An Investigation on Students’ Level of Adjustment to Online Education with C5.0 Decision Tree Algorithm

Nesrin Hark Söylemez
Dicle University, Turkey
https://orcid.org/0000-0002-6306-5595

Abstract
This study aims to determine the independent variables that have a significant effect on the level of students’ adaptation to online education and their order of importance. Relational screening model was used in the study. Adaptability Level in Online Education dataset provided by Kaggle repository constitutes the main data source for this study. C5.0 decision tree algorithm was used to analyze the data. It was found that that 51,867% of the students had a moderate level of adaptation to online education, while 39,834% had low, and 8,299% had a high level adaptation. The findings indicate that students’ level of adaption to online learning is insufficient. We also found that the variable that best explains the online education compliance levels is daily class duration. “Financial condition” was found as the best explanatory variable of the cluster formed by the students whom “daily class duration” was between “1-3 hours”. “Age” was found as the best explanatory variable of the cluster formed by the students whom “daily class duration” was between “1-3 hours” and has “financial condition” of “poor”, “mid” and “rich”.

Keywords: Online Education, C5.0 Algorithm, Decision Tree.

Introduction
With the developments in communication and information technology, developments and changes occur in many fields. The field of education is one of the areas were changes are most noticeable. The way we lean has changed due to technological developments; learning and teaching are now feasible outside the traditional classroom setting due to new technology and the Internet. Particularly with the Covid-19 pandemic, learning and teaching activities have rapidly moved to online environments. Online education is embedded in the education system with an inevitable trend.

Due to its accessibility from any location, cheap cost, and flexibility for both students and teachers, online education has become commonplace in many educational institutions. (O’Lawrence, 2005; Oliveira et al., 2018). It is the most often used term to refer to communication and information technology-based learning methodologies, along with the terms online education, distance education, e-learning, online learning, and distance learning (Lee, 2010). Online education is to convey information to students with the help of web technologies and computers developed through the internet (Noe et al., 2019). In this system, teachers and learners can interact together, regardless of time and place, thanks to the communication provided through web technologies and computers (Fiş Erimi, 2011). Online education acts as a bridge connecting education stakeholders who are not physically in the same environment.

An increasingly common substitute for traditional classroom instruction is online learning. It offers educational opportunities to individuals with space, time or other constraints that make it difficult or almost impossible to continue traditional education, and it is an option for those who prefer the flexibility of online learning (Crawford-Ferre & Wiest, 2012).
Developing technologies provide convenience to individuals in online education (Aoki, 2012) and offer students the opportunity to choose in the online education process (Sumuer, 2018). Students can participate in online settings using their tablets and mobile devices (Joanne & Michael, 2013). Online education has advantages such as interaction (Leszczyński et al., 2018; Wagner et al., 2008), flexibility (Smedley, 2010), and self-pace (Amer, 2007). Also, it can be used to eliminate educational inequalities between age groups, expand educational access geographically, provide education to large audiences, combine education with work or family life, improve students’ self-motivation, self-discipline and critical thinking skills, allowing students to progress at their own rate, and improve student’s technical skills (Buselic, 2012).

Studies on online education can be found in the literature under the categories of efficacy, advantages, and difficulties (Afrouz & Crisp, 2020), feedback gathered from LMS courses (Cavalcanti et al., 2019), improvement of the online education model (Wiliam, 2008), examination of course content on online and digital platforms (Ciolacu et al., 2017), estimation of student passing rates in online education (Ma et al., 2018), reduction of dropout rate in online education or e-learning courses (Tan & Shao, 2015), students’ readiness for online learning (Latheef et al., 2021), strategies to encourage student participation in online environments (McKeithan et al., 2021), and student success in online courses (Vayre & Vonthron, 2019).

However, no study has been found to assess the independent variables that have a significant effect on students’ adaptation to online education and their order of importance. For this reason, we think our research will add to the body of currently existing literature.

**Aim of the Research**

In the current research, we sought to identify the independent variables that have a significant effect on the level of students’ adaptation to online education and their order of importance.

**Method**

**Research Design**

We preferred the relational screening model in this study. Relational research is a form of analysis in which parameters and variables are interrelated and information is systematically integrated (Cohen et al., 2007). The relationships between two or more variables are found via relational research, and it is aimed to determine the effects of these relationships on a ‘cause and effect’ basis (Fraenkel et al., 2012). The relationship revealed in the relational screening model shows that part of the change seen in one of the two variables may be due to the other variable (Christensen et al., 2010).

**Participants**

The participants consist of 1205 students enrolled in universities, colleges and schools. Table 1 contains demographical details of the participants.

<table>
<thead>
<tr>
<th>Variable</th>
<th>f</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>663</td>
<td>55</td>
</tr>
<tr>
<td>Female</td>
<td>542</td>
<td>45</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-5</td>
<td>81</td>
<td>6.7</td>
</tr>
<tr>
<td>6-10</td>
<td>51</td>
<td>4.2</td>
</tr>
<tr>
<td>11-15</td>
<td>353</td>
<td>29.3</td>
</tr>
<tr>
<td>16-20</td>
<td>278</td>
<td>23.1</td>
</tr>
<tr>
<td>21-25</td>
<td>374</td>
<td>31</td>
</tr>
<tr>
<td>26-30</td>
<td>68</td>
<td>5.6</td>
</tr>
<tr>
<td>Education Level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>School</td>
<td>530</td>
<td>44</td>
</tr>
<tr>
<td>College</td>
<td>219</td>
<td>18.2</td>
</tr>
<tr>
<td>University</td>
<td>456</td>
<td>37.8</td>
</tr>
</tbody>
</table>
Data Collection

The dataset for this research was obtained from the Kaggle depository titled “Adaptability Level in Online Education.” (Suzan & Samrin, 2022). The dataset includes survey data conducted with students enrolled in universities, schools and colleges. Data were collected between 10 December 2020 and 5 February 2021. Table 2 contains details regarding the data set.

<table>
<thead>
<tr>
<th>Features</th>
<th>Possible Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Female {0}, Male {1}</td>
</tr>
<tr>
<td>Age</td>
<td>Around 1-5 {0}, 6-10 {1}, 11-15 {2}, 16-20 {3}, 21-25 {4}, 26-30 {5}</td>
</tr>
</tbody>
</table>

Table 2 Information about the Dataset

Data Collection

The dataset for this research was obtained from the Kaggle depository titled “Adaptability Level in Online Education.” (Suzan & Samrin, 2022). The dataset includes survey data conducted with students enrolled in universities, schools and colleges. Data were collected between 10 December 2020 and 5 February 2021. Table 2 contains details regarding the data set.

Table 2 Information about the Dataset

<table>
<thead>
<tr>
<th>Features</th>
<th>Possible Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Female {0}, Male {1}</td>
</tr>
<tr>
<td>Age</td>
<td>Around 1-5 {0}, 6-10 {1}, 11-15 {2}, 16-20 {3}, 21-25 {4}, 26-30 {5}</td>
</tr>
</tbody>
</table>

Analysis of Data

The data analysis program used was IBM SPSS Modeler. The research data was applied to the C5.0 decision tree model. Decision trees are used by dividing large structured data into smaller data groups using easy decision-making procedures. The members of the result groups resemble one another after each successful division (Sun & Li, 2008). Decision trees can be used in both predictive and descriptive tasks. The decision maker can use trees to decide the factors to be considered and how each one relates to the various outcomes of the decision (Bounsaythip & Rinta-Runsala, 2001). Classification is made by using the property values of the samples in decision trees. Each node in the decision tree represents a feature of the classified sample.

By doing numerous experiments as part of the knowledge discovery process, decision trees attempt to determine the appropriate order while estimating the target. Each test adds branches to the decision tree, and these branches allow for the execution of other tests. This keeps happening up until a leaf node’s test process is finished. The “rule” that categorizes the target is the route from the root to the target leaf. The “if-then” structure establishes the rules. (Bounsaythip & Rinta-Runsala, 2001).
C5.0 algorithm uses boosting for enhancing model accuracy. Initially, a model is built as usual. A second model is then created by focusing on the misclassifications of the first model. A third model is produced by focusing on the misclassifications of the second model, and so on. Finally, a weighted voting is used to aggregate all predictions of different models into final prediction. By this voting process, a robust classifier is achieved.

When the C5.0 algorithm is used with categorical variables, large trees can be obtained because a separate branch is created for each category. To prevent this situation, some categories can be combined (Larose, 2005). Since there is a simple interpretation of the rules obtained with this model, the model is easier to understand than other model types (Pang & Gong, 2009). Categorical target variables are estimated with the C5.0 decision tree algorithm (IBM, 2021). In this study, the categorical version (low-moderate-high) of the predicted (dependent) variable “the level of adaptation of students to online education” was used.

Results
What is the Order of Importance of the Predictive Variables that have a Significant Effect on Students’ Adaptation Levels to Online Education?

The order of importance of the predictive variables that have a significant effect on students’ adaptation to online education is shown in Figure 1.

![Figure 1 Order of Importance of Predictive Variables](image)

As seen in Figure 1, “financial condition” is the predictive variable that has the highest effect on students’ adaptation to online education. Other predictive variables that have a significant effect on students’ adaptation to online education are “class duration”, “education level”, “age”, “self lms”, “network type”, “institution type”, “internet type”, “load shedding”, “device”, “location”, “IT student”, and “gender”. For the C5.0 decision tree algorithm in the study, 4 predictive variables (“financial condition”, “class duration”, “education level”, “age”) that have the highest effect on students’ adaptation to online education were used.

How is the C5.0 Decision Tree Algorithm Related to the Level of Students’ Adaptation to Online Education?

The decision tree regarding the students’ adaptation levels to online education is given in Figure 2.

![Figure 2 Decision Tree](image)

Figure 2 shows 16 nodes that collectively explain the degree of adaptability of online education. The variable that best explains students’ adaptation levels to online education is “class duration”. We found that that the majority of students (93,506%), with “daily class duration” “less than an hour” have a low level of adaptation to online education. Students with “daily class duration” between “1-3 hours” (55,714%) and between “3-6 hours” (69,668%) have a moderate level of adaptation to online education.

“Financial situation” is found as the variable that best explains the group with the “daily class duration” between “1-3 hours”. We found that 49,180% of the students with “rich” “financial condition” have a “high” level of adaptation and 45,763% of the students with “poor” “financial condition” have a “poor” level of adaptation to online education. On the other hand, 61,130% of the students with “mid” “financial condition” have a “moderate” level of adaptation to online education.

“Age” is found as the variable that best explains the group with the “daily class duration” between “1-3 hours” and “financial status” “poor”. Having “daily class duration” between “1-3 hours” with “financial condition” “poor”, 65.278% of students between the ages “11-15” have “low” level of adaptation, 71.429% of student between the ages “1-5” “16-20” and “26-30” have “moderate” level of adaptation, and 52.381% of the students between the ages “6-10” and “21-25” have “high” level of adaptation to online education.
“Age” is found as the variable that best explains the group with the “daily class duration” between “1-3 hours” and “financial status” “mid”. Having “daily class duration” between “1-3 hours” with “financial condition” “mid”, 55.056% of students between the ages “16-20” have “low” level of adaptation, and 67.925% of student between the ages “1-5”, “6-10”, “11-15”, “21-25” and “26-30” have “moderate” level of adaptation to online education.

“Age” is found as the variable that best explains the group with the “daily class duration” between “1-3 hours” and “financial status” “rich”. Having “daily class duration” between “1-3 hours” with “financial condition” “rich”, 100% of students between the ages “1-5” have “low” level of adaptation, 59.091% of student between the ages “6-10”, “11-15” and “16-20” have “moderate” level of adaptation, and 100% of the students between the ages “21-25” and “26-30” have “high” level of adaptation to online education.

“Education level” is found as the variable that best explains the group with the “daily class duration” between “1-3 hours”, “financial status” “mid”, and “age” “16-20”. Having “daily class duration” between “1-3 hours” with “financial condition” “mid”, and “age” “16-20”, 58.065% of students with “education level” as “college” and “university” have “low” level of adaptation, and 65.217% of student with “education level” as “school” have “moderate” level of adaptation to online education.

What are the Rule Sets Obtained as a Result of the C5.0 Decision Tree Algorithm Applied to the Students’ Adaptation Levels to Online Education?

The important rule sets obtained as a result of the C5.0 decision tree algorithm applied to the students’ adaptation levels to online education are given in Figure 3.
When the C5.0 decision tree method is applied to the level of student adaptation to online education, the following significant rule sets are revealed in Figure 3.

- Students with “daily class duration” “less than an hour” have a “low” adaptation level to online education.
- Students with “education institution level” “school” and “financial condition” “mid” have “low” adaptation level to online education.
- Students with “age” between “21-25” and “financial condition” “mid” have a “low” adaptation level to online education.
- Students with “age” between “16-20” and “educational institution level” “school” and “daily class duration” between “3-6 hours” have a “low” level of adaptation to online education.
- Students with “daily class duration” between “1-3 hours” have a “moderate” level of adaptation to online education.
- Students with “daily class duration” between “1-3 hours” have a “moderate” level of adaptation to online education.
- Students with “daily class duration” between “3-6 hours” have a “moderate” level of adaptation to online education.
- Students with “age” between “26-30” and “education institution level” “college” and “financial condition” “mid” and “daily class duration” between “3-6 hours” have a “high” level of adaptation to online education.

**Discussion**

Current research aimed to find the independent variables that have a significant effect on the level of students’ adaptation to online education and their order of importance. We found that 51,867% of the students had a moderate level of adaptation to online education, while 39,834% had a low, and 8,299% had a high level of adaptation to online education. The findings indicate that students’ level of adaption to online learning is insufficient. Therefore, it is necessary to examine the variables that have a positive effect on students’ adaptation levels to online education and to make arrangements that will increase students’ adaptation levels to online education in line with the results obtained.

In the study conducted by Alper (2020), teachers working at the K-12 level stated that most students can easily adapt to the distance education process and their participation in the course is sufficient. However, some students stated that they have problems both in attending the course and doing homework, and this situation is not different from traditional education. The adaptation levels of the stakeholders of distance education to distance education were examined in the research conducted by Kaysi (2020). The participants of that study conducted with university students, stated that their friends’ adaptation level to distance education was medium and their level of adaptation to distance education was high. In another study, approximately 65% of secondary school students said that they could adapt to the distance education system (Kaynar et al., 2020). When the results of the studies are evaluated in general, it is possible to say that the adaptation is higher in the studies conducted with direct participant views. However, in the current study carried out on a larger sample, and aimed to measure the adaptation levels, we found that adaptation levels were not sufficient. We can attribute this to the differences in the methods used to determine the level of compliance with online education.

We found that the variable that best explains the adaptation level with online education is “daily class duration”. We also found that the majority of students whose “daily course duration” is “less than an hour” have “low” adaptation level to online education. The majority of students who state that the “daily course duration” is between “1-3 hours” and “3-6 hours” have a “moderate” level of adaptation to online education. This result is also among the rule sets obtained as an outcome of the C5.0 decision tree algorithm applied to the students’ adaptation levels to online education. We can interpret this as students who have less than an hour daily class duration cannot adapt to online education. The explanation for this is that students who receive online education regularly experience the process and can adapt as they gain experience. Students who have experience in the utilization of technology in the online education will adapt to the system without difficulty, while students who do not have experience in using technology will have difficulty adapting to the system.
Each student in online education may encounter different difficulties according to their experiences. The difficulties encountered may cause students to move away from the system while creating a motivation problem. To ensure that students can continue to work and be a part of the learning environment despite the challenges they confront, it is crucial to improve student involvement in learning environments in online education (Ergün & Kurnaz, 2017).

Online education is largely dependent on the internet and digital equipment (Adedoyin & Soykan, 2020). Fishbane and Tomer’s (2020) research findings indicate that as the poverty level increases in the society, the rate of access to the internet decreases and as a result, students who have low socio-economic power tend to fall behind to access other students in online learning or that they are most open to encountering additional challenges.

We found “financial condition” as the variable that best explains the group formed by the students who had “daily class duration” between “1-3 hours”. We also found that a significant majority (49.180%) of the students who has “daily class duration” between “1-3 hours” and has “high” “financial status” has a “high” level of adaptation to online education. It was revealed that an important part of the students (45.763%) who has “poor” “financial status” had “low” adaptation level to online education, and the majority of students (61.130%) who had “moderate” “financial status” had “moderate” adaptation level to online education. The cause of this could be connected to the fact that students with good financial status can more easily meet the infrastructure and equipment needed to access online learning. In fact, in order for a student to adapt to online education, the necessary infrastructure and equipment needs to be met. In addition, it can be said that students with good financial status have knowledge about the use of most digital equipment used in the online education, and therefore, their high digital literacy levels can have a positive effect on their level of adaptation to online education.

Digitalization in the field of education has made digital literacy a skill that students should acquire (Stripling, 2010). Individuals need this skill in order to use technology critically and effectively (Buckingham, 2010). Students should be well digitally literate to fulfill their responsibilities in online environments such as problem solving, knowledge management, efficiency and productivity. Students with low digital literacy may lag behind in online learning.

We found “age” as the variable that best explains the group formed by students who had “daily course duration” between “1-3 hours” and had “poor”, “mid” or “rich” “financial status”. Along with this, having “daily class duration” between “1-3 hours” with “rich” “financial condition”, 100% of students between the ages “1-5” have “low” level of adaptation, 59.091% of student between the ages “6-10”, “11-15” and “16-20” have “moderate” level of adaptation, and 100% of the students between the ages “21-25” and “26-30” have “high” level of adaptation to online education. This can be explained by the fact that individuals between the ages of 1-5 cannot fully concentrate their attention in front of the screen during the online education process and do not have the self-regulation skills to manage their learning processes.

Children’s ability to plan and cooperate, focus on a subject, control their impulses and follow directions depends on their self-regulation skills (McClelland & Cameron, 2011). Self-regulation is not an academic performance or a cognitive ability. It is the self-management process in which students convert their mental abilities into academic skills (Zimmerman, 2002). As students get older, it will be easier for them to gain self-regulation skills for the online education process and to take responsibility for their own learning-teaching processes.

Concluding Remarks and Suggestions

In this study, we found that students’ levels of online education adaption are insufficient. Therefore, before starting online education practices, studies should be carried out to adapt students to the requirements of distance education. C5.0 decision tree algorithm was preferred in this study. Different decision tree algorithms can be used and comparisons can be made among themselves. When the “daily class duration” is “less than an hour”, level of adaptation to online education is low. It is recommended to carry out studies to ensure that
students participate in online education on a daily basis. The reasons why the class duration variable is included in the rule sets created at all levels related to level of adaptation to online education can be investigated. Other rule sets can be reached by using different methods and techniques, and new rules can be discovered to increase effectiveness and efficiency in education. By using different variables, different rule sets can be reached in the further studies.

References


**Author Details**

**Nesrin Hark Söylemez**, Dicle University, Turkey, *Email ID*: nesrin_hark@hotmail.com