering education : The case study of financial modelling using games, *International Journal of Engineering Education*, **29**(3), 2013, pp. 634–643.
10. B. D. Coller and D. J. Shernoff, Video game-based education in mechanical engineering: A look at student engagement, *International Journal of Engineering Education*, **25**(2), 2009, pp. 308–317.
11. M. Mortara, E. C. Catalona, F. Bellotti, G. Fiucci, M. Houry-Panchetti, and P. Petridis, Learning cultural heritage by serious games, *Journal of Cultural Heritage*, **15**(3), 2014, pp. 318–325.
12. A. M. Ross, M. E. Fitzgerald and D. H. Rhodes, Game-Based Learning for Systems Engineering Concepts, *Procedia—Computer Science*, **28**, pp. 430–440.
13. B. A. Foss and T. I. Eikaas, Game play in engineering education-concept and experimental results, *International Journal of Engineering Education*, **22**(5), 2006, pp. 1043–1052.
14. K. Browne, C. Anand and E. Gosse, Gamification and serious game approaches for adult literacy tablet software, *Entertainment Computing*, **5**(3), 2014, pp. 135–146.
15. U. Faghihi, A. Brautigam, K. Jorgenson, D. Martin, A. Brown, E. Mesaures and S. Maldonado-Bouchard, How gamification applies for educational purpose specially with college algebra, *Procedia—Computer Science*, **41**, 2014, pp. 182–187.
16. S. Juzeleniene, J. Mikelioniene, P. Escudeiro and C. V. Carvalho, GABALL project: Serious game based language learning, *Procedia—Social and Behavioral Sciences*, **136**, 2014, pp. 350–354.
17. E. Z. F. Liu and P. Chen, The effect of game-based learning on students learning perfomance in science learning—a case of "Conveyance Go"', *Procedia-Social and Behavioral Sciences*, **103**, 2013, pp. 1044–1051.
18. M. M. Ariffin, A. Oxley and S. Sulaiman, Evaulating game-based learning effectivness in higher education, *Procedia—Social and Behavioral Sciences*, **123**, 2014, pp. 20–27.
19. V. Cojocariu and I. Boghian, Teaching the relevance of game-based learning to preschool and primary teachers, *Procedia—Social and Behavioral Sciences*, **142**, 2014, pp. 640–646.
20. S. Ucus, Elemantary school teachers' views on game-based learning as a teaching method, *Procedia—Social and Behavioral Sciences*, **186**, 2015, pp. 401–409.
21. M. Li and C. Tsai, Game-based learning in science education: A review of relevant research, *Journal of Science Education and Technology,* **22**(6), 2013, pp. 877–898.
22. F. H. Tsai, C. Kinzer, K. H. Hung, C. C. Chen and I. Y. Hsu, The importance and use of targeted content knowledge with scaffolding aid in educational simulation games, *Interactive Learning Environments*, Taiwan, **21**(2), 2013, pp. 116–128.
23. M. Soflano, T. M. Connolly and T. Hainey, An application of adaptive games-based learning based on learning style to teach SQL, *Computers and Education*, **86**(C), 2015, pp. 192–211.
24. I. Ouahbi, F. Kaddari, H. Darhmaoui, A. Elachqar and S. Lahmine, Learning basic programming concepts by creating games with scratch programming environment, *Procedia—Social and Behavioral Sciences*, **191**, 2015, pp. 1479–1482.
25. R. Bourbia, N. Gouasmi, M. Hadjeris and H. Seridi, Development of serious game to improve computer assembly skills, *Procedia—Social and Behavioral Sciences*, **142**, 2014, pp. 96–100.
26. H. Hou and M. Li, Evaulating multiple aspects of a digital educational problem-solving-based adventure game, *Computer in Human Behavior*, **30**, 2014, pp. 29–38.
27. D. Rodriguez-Cerezo, A. Sarasa-Cabezuelo, M. Gomez-Albarran and J. Sierra, Serious games in tertiary education: A case study concerning the comprehension of basic concepts in computer language implementation courses, *Computer in Human Behavior*, **31**, 2014, pp. 558–570.
28. J. M. R. Corral, A. C. Balcells, A. M. Estevez, G. J. Moreno and M. J. F. Ramos, A game-based approach to the teaching of object-oriented programming language, *Computers and Education*, **73**, 2014, pp. 83–92.
29. W. Y. Seng and M. H. M. Yatim, Computer game as learning and teaching tool for object oriented programming in higher education institution, *Procedia-Social and Behavioral Sciences*, **123**, 2014, pp. 215–224.
30. E. Nunohiro, K. Matsushita, K. J. Mackin and M. Ohshiro, Development of game-based learning features in programming learning support system, *Artif Life Robotics*, **17**(3), 2013, pp. 373–377.
31. C. Kazimoglu, M. Kiernan, L. Bacon and L. Mackinnon, A serious game for developing computational thinking and learning introductory computer programming, *Procedia—Social and Behavioral Sciences*, **47**, 2012, pp. 1991–1999.
32. B. Schmitz, A. Czauderna, M. Specht and R. Klemke, Game based learning for computer science education, *Computer Science Education Research Conference*, The Netherlands, 2011, pp. 81–86.

33. M. J. Lee and A. J. Ko, Personifying programming tool feedback improves novice programmers' learning, *Proceedings of the seventh international workshop on Computing education research*, Providence, RI, US, 8-9 August 2011, pp. 109–116.
34. M. Papastergiou, Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation, *Computers and Education*, **52**(1), 2009, pp. 1–12.
35. K. Kuk and D. Jovanovic, Design and implementation of CoAeLearn modules for personalized game based-learning within computer architecture course, *International Journal of Engineering Education*, **29**(3), 2013, pp. 620–633.
36. M. Minovic, M. Milovanovic, I. Kovacevic and D. B. Starcevic, Game design as a learning tool for the course of computer networks, *International Journal of Engineering Education*, **27**(3), 2011, pp. 498–508.
37. S. Mladenovic, D. Krpan and M. Mladenovic, Using games to help novices embrace programming: From elementary to higher education, *International Journal of Engineering Education*, **32**(1), 2016, pp. 521–531.
38. D. H. Jonassen, J. Howland, J. Moore, and R. M. Marra, *Learning to Solve Problems with Technology: A Constructivist Perspective (2nd ed.)*. Prentice Hall, 2002.

Appendix

Algorithm of score from interaction with objects

```
Algorithm Interaction_Score()
  // Scoring user interactions with objects
  Input: A set of interaction objects O={o1, o2, ...on}
         A set of data mining objects M={m1, m2, ..., mk}, where M⊆O
         Desired point scale, such as high=10; mid=5; low=2;
  Output: Score of interactions (γ)
  foreach obj ∈ O do
    obj.interactionFlag = false;
    if obj ∈ M
      obj.miningFlag = true;
    else
      obj.miningFlag = false;
    endif
  endfor
  while (true)   // GameLoop()
    Event e = get_next_event();
    if e.eventType = QUIT or count =M.Length
      exit();
    endif
    if e.eventType = OBJECT_CLICK
      obj = e.eventSource;
      if obj.miningFlag & !obj.interactionFlag
        γ += high
        count++
      else if !obj.miningFlag & !obj.interactionFlag
        γ += mid
      else if obj.miningFlag & obj.interactionFlag
        γ -= low
      endif
      obj.interactionFlag = true;
    endif
  endwhile
  return γ
```

Mehmet Cengiz received his B.Sc. and M.Sc. degrees in computer engineering at Pamukkale University, Denizli, Turkey, in 2011 and 2013, respectively. He began working on a Ph.D. in computer engineering at Dokuz Eylul University in 2014. He is currently working as a research assistant in the Department of Computer Engineering at Dokuz Eylul University. His research interests include game-based learning, educational systems, artificial intelligence and data mining. He has been involved in several projects on education.

Kokten Ulas Birant received his B.S., M.S. and Ph.D. degrees in computer engineering from Dokuz Eylul University, Turkey, in 2000, 2002 and 2006, respectively. From 2000 to 2007, he was a Research Assistant; from 2007 to 2011 he was a Lecturer; and since 2011 he has been an Assistant Professor at the Computer Engineering Department of Dokuz Eylul University. His research interests include software engineering and gamification.

Pelin Yıldırım received her B.Sc. and M.Sc. degrees in computer engineering at Dokuz Eylul University, Izmir, Turkey, in 2013 and 2015, respectively. She began working on a Ph.D. in computer engineering at Dokuz Eylul University in 2015. She is currently working as a research assistant in the Department of Software Engineering at Manisa Celal Bayar University. Her research interests include data mining and intelligent learning systems.

Derya Birant received her B.S., M.S. and Ph.D. degrees in computer engineering from Dokuz Eylul University, Izmir, Turkey, in 2000, 2002 and 2006, respectively. From 2000 to 2007, she was a Research Assistant; from 2007 to 2009, she was a Lecturer; and since 2009, she has been an Assistant Professor in the Computer Engineering Department of Dokuz Eylul University. She was a Visiting Lecturer at the South East European University in 2006. She is the author of three book chapters and more than 40 publications. She serves on the editorial boards of several journals and is a reviewer for major scientific journals. Her research interests include data mining and intelligent systems. She has been involved in various long-term projects on education. Dr. Birant has several "Most Downloaded Article" certifications because her publications are ranked 1st in several books and journals in 2013, 2014 and 2015. She has graduation awards, ranking 2nd both in the Faculty and the Department. She was also the recipient of the "Outstanding Achievement Award" in 2010.