Araştırma Makalesi / Research Article

An Application on Technology Addiction with C4.5 Classification Algorithm

Akıner KAÇMAZ[1], Kazım YILDIZ[2*], Ali BULDU[2]

[1]Marmara University, Computer Control Education, İstanbul
[2]Marmara University, Department of Computer Engineering, İstanbul
(ORCID: 0000-0002-2364-3346)(ORCID: 0000-0001-6999-1410) (ORCID: 0000-0002-8508-3065)

Abstract
The increment of the usagethe technological applications and tools, has led to an increase towards technology regarding for communication in the way of individuals' desire.With the development of technologies that facilitate our lives, rapid increases in individuals' use of technological tools have been observed.This has led to the fact that data mining and technology dependence concepts, which are expressed as the process of obtaining confidential information, are hidden in the large scale data in the digital environment and have become more involved in scientific studies. This study was carried out to determine the technology dependence of the students who continue their education in vocational and technical education based on C4.5 decision tree algorithm which is one of the data mining classification methods.In the questionnaire, the students were asked whether there is any positive or negative relationship between technology addiction and course success and demographic characteristics.Demographic information in the data set shows the characteristics and the question shows the class.The decision tree model of 411 observations in the data set was created using R programming language.The effect of demographic information such as gender, profession, number of siblings, date of birth on the class was determined according to the model made with C4.5 decision tree algorithm.As a result of the research, the most important feature that determines the answers is gender and the accuracy rate of the model is determined as 84.42%.

Keywords: Data Mining, Technology Addiction, C4.5 Algorithm, RStudio, Decision Tree.

C4.5 Sınıflandırma Algoritması ile Teknoloji Bağımlılığı Üzerine Bir Uygulama

Öz
Teknolojik uygulamaların ve aletlerin kullanımının artmasıyla birlikte, bireylerin ortamdan bağımsız olarak her yerde iletişim içinde olma istekleri, insanların teknolojiye karşı olan ihtiyaçlarının ve ilgilerinin artmasına sebep olmuştur. Yaşamıkolaylaştıran teknolojilerin gelişmesiyle birlikte, bireylerin teknolojik aletleri kullanma sürelerinde hızlı artışlar gözlemlenmiştir. Bu durum, dijital ortamdaki büyük ölçekli veride saklı kalmış, gizli bilgiyi elde etme süreci olarak ifade edilen veri madenciliği ve teknoloji bağımlılığı kavramlarının, bilimsel çalışmalarda daha fazla yer almasına neden olmuştur. Bu çalışma, veri madenciliği sınıflandırma yöntemlerinden, C4.5 karar ağacı algoritmasına dayalı olarak mesleki ve teknik eğitimde, eğitimlerini sürdüren öğrencilerin, teknoloji bağımlılıklarının belirlenmesi amacıyla yürütülmüştür. Ankette öğrencilere teknoloji bağımlılığı ile ders başarısı arasında olumlu ya da olumsuz herhangi bir bağ bulunmamaktadır ifadesi ile birlikte demografik özellikler sorulmuştur. Veri setinde yer alan demografik bilgiler özellikleri, soru ise sınıfı göstermektedir. Veri setindeki 411 adet gözleme ait karar ağacı modeli, R programlama dili kullanılarak oluşturulmuştur. C4.5 karar ağacı algoritmasıyla oluşturulan modele göre cinsiyet, meslek alanı, kardeş sayısı, doğum tarihi gibi demografik bilgilerin sınıf üzerindeki etkisi belirlenmiştir. Araştırma sonucunda cevapları belirleyen en önemli özelliğin cinsiyet olduğu ve oluşturulan modelin doğruluk oranı ise %84,42 olarak belirlenmiştir.

Anahtar kelimeler: Veri madenciliği, Teknoloji Bağımlılığı, C4.5 Algoritması, RStudio, Karar ağacı.

[*]Corresponding author: kazim.yildiz@marmara.edu.tr
Received: 14.01.2020, Accepted: 18.05.2020

1. Introduction

The rapid circulation of information leads to the rapid proliferation of technological developments arithmetically. The human who captures technology to meet their needs, has been the slave of technology [1]. The digital age is defined as the era in which social workers use social networks smartly and efficiently, and those who find more space in digital platforms [2]. While technological advances, which can easily and quickly guide all areas of our lives without being aware, contribute positively to lifelong learning processes, problems caused by the use of uncontrolled technology have occurred today [3].

With the increasing effects of digitalization on our lives, the process of recording our lives with technological instruments is increasing rapidly and everything is expressed numerically [4].The need to obtain meaningful, usable and valuable information from the ever-increasing data stacks in the digital environment has increased individuals' interest in data mining [5].Today, there are many methods for people to access information. The search for applications is ongoing, which will enable individuals to obtain valuable information, gain experience through real data frames and reveal the relationship between data and information. The concept of data mining comes into our mix as a promising technology that will end individuals' searches [6].Concepts such as smart life, artificial intelligence, machine learning, cloud technology, online learning, technology addiction are becoming more and more used in our lives day by day and everything is shaped based on data. Decision-making processes are now based on valuable information obtained by data mining methods. Scientific studies based on decision tree algorithms are widely used. For this purpose, some case studies are given below.

Baydir et al. determine C4.5 decision tree preferences of voters in Turkey. In this study, RStudio was used as a software environment. As a result of the study, the model revealed that decision trees are important in revealing political tendencies and that the parties can direct the election studies by using decision trees structures used in data mining [7]. In 2017, Aksu and Karaman conducted a study to determine the link analysis on a website with decision trees. In this study, Google analytics data belonging to a website is used. Based on the data obtained, they tried to present the links that should be on the main page with the decision trees. With the model created, it was thought that the main page would be more usable, faster and functional [8]. Aydın carried out a study on receiving student views on the problems caused by technology addiction in the classroom. As a result of the study, it was observed that the technology dependency levels of the participants were not different in terms of age group, socio-economic level and class variables, but they had a significant difference between the department and gender variables. It has been determined that as the time spent on technological applications increases, the level of technology addiction increases and the level of technology addiction increases as the level of technology addiction increases [9]. Arik conducted a study in 2018 to determine the technology dependence and social identity formation: the Y generation. According to this study, it was observed that the problems arising from being online differed by gender and it was determined that men experienced the problems arising from being online more [10]. In another article titled “Technology Addiction, Our Family and Our Children”, Gürmen et al. determined that the technology addiction reduced their social communication skills. He found that technology addiction can damage friendship, parent-child relationships and bring it to the breaking point. He suggested that in the solution of the technology addiction problem, people should be contacted openly and intervention techniques should be developed in the areas that cause the problem [11].

2. Material and Method

2.1. Addiction and Technology Addiction

Addiction is defined as the effect of a separate decision-maker's judgement on an individual or an irresistible demand against an asset, object or person, and reflects the individual's pathological behaviors related to mental activities [12]. According to the previous studies, although the separate criteria for technology addiction have been put forward; it is the health problems that occur due to the fact that the internet and technology addict gets the pleasure ones desires’ by increasing the time spent at the screen. These obsessive ideas and dreams about the internet continue for periods that are not connected to the internet and remain for hours each sitting [13]. There are four levels of technology addiction. The first

level is experimental usage. The second one isthe social usage and the third one is the targeted operational usage. The last one is dependent usage. In the dependent usage phase, one can use the technology without the need for a reason. In this case, the individual will continue to use the technological tool to which he is dependent, regardless of the problems in his life, whether he is curious or not, whether or not his social environment is dependent [14].

2.2. Method of Research

In the research, decision tree was created by using C4.5, one of the data mining classification algorithms.

2.2.1. Decision Trees and C4.5 Algorithm

Decision tree, one of the data mining classifier, is easy to interpret and understand. It enablesfor quick processing of data [15].The most important step in the creation of decision trees is to determine the criteria by which the branch of the tree will be made. Thus, according to which property values the tree structure will be created is determined [16]. The C4.5 classification learning algorithm, developed by J. Ross Quinlan in 1993, is one of the most widely used efficient and well known algorithms [17]. The C4.5 algorithm, which has more diverse and newer learning algorithms compared to ID3, emerged as the current version of the ID3 algorithm [18]. The C4.5 algorithm uses the gain ratio to solve the limitation that the ID3 algorithm is highly sensitive to properties with multiple values [19]. Entropy is used to quantify the uncertainty of a variable. If the Y variable has a k level, the entropy value is calculated as Equation 1:

$$H(Y) = H(p_1, p_2, \dots\dots, p_k) = \sum_{j=1}^{k} (p_j \log_b(1/p_j)) \quad (1)$$

Here, p_j notation j. indicates the probability of the level. The value range of entropy is obtained as Equation 2:

$$0 \leq H(Y) \leq log_b(k) \quad (2)$$

Decreased uncertainty in the data means that the entropy value approaches zero.
Gain value is calculated in Equation 3;

$$\mathrm{Gain}(X_i) = H(Y) - \sum_{j=1}^{k_i} P(X_{ij}) H\left(\frac{Y}{X_{ij}}\right) ; i = 1,2,\dots,m \quad (3)$$

$\mathrm{Gain}(X_i)$: Gain value for i. independent variable

The selection of the property that has the most decisive influence on the class is determined by the magnitude of the earnings ratio. The feature with the highest yield rate was chosen for branching in decision trees [20].

2.3. Data Collecting

A questionnaire was used to collect data. The questionnaire forms were applied to 411 randomly selected students in vocational and technical education schools in Küçükçekmece district of İstanbul. For reliability of the questionnaire, Reliability Analysis test was applied in RStudio program and reliability coefficient was determined as 0.71.The reliability coefficient is between 0.7 and 1.00 indicating acceptable reliability [21].

2.4. Data Analysis

RStudio program was used for data analysis. R is a language and programming environment for statistical analysis and graphs [22]. The R is used for handling, clearing and developing new statistical methods, for creating simulations, for visualization, and for all commonly used data processing tools for statisticians and data scientists [23].

3. Results and Discussion

The data set of the research consists of 411 observations, 17 characteristics and a class. The read.csv () function is used to transfer the data frame to the RStudio environment. The answers and percentages of the students participating in the study are given in Table 1. In order to apply the C4.5 algorithm to the data frame in the RStudio environment, the "RWeka" and "partykit" packages must be installed and enabled. The install. packages () function was used to install the packages, and the library () function was used to enable the packages. After the packages were installed, the decision tree model in Figure 1 was created with the plot () function.

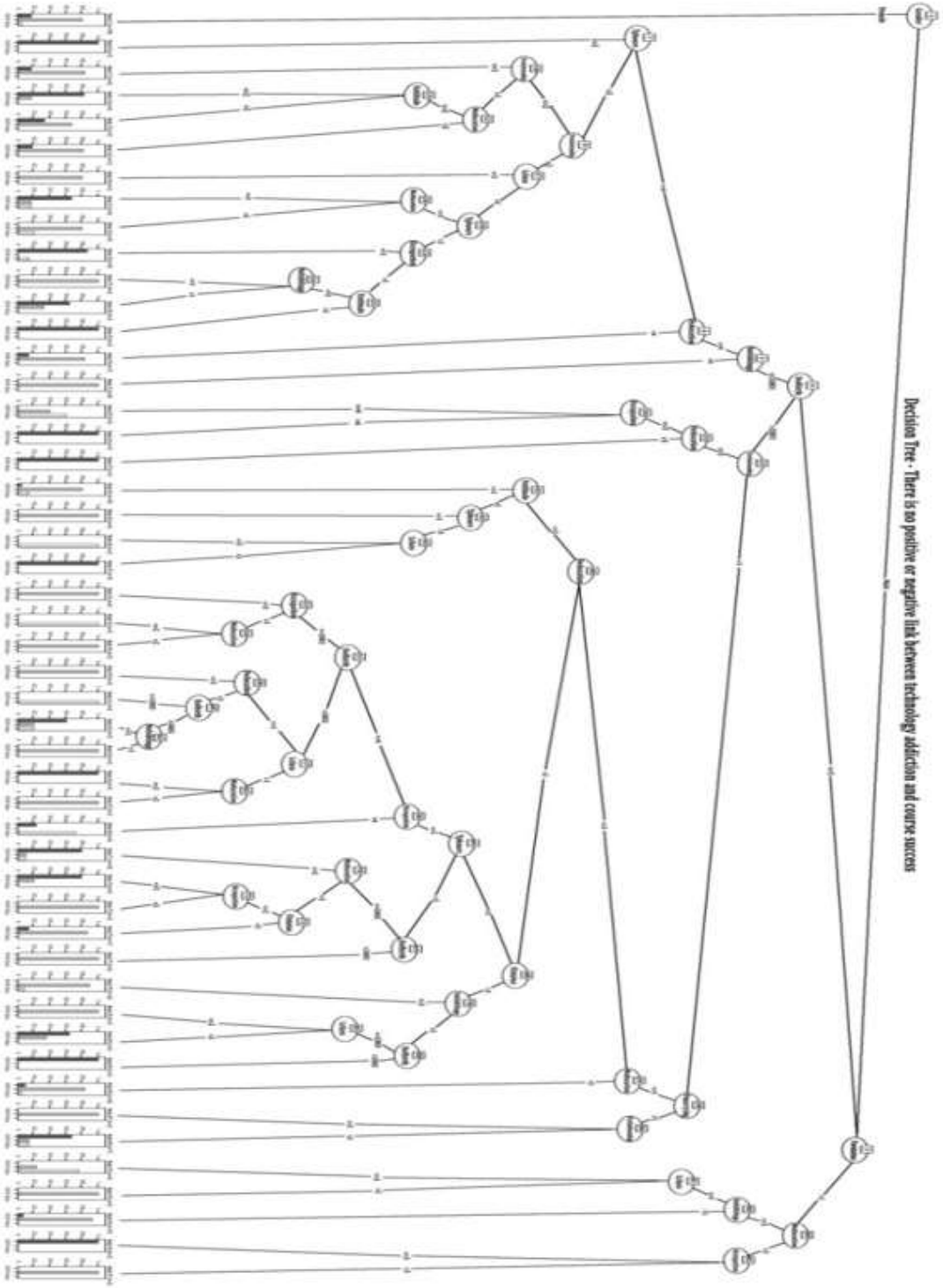

Figure 1. Decision tree based on C4.5 algorithm for dataset

Table 1. Statistical results of the data frame

Features	Frequency	Percent
Gender		
Female	130	31.63
Male	281	68.36
Enter your date of birth (year)		
1999	5	1.21
2000	31	7.54
2001	81	19.70
2002	141	34.30
2003	113	27.49
2004	40	9.73
Number of siblings		
0	16	3.89
1	144	35.03
2	121	29.44
3	59	14.35
4	34	8.27
5	14	3.40
6	12	2.91
7	5	1.21
8	2	0.48
10	2	0.48
11	1	0.24
12	1	0.24
Current learning class		
9th grade	67	16.30
10th grade	130	31.63
11th grade	153	37.22
12th grade	61	14.84
What is the occupational field of education?		
No Area	67	16.30
İnformation Technologies	126	30.65
Office Management	65	15.81
Accounting and Finance	93	22.62
Marketing and Retail	60	14.59
Are your parents alive?		
Mother - Yes	410	99.75
Mother - No	1	0.24
Father - Yes	403	98.05
Father - No	8	1.94
Are your parents stil married?		
Yes	357	86.86
No	54	13.13
Income status of your family		
0-2000	91	22.14
2001-3000	156	37.95
3001-4000	92	22.38
4001 and more	72	17.51

Working status of your parents		
Mother works	157	38.19
Mother does not work	254	61.80
Father work	371	90.26
Father does not work	40	9.73
Your mother's education		
Uneducated	33	8.02
Primary school	177	43.0
Middle school	133	32.3
High school	60	14.5
University	8	1.94
Your father's education		
Uneducated	5	1.21
Primary school	148	36.0
Middle school	154	37.4
High school	87	21.1
University	17	4.13
How much time do you spend on average with technological products?		
Less than 1 hour	16	3.89
1-2 hour	58	14.11
3-4 hour	120	29.19
4 hours and more	217	52.79
What is the attitude of your family towards you?		
Irrelevant	7	4.13
Tolerant	282	68.61
Democratic	36	8.75
Prim	76	18.49
How is your relatonship with your mother?		
Bad	6	1.45
Middle	53	12.89
Good	352	85.64
How is your relatonship with your father?		
Bad	30	7.29
Middle	97	23.60
Good	284	69.09
There is no positive or negative link between technology addiction and course success		
I find it right**(FR)**	97	23.60
I find it wrong**(FW)**	280	68.12
I have no idea**(Nidea)**	34	8.27

The summary () function was used to obtain the image in Figure 2 of the model, which summarizes information.

```
=== Summary ===
Correctly Classified Instances         347               84.4282 %
Kappa statistic                          0.6368
Mean absolute error                      0.1677
Root mean squared error                  0.2895
Relative absolute error                 52.9798 %
Root relative squared error             72.8941 %
Total Number of Instances              411
=== Confusion Matrix ===
   a   b   c   <-- classified as
  63  33   1 |  a = FR
   8 269   3 |  b = FW
   5  14  15 |  c = Nidea
```

Figure 2. Summary informations about the model

According to Figure 2, the following information was obtained;

- ✓ 411 rows were processed in the dataset.
- ✓ The accuracy rate of the classification process was 84.42% and as a result of the classification process, it was seen that 347 data were estimated correctly.

When the Confusion Matrix table is examined;

- ✓ FW class instead of 33 FR and Nidea class instead of 1 FR
- ✓ FR class instead of 8 FW and Nidea class instead of 3 FW
- ✓ FR class instead of 5 Nidea and FW class instead of 14 Nidea, incorrect predictions were made.

When the decision tree in Figure 1 is examined, some of the interpretation that can be reached in determining the answers given to the question that there is no positive or negative link between technology addiction and course success are given below.

- ✓ There is no positive or negative link between technology addiction and course achievement, and gender is the most important feature affecting the answers to the statement.
- ✓ If the gender variable is female, the percentage of FW response is 77.69% (101/130), while the percentage of non-FW is 22.30% (29/130).
- ✓ Gender variable is male, working status of father is working, birth date is before 2002, number of siblings is less than five, father education level is not university, while the usage time of technological products is less than 3-4 hours, , the answer is FR. The percentage of being FR is 100% (3/3).
- ✓ Gender variable is male, working status of the father is working, birth date is before 2002, the number of siblings less than five, the education level of the father is not university, the use of technological products more than 1-2 hours, the father's relationship status is not good, the father relationship status is poor, the answer FW % 83.33 (5/6). The non-FW percentage is 16.66% (1/6).
- ✓ Gender variable is male, working status of father is working, birth date is before 2002, number of siblings is less than five, father's education level is not university, technology products use time is more than 1-2 hours, father's relationship status is good, education class is less than 11, the answer is FW. The percentage of being FW is 100% (2/2).
- ✓ Gender variable is male, working status of father is working, birth date is before 2002, the number of siblings is less than five, while the education level of the father is university, the percentage of FW being the answer is 85.71% (6/7). The percentage of non-FW is 14.28% (1/7).
- ✓ The gender variable is male, the working status of the father is working, birth date is before 2002, while the number of siblings is more than four, the FW percentage of response is 100% (8/8).
- ✓ The gender variable is male, the working status of the father is working, the date of birth is greater than 2001, the relationship status of the father is poor, the education level of the father is high school or university, the answer is FR. The percentage of being FR is 100% (3/3).
- ✓ Gender variable male, working status of father working, date of birth greater than 2001, father's relationship status is medium or good, marital status of the family yes, mother's education level is lower than high school, father's education level is middle school and more, family's income level is 3001-4000 or smaller While the use of technological products is 4 hours and less, occupational area is not marketing and retail, the birth date variable is less than 2003, while occupational area is not accounting and financing or marketing and retail, , the answer is FW. The percentage of being FW is 100% (6/6).

- ✓ Gender variable is male, father's working status is not working, mother's education level is less than high school and the number of siblings is more than 1, the percentage of FW response is 94.73% (18/19). The non-FW percentage is 5.26% (1/19).
- ✓ Gender variable is male, the working status of the father is not working, the education level of the mother is high school or university, the occupational area is information technology or office management or no area, the answer is FR. The percentage of being FR is 100% (2/2).
- ✓ Gender variable is male, the working status of the father is not working, the mother's education level is high school or university, and the occupation area is accounting and financing or marketing and retail, the answer is FW. The percentage of being FW is 100% (2/2).
- ✓ Gender variable is male, father's working status is working, date of birth is greater than 2001, father's relationship status is medium or good, marital status of the family is yes, mother's education level is high school or university, the percentage of FW response is 88.46% (23/26). The non-FW percentage is 11.53% (3/26).
- ✓ Gender variable is male, working status of father is working, date of birth is greater than 2001, father's relationship status is medium or good, marital status of the family is yes, mother's education level is less than high school, father's education level is lower than middle school, your family's attitude towards the individual is indifferent or tolerant. The percent FW was 82.85% (29/35). The percentage of non-FW is 17.14% (6/35).
- ✓ If the gender variable is male, the working status of the father is working, the date of birth is greater than 2001, the relationship status of the father is poor, the education level of the father is less than high school, and the profession is marketing and retail, the answer is FR. The percentage of being FR is 100% (2/2).
- ✓

4. Conclusion

This study was carried out to determine the technology dependence of the students who continue their education in vocational and technical education according to demographic characteristics by using C4.5 decision tree algorithm which is one of the data mining classification methods.Regardingto the question of whether there is any positive or negative relationship between technology addiction and course success, 68.12% of the respondents find it wrong, 23.60% find it right, 8.27% have no idea.According to the decision tree model, the most important feature affecting the answers given to the question is that there is no positive or negative link between the technology addiction and the course success in terms of gender.“Are your parents alive? “ demographic characteristics do not have a significant effect on the decision tree model.It was determined that the classification model created by C4.5 algorithm showed a 84.42% prediction accuracy.It was determined that 347 of the 411 observation values were estimated with the correctly prediction.

This research, at the micro level, is thought that developing students will recognize the concept of technology addiction, become aware of technology usage habits and review their habits. At the macro level, it is thought that increasing the use of data mining in the field of education will contribute to the development of effective education policies on technology addiction and contribute to the education and training processes that can be applied in this direction.

Acknowledgements

The authors would like to thank Green Crescent (Yeşilay) for the usage of a high school survey (Living Independently from Technology) as a part of Turkey Campaign with Addiction Education Programme. Turkey for Addiction Education Program to Combat Survival Dependent on Technology established under the (high school) Temperance Society for the use of the survey, thank you.

Author’s Contributions

Ali BULDU determined the programming language and algorithm type which is used in the research. Akıner KAÇMAZ completed the literature search and created the data set. Akıner KAÇMAZ and Kazım YILDIZ analyzed the data set and evaluated the results. All authors wrote the article together and approved the final version.

Statement of Conflicts of Interest

There is no conflict of interest among the authors.

Statement of Research and Publication Ethics

The authors declares that this study complies with Research and Publication Ethics.

References

[1] Açıkgenç A. 2017. Teknoloji ve Değerler. Uluslararası Teknolojik, Ekonomik ve Sosyal Araştırmalar Vakfı, vs. İstanbul.
[2] Soruhan O. 2018. Sosyal Medya Canavarı Olmak İster Misin?, MediaCat Kitapları, 9s. İstanbul.
[3] Güçlü G. 2015. Yaşam Boyu Öğrenme Argümanı Olarak Teknoloji Bağımlılığı Ve Yaşama Yansımaları. Yüksek Lisans Tezi, Cumhuriyet Üniversitesi, Eğitim Bilimleri Enstitüsü, Sivas.
[4] Şimşek Gürsoy U.T. 2017. Veri Madenciliğinde Güncel Yaklaşımlar. Çağlayan Yayınları, 2s. İstanbul.
[5] Kılınç Ç. 2015. Üniversite Öğrenci Başarısı Üzerine Etki Eden Faktörlerin Veri Madenciliği Yöntemleri İle İncelenmesi. Yüksek Lisans Tezi, Eskişehir Osmangazi Üniversitesi, Fen Bilimleri Enstitüsü, Eskişehir.
[6] Žalik K.R. 2004. Learning Through Data Mining. Computer Applications in Engineering Education, 13 (1): 60-65.
[7] Bayır A., Özdemir Ş., Gülseçen S. 2016. Türkiye'deki Seçmen Eğilimlerinin C4.5 Karar Ağacı Algoritması İle Belirlenmesi. Yönetim Bilişim Sistemleri Dergisi, 1 (3): 223-233.
[8] Aksu M.Ç., Karaman E. 2017. Karar Ağaçları ile Bir Web Sitesinde Link Analizi ve Tespiti. Acta Infologıca, 1 (2): 84-91.
[9] Aydın F. 2017. Teknoloji Bağımlılığının Sınıf Ortamında Yarattığı Sorunlara İlişkin Öğrenci Görüşleri. Yüksek Lisans Tezi, Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
[10] Avşar Arık I. 2018. Teknoloji Bağımlılığı Ve Sosyal Kimlik Oluşumu: Y Nesli. Yüksek Lisans Tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
[11] Gürmen M.S. 2016. Teknoloji Bağımlılığı, Ailemiz Ve Çocuklarımız. 3.Uluslararası Teknoloji Bağımlılığı Kongresi, pp: 161-171, 03-04 Mayıs, İstanbul.
[12] Uzbay İ.T. 2015. Madde Bağımlılığı: Tüm Boyutlarıyla Bağımlılık Ve Bağımlılık Yapan Maddeler. İstanbul Tıp Kitabevi, İstanbul.
[13] Martı H. 2016. Teknoloji Bağımlılığının Dini ve Ahlaki Etkileri. 3.Uluslararası Teknoloji Bağımlılığı Kongresi, pp: 273-284, 03-04 Mayıs, İstanbul.
[14] Dinç M. 2015. Teknoloji Bağımlılığı ve Gençlik. Gençlik Araştırmaları Dergisi, 2147 (8473): 31-65.
[15] Küçükönder H., Vursavuş K.K., Üçkardeş F. 2015. K-Star, Rastgele Orman ve Karar Ağacı (C4.5) Sınıflandırma Algoritmaları ile Domatesin Renk Olgunluğu Üzerinde Bazı Mekanik Özelliklerin Etkisinin Belirlenmesi. Türk Tarım–Gıda Bilim ve Teknoloji Dergisi, 3 (5): 300-306.
[16] Kavzoğlu T., Çölkesen İ. 2010. Karar Ağaçları İle Uydu Görüntülerinin Sınıflandırılması: Kocaeli Örneği. Harita Teknolojileri Elektronik Dergisi, 2 (1): 36-45.
[17] Andrienko G., Andrienko N. 1999. GIS Visualization Support To The C4. 5 Classification Algorithm of KDD. In Proceedings of the 19th International Cartographic Conference, pp: 1-7, 14-21 August, Ottawa.
[18] Kılıçalan M.B. 2018. Hanehalkı İşgücü Araştırma Verileri İle Veri Madenciliği Yöntemlerinin Uygulanması Ve Modellerin Karşılaştırılması. Yüksek Lisans Tezi, Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.
[19] Hssina B., Merbouha A., Ezzikouri H., Erritali M. 2014. A Comparative Study Of Decision Tree ID3 and C4.5. International Journal of Advanced Computer Science and Applications, 4 (2): 13-19.
[20] Altunkaynak B. 2019. Veri Madenciliği Yöntemleri ve R Uygulamaları. Seçkin Yayıncılık, Ankara, 38-52.

[21] Yurdakul I.K. 2013. Veri Toplama Araçlarında Bulunması Gereken Nitelikler, Bilimsel Araştırma Yöntemleri. Anadolu Üniversitesi Web Ofset Tesisleri, Eskişehir, 117-138.
[22] Logan M. 2010. Biostatistical Design and Analysis Using R: A Practical Guide. Wiley-Blackwell, p1., Hoboken.
[23] Buttrey S., Whitaker L.R. 2018. A Data Scientist's Guide to Acquiring, Cleaning, and Managing Data in R. Wiley, p2. Hoboken.