CHAPTER 8

BIG DATA IN EDUCATION: A CASE STUDY ON PREDICTING E-LEARNING READINESS OF LEARNERS WITH DATA MINING TECHNIQUES

Zeki ÖZEN*, Elif KARTAL, İlkim Ecem EMRE*****

*Ph.D., İstanbul University, Informatics Department, İstanbul, Turkey
E-mail: zekiozen@istanbul.edu.tr

**Ph.D., İstanbul University, Informatics Department, İstanbul, Turkey
E-mail: elifk@istanbul.edu.tr

***Res. Assist., Marmara University, Faculty of Business Administration,
Business Informatics Department, İstanbul, Turkey
E-mail: ecem.emre@marmara.edu.tr

DOI: 10.26650/B/ET06.2020.011.08

Abstract

Since the term "personalized learning" became popular, smart features have begun to be integrated into the e-learning environment. Data mining and machine learning algorithms are used to analyze big data stored in an e-learning system to make predictions to improve course quality or learners' performance. From the learners' perspective, it might now be considered possible for everybody to benefit from e-learning by considering their personal interests or their own specific development plan as long as the course contents are available in the system. In addition, in an e-learning environment, there is no limitation on the time and place where a course can be attended and a program completed. However, it is just not that simple. Today not the only, but by far the most important, the requirement is still the readiness of the learners to study in an e-learning system. The aim of this chapter is to predict the e-learning readiness of learners using data mining techniques. This chapter aims to provide feedback for institute managers and admin staff of e-learning systems which are intended to be used in an institution. Moreover, this section of the book contains one of the applications of big data analysis in education. Therefore, the main topic of this study is examined in terms of both classification and clustering techniques in order to provide a wider perspective to readers while using the sample application.

According to the results of this study, the highest accuracy value (0.831) is obtained with C4.5 Decision Tree Algorithm. While students, who agree and strongly agree with the statement "My studying/research area is appropriate for e-learning" are classified as ready to attend an e-learning course, students who disagree with the same statement are classified as not ready to attend an e-learning course. Students who strongly disagree with the statements "My studying/research area is appropriate for e-learning" and "E-learning is better than face to face learning", are also classified as not ready to attend an e-learning course. Furthermore, the statement "My studying/research area is appropriate for e-learning" is at the top of the obtained decision tree which indicates that it is an effective and directly related attribute which expresses student opinions about attending an e-learning course.

Keywords: Big data, Classification, Clustering, Data mining, Education

1. Introduction

E-learning is defined as "*instruction delivered on a digital device that is intended to support learning*" (Clark & Mayer, 2016). The e-learning concept can quite simply be argued on the basis of its two main pillars: technology and learners. From a technology perspective, e-learning is a serious investment. Technology directly affects techniques, tools, and applications that are used in e-learning. For training or educational purposes, e-learning management systems are used. There are many different open source and commercial alternatives for learning management systems (LMS), content management systems (CMS), virtual classrooms (VC), etc., that help to build and organize an e-learning environment. Course contents can be in various forms such as videos, text documents, podcasts, presentations, and such like, as long as they are compatible with the system. Learners can interact with each other and instructors via video and messaging features of the system. In addition, they can use their social media accounts or e-mails on the system for communication. Today, there are many other areas such as virtual reality and robotics that help instructors to improve the course contents and also the e-learning environment. The e-learning environment is also expected to work properly on desktop PCs, laptops, tablets, and mobile phones. Furthermore, any limitation of e-learning environment design mostly depends on the demands of the users. Improvements can be carried out using eye-tracking tools in terms of human-computer interaction for the evaluation phase of the system.

Since the term "personalized learning" became popular, smart features have been integrated into the e-learning environment as well. Data mining and machine learning algorithms are used to analyze big data stored in the system to make predictions to improve course quality or learners' performance. Big data means not only data with high volume, but it also indicates the production speed (velocity) of the data and the variety of the data resources (Zikopoulos et al., 2011). Big data in education includes courses, course scores of the students, content information (time to complete, repetition time, pause points, last access time, etc.), system information (the most frequent usage time, the most frequently used browser, tools to access the system, etc.), face recognition data / keystroke dynamics data of the users, social media shares, etc. (Özen, Kartal, & Emre, 2017).

From the learners' perspective, it might be thought that today everybody can benefit from e-learning by considering personal interests or development plans as long as the course content is available in the system. In addition, in an e-learning environment there is no limitation on the time and place where a course can be attended and a program completed. However, it is just not that simple. Today not the only, but by far the most important, requirement is still the readiness of the learners to study in an e-learning system. The aim of this chapter is to predict the e-learning readiness of learners using data mining techniques. This chapter aims to provide feedback for institute managers and admin staff of e-learning systems which are intended to be used in an institution. . Moreover, this section of the book contains one of the applications of big data analysis in education. Therefore, the main question examined by this study is addressed using both classification and clustering techniques in order to provide a wider perspective to readers while demonstrating the sample application.

The following section gives a quick overview of e-readiness and e-learning readiness. The third section will focus on the study method (understanding the dataset, data pre-processing stage, and data mining techniques and models). The results obtained from the models are given in the findings section and an interpretation of the results and intended future research are discussed in the last section.

2. E-readiness and E-learning Readiness

E-readiness is "*a measure of the quality of a country's Information and Communications Technology (ICT) infrastructure and the ability of its consumers, businesses and governments to use ICT to their benefit*" (Economist Intelligent Unit, 2009). In other words, it is not enough merely to have perfect software and ICT infrastructure in a country, but also the citizens of that country should have the necessary skills to use ICT. One of the most noticeable research studies in the literature is "E-readiness Rankings". These reports were published by The Economist, Economist Intelligence Unit with the collaboration of IBM Institute for Business Value. In these reports, the selected countries were compared and ranked in terms of the e-readiness concept. The following definition from the Economist Intelligent Unit (2005) also helps to define the concept of e-readiness:

> *"A country's e-readiness is essentially a measure of its e-business environment, a collection of factors that indicate how amenable a market is to Internet-based opportunities". ... "E-readiness is not simply a matter of the number of computer servers, websites and mobile phones in the country (although these naturally form a core component of the rankings), but also such things as its citizens' ability to utilize technology skillfully, the transparency of its business and legal systems, and the extent to which governments encourage the use of digital technologies."*

This description refers to the concept that e-readiness is not only about technological infrastructures or opportunities, but also the ability/skills of the users and the other fields that support or affect these technologies. These reports have been published as "E-readiness Rankings" since 2000. However, in 2010 the institutions decided to change the name of the research and called it "Digital Economy Ranking". The explanation about this change is reported in the 2010's rankings as quoted below (Economist Intelligent Unit, 2010):

> *"Since 2000, the Economist Intelligence Unit has assessed the world's largest economies on their ability to absorb information and communications technology (ICT) and use it for economic and social benefit. Previously titled the "e-readiness rankings", in 2010 the study is being renamed as the "digital economy rankings", to reflect the increasing influence of ICT in economic (and social) progress". … "Given the prevalence of Internet-connected consumers, businesses and governments, and the indispensable role that digital communications and services now play in most of the world's economies, we believe that the countries in our study have achieved, to one degree or another, a state of e-readiness. The study's new title, the "digital economy rankings", captures the challenge of maximizing the use of information and communications technology (ICT) that countries face in the years ahead".*

It seems that the concept of e-readiness was considered in the light of its technical and technological aspects from 2000 to 2010. In 2010, it was decided to consider the effects of technological developments on the economies, and thus the name of the report changed to "Digital Economy Rankings". Table 1 shows the rankings of Turkey between 2002-2010. In these reports, all countries are given a score of e-readiness and a ranking is made accordingly. Between 2002 and2008, the e-readiness score of Turkey steadily increased , however in 2008 it started to decrease. Furthermore, it can be seen on the table that through the years there has not been a major change in Turkey's ranking.

Table 1. E-readiness rankings of Turkey between 2002-2010

	E-Readiness Rankings								***Digital Economy Ranking***
	2002	***2003***	***2004***	***2005***	***2006***	***2007***	***2008***	***2009***	***2010***
E-readiness score (of 10)	4.37	4.63	4.51	4.58	4.77	5.61	5.64	5.34	5.24
E-readiness rank	40	39	45	43	45	42	43	43	43
Total number of countries	60	60	64	65	68	69	70	70	70

Source: *(Economist Intelligent Unit, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010)*

In a report on the most recent ranking study of The International Institute for Management Development (IMD) World Competitiveness Center (2019), The Digital Competitiveness Ranking Results of Turkey among 63 countries can be seen in terms of overall, knowledge, technology, and future readiness between 2015 and 2019 (Table 2). The report explains knowledge as the "*know-how necessary to discover, understand and build new technologies*", technology as the "overall context that enables the development of digital technologies", and future readiness as the "*level of country preparedness to exploit digital transformation*".

Table 2. IMD World Digital Competitiveness Ranking 2019 Results

	2015	***2016***	***2017***	***2018***	***2019***
overall	52	50	52	52	52
knowledge	59	58	60	59	60
technology	48	48	49	45	48
future readiness	42	42	40	42	41

Source: (IMD World Competitiveness Center, 2019)

With the help of e-readiness reports and different studies which focus on the e-readiness of specific groups or countries (Purcell & Toland, 2004; Rizk, 2004; Al-Solbi & Mayhew, 2005; Ifinedo & Davidrajuh, 2005; Princely Ifinedo, 2005; Beig, Montazer, & Ghavamifar, 2007; Zaied, Khairalla, & Al-Rashed, 2007), the e-learning readiness of learners can be defined as the grasp of essential know-how in the field of ICT and the ability to use e-learning systems sufficiently. Furthermore, e-learning readiness includes the psychological acceptance and demand of the learners, as well. According to studies in Turkey and around the world, e-learning readiness is a concept that has captured the attention of industries, schools, and universities. Studies in the literature can be classified according to the focus on their content. Some of them investigate the e-learning readiness from the perspective of the instructor and others from that of the learners.

There are many studies that have been conducted at the higher education level. Eslaminejad, Masood, & Ngah (2010) examined e-learning readiness from the perspective of the instructors. Hung et al. (2010) searched for the underlying dimensions of students' readiness for online learning. Mafenya (2013) conducted an observation of the pedagogical readiness of first-year students at the University of South Africa. A model was developed by Ha, Jm, & An (2014) in order to assess the e-learning readiness of lecturers in higher education institutions and the study was carried out with lecturers at the University of Nairobi. Parkes et al. (2015) investigated the readiness of students by surveying staff and students with previous learning experiences with e-learning. Paturusi et al. (2015) investigated lecturers' and students' readiness for e-learning at the University of Sam Ratulangi in Indonesia. Rasouli et al. (2016) investigated the readiness of art students in public Iranian Universities (Alzahra, Tarbiat Modares, and Tehran). Rohayani, Kurniabudi, & Sharipuddin (2015) reviewed the literature in order to investigate the factors that affect e-learning readiness in higher education by other researchers.

The study of Bonanno (2011) examines the usage of ICT technologies in education and this includes e-learning. The readiness in this study is investigated from a larger perspective. Akaslan & Law (2011a, 2011b) investigated the e-learning readiness of academic staff and students in the departments associated with the subject of electricity in different universities in Turkey. Soydal et al. (2011) made an assessment of e-learning readiness of academic staff at Hacettepe University. Ünal et al. (2013) conducted a study to assess the e-readiness of students at Hacettepe University. Doğan (2013) tried to observe the readiness in e-learning of lecturers at Osmangazi University. Korkmaz et al. (2015) investigated readiness and satisfaction in e-learning and their impact on academic achievement at Amasya University. Yurdugül (2016) concluded that the concept of e-learning is now more involved in education and thus the issue of eliminating deficiencies in the e-learning readiness level of university students has come to the fore. In the aforementioned study, the authors conducted a case study on university students and developed a scale for measuring e-learning readiness. As stated in that study there are other scales in the literature. Yilmaz (2017) conducted a study in a state university in order to observe the e-learning readiness of the students specifically in the flipped classroom model. Table 3 gives a brief summary of e-learning readiness studies in Turkey in terms of education levels and instructor/learner perspectives.

Table 3. Studies about e-learning readiness in education in Turkey

Title of the Study	***Reference***	***Education Level***	***Perspective***
Developing an instrument to assess teachers' readiness for technology-enhanced learning	(Bonanno, 2011)	-	Readiness of instructors
Measuring teachers' readiness for e-learning in higher education institutions associated with the subject of electricity in Turkey	(Akaslan & Law, 2011b)	University	Readiness of instructors
Measuring Student E-Learning Readiness: A Case about the Subject of Electricity in Higher Education Institutions in Turkey	(Akaslan & Law, 2011a)	University	Learner's readiness
Are Turkish universities ready for e-learning: A case of Hacettepe University Faculty of Letters	(Soydal et al., 2011)	University	Readiness of instructors
The Scale of Online Learning Readiness: A Study of Validity and Reliability	(Yurdugül & Sırakaya, 2013)	University	Learner's readiness
Evaluating E-Learning Readiness of Faculty of Letters of Hacettepe	(Moftakhari, 2013)	University	Readiness of instructors and learners
Students Readiness for E-Learning: An Assessment on Hacettepe University Department of Information Management	Ünal et al., (2013)	University	Readiness of learners
The examination of E-readiness levels of academicians	(Doğan, 2013)	University	Readiness of instructors
Two Concepts that Have to be Considered in the Transition of Vocational Colleges to E-Learning Model: Student's Computer Self Efficacy and E-Learning Readiness	(Pınar, Selçuk, & Dağ, 2014)	University, vocational college	Readiness of learners
E-Learning Readiness among Academic Staff in the Department of Information Science at the University of South Africa	(Ncube, Dube, & Ngulube, 2014)	University	Readiness of instructors
Assessing E-learning Readiness of Learners in Turkey	(Sharma, Gülseçen, Özen, & Kartal, 2014)	University	Readiness of learners
Assessing E-learning Readiness of Instructors in Turkey	(Sharma, Gülseçen, Özen, & Kartal, 2015)	University	Readiness of instructors
Students E-learning Readiness and Satisfaction Levels and Effects on the Academic Achievement	(Korkmaz et al., 2015)	University, vocational college	Readiness of learners
An investigation of Pre-service Teachers' Readiness for E-learning at Undergraduate Level Teacher Training Programs: The Case of Hacettepe University	(Yurdugül, 2016)	University	Readiness of learners
Iraqi Nursing Faculty Attitudes toward E-Learning a National Survey	(AL-Fayyadh & Mohammad, 2016)	University	Readiness of instructors
E-Learning readiness amongst nursing students at the Durban University of Technology	(Coopasami, Knight, & Pete, 2017)	University	Readiness of learners
Exploring the role of e-learning readiness on student satisfaction and motivation in the flipped classroom	(Yilmaz, 2017)	University	Readiness of learners

E-learning readiness is not only observed for educational purposes, but also for organizations/institutions (Aydin & Tasci, 2005; Lopes, 2007; Mercado, 2008; Schreurs, Ehlers, & Sammour, 2008; Schreurs, Moreau, & Ehlers, 2008; Darab & Montazer, 2011; Schreurs & Al-Huneidi, 2012; Azimi, 2013; Okinda, 2014; Kuruliszwili, 2015; Cathy, 2016; Doculan, 2016) or countries (Abas, Kaur, & Harun, 2004; Minges, 2005).

2. Method

This section covers the understanding of data, data preparation, and modeling steps of CRoss-Industry Standard Process for Data Mining (CRISP-DM) (Shearer, 2000). The evaluation step of CRISP-DM, which is related to evaluating the performance of the data mining models obtained, is also given in the Findings section.

2.1. Data Understanding and Data Preparation

In this study, data mining techniques are performed on “readiness” dataset, which is a subset of data that was used earlier for assessing the e-learning readiness of the learners in Turkey by Sharma, Gülseçen, Özen, & Kartal (2014). There are 667 observations (between the ages of 18 and 69, 330 male and 337 female) and 20 attributes in the “readiness” dataset. The statement “I am ready to attend an e-learning course” is taken as the target attribute of the survey (e-learning readiness status of a student); other statements are taken as predictive attributes. Age is numeric, gender, and e-learning readiness are binary. The rest of the attributes are coded according to the 5-Likert Scale (1=Strongly Disagree, 2=Disagree, 3=Have No Idea, 4=Agree, 5=Strongly Agree). Predictive attributes (except age) are treated as numeric in clustering analysis and as categorical in classification analysis. In addition, survey items are divided into three groups, namely “ICT skills” (S[1-5]), “e-learning experience” (E[1-6]), and “personal e-learning assessment” (A[1-8]). Figure 1 shows the class distribution of the target attribute which is designed based on the status of students regarding e-learning readiness. Table 4, Table 5, and Table 6 indicate the frequency (f) and the percentage (%) distribution of 5-Likert Scale attributes in terms of ICT skills, e-learning experience, and personal e-learning assessment. In this study, the authors focused particularly on predicting participants who are not ready for e-learning.

Dataset is normalized using min-max normalization technique at the data pre-processing stage. Data mining analyses are performed with C4.5 Decision Tree Algorithm for classification purpose and k-Means Algorithm for clustering purposes.

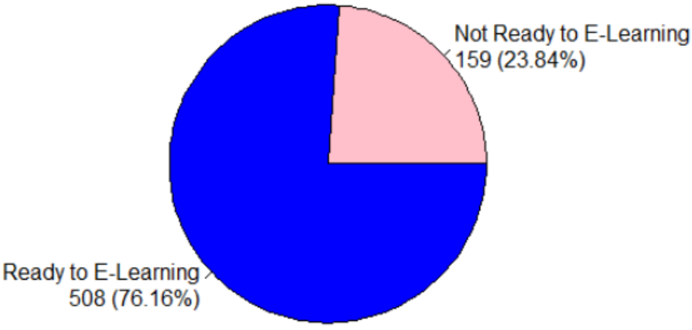

Figure 1: Class distribution of the target attribute

Table 4. Frequency (f) and percentage (%) distribution of 5-likert scale skill attributes

Attributes		***Strongly Disagree***		***Disagree***		***Have No Idea***		***Agree***		***Strongly Agree***	
		f	***%***	***f***	***%***	***f***	***%***	***f***	***%***	***f***	***%***
S1	I am good at using computer/internet	19	2.8	22	3.3	5	0.7	224	33.6	397	59.5
S2	I use my smartphone to communicate with my instructors outside the classroom	142	21.3	139	20.8	80	12	149	22.3	157	23.5
S3	I use social media for my courses	52	7.8	81	12.1	41	6.1	254	38.1	239	35.8
S4	I have a good e-learning background	93	13.9	133	19.9	178	26.7	176	26.4	87	13
S5	I have required IT infrastructure for e-learning	77	11.5	59	8.8	109	16.3	180	27	242	36.3

Table 5. Frequency (f) and percentage (%) distribution of 5-likert scale experience attributes

Attributes		***Strongly Disagree***		***Disagree***		***Have No Idea***		***Agree***		***Strongly Agree***	
		f	***%***	***f***	***%***	***f***	***%***	***f***	***%***	***f***	***%***
E1	I have instructors which live in different cities/ countries	231	34.6	165	24.7	101	15.1	96	14.4	74	11.1
E2	I ask questions to instructors by e-mail	118	17.7	78	11.7	64	9.6	198	29.7	209	31.3
E3	I have joined a video conference before	245	36.7	151	22.6	65	9.7	93	13.9	113	16.9
E4	I want to join different courses at different universities	30	4.5	30	4.5	42	6.3	159	23.8	406	60.9
E5	I have used smartboard before	216	32.4	116	17.4	54	8.1	105	15.7	176	26.4
E6	I have attended an online course before	217	32.5	109	16.3	47	7	124	18.6	170	25.5

Table 6. Frequency (f) and percentage (%) distribution of 5-likert scale assessment attributes

Attributes		***Strongly Disagree***		***Disagree***		***Have No Idea***		***Agree***		***Strongly Agree***	
		f	***%***	***f***	***%***	***f***	***%***	***f***	***%***	***f***	***%***
A1	I want lecture notes be shared electronically	35	5.2	27	4	20	3	147	22	438	65.7
A2	I prefer online exams because they are time saving and secure	107	16	95	14.2	92	13.8	143	21.4	230	34.5
A3	E-learning course content is different to face to face course content	59	8.8	71	10.6	99	14.8	232	34.8	206	30.9
A4	E-learning is better than face to face learning	174	26.1	186	27.9	176	26.4	69	10.3	62	9.3
A5	My studying/research area is appropriate for e-learning	84	12.6	101	15.1	124	18.6	221	33.1	137	20.5
A6	I prefer e-learning instead of face to face learning	164	24.6	185	27.7	121	18.1	102	15.3	95	14.2
A7	My instructors have enough IT skills for e-learning	86	12.9	107	16	258	38.7	131	19.6	85	12.7
A8	My university has the required IT infrastructure for e-learning	115	17.2	81	12.1	231	34.6	128	19.2	112	16.8

2.2. Modeling

In this study, our first stage of research involved the use of C4.5 Decision Tree Algorithm, developed by Quinlan (1993). It is one of the supervised learning algorithms that is used for classification problems. The aim of the C4.5 Decision Tree Algorithm is to find the best split node (attribute) using information gain and gain ratio and to create a set of rules in a tree form (Han & Kamber, 2006). These rules are used to predict the class label of an unlabeled observation. In addition, since it is the variable that provides the best partition, the algorithm also reveals the order of importance of the predictive attributes.

Stratified 5-fold cross-validation is used as a model performance evaluation technique for C4.5 Algorithm. In other words, the dataset is randomly divided into five pieces and each piece is used as testing dataset and the rest of the pieces are used as training dataset recursively. Each time accuracy value, error rate, sensitivity, specificity, positive predictive value, negative predictive value, and F-Measure are used as model performance evaluation metrics. At the end of the analysis, the mean of these metrics is used to evaluate the total model performance. In addition, the ratio of the class values of the target attribute is preserved in training and test datasets by means of stratified sampling.

The target attribute used in classification was based on the self-assessment of the respondents about their e-learning readiness. However, in similar analyzes to be performed with data collected through an e-learning system, a target attribute may not always be available in the dataset. Therefore, k-Means Algorithm, which is one of the clustering techniques, was used. This algorithm was also helpful to see how coherent the self-assessment of the participants was.

k-Means Algorithm is performed in order to cluster instances according to their similarities. This algorithm is one of the unsupervised learning algorithms that can be used when the target attribute does not exist in the dataset. The main aim of the algorithm is to cluster instances by minimizing the distance within clusters and maximizing distances between clusters. Since the dataset is intended to split into two groups, namely ready or not ready for e-learning, the cluster number k is determined as 2 for the k- Means Algorithm. After the k-Means Algorithm is performed, the class labels (ready or not ready for e-learning) are matched with the clusters according to the correlation between predictive attributes and the target attribute. Therefore, it can be said that all observations are labeled using k-Means Algorithm. Finally, new class labels are compared with the real class labels of the target attribute (Balaban & Kartal, 2015).

All the analyses are performed using R programming language (cran.r-project, 2019) on RStudio (RStudio, 2019). The following R packages are used: caret (Kuhn, 2018), clusterSim (Walesiak & Dudek, 2016), RWeka (Hornik, Buchta, & Zeileis, 2009; Witten & Frank, 2005), stats (R Core Team, 2018), and xlsx (Dragulescu & Arendt, 2018).

3. Findings

The performance of C4.5 Decision Tree Algorithm is given in Table 7. Performance evaluation metrics are calculated for each fold, then average performance results are given in the last column.

Table 7. Performance of C4.5 Decision Tree Algorithm

	Fold 1	*Fold 2*	*Fold 3*	*Fold 4*	*Fold 5*	*Mean*
Accuracy	0.836	0.806	0.865	0.820	0.827	0.831
Error Rate	0.164	0.194	0.135	0.180	0.173	0.169
Sensitivity	0.469	0.469	0.548	0.469	0.469	0.485
Specificity	0.951	0.912	0.961	0.931	0.941	0.939
Positive Predictive Value	0.750	0.625	0.810	0.682	0.714	0.716
Negative Predictive Value	0.851	0.845	0.875	0.847	0.848	0.853
F-Measure	0.577	0.536	0.654	0.556	0.566	0.578

Figure 2 shows the C4.5 decision tree obtained from the best cross-validation fold (Fold 3). Decision rules can be extracted from this decision tree such as:

- **Rule 1:** IF a student “strongly disagrees” with the statement “My studying/research area is appropriate for e-learning” AND “E-learning is better than face to face learning”, the E-learning Readiness Status of the Student is NO.
- **Rule 2:** IF a student has “no idea” about the statement “My studying/research area is appropriate for e-learning” and “E-learning is better than face to face learning” AND strongly agrees with the statement “I want to join different courses at different universities”, the E-learning Readiness Status of the Student is YES.
- **Rule 3:** IF a student “agrees” with the statement “My studying/research area is appropriate for e-learning”, the E-learning Readiness Status of the Student is YES.
- **Rule 4:** IF a student “strongly agrees” with the statement “My studying/research area is appropriate for e-learning”, the E-learning Readiness Status of the Student is YES.

Two models were performed using C4.5 Decision Tree Algorithm. C4.5 is first performed using real class labels as target attribute and secondly using cluster labels obtained using k-Means Algorithm as real target attribute. As seen in Table 8, specificity and negative predictive values are better than sensitivity and positive predictive values, because of the imbalanced dataset used in the analyses.

Table 8. Performance comparison of C4.5 decision tree models

Performance Evaluation Metric	***Classification with real class labels as target attribute***	***Classification with cluster labels as target attribute***
Accuracy	0.831	0.700
Error Rate	0.169	0.300
Sensitivity	0.485	0.868
Specificity	0.939	0.648
Positive Predictive Value	0.716	0.435
Negative Predictive Value	0.853	0.940
F-Measure	0.578	0.580

```
J48 pruned tree
------------------

My_studying/research_area_is_appropriate_for_e-learning = Strongly_Disagree
|   E-learning_is_better_than_face_to_face_learning = Strongly_Disagree: No (49.0/7.0)
|   E-learning_is_better_than_face_to_face_learning = Disagree: Yes (11.0/4.0)
|   E-learning_is_better_than_face_to_face_learning = Have_No_Idea
|   |   I_want_to_join_different_courses_at_different_universities = Strongly_Disagree: No (2.0)
|   |   I_want_to_join_different_courses_at_different_universities = Disagree: Yes (0.0)
|   |   I_want_to_join_different_courses_at_different_universities = Have_No_Idea: Yes (0.0)
|   |   I_want_to_join_different_courses_at_different_universities = Agree: Yes (1.0)
|   |   I_want_to_join_different_courses_at_different_universities = Strongly_Agree: Yes (4.0/1.0)
|   E-learning_is_better_than_face_to_face_learning = Agree: Yes (3.0/1.0)
|   E-learning_is_better_than_face_to_face_learning = Strongly_Agree: No (1.0)
My_studying/research_area_is_appropriate_for_e-learning = Disagree
|   I_want_to_join_different_courses_at_different_universities = Strongly_Disagree: No (5.0/1.0)
|   I_want_to_join_different_courses_at_different_universities = Disagree: No (5.0)
|   I_want_to_join_different_courses_at_different_universities = Have_No_Idea
|   |   Gender = Male: Yes (2.0)
|   |   Gender = Female: No (6.0/1.0)
|   I_want_to_join_different_courses_at_different_universities = Agree
|   |   I_have_a_good_e-learning_background = Strongly_Disagree: No (4.0)
|   |   I_have_a_good_e-learning_background = Disagree: Yes (11.0/2.0)
|   |   I_have_a_good_e-learning_background = Have_No_Idea: No (2.0/1.0)
|   |   I_have_a_good_e-learning_background = Agree
|   |   |   I_have_used_smart_board_before = Strongly_Disagree: No (0.0)
|   |   |   I_have_used_smart_board_before = Disagree: Yes (2.0)
|   |   |   I_have_used_smart_board_before = Have_No_Idea: No (0.0)
|   |   |   I_have_used_smart_board_before = Agree: No (3.0)
|   |   |   I_have_used_smart_board_before = Strongly_Agree: No (0.0)
|   |   I_have_a_good_e-learning_background = Strongly_Agree: Yes (1.0)
|   I_want_to_join_different_courses_at_different_universities = Strongly_Agree: Yes (41.0/9.0)
My_studying/research_area_is_appropriate_for_e-learning = Have_No_Idea
|   I_want_to_join_different_courses_at_different_universities = Strongly_Disagree: No (3.0/1.0)
|   I_want_to_join_different_courses_at_different_universities = Disagree: No (5.0/1.0)
|   I_want_to_join_different_courses_at_different_universities = Have_No_Idea
|   |   I_have_instructors_which_live_in_different_cities/countries = Strongly_Disagree
|   |   |   My_university_has_required_IT_infrastructure_for_e-learning = Strongly_Disagree: No (2.0)
|   |   |   My_university_has_required_IT_infrastructure_for_e-learning = Disagree: Yes (1.0)
|   |   |   My_university_has_required_IT_infrastructure_for_e-learning = Have_No_Idea: Yes (4.0)
|   |   |   My_university_has_required_IT_infrastructure_for_e-learning = Agree: Yes (0.0)
|   |   |   My_university_has_required_IT_infrastructure_for_e-learning = Strongly_Agree: Yes (0.0)
|   |   I_have_instructors_which_live_in_different_cities/countries = Disagree: Yes (4.0/1.0)
|   |   I_have_instructors_which_live_in_different_cities/countries = Have_No_Idea: No (2.0)
|   |   I_have_instructors_which_live_in_different_cities/countries = Agree: No (1.0)
|   |   I_have_instructors_which_live_in_different_cities/countries = Strongly_Agree: Yes (0.0)
|   I_want_to_join_different_courses_at_different_universities = Agree
|   |   Age <= 21: No (7.0/2.0)
|   |   Age > 21: Yes (18.0/2.0)
|   I_want_to_join_different_courses_at_different_universities = Strongly_Agree: Yes (60.0/8.0)
My_studying/research_area_is_appropriate_for_e-learning = Agree: Yes (169.0/14.0)
My_studying/research_area_is_appropriate_for_e-learning = Strongly_Agree: Yes (105.0/3.0)

Number of Leaves  :     39

Size of the tree :      50
```

Figure 2: C4.5 decision tree obtained from the best cross-validation fold

4. Discussion and Conclusion

This study aimed to predict the e-learning readiness of learners using data mining techniques. The intention was to provide feedback for institute managers and admin staff of e-learning systems planned to be used in an institution.

First, C4.5 Decision Tree Algorithm was used to predict the e-learning readiness status of the learners in the classification analysis. An approximate accuracy value of 83% was obtained. Classification results showed us that this study method was very effective and directly related to student opinions about e-learning readiness. In the Findings Section, Rule 3 and Rule 4 (if a student "agrees" with the statement "My studying/research area is appropriate for e-learning", the e-learning readiness status of the student was labeled as YES, if a student "strongly agrees" with the statement "My studying/research area is appropriate for e-learning", the e-learning readiness status of the student was labeled as YES.) can be seen a simple proof of this.

Moreover, from the beginning of studies on e-learning, a lot of work has been done regarding the comparison of face to face learning with e-learning. In this study, when this kind of comparison is considered with studying area of a student, in other words, if a student "strongly disagrees" with the statement "My studying/research area is appropriate for e-learning" and "E-learning is better than face to face learning", the e-learning readiness status of a student is predicted as NO.

Furthermore, the preference of a student about to join different courses in different universities is a significant factor. The possibility of studying together with faculty members from different universities, even from different countries, makes e-learning more attractive. If a student has "no idea" about the statement "My studying/research area is appropriate for e-learning" and "E-learning is better than face to face learning" and strongly agrees with the statement "I want to join different courses at different universities", the e-learning readiness status of a student is predicted as YES.

The sensitivity value in classification (0.485) is lower than the specificity (0.939) in the other model. The frequency difference between class labels of the target attribute is seen as the reason for a low sensitivity value in classification. Class labels of the target attribute should be taken on balance for further studies in order to obtain better performance results.

Secondly, k-Means Algorithm was used to cluster dataset without the target attribute based on the self-assessment of the respondents about e-learning readiness. Samples were grouped by considering the similarities (in other words dissimilarities or distances) between them. Clusters obtained from the k-Means Algorithm were replaced with the real class labels (ready or not ready for e-learning). After that, performance evaluation metrics were calculated using cluster labels as target attributes. The accuracy value 70% was obtained. At this point, it can be seen that there was no such big difference between the results of the two models.

The authors believe that the results of the study will be beneficial for the feasibility of further study of an e-learning project and that this present study is a good example of big data analyses in the education field. Also, the dataset is only limited to the higher education students in this study. Dataset can be extended with students in primary and secondary schools or employees in the public and private sectors. In addition, some resampling methods such as oversampling, undersampling, and other such different methods can be used to eliminate imbalanced data problems just before the data mining analyses. Furthermore, other data mining techniques such as Naive Bayes Classifier, Binary Logistic Regression, k-Nearest Neighbor Algorithm, Support Vector Machines, etc. can be used to improve the performance of the prediction model.

5. Acknowledgments

This study was supported by the Scientific Research Projects Coordination Unit of Istanbul University. Project number 26089.

References

Abas, Z. W., Kaur, K., & Harun, H. (2004). E-learning readiness in Malaysia. *A National Report Submitted to the Ministry of Energy, Water and Communications*.

Akaslan, D., & Law, E. L.-C. (2011a). Measuring Student E-Learning Readiness: A Case about the Subject of Electricity in Higher Education Institutions in Turkey. In H. Leung, E. Popescu, Y. Cao, R. W. H. Lau, & W. Nejdl (Eds.), *Advances in Web-Based Learning—ICWL 2011* (pp. 209–218). https://doi.org/10.1007/978-3-642-25813-8

Akaslan, D., & Law, E. L.-C. (2011b). Measuring teachers' readiness for e-learning in higher education institutions associated with the subject of electricity in Turkey. *2011 IEEE Global Engineering Education Conference (EDUCON)*, 481–490. https://doi.org/10.1109/EDUCON.2011.5773180

AL-Fayyadh, S. A., & Mohammad, Q. Q. (2016). Iraqi Nursing Faculty Attitudes toward E-Learning a National Survey. *IOSR Journal of Nursing and Health Sciences*, *5*(3), 57–63.

Al-Solbi, A., & Mayhew, P. J. (2005). Measuring e-readiness assessment in saudi organisations preliminary results from a survey study. *From E-Government to m-Government*, 467–475.

Aydin, C. H., & Tasci, D. (2005). Measuring readiness for e-learning: Reflections from an emerging country. *Journal of Educational Technology & Society*, *8*(4). Retrieved from http://www.jstor.org/stable/jeductechsoci.8.4.244

Azimi, H. M. (2013). Readiness for implementation of e-learning in colleges of education. *Journal of Novel Applied Sciences*, *2*(12), 769–775.

Balaban, M. E., & Kartal, E. (2015). *K-Ortalamalar Algoritmasıyla Ülkelerin Bilişim Alanında Kümelenmesi [Clustering of Countries in Informatics with k-Means Algorithm]*. 112–117. ATO Congresium, Ankara, Turkey: Türkiye Bilişim Derneği.

Beig, L., Montazer, A. P. G. A., & Ghavamifar, A. (2007). Adoption a Proper Tool For E-Readiness Assessment in Developing Countries (Case Studies: İran, Turkey and Malaysia). The *Journal of Knowledge Economy & Knowledge Management (JKEM)*, *2*(1). Retrieved from http://dergipark.ulakbim.gov.tr/beyder/article/view/5000098823

Bonanno, P. (2011). Developing an instrument to assess teachers' readiness for technology-enhanced learning. *2011 14th International Conference on Interactive Collaborative Learning*, 438–443. https://doi.org/10.1109/ICL.2011.6059622

Cathy, J.-S. (2016). *Building a Tool for Determining E-learning Readiness of Organizations: A Design and Development Study* (Doctoral Thesis, Virginia Polytechnic Institute and State University). Retrieved from https://vtechworks.lib.vt.edu/bitstream/handle/10919/70912/James-Springer_CD_D_2016.pdf?sequence=2

Clark, R. C., & Mayer, R. E. (2016). *E-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning*. John Wiley & Sons.

Coopasami, M., Knight, S., & Pete, M. (2017). E-Learning readiness amongst nursing students at the Durban University of Technology. *Health SA Gesondheid*, *22*, 300–306. https://doi.org/10.1016/j.hsag.2017.04.003

cran.r-project. (2019). The Comprehensive R Archive Network. Retrieved October 3, 2019, from https://cran.r-project.org/

D. Doculan, J. A. (2016). E-Learning Readiness Assessment Tool for Philippine Higher Education Institutions. *International Journal on Integrating Technology in Education*, *5*(2), 33–43. https://doi.org/10.5121/ijite.2016.5203

Darab, B., & Montazer, Gh. A. (2011). An eclectic model for assessing e-learning readiness in the Iranian universities. *Computers & Education*, *56*(3), 900–910. https://doi.org/10.1016/j.compedu.2010.11.002

Doğan, Ş. (2013). *The examination of E-readiness levels of academicians* (Master's Thesis, Eskişehir Osmangazi Üniversitesi). Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp

Dragulescu, A. A., & Arendt, C. (2018). *xlsx: Read, Write, Format Excel 2007 and Excel 97/2000/XP/2003 Files*. Retrieved from https://CRAN.R-project.org/package=xlsx

Economist Intelligent Unit. (2002). *The 2002 e-readiness rankings*. Retrieved from https://www.westerncape.gov.za/text/2004/2/ereadiness_2002.pdf

Economist Intelligent Unit. (2003). *The 2003 e-readiness rankings*. Retrieved from http://graphics.eiu.com/files/ad_pdfs/eready_2003.pdf

Economist Intelligent Unit. (2004). *The 2004 e-readiness rankings*. Retrieved from http://graphics.eiu.com/files/ad_pdfs/err2004.pdf

Economist Intelligent Unit. (2005). *The 2005 e-readiness rankings*. Retrieved from http://graphics.eiu.com/files/ad_pdfs/2005ereadiness_ranking_wp.pdf

Economist Intelligent Unit. (2006). *The 2006 e-readiness rankings*. Retrieved from http://graphics.eiu.com/files/ad_pdfs/2006ereadiness_ranking_wp.pdf

Economist Intelligent Unit. (2007). *The 2007 e-readiness rankings Raising the bar*. Retrieved from http://graphics.eiu.com/files/ad_pdfs/2007ereadiness_ranking_wp.pdf

Economist Intelligent Unit. (2008). *E-readiness rankings 2008 Maintaining momentum*. Retrieved from http://graphics.eiu.com/upload/ibm_ereadiness_2008.pdf

Economist Intelligent Unit. (2009). *E-readiness rankings 2009 The usage imperative*. Retrieved from http://graphics.eiu.com/pdf/e-readiness%20rankings.pdf

Economist Intelligent Unit. (2010). *Digital Economy Rankings 2010 Beyond E-readiness*. Retrieved from https://www-935.ibm.com/services/us/gbs/bus/pdf/eiu_digital-economy-rankings-2010_final_web.pdf

Eslaminejad, T., Masood, M., & Ngah, N. A. (2010). Assessment of instructors' readiness for implementing e-learning in continuing medical education in Iran. *Medical Teacher*, *32*(10), e407-412. https://doi.org/10.3109/0142159X.2010.496006

Ha, O., Jm, N., & An, W. (2014). E-Learning Readiness Assessment Model In Kenyas' Higher Education Institutions: A Case Study Of University Of Nairobi. *International Journal of Scientific Knowledge*, *5*(6), 29–42.

Han, J., & Kamber, M. (2006). *Data mining: Concepts and techniques* (2nd ed.). San Francisco, CA, USA: Morgan Kaufmann.

Hornik, K., Buchta, C., & Zeileis, A. (2009). Open-Source Machine Learning: R Meets Weka. *Computational Statistics*, *24*(2), 225–232. https://doi.org/10.1007/s00180-008-0119-7

Hung, M.-L., Chou, C., Chen, C.-H., & Own, Z.-Y. (2010). Learner readiness for online learning: Scale development and student perceptions. *Computers & Education*, *55*(3), 1080–1090. https://doi.org/10.1016/j.compedu.2010.05.004

Ifinedo, P., & Davidrajuh, R. (2005). Digital divide in Europe: Assessing and comparing the e-readiness of a developed and an emerging economy in the Nordic region. *Electronic Government, an International Journal*, *2*(2), 111–133. https://doi.org/10.1504/EG.2005.007090

Korkmaz, Ö., Çakır, R., & Tan, S. S. (2015). Students E-learning Readiness and Satisfaction Levels and Effects on the Academic Achievement. *Journal of Kirsehir Education Faculty*, *16*(3). Retrieved from http://kefad2.ahievran.edu.tr/archieve/pdfler/Cilt16Sayi3/JKEF_16_3_2015_219-241.pdf

Kuhn, M. (2018). *caret: Classification and Regression Training*. Retrieved from https://CRAN.R-project.org/package=caret

Kuruliszwili, S. (2015). E-learning Readiness of Organization and Employees. *International Journal of Electronics and Telecommunications*, *61*(3), 245–250. https://doi.org/10.1515/eletel-2015-0032

Lopes, C. T. (2007). Evaluating E-Learning Readiness In A Health Sciences Higher Education. *El2007 Proceedings of the IADIS International Conference on E-Learning*. Presented at the IADIS International Conference on e-Learning, Portugal.

Mafenya, P. N. (2013). An Investigation of First-Year Students' Pedagogical Readiness to E-Learning and Assessment in Open and Distance Learning: An University of South Africa Context. *Mediterranean Journal of Social Sciences*, *4*(13), 353.

Moftakhari, M. M. (2013). *Evaluating E-Learning Readiness of Faculty of Letters of Hacettepe* (Master's Thesis). Hacettepe University, Ankara.

Mercado, C. (2008). Readiness Assessment Tool for an E-Learning Environment Implementation. *Fifth International Conference on E-Learning for Knowledge Based Society*, 183–187.

Minges, M. (2005). *Evaluation of e-readiness indices for Latin America and the Caribbean*. Retrieved from http://repositorio.cepal.org//handle/11362/31929

Ncube, S., Dube, L., & Ngulube, P. (2014). E-Learning Readiness among Academic Staff in the Department of Information Science at the University of South Africa. *Mediterranean Journal of Social Sciences*, *5*(16), 357–366. https://doi.org/10.5901/mjss.2014.v5n16p357

Okinda, R. A. (2014). Assessing E-Learning Readiness at the Kenya Technical Teachers College. *Journal of Learning for Development-JL4D*, *1*(3).

Özen, Z., Kartal, E., & Emre, İ. E. (2017). Eğitimde Büyük Veri [Big Data in Education]. In H. F. Odabaşı, B. Akkoyunlu, & A. İşman (Eds.), *Eğitim Teknolojileri Okumaları 2017* (1st ed., pp. 183–204). Retrieved from http://www.tojet.net/e-book/eto_2017.pdf

Parkes, M., Stein, S., & Reading, C. (2015). Student preparedness for university e-learning environments. *The Internet and Higher Education*, *25*, 1–10. https://doi.org/10.1016/j.iheduc.2014.10.002

Paturusi, S., Chisaki, Y., & Usagawa, T. (2015). Assessing Lecturers and Student's Readiness for E-Learning: A preliminary study at National University in North Sulawesi Indonesia. *GSTF Journal on Education (JEd)*, *2*(2). Retrieved from http://dl6.globalstf.org/index.php/jed/article/view/1160

Pınar, İ., Selçuk, A. G., & Dağ, B. (2014). Two Concepts that Have to be Considered in the Transition of Vocational Colleges to E-Learning Model: Student's Computer Self Efficacy and E-Learning Readiness. *Electronic Journal of Occupational Improvement and Research*, *2*(3), 50–60.

Princely Ifinedo, U. of J. (2005, March 26). Measuring Africa's e-readiness in the global networked economy: A nine-country data analysis. Retrieved June 4, 2017, from International Journal of Education and Development using ICT, Vol. 1, No. 1, 2005 website: http://ijedict.dec.uwi.edu/viewarticle.php?id=12.&layout=html

Purcell, F., & Toland, J. (2004). Electronic Commerce for the South Pacific: A Review of E-Readiness. *Electronic Commerce Research*, *4*(3), 241–262. https://doi.org/10.1023/B:ELEC.0000027982.96505.c6

Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

R Core Team. (2018). *R: A Language and Environment for Statistical Computing*. Retrieved from https://www.R-project.org/

Rasouli, A., Rahbania, Z., & Attaran, M. (2016). Students' Readiness for E-Learning Application in Higher Education. *Malaysian Online Journal of Educational Technology*, *4*(3), 51–64.

Rizk, N. (2004). E-readiness assessment of small and medium enterprises in Egypt: A micro study. *Topics in Middle Eastern and North African Economies*, *6*. Retrieved from http://ecommons.luc.edu/cgi/viewcontent.cgi?article=1055&context=meea

Rohayani, A. H. H., Kurniabudi, & Sharipuddin. (2015). A Literature Review: Readiness Factors to Measuring e-Learning Readiness in Higher Education. *Procedia Computer Science*, *59*(Supplement C), 230–234. https://doi.org/10.1016/j.procs.2015.07.564

RStudio. (2019). RStudio—RStudio. Retrieved October 3, 2019, from https://rstudio.com/

Schreurs, J., & Al-Huneidi, A. M. (2012). E-Learning Readiness in Organizations. *International Journal of Advanced Corporate Learning (IJAC)*, *5*(1), 4–7.

Schreurs, J., Ehlers, U.-D., & Sammour, G. (2008). E-learning Readiness Analysis (ERA): An e-health case study of e-learning readiness. *International Journal of Knowledge and Learning*, *4*(5), 496–508.

Schreurs, J., Moreau, R., & Ehlers, U. (2008). *Measuring e-learning readiness*. Retrieved from https://doclib.uhasselt.be/dspace/handle/1942/8740

Sharma, S. K., Gülseçen, S., Özen, Z., & Kartal, E. (2014, May 5). *Assessing E-learning Readiness of Learners in Turkey* (S. Gülseçen, Z. Ayvaz Reis, & Ç. Selçukcan Erol, Eds.). İstanbul Üniversitesi, İstanbul, Türkiye: İstanbul Üniversitesi Yayınları.

Sharma, S. K., Gülseçen, S., Özen, Z., & Kartal, E. (2015). Assessing E-learning Readiness of Instructors in Turkey. *İstanbul Journal of Innovation in Education*, *1*(3), 13–28.

Shearer, C. (2000). The CRISP-DM model: The new blueprint for data mining. *Journal of Data Warehousing*, *5*(4), 13–22.

Soydal, İ., Alır, G., & Ünal, Y. (2011). Are Turkish universities ready for e-learning: A case of Hacettepe University Faculty of Letters. *Information Services & Use*, *31*(3–4), 281–291. https://doi.org/10.3233/ISU-2012-0659

The International Institute for Management Development (IMD) World Competitiveness Center. (2019). *IMD World Digital Competitiveness Ranking 2019 Results*. Retrieved from https://www.imd.org/globalassets/wcc/docs/release-2019/digital/imd-world-digital-competitiveness-rankings-2019.pdf

Ünal, Y., Alır, G., & Soydal, İ. (2013). Students Readiness for E-Learning: An Assessment on Hacettepe University Department of Information Management. *International Symposium on Information Management in a Changing World*, 137–147. Retrieved from https://link.springer.com/chapter/10.1007/978-3-662-44412-2_13

Walesiak, M., & Dudek, A. (2016). *clusterSim: Searching for Optimal Clustering Procedure for a Data Set*. Retrieved from https://CRAN.R-project.org/package=clusterSim

Witten, I. H., & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques* (2nd ed.). San Francisco, CA: Morgan Kaufmann.

Yilmaz, R. (2017). Exploring the role of e-learning readiness on student satisfaction and motivation in flipped classroom. *Computers in Human Behavior*, *70*, 251–260. https://doi.org/10.1016/j.chb.2016.12.085

Yurdugül, H. (2016). An investigation of Pre-service Teachers' Readiness for E-learning at Undergraduate Level Teacher Training Programs: The Case of Hacettepe University. *Hacettepe University Journal of Education*, 1–1. https://doi.org/10.16986/HUJE.2016022763

Yurdugül, H., & Sırakaya, D. A. (2013). The Scale of Online Learning Readiness: A Study of Validity and Reliability. *Education and Science*, *38*(169).

Zaied, A. N. H., Khairalla, F. A., & Al-Rashed, W. (2007). Assessing e-readiness in the Arab countries: Perceptions towards ICT environment in public organisations in the State of Kuwait. *The Electronic Journal of E-Government*, *5*(1), 77–86.

Zikopoulos, P., Eaton, C., deRoos, D., Deutsch, T., & Lapis, G. (2011). *Understanding big data: Analytics for enterprise class hadoop and streaming data*. New York, USA: McGraw-Hill Osborne Media.