Electronic Letters on Science & Engineering 15(3) (2019)
Available online at dergipark.org.tr/else

Activity Suggestion Decision Support System Design In Online Learning Environment

Hakan KÖR[1]* , Hasan ERBAY[2] , Melih ENGİN[3]

[1]Department of Computer Programming, Sungurlu Vocational School, Hitit University
[2]Departmet of Computer Engineering, Faculty of Engineering, Turkish Aeronautical Association University
[3]Department of Management Information Systems, Faculty of Inegol Business, Uludag University

Abstract: Decision support systems are created for organizations to enable decision-makers to have correct and more reasonable actions. These systems are made available to students and administrators in online education environments to achieve higher success rates. In online learning environments, students utilize different types of course materials and interaction tools, which provides reaching a higher success rate in a considerable amount. However, students often difficult to choose suitable course content and activities that will positively affect their academic performance. In this study, the decision support system model is constituted for students and lecturer in terms of online learning environments. The model helps students choose the best activity by processing their previous data. Data mining methods have been used in the decision-making process. Possible features and data for the data warehouse are obtained through moodle learning management system. Then, the attributes that contributed to improving the performance of the model were filtered to implement the data mining process. In the data mining process of the research, many decision tree algorithms have been used for successful predictions. However, it has been seen that the C5 algorithm performs better than other decision tree algorithms. In addition to the data mining process, the demographic structure of the sample, weekly success rates, and number of course document usage were added to the model to improve performance in various statistical information. In order to create this model, the data were obtained in the moodle learning management system. So that increase students' academic success, the activity selection decision support model is suitable for integrating into the moodle system and can be included as an add-on to the system.

Keywords: Online Learning; Educational Data Mining; Classification Methods; Decision Tree Algorithms; C5 Decision Tree Algorithm

Reference to this paper should be made as follows :
Hakan KÖR , Hasan ERBAY , Melih ENGİN, 'Activity Suggestion Decision Support System Design In Online Learning Environment', Elec Lett Sci Eng , vol. 15(3) , (2019), 8-22

1. Introduction

As a result of the developments in information technology in recent years, computer science has also positively affected the education and training activities. In addition, today's mathematical and statistical algorithms and computers' high computing and visualization capabilities are integrated into the solution of real-life problems, especially in education and training [1]. Distance education has also taken its share from these developments and methods and techniques used in distance education have been developed and diversified. Academic studies evaluating distance education performance positively affected the distance education preference rate with these developments. Many hardware such as smartphones, IP based remote control devices, and smart meter systems, which have become an indispensable part of our lives, transfer the data they produce through their sensors to various applications. Thus, there is an exponential increase in the amount of data produced, stored and transmitted. In a study called "Digital Universe Study" prepared by IDC, a research organization, it is estimated that the amount of data to be reached in 2020 will be 44 times that of 2009 and the annual data volume will reach 35 zettabytes [2]. Raw data is likened to raw

This paper is based on a PhD study titled "Activity Suggestion System Design In Cloud Based Online Learning Environment: Educational Data Mining Application".

* *Corresponding author;hakankor19@gmail.com*
ISSN 1305-8614

oil and has little value. However, just as raw oil is used in different fields by processing, it can be converted into meaningful information by processing raw data. In other words, meaningful relationships, clusters, and patterns that are previously unknown are discovered through data mining techniques to make the decision-making process more efficient. In many important applications, data must be processed quickly and rendered meaningful. Thanks to concurrency, the old and new data are combined instantly, using the ones that contribute positively to the result from the data, while those that contribute negatively are excluded [3].

Contrary to the foregoing, the current student information systems used in many universities follow only the basic records of students, such as some demographic information, course follow up. However, these systems can be made more effective thanks to online learning activities follow for students.

On the other hand, in today's online learning management systems, there are many types of activities prepared using web 2.0 technology. However, students are not aware of which of these activities is most effective in terms of academic achievement. Considering that these activities are directly related to the general profiles and qualifications of the students, many variables such as age, gender, type of school, level of education, place of residence and attitudes towards online learning are effective.

The aim of this study is to create an educational data mining application that uses online learning activity data of students together with their general characteristics. With this application, classification was made on the data collected according to predetermined variables. Possible end of-term achievements of the students included in the system are predicted on a subject-by-subject basis, and then the students are offered activities that they need to do online to succeed. In short, this study aims to establish an online activity suggestion system that increases students' academic achievement.

1.1. Research Purposes

In this study, the objectives can be listed as follows; to predict students' academic achievement from student profiles and activities on the online learning environment systems. To determine the effect of different classification algorithms on predicting success in data mining process. To determine the effect of all variables used on educational data mining process on the suggestion success of the system. To determine the variables that contribute the most from the variables used in the suggestion system as a result of the analysis of educational data. Creating a web-based activity suggestion model with obtained results.

2. Literature Review

Many methods in decision support systems are used. Each of methods helps as long. In order to move educational institutions forward and universities need fast, accurate decision-making support systems.

Different methods are used in many distinct Decision Support System (DSS) systems. The DSS makes decisions based on algorithms appropriate to the field of application. In order to have rapid and healthy decisions, it is required for universities that support systems to move forward with their training schemes. For decision-making functions, these systems need tools that collect, organize, and evaluate the necessary information. It is significant for foundations to utilize from these tools in order to create a competitive environment.

In his study titled Decision Support System (DSS) in Higher Education System, Khalid A. Fakeeh developed the RAPID decision model by using a method of evaluation of the activities of students, researchers, and professors. There exists five different meaning within the term of RAPID, which is regarded as to research, accept, propose, implement and decide.

The literature illustrates that the main ingredients of the EDM are students' demographical characteristics, attitudes, and/or their activities that what they have acted during the teaching process. Guruler and others [4] classified the students according to their demographic characteristics and the exam scores of those through the MUSKUP (Mugla University Student Knowledge Discovery Unit Program) program which is developed at Mugla University. The study explores the relationship between student characteristics and achievements by using the Microsoft Decision Tree Classification technique. However, the study does not take students' activities according to their exams into account. In fact, the MUSKUP classifies the students.

In addition to the periodic test performance of the students, if the constantly monitoring and analyzing their academic activities are provided, then the instructors could choose the most effective method for the current students. In short, by focusing on the data analysis, the instructors can provide better learning process in many different ways [5]. At this point, online tools are the backbone of that enable instructors to assess a much broader range of student actions, such as how much time those spend time to read, where students reach the references to comprehend the topic [6].

Educational Data Mining is a field of knowledge in which the educational institution has a limited relationship with mining data. EDM basically uses educational data in computational approaches to analyze some information about the context of learning and teaching activities. EDM can manage teaching course contents and learning environment through a Learning Management System [7]. Recently, different learning platforms, such as Moodle, have been used to meet the needs of students, educators and administrators. These platforms have gained tremendous value for educators; The mining of large data is still necessary for various interesting patterns and facts to emerge in the decision-making process for the benefit of the students [8] [9].

In another study, under the title of 'Data Mining and Learning Analytics', MOODLE 2.0 course data analysis was conducted from 84 undergraduate students in psychology undergraduate program. The program consists of 11 different units sent to students once a week and each student was able to work in each unit for a period of 15 days. In this study, clustering method was applied on MOODLE activities. Examined the behavior of the unsuccessful student and the successful student. Therefore, from an educational and practical point of view, these models can easily be used to indicate which new students are at risk of failing a course in order to be able to use this information to provide feedback to instructors about student learning [10].

Romero and Ventura focused on educational data science. Educational data mining, learning analysis, academic, institutional analysis system, learning analysis, personalization, systematic and instructional analysis are the topics examined in educational data science. Analysis of students' MOOC interaction data can provide useful information to teachers, resource creators, and members of the organization that are working to improve their MOOCs by pointing out important topics in the lesson [11].

Educational data mining includes data mining methods and tools for processing education-related data, usually collected using an e-learning platform. When the student interacts with the platform an e-learning platform focuses on data stored in the database, per session, or in total, between users' interactive activity platform access times, between events, and the time between the student's actions, such as the exam, session, or exam notes. These time values hold information about a student's interaction with the platform. In addition, several optimally placed PDF parameters used as a feature vector result in the division of learning content pieces into similar sets of "characters". Clustering results can then be used as a recommendation to the course designer and trainer to improve the content structure or distribute portions of the course content in an optimal way [12].

In the article which is published by Wanli et al. (2015), the authors developed the Genetic Programming-Interpretable Classification Rule Mining Model (GP-ICRM). They observed students' manners during a certain period of time [13]. Finally, each student is presented with concrete and individual suggestions for better learning motivation. In this way, the awareness of

the learning process increases regularly. Identifying similar groups of students from different dimensions is another key concept for EDM researchers, such as individual differences, learning preferences, personal characteristics, and the usage of data. Classification and clustering are the most preferred DM methods by researchers for this purpose [14].
From the perspective of higher education, there are many discrete and independent teaching methods which include online, mixed, and web-based techniques. According to Sloan's United States online report, in the fall of 2009, 5.6 million students willingly took at least one online course. This number has increased by over one million in comparison with the previous year. The rate of increase is approximately more than 20%. It is also stated that 30% of the higher education student potential receives at least one course online [15].
The access of course content is often provided via the Internet in terms of online learning. However, it is difficult to produce an effective and essential teaching implementation by disregarding the dimensions of education by paying attention only to the technology [16]. The concept of the online learning is considered as a system which is based on learning and it is associated with a few learning principles and web-based learning by researchers [17].
There are plenty of tools in terms of online learning environments to contribute to the learning and teaching process. These tools could be utilized to prepare homework and tests, share course materials, and conduct discussions by instructors [18]. The online environment enables students to blog, wiki, forum, message, etc. in a very large scale [19]. In other words, virtual internet platforms have been established and designed in order to allow creating and sharing information and experiences of individuals through online learning environments [20].
Registered group members in online environments are fundamentally interested in activities, such as acquiring new information, reaching more useful resources, facilitating problem solving skills, and promoting professional skills to enhance interpersonal communication found that team learning has a positive relationship with knowledge sharing, and social capital [21, 22]. Both have been positively related with information sharing and team learning in their study of information sharing, team learning, social capital, and digital loyalty in virtual communities.
Educational support systems are often considered as cognitive tools though those are precisely operated as tools that have been presented in the form of topic maps. For instance, course content is offered modularly online in order to empower individuals to learn on their own pace. Since the previous knowledge, understanding, perception and learning manners of an individual are not identical with each other, it would not be appropriate to educate every individual in the same manner.

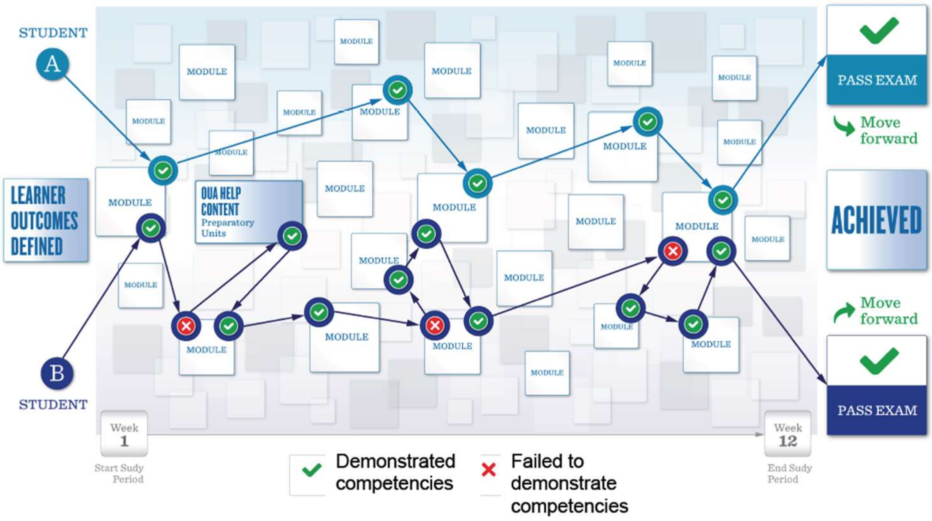


Figure 1. Personalized learning model

In this context, as shown in Figure 1 [23], it could be declared that the provision of modules in the consideration of the speed of every single individual would increase and enlarge the effectiveness of learning process. In some cases it is observed that individuals who have faced with problems during learning undertaking have effortlessly progressed by passing through the modules, after those are guided to struggle with recent module in order to solve the difficulty.

DSS is the educational support of each person by the unit manager and expected to reduce the pressure on the person. Investigation of decision-making systems shows a proportion of variables that can contribute to productive decisions [24]. Educational Institutes and Universities should not aim to produce and present not only a fair environment but also creative and profitable data related to the financial, administrative and social environment. People at different levels, such as senior managers, rectors, deans, experts and different teammates, are effective in helping decision-makers make decisions [25]. It is equally important to add assessment tools to an application, along with the use of a talented organization, in the progress of the activities of the top educational institutions. In DSS systems, there are several methods, each of which helps the process in different ways. The DSS then makes decisions based on algorithms derived from an understanding of the field of application. As a result of creating the highest level of competition in educational conditions, higher education institutions endeavor to improve the learning environment, equip them with appropriate information and services, and develop newly developed tools. Decisions in all areas are faced with dynamic, dangerous information and uncertainty, as well as excessive information and overly intensive conditions. Educational institutions or universities need a convincing decision to strengthen the tools to help the management system. One of the first attempts to use a DSS in state-of-the-art education would be to make a valid academic logical tool for collecting, organizing, and evaluating compulsory decision-making data and information [26]. Such tools provide an effective Management Information System (MIS) [25].

The techniques of data mining and data analysis aid to advance immediate feedback for both students and teachers about the academic performance of students. Therefore, the underlying reasons can be predicted and anticipated whether they attempt to drop a course, needing extra help, or capability of facing with challenging tasks. Moreover, the most effective pedagogical approaches for certain students can be smoothly determined [27].

Although the history of data mining procedure is bounded on 1980's, the possible development position is not placed very well, it could not improve itself; that is why, the procedure is still ongoing on to be advanced and upgraded. Therefore, it could not be denied that there have been collapse and disintegration of the group of data mining and the analysis of algorithm. An automobile designer, the provider of insurance and the manufacturer of the software and hardware came together in order to consolidate and reform the data mining as well as constitute a standard in 1999. The CRISP-DM was created as final results of researches which had been arranged in order to empower the data mining process as a cross institute standard.

In addition, the ability to predict students' end-of-term performance has acquired an increasing importance in education [28]. Previous prediction models have principally focused on the statistical modeling and data mining techniques [29]. These modeling techniques may have their own limitations. And the EDM have emphasized the development of models and algorithms to strengthen estimates of learning consequences [30]. Nevertheless, there is a lack of existing statistical and EDM methods in the light of optimizing the prediction success rate in terms of. Poor or misinterpretations arise when the basic requirements of regression models are ignored or disregarded [31].

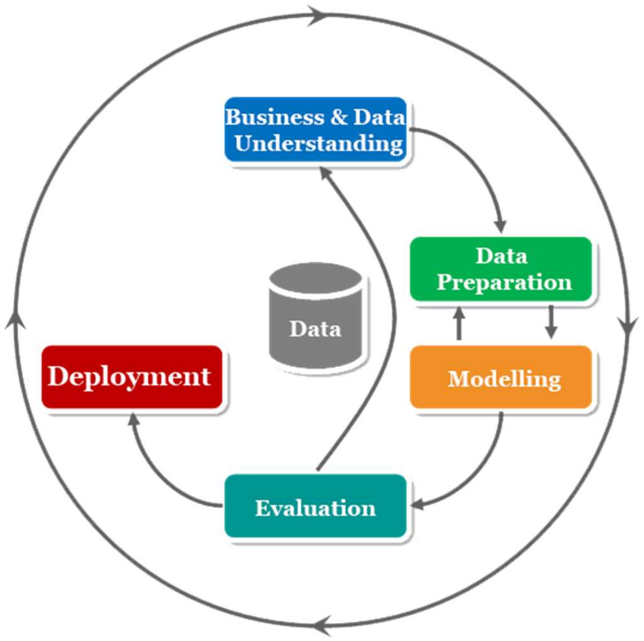


Figure 2. CRISP-DM operational model

Traditional data mining process contains six steps as visualized in Figure 2 [32]. When the figure is examined and analyzed, there is no doubt that the "what is to be wished?" question is replied comprehensively. The rest is existed and consisted by the informed data of creating the model, evaluation, and deployment.

The abstraction of online learning performance has a quite disparate interpretation from the traditional classroom settings in the case of online learning environments allow to record students' traces as they leave data behind while performing learning activities. The data may be the students' logging onto and off from the online environment, interaction with course materials, and answering a question in the discussion forum. Nowadays, in the consideration of the increase in the number of students enrolled in online environments, it is uttered that the serious amounts of learning processes are recorded in online databases. However, the use of these recorded data is limited to only simple graphics and descriptive statistics which ought to aid researchers to improve education [33]. In addition to this, the DM methods which have been used to identify hidden patterns and a tendency in databases in different fields have an important potential for the analysis of such datum from educational environments. The DM techniques could be effective in terms of support, encouragement and feedback for students, assessment of learning theories, early warning systems, performance estimates, learning techniques, and the development of future learning applications in accordance with educational environments [34]. Detection of learning groups that present similar behavior patterns is an example of this. In this context, similar student groups can be classified by considering individual characteristics, learning preferences, and individual dissimilarities [18]. In this way, instructors and teachers can operate data base to observe and record the development of students. Furthermore, in order to produce valuable information in terms of developing appropriate and accurate intervention methods for the students who have difficulties? This data can also be used in adaptive learning environments to automatically classify students and automatically adapt similar student groups.

Decision Tree, Nearest Neighbor, and Bayesian are counted as the most common and frequent data mining algorithms in the EDM studies. Different arrangement methods are perceived as the backbone of the decision trees, which are ordered as ID3, C4.5, C5.0, CART, CHAID, and QUEST.

3. Method

In this section of the research, it is going to be presented research model objectives along with data collection tools, research universe, sample selection, and data mining steps.

3.1. Sample

The universe of the research is composed of 3,988 students enrolled in one of the undergraduate and associate degree programs of the Distance Education Center at Hitit University during the 2015–2016 Spring semester. Among these, a data set was constituted by recording the online activity data of 1,760 students who took an online course named as Atatürk's Principles and Revolution History II. Having cleaned the data formation, 881 students who had attended online courses were selected as the sample. The objective sampling method has been operated in the selection of the sample. It could be noted that the objective sampling is the reflection of the universe to any simplification of the problem [35]. In descriptive research, at least a sample of 10% ought to be taken into consideration, 20% is required in small universes [36]. In this case, the sample is acceptable with a ratio of 24%.

3.2. Conceptual Model of Study

The main dimensions that have been included in the research could be arranged as demographic structures, academic averages, and online activities. In the same manner, a research model and research objectives have been generated objectively. The conceptual model of study can be smoothly perceived at shape. When the Figure 3 is examined carefully, online activity suggestion system could be revealed well.

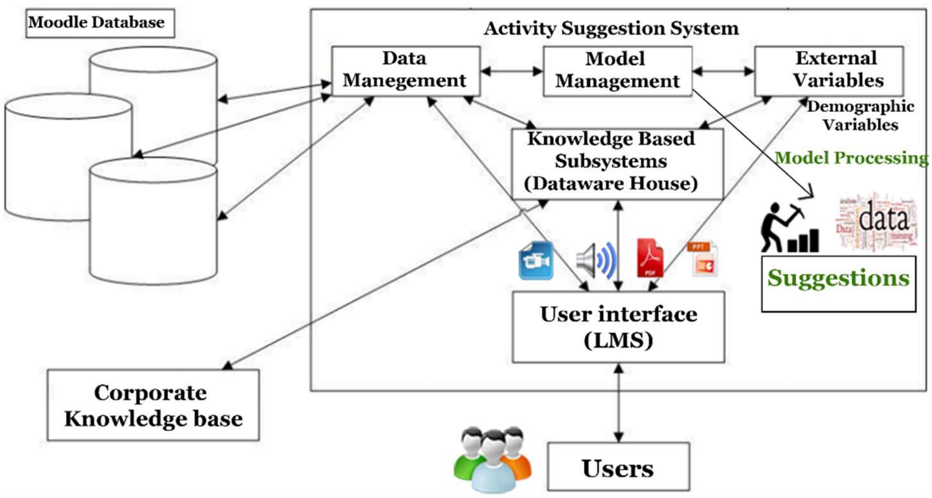


Figure 3. Conceptual model of study

3.3. Data Collection Tools

In this study, online learning software Moodle was designed as the data collection source. Student activities conducted in online learning environments, and profiles of students have been recorded. In the spring period of 2015–2016, the material and objective number of individuals who are keen on pursuit the Atatürk's Principles and Revolution History II lecture online is reached via data mining tool. In this study, the OYS program has been served as online learning environment platform towards learners. The policy of PHP is exerted as a programming language and MYSQL

is used as the data mining. The course material variety in accordance with the number of messages written of the students and the number of user reports (e.g., video, PDF, audio file, or presentation) is obtained through various queries through the database.
In this process, it is mentioned that the steps of students' profile datum and the activity records of students are deleted and compromised.

3.4. Examination of Student Profile Datum

The demographic features of students who have attended regularly can be easily distinguished and presented in Table1.

Table 1. Demographic Variables

Variable Name	Explanation
d1-gender:	Student gender is coded with 1 for male and 2 for female.
d2-place-of-residence:	The city you live in is coded with cities 1 and 2.
d3-mother-job:	Mother's job is coded with housewife 1, employee 2.
d4-father-job:	The father's job is coded with officer 1, self-employed 2.
d5-number-of-siblings:	The number of siblings is coded as 1 for a sibling, 2 for two and three siblings, and 3 for more than three siblings.
d6-mother-graduation:	The school where her mother graduated last was coded 1 for primary school, 2 for upper secondary school and above.
d7-father-graduation:	The school where her father graduated last was coded 1 for primary school, 2 for high school and above.
d8-family-income:	The monthly income of the family is coded with 1 up to 1500 TL, 2 between 1500 and 3000 TL, and 3 over 3,000 TL.
d9-high-scholl-type:	High school type where the student graduated was coded with 1 for vocational high school and 2 for normal high school

3.5. The Examination of the Online Learning Environment Data

The titles of interchangeable variables and expressions of students who have been willing to attend as a result of an activity under the name of OLS are considerably portrayed on Table2.

Table 2. Variables obtained from online learning environment

Variable Name	Explanation
d10-department:	Education unit where the student is registered, vocational school 1, license 2.
d11-email:	Does the student have e-mail?, yes 1, no 2 is coded.

d12-university-entrance-point:	The score of the student's placement in the university is 1 if below average, 2 if average and above.
d13-acceptance-point-type:	The type of placement in the unit is coded with numerical 1, verbal 2 and equal weight 3.
d14-PDF-follow-up-count:	Number of students tracking PDF material.
d15-audio-follow-up-count:	Number of followers of audio material.
d16-video-follow-up-count:	Number of follow-up videos of students.
d17-presentation-follow-up-count:	The number of follow-up of the presentation material of the students.
d18-login-count:	The number of students logged in.
d19-message-count:	Number of messages written by students.
d20-forum-views-count:	Number of forum views of learners.
d21-reporting-count:	The number of reviews of students' development reports.
d22-achievement-score:	Exam score of the students.

3.6. Preparing the Data

In this section, the implementation of obtaining data process is mentioned. The integration and merge of data commands are analyzed and uttered, which have been utilized to filter only the students who follow the online learning activities rather than all the other students in the study universe.

3.7. **Establishing the Model**

In this part of the study, it is described that the process of constitution of the model. Problem identification process that could be regarded as the first step of the data mining process is described as one of the goals and aims of this study objectively. The created model here is designed to answer research purposes. Existing datum that just before the model has been installed has been obtained and processed through the Learning Management System (LMS). Profile information, the number of the course material pursuit, forums, sending messages and the statistics of wiki usage of students have been acquired with SQL database from the Moodle LMS. Decision tree algorithms are meticulously used in the data of mining processes that have been obtained the data of online learning environment. Decision tree algorithms are perceived as it has a common usage for classification, also it is admitted to be preferred ordinarily. The results that have been gathered from objective model are illustrated on the "variables" section.

3.7.1. **Determination of the Most Appropriate Decision Tree Algorithm**

The obtained data from the sample is processed by SPSS Modeler which allows for free usage for 30 days in terms of data mining software. The data is sent to software in Excel format. After completed the type adjustment of data, it is processed through the automatic classification modeling algorithms.

It is observed that the data is revealed itself on 12 different classification methods. The most successful methods are emphasized comprehensively.

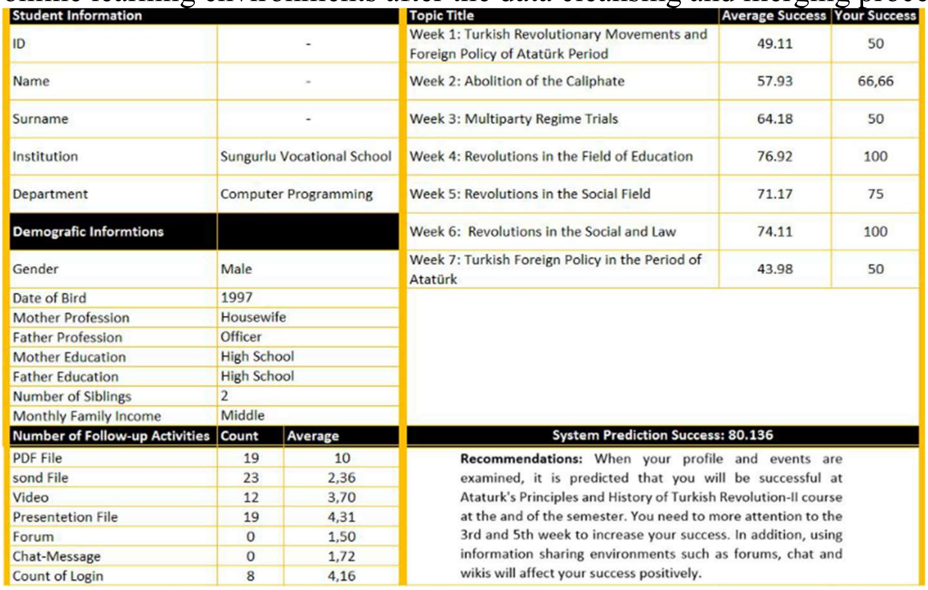
Table 3. The Rule Structure of Obtained Decision Tree

Model	Build Time (mins)	Max. Profit	Max. Profit Occurs in (%)	Lift Top(%30)	Overall Accuracy	No. Fields Used	Area Under Curve
C5	<1	417,273	14	1,87	80,136	16	0,731
Bayesian Network	<1	185.0	20	1,803	73,893	21	0,758
Neural Network	<1	135.0	16	1,638	72,304	21	0,72

When the Table 3 is analyzed, it ought to be emphasized that the average success of the C5, Bayesian Network and Neural Network has a higher performance than 70%, which has been recorded as the most successful ratio ever. According to this consequence, C5 algorithm has accomplished the best performance than others; the ratio has been noted as 80,136 %.

3.8. **Evaluation**

Data mining operations have been carried out with 22 variables that have been obtained from the online learning environments after the data cleansing and merging processes.



Student Information		
ID	-	
Name	-	
Surname	-	
Institution	Sungurlu Vocational School	
Department	Computer Programming	
Demografic Informtions		
Gender	Male	
Date of Bird	1997	
Mother Profession	Housewife	
Father Profession	Officer	
Mother Education	High School	
Father Education	High School	
Number of Siblings	2	
Monthly Family Income	Middle	
Number of Follow-up Activities	Count	Average
PDF File	19	10
sond File	23	2,36
Video	12	3,70
Presentetion File	19	4,31
Forum	0	1,50
Chat-Message	0	1,72
Count of Login	8	4,16

Topic Title	Average Success	Your Success
Week 1: Turkish Revolutionary Movements and Foreign Policy of Atatürk Period	49.11	50
Week 2: Abolition of the Caliphate	57.93	66,66
Week 3: Multiparty Regime Trials	64.18	50
Week 4: Revolutions in the Field of Education	76.92	100
Week 5: Revolutions in the Social Field	71.17	75
Week 6: Revolutions in the Social and Law	74.11	100
Week 7: Turkish Foreign Policy in the Period of Atatürk	43.98	50

System Prediction Success: 80.136

Recommendations: When your profile and events are examined, it is predicted that you will be successful at Ataturk's Principles and History of Turkish Revolution-II course at the and of the semester. You need to more attention to the 3rd and 5th week to increase your success. In addition, using information sharing environments such as forums, chat and wikis will affect your success positively.

Figure 4. Web based design of model

It is noticed that the C5 algorithm which is concerned as one of the classification algorithms after the process of data mining has the considerably accurate prudence with the success ratio of %80.136. Since the prediction is over 70%, this result has an acceptable classification performance.

3.9. Model Propagation (Application)

In this section, the applicability of the web based model and the example design are reflected in figure 4. When it is fussily examined, the fundamental information of students, demographic information and the number of activity following rates of students could be observed apparently. The purpose of this design is to inform the students in accordance with the created model in terms of the success rate of those as well as it is designed to inform the students in terms of their success rates in order to raise the awareness of the increase of success.

4. Findings

In this part of the study, findings obtained are outlined in charts. In the first part of the findings section, the following student information is introduced as; demographic characteristics, gender, city of residence, income level, high school type graduated, parents' occupation, and school where mother and father studied. In the second part of the findings section, the results of the data mining model are exhibited.

4.1. The Introductory Information of Participants

This section includes that the demographic information of the students who perform activities online and the types and amount of the material objectives that are followed by students regularly are illustrated. Figure 5 reveals that when the performing activities of students are examined, it is observed that 55.8% of those are male and 44.2% are female. Additionally, it is recorded that 91.3% of mothers are housewives, 75.3% of the education levels of fathers is limited by high school. For primary education, 59.7% of the students graduated from vocational high school, and 51.4% are signed up for an associate degree. Almost all of the students have e-mail accounts, whereas the number of placement and numeric equivalence points in university placement types is closed to students with verbal ability. In addition, it has been consistently observed that there is an absence of a huge gap among the ratio of the course materials, the participants who are willing to use PDF is on the top of the list with the percentage of 31.4%, 29.7% ratio has the presentation material, and 24.4% are in video material. Additionally, it is expressed that the rate of audio material is less than 19.5%.

Variable Type		N	%
Gender	Male	492	55,8
	Female	389	44,2
Place of Residence	City	445	50,5
	Metropolis	436	49,5
Mother's Profession	Housewife	777	88,2
	Working	104	11,8
Father's Profession	Government Official	259	29,4
	Self-employment	622	70,6
Family Income	Low	650	73,8
	Middle	198	22,5
	High	33	3,7
Mother's Education	Primary Education	804	91,3
	High School and Ove	77	8,7
Father's Education	Primary Education	663	75,3
	High School and Ove	218	24,7
Number of Siblings	One	21	2,4
	Two and Three	615	69,8
	Three and Over	245	27,8
Previous Graduation	High School	355	40,3
	Vocational School	526	59,7
Registered Department	Associate Degree	496	51,4
	Undergraduate	469	48,6
E-mail	Existent	879	99,8
	Absent	2	0,2
Score Type	Science Score	329	37,3
	Social Score	200	22,7
	Equal Weight Score	352	40,0
Entry Score	Low	529	60,0
	High	352	40,0
Type of Material	PDF	1707	31,4
	Audio	1062	19,5
	Presentation	1341	24,7
	Video	1329	24,4
Interaction	Login	2471	43,5
	Message	939	16,5
	Forum	1013	17,8
	Reporting	1261	22,2
Exam Score	Successful	622	70,6
	Unsuccessful	259	29,4
Achievement Average			60,7

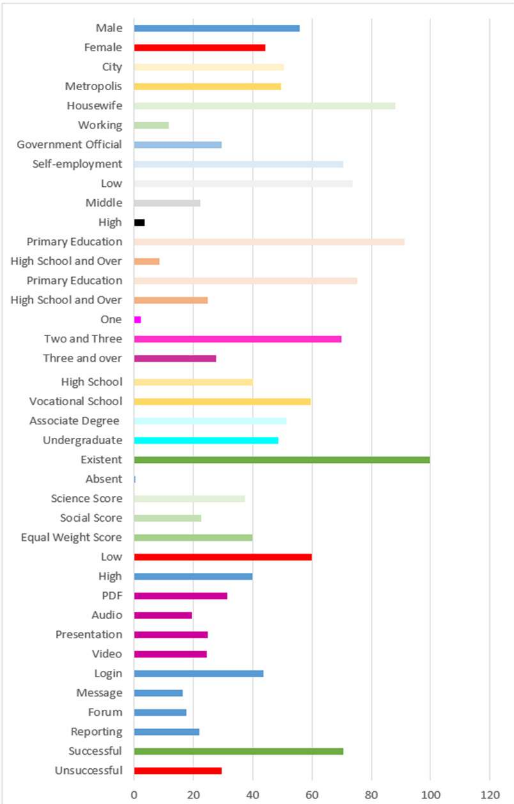


Figure 5. The demographic Characteristics of Students Who Follow the Topic Regularly

4.2. Model Results

In this section, the decision tree structure which is formed as a result of the established model, the success of the ability of prediction and the results are emphasized. Decision tree classification algorithms can be apparently indicated. The first purpose is tested with the obtained decision tree and the optimum success rate is estimated as 80.136%. The influence on the success rate of prediction of different classification algorithms is adequately examined and analyzed for the second purpose test during the period of the data mining process.

Table 4. Weekly Student Achievements

Sections	Number of Exam	Average Success
1st Section	2	49.11
2nd Section	3	57.93
3rd Section	2	64.18
4th Section	3	76.92
5th Section	4	71.17
6th Section	2	74.11
7th Section	4	43.98

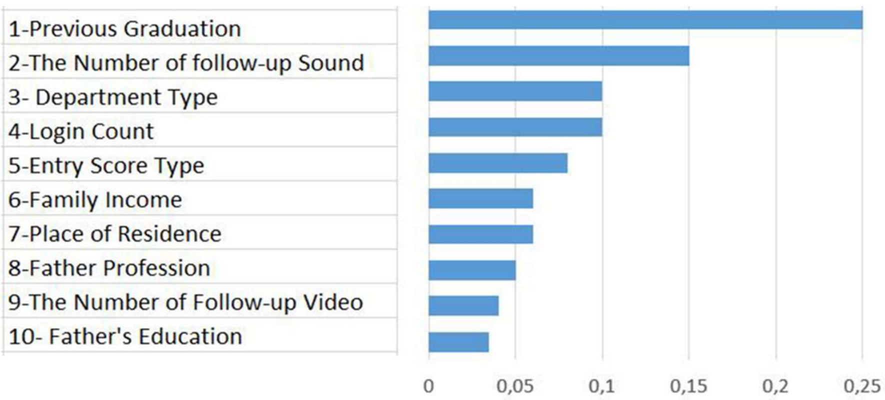


Figure 6. Influence of Variables on the Proposal Success

For the third purpose P3, it is conversely seen that the success differences of students in terms of topics on online learning environments. In addition to this, it is notably discovered that the lowest achievement occurred at the 7th topic while the highest achievement is at the 4th topic.

While the C5 algorithm is the first one, it has been determined that the second and third important effects have Bayesian and Neural Network algorithms.

For the purposes 4th and 5th, the influence of the variables used in the training data mining process on the proposal success has been examined with the C5 algorithm, and the result is presented in the table 3.

When Figure 6 is distinguished, it is noted that the prediction success has an effect of high school graduation level, the amount of follow-up sound material, department type, the number of logins, the type of settlement score in the university, monthly income of family, place of residence, the profession father, the amount of follow-up video material, and the education level of father. Therefore, it was determined that the most influential variable is the type of school where participants have graduated.

For the purpose P6, the application of the model to the process of designing the web-based application of the generated proposal model is included. Reporting and instant suggestions are presented using warehouse data generated from the existing Moodle LMS database.

5. Conclusions and Suggestions

In this section, conclusions and suggestions are categorized in the direction of the obtained findings. The findings and variables that have been obtained from the selected sample are examined interpreted significantly and suggestions are enabled to be served to the both students and institutions where students have been educated. Furthermore, recommendations are constituted for researchers who will conduct research in the future by making generalizations from sampling data.

The design and implementation of the online learning suggestion system according to the findings of the study indicate the following conclusions:

- It is seen that the variables that are considered as to contribute to the study at the beginning of the research are high school graduation type, audio material type, unit type, number of logins, type of settlement in the university, monthly income of the family, place of residence, the profession of father, video material type, and paternal education status. This result is included in the limitations section whether other factors have the impact or not. As the number of new sample students increases in the following process, more efficient data will be gathered from the LMS in this context. Then the system will acquire a more convincing learning ability and a better classification result.
- The obtained decision-making tree has an 80.136% ratio of predicting success. This result is acceptable and appropriate because the result is successfully higher than 70%.
- It has been determined that there is a lack of the tendency of participants in terms of the usage of the online courses in terms of Web 2.0 technologies such as forums, chats and wikis. The conception of Web 2.0 technologies are perceived as the abstract technological environments that provide information sharing among individuals. It ought to be implemented as solutions that encourage individuals to benefit from such technological assistances considerably.
- When student achievement is examined on the base subject, it appears that there is an existence of significant differences so that the record of subject success environments by calculating aids the preparation of study plan for students. In this way, the lecturer and professor can add supporting lecture materials, projects, or extra lecture hours to overcome the failure.
- In order to increase students' academic success, the activity selection decision support model is suitable for integrating into the MOODLE system and can be included as an add-on to the system.

5.1. Limitations of the Study

This study is limited to 881 students who take online courses from the Hitit University Distance Education Center. In this research, the classification of learning groups for 7 weeks in terms of the course follow-up process. The more the number of research wideness rises, the better results of classification success will be recorded.

For future research, it would be more beneficial for the students to present the subject-based classification and conclusions in terms of the subjects that are determined as online activities for seven weeks. Based on the online activities that were completed in the seven-week period, the learning group in the sample performed fewer activities than expected.

In the study, Moodle has been chosen as the LMS. The Moodle Learning Management System was introduced in 1998 and has become the most worldwide preferred system because of the wide range of language option. Nonetheless, the fact that the database structure that only belongs to

Moodle has brought some difficulties in terms of acquiring data. The detected variables and the database inquiries are combined with each other in order to examine previous of the data mining process.
This research is only limited to Hitit University. Different results may be acquired in terms of researches with samples from distinct universities.

References

[1] Cao, J., & Xiong, L. Protein Sequence Classification with Improved Extreme Learning Machine Algorithms, *BioMed Research International*, (2014), 1-12.
[2] Ohlhorst, F. Big Data Analytics Turning Big Data into Big Money. Hoboken, NJ, U.S.A.: J. Wiley and SAS Business Series, John Wiley & Sons, (2013).
[3] Sayad, S. *Real time data mining*. Cambridge: Self-Help Publishers, (2011).
[4] Guruler H., Istanbullu, & A., Karahasan, M. A. New student performance analysing system using knowledge discovery. *Computers & Education*, **55(1)** (2010), 247-254.
[5] Manyika J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh,C., & Byers, A. *Big Data: The Next Frontier for Innovation, Competition, and Productivity*, McKinsey Global Institute, 2011.
[6] Castro F., Vellido, A., Nebot, A., & Mugica, F. Applying Data Mining Techniques to e-Learning Problems. *Studies in Computational Intelligence*, **62(1)** (2007), 183-221.
[7] N., Hidayat, R., Wardoyo, S.N., Azhari. "Educational Data Mining (EDM) as a Model for Students' Evaluation in Learning Environment." *2018 Third International Conference on Informatics and Computing* (ICIC). IEEE, (2018).
[8] A.,Verma, S., Rathore and S.K., Vishwakarma. "Mining CMS Log Data for Students' Feedback Analysis." *Third International Congress on Information and Communication Technology*. Springer, Singapore, (2019).
[9] Villegas-Ch, W., and S. Luján-Mora. "Analysis of data mining techniques applied to LMS for personalized education*." 2017 IEEE World Engineering Education Conference (EDUNINE). IEEE*, 2017.
[10] Romero, C., Cerezo, R., Bogarín, A. and Sánchez-Santillán, M. "Educational process mining: a tutorial and case study using Moodle data sets." *Data mining and learning analytics: Applications in educational research* (2016), 1-28.
[11] C Romero and S Ventura. "Educational data science in massive open online courses." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **7(1)** (2017): e1187.
[12] Charitopoulos, A., Rangoussi, M. and Koulouriotis, D. "Educational data mining and data analysis for optimal learning content management: Applied in moodle for undergraduate engineering studies." *2017 IEEE Global Engineering Education Conference (EDUCON). IEEE*, 2017.
[13] Wanli, X., Rui, G., Eva, P., & Sean, G., Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory, *Computers in Human Behavior*, **47** (2015)**,** 168-181.
[14] Peña-Ayala A. Educational data mining: A survey and a data mining-based analysis of recent works. *Expert Systems with Applications*, **41(4)** (2014), 1432-1462.
[15] Diaz, V. Cloud-Based Technologıes: Faculty Development, Support, And Implementatıon, *Journal of Asynchronous Learning Networks*, **15(1)** (2011), 95-102.
[16] Horzum M. B. Interaction, Structure, Social Presence, and Satisfaction in Online Learning, *Eurasia Journal of Mathematics, Science & Technology Education*, **11(3)** (2015), 505-512.
[17] Pala F. K., & Erdem, M. Öğretmen Adaylarının Çevrimiçi Tartışma Ortamlarına Yönelik Görüşleri, *Turkish Online Journal of Qualitative Inquiry*, **6(2)** (2015), 24-47.

[18] Cristóbal, C.S. e-Learning as an Object of Study, *Universities and Knowledge Society Journal*, **7(2)** (2010), 1-11.
[19] Moreno L., Gonzalez, C., Castilla, I., Gonzalez, E., & Sigut, J. Applying a constructivist and collaborative methodological approach in engineering education. *Computers & Education*, **49(3)** (2007), 891-915.
[20] Cheung Christy, M.K., Lee, Matthew, K.O., Lee, & Zach, W.Y. Understanding the continuance intention of knowledgesharing in online communities of practice through the post-knowledge-sharing evaluation processes. *Journal of the American Society for Information Science & Technology*, **64(7)** (2013), 1357-1374.
[21] Tseng F. C., & Kuo, F.Y. A. Study of social participation and knowledge sharing in the teachers' online professional community of practice. *Computers & Education*. **72(1)** (2014), 37-47.
[22] Yao C.Y., Tsai, C.C. & Fang, Y.C. Understanding social capital, team learning, members' e-loyalty and knowledge sharing in virtual communities. *Total Quality Management & Business Excellence*. **26(6) (2015)**, 619–631.
[23] Ifenthaler, D. From Educational Data Mining To Automated Analysis Of Semantics, *Applied Teaching and Learning Research*. Retiered from http://documents.epfl.ch/groups /m/mo/mooc-admins/www/documents/Ifenthaler-LAW.pdf, 2014.
[24] Siemens G., & Baker, R. S. *Learning analytics and educational data mining: towards communication and collaboration*, 2nd international conference on learning analytics and knowledge, Vancouver, British Columbia, Canada, 252-254, 2012.
[25] Ghisoiue et al. *Designing a DSS for Higher Education Management*, Proceedings of CSEDU2009, Lisbon, Portugal, (2009), 335-340.
[26] Bresfelean, V. P., & Ghisoio N.(2010), Higher Education Decision Making and Decision Support Systems, *Wseas Transactions On Advances In Engineering Education*, **2(7)** (2010), 43-52.
[27] Bienkowski M., Feng, M., & Means, B.(2012). *Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief*, U.S. Department of Education, Office of Educational Technology. Washington, D.C, 2012.
[28] Baker R., S., & Inventado, P.S. *Educational Data Mining and Learning Analytics, Learning Analytics: From Research to Practice*. Ed: White J.A. Larusson and White. B. New York : Springer Science+Business Media, 2014.
[29] Gunnarsson B. L., & Alterman, R. Predicting failure: A case study in coblogging, Computers in Human Behavior. 2nd international conference on learning analytics. Vancouver, 29 Nisan - 2 May, 2012, British Columbia, Canada, 263-266, 2012.
[30] Siemens G., & Baker, R. S. *Learning analytics and educational data mining: towards communication and collaboration*, 2nd international conference on learning analytics and knowledge, 29 April-2 May, Vancouver, British Columbia, Canada, 252-254, 2012.
[31] Harrell F. E. *Regression Modeling Strategies: With Applications to Linear Models*, Springer, 2001.
[32] North M., *Data Mining for the Masses*. A Global Text Project Book, 2012.
[33] Ali L., Asadi, M., Gašević,D., Jovanović, J., & Hatala, M. *Factors Influencing Beliefs For Adoption of Alearning Analyticstool*: An Empirical Study. Computers & Education, **62(1)** (2013), 130-148.
[34] Greller W., & Drachsler, H.*Translating Learning into Numbers: A Generic Framework for Learning Analytics*, Educational Technology & Society, **15(3)** (2012), 42-57.
[35] Sencer M., *Toplum bilimlerinde yöntem*, Beta Basım, İstanbul, 1989.
[36] Gay L. R. , *Educational research: Competencies for analysis and application (3rd ed.)*. Columbus: OH: Merrill, 1987.