The Physical Educator Vol. 80 • pp. 235–254 • 2023
https://doi.org/10.18666/TPE-2023-V80-I2-11471

PHYSICAL FITNESS

Classifying High School Students' Health-Related Physical Fitness Report Cards With Data Mining

Hande Busra Eren and Gokhan Caliskan

Abstract

In this study, classifications were made from the data obtained from the Health-Related Physical Fitness Report cards and BMIs of students through data mining methods, artificial neural networks, and decision trees models. Then the classification performances of both models were compared. The body weight and height measurements of the students in the Health-Related Physical Fitness Report were formulated, and their BMI classification was made. In addition, it was investigated whether other parameters such as shuttle, push-up, and sit-and-stretch flexibility test values had an effect on BMI classification. The study group comprised 1,050 secondary school students studying in the Cihanbeyli district of Konya in 2017. In conclusion, it was determined that artificial neural networks had more correct classifications than decision trees analysis. In the Health-Related Physical Fitness Report, *shuttle and push-up stood out among the variables affecting BMI classification.*

In the developing and globalizing world, great advances have been made in information technologies. In parallel with this development, storing every desired data has caused billions of bytes of data to be accumulated in electronic media in a very short time. New mechanisms have emerged that meet the requirements of today

Hande Busra Eren, Faculty of Sport Science, Ankara University. Gokhan Caliskan, Faculty of Sport Science, Igdir University.

The authors would like to thank the athletes for their participation in the study.

Please send author correspondence to gokhan.caliskan@igdir.edu.tr

and go beyond data collection. From the aggregated data accumulated in the databases, the need to access meaningful information was felt, and the process of accessing the gold data from raw data was initiated. This process is called "data mining." Data mining is the process of discovering interesting models and information from a large amount of data in databases, enabling automatic retrieval of previously unknown, hidden, meaningful, and useful patterns from large-scale databases (Sever & Oğuz 2002).

Data mining has been successfully applied in the field of educational sciences, as well as in many sectors such as health, economy, and telecommunications for years. Data mining in the field of education is known as educational data mining. The increasing use of technology in education systems has led to the storage of large amounts of student data. This has increased the importance of the use of educational data mining to improve teaching and learning processes (Algarni, 2016). Research using educational data mining has aimed to receive feedback from students about lessons, to determine the factors that affect their success and increase student achievement, to identify deficiencies in the content of education, and to create a more effective education process. In addition, research has aimed to guide students according to their needs and abilities and to offer teaching styles with customizable curricula by clustering students with similar profiles (Algarni, 2016; Altun et al., 2019; Aybek, 2016; J. Chen et al., 2014; Demiralay et al., 2017; Çırak & Çokluk, 2013; Romero & Ventura, 2007). The use of educational data mining has provided benefits to teachers and managers in many branches, as well as to physical education and sports teachers in their lessons.

The literature has included studies on the use of educational data mining in physical education and sports lessons. These studies have explored detailed information from the sports performance data of students, have determined the variables that affect their physical performance, have mentored students by creating prediction models, and have predicted the number of future athletes (Atasoy et al., 2017; Fan et al., 2019; Pan, 2019; Wang, 2020; Zhu, 2018). Moreover, the parameters affecting students' physical fitness measurements have been determined through data mining methods (Rani et al., 2012); footballers have been classified according to their physical condition through cluster analysis, a data mining method (Jiang et al., 2018);

and future obesity situations have been predicted through data mining methods, students' existing BMI, and physical fitness measurement data (YoussefAgha et al., 2013).

One of the objectives of the Republic of Turkiye Ministry of National Education, physical education and sports course is to enable students to adopt these skills by improving their movement, knowledge, and skills through physical activity (Republic of Turkiye Ministry of National Education, 2018). It is known that individuals who regularly perform physical activity are less sick, have higher energy, and feel stronger psychologically (Corbin & Pangrazi, 1993). Individuals who are not physically active enough are known to face health problems such as obesity, high blood pressure, diabetes, and joint disorders (Chinn & Rona, 2001; Doinne et al., 2000).

Countries have been developing policies to solve these problems, which have reached a level that threatens individuals of all ages and has led to serious health expenditures and loss of labor. Schools are the most suitable places to find solutions to social problems because they are the places where children and adolescents spend most of their time. The physical education and sports course, which is an integral part of education programs, is the only class in which students can be physically active. Therefore, it provides an important opportunity to produce solutions to health problems that may occur due to a lack of physical activity (McGinnis et al., 1991). In this context, the *Health-Related Physical Fitness Report* is included in the curriculum of Physical Education and Sports (Republic of Turkiye Ministry of National Education, 2017). The purpose of the *Health-Related Physical Fitness Report* is to develop the knowledge, skills, and attitudes required to lead a healthy lifestyle, by creating the skills necessary for individuals to participate in physical activities (Republic of Turkiye Ministry of National Education, 2017). In this context, height, weight, push-ups, sit-ups, and sit-and-reach flexibility test (right and left) measurements of the students are taken twice a year, at the beginning (September 15–October 15) and end (April 15–May 15) of the academic year. With these measurements, the students are classified as underweight, normal weight, overweight, or obese according to the World Health Organization classifications of BMI. Students and their parents can access the measurement results from the relevant module in the e-school system (Republic of Turkiye Ministry of National Education, 2017). According to the *Health-Related Physical Fitness Report* published by the Republic of Turkiye Ministry of Health in 2018, 16.5% of male students in secondary education are overweight and 6.8% are obese; 13.5% of female students are overweight and 4.3% are obese. Determining the parameters affecting BMI classification plays an important role in reducing the proportion of students classified outside the "normal weight" category in the BMI classification of the *Health-Related Physical Fitness Report*. However, it is necessary to provide the most accurate estimation and classification for this purpose. Artificial neural networks determine the information they receive from the given examples, which constitute the experiences of their own information processes and, thus, in which classification the student belongs (Öztemel, 2012). Another widely preferred classification model with easy installation and interpretation is the decision trees method, which is easy to adapt to databases (Kuyucu, 2012).

The literature review found a limited number of studies that use data mining, especially in physical education and sports. For this reason, it is thought that this study will contribute to the field. In this context, it will be in the community's best interest for physical education and sports to better tackle obesity in students.

For this reason, data mining is an important issue for which to discover and apply new methods that will enable existing programs to better serve their intended goals. It is desired that students meet the classification of "normal weight" rather than "obese." The main purpose of the *Health-Related Physical Fitness Report* is to raise awareness among individuals. Besides revealing which of the *Health-Related Physical Fitness Report* parameters are effective in BMI classification, the use of data mining methods will provide physical education and sports teachers with valuable information about reducing obesity. In this context, the general aim of this study was to classify the measurement data of the students' *Health-Related Physical Fitness Report* through the methods of artificial neural networks and decision trees and to compare their classification performance. In addition, the BMI classification in the study came from the body weight and height measurements of the students in the *Health-Related Physical Fitness Report*. The study investigated whether sit-ups, push-ups, and sit-and-reach flexibility test values affected this BMI classification. In this context, answers to these questions were sought:

1. What accuracy rate does the model, obtained by artificial neural network analysis, have in classifying students in the classification of the *Health-Related Physical Fitness Report*?
2. What accuracy rate does the model, obtained by decision trees analysis, have in classifying students in the classification of the *Health-Related Physical Fitness Report*?
3. What are the results regarding the comparison of the general correct classification rates of artificial neural networks and decision trees methods according to the classification of students in the *Health-Related Physical Fitness Report*?
4. What is the importance of independent variables in the BMI classification of the *Health-Related Physical Fitness Report*?

Method

Participants

One thousand fifty students (460 [43.8%] females, 590 [56.2%] males) studying at secondary schools in the Cihanbeyli district of Konya were included. An appropriate sampling method was used. Table 1 shows the class levels of the students in the study.

Data Collection Tools

Data on the *Health-Related Physical Fitness Report* were obtained from the Republic of Turkiye Ministry of National Education e-school system database. For the data to be used in the research process, ethical principles to be followed during the use of data were determined and permission was obtained from the Konya Provincial Directorate of National Education for research. After all the permissions were completed, data from each student's *Health-Related Physical Fitness Report* were obtained via e-school. Table 2 shows the variables used in the research as a data collection tool and included in the *Health-Related Physical Fitness Report*.

Table 1
Class Levels of the Students in the Study

	Grade				
Students	**9**	**10**	**11**	**12**	**Total**
N	380	272	255	143	1050
%	36.1	25.9	24.3	13.7	100

Table 2
Health-Related Physical Fitness Report Variables

Independent variable	**Tests**
Anthropometric measurements	Weight
	Height
Muscle endurance	Push-ups
	Sit-ups
Flexibility	Sit-and-reach flexibility test (right)
	Sit-and-reach flexibility test (left)

In addition, BMI classification according to the World Health Organization was made for the variables in the *Health-Related Physical Fitness Report*. Students were classified as underweight, normal weight, overweight, or obese.

Statistical Analysis

Artificial neural network analysis and decision trees analysis were used in the data analysis. BMI classification was determined as a dependent variable for decision trees analysis and as an output variable for artificial neural network analysis. The variables of height, weight, push-ups, sit-ups, and sit-and-reach flexibility test (right and left) were included as independent variables for decision trees analysis and as input variables for artificial neural network analysis.

CHAID analysis method was used for decision trees analysis, and multilayer perceptron method was used for artificial neural networks. The data were tested with different algorithms in SPSS 23 software. The model with the highest accuracy was chosen.

Artificial Neural Networks

Artificial neural networks are computer systems that have emerged as a result of the artificial simulation efforts of the human brain, and the ability to create, derive, and discover new information spontaneously through learning, one of the features of the brain (Kantardzic, 2011). They give easier and more accurate results in predicting nonlinear problems in comparison to traditional analysis methods. They consist of artificial cells that are hierarchically linked at the same time and can operate in parallel with each other. The main task of an artificial neural network is to produce an output set that can correspond to an input set shown to it. To do this, the artificial neural network is trained with examples of the same event, and the learning of the network is provided. In this way, the network is given the ability to generalize. It generates its own output values in the face of similar events with its generalization capability. This situation provides the most effective benefit in decision-making processes when there is limited information about the situations but there are examples (Öztemel, 2012). Artificial neural networks consist of many layers and many artificial neural cells. They have three parts: input layer, hidden layer, and output layer. Input variables that enter from the input layer pass through here and come to the hidden layer and are then passed to the output layer (Silahtaroğlu, 2013). There are many artificial neural network structures in the literature, one of which is the multilayer perceptron networks used in this study.

Multilayer perceptron networks have a feed-forward network structure and use the method of teacher learning. Their operating principles are based on reconnecting weights until the mean of error squared is minimized. This is called the "generalized delta rule" (Haykin, 2009).

Decision Trees

On the basis of historical data, decision trees decide which classes the new data belong to by establishing certain rules. Many approaches are used in decision trees, and CHAID is the most widely used approach among them.

CHAID (chi-squared automatic interaction detection) analysis was developed by Kaas in 1980. One advantage of CHAID analysis is that it can work with categorical and continuous variables (Mattison, 1997). In CHAID analysis, there is one dependent (predicted) variable and multiple independent variables (predictor). In this analysis method, all independent variables are compared and the variable that best describes the dependent variable is determined. Then, the data set is divided into nodes, that is, subgroups according to the independent variable that best describes the dependent variable. Nodes continue to create new nodes for other significant independent variables (Avşar & Yalçın, 2015).

Results

Artificial Neural Network Analysis

Table 3 presents the data on students' *Health-Related Physical Fitness Report* classified with the artificial neural network. The artificial neural network model correctly classified the students in the underweight category at 92.9%, the normal weight category at 95.9%, and the overweight category at 100%. Students in Class 1, 2, and 3 obesity categories could not be classified correctly. Total correct classification rate of the training set was 91.5%.

The test set was correct for 91.8% of students in the underweight category, 95.1% of students in the normal weight category, and 95.4% of students in the overweight category. Students in Class 1, 2 and 3 obesity categories could not be classified correctly. The total correct classification rate of the test set was 90.6%.

The training data set was used in the training of the network, and the test data set was used in the measurement of the performance of the training data set. When the findings of the test data set were examined to measure the performance of the training application, on the test data set, the training application performed well with a high correct classification rate.

Table 4 shows the findings regarding the importance order and percentages of the independent variables according to the classification status of the students from the *Health-Related Physical Fitness Report*.

Table 3
Classification Table Resulted From Artificial Neural Network Analysis

World Health Organization's BMI classification categories	Training dataset			Test dataset		
	Correct	Incorrect	Correct classification %	Correct	Incorrect	Correct classification %
Underweight	65	5	92.9	78	7	91.8
Normal weight	301	13	95.9	352	18	95.1
Overweight	76	0	100	83	4	95.4
Class 1 obesity	0	19	0	0	19	0
Class 2 obesity	0	3	0	0	4	0
Class 3 obesity	0	1	0	0	1	0
Total	442	41	91.5	513	532	90.6

Table 4
Artificial Neural Network Analysis Percentage of Importance of Independent Variables

Independent variable	Importance	Normalized importance %
Body Mass Index	0.244	100
Weight	0.240	98.1
Push-ups	0.109	44.5
Height	0.096	39.4
Sit-ups	0.093	38.2
Age	0.090	36.7
Sit and reach flexibility test (left)	0.051	21
Sit and reach flexibility test (right)	0.042	17.2
Sex	0.035	14.4

The most important variable determining the classification was body mass index (100%), followed by weight (98.1%) and push-ups (44.5%). These were followed by variables such as height (39.4%), sit-ups (38.2%), age (36.7%), sit-and-reach flexibility test (left; 21%), sit-and-reach flexibility test (right; 17.2%), and sex (14.4%).

Decision Trees Analysis

Table 5 shows the results of the classification made with the decision trees model. Correct classification was done for 77.4% of students in the underweight category, 90.9% of students in the normal weight category, 79.1% of students in the overweight category, and 31.6% of students in the Class 1 obesity category. Students in Class 2 and 3 obesity categories could not be classified correctly. The total correct classification rate of the test set was 84.2%.

Table 5
Classification Table Resulted From CHAID Analysis

	Classification						
	Estimation						
Observed	**Underweight**	**Normal weight**	**Overweight**	**Class 1 obesity**	**Class 2 obesity**	**Class 3 obesity**	**Correct classification %**
Underweight	120	35	0	0	0	0	77.4
Normal weigh	34	622	28	0	0	0	90.9
Overweight	0	33	129	1	0	0	79.1
Class 1 obesity	0	0	26	12	0	0	31.6
Class 2 obesity	0	0	4	3	0	0	0
Class 3 obesity	0	0	2	0	0	0	0
Correct Classification %	14.7	65.8	18	1.5	0	0	84.2

Comparison of Artificial Neural Network Analysis and Decision Trees Analysis

The performances of artificial neural network analysis and CHAID analysis methods were compared through classification tables of both methods. As a result of artificial neural network analysis, separate classification tables for the test data set and the training data set were created. The average of these two tables was calculated for the general accuracy classification. Table 6 shows the comparative correct classification percentages obtained through artificial neural network and CHAID analyses.

The artificial neural network analysis achieved the highest classification success in terms of the overweight category at 97.7%, whereas the CHAID analysis achieved the highest classification success in terms of the normal weight category at 90.9%.

The artificial neural network analysis made the most accurate classification for students in the underweight, normal weight, and overweight categories, whereas the CHAID analysis made the most correct classification for students in the Class 1 obesity category. The overall correct classification rates were 91.1% for artificial neural network analysis and 84.2% for CHAID analysis.

Table 6
Artificial Neural Networks and CHAID Analyses Classification Percentage Comparison

	Estimation	
Observed	**Artificial neural networks %**	**CHAID %**
Underweight	92.4	77.4
Normal weight	96	90.9
Overweight	97.7	79.1
Class 1 obesity	0	31.6
Class 2 obesity	0	0
Class 3 obesity	0	0
Total correct classification %	91.1	84.2

Discussion

"If you know the enemy and know yourself, you need not fear the result of a hundred battles. If you know yourself but not the enemy, for every victory gained you will also suffer a defeat. If you know neither the enemy nor yourself, you will succumb in every battle."
—Sun Tzu Wu, The Art of War, 450 BC

Meaningful information can be obtained from various data during the education and training process. In this way, educators can get to know themselves, the methods they use, and their students better. Thus, they have the opportunity to produce more effective solutions to their problems. Educators are always striving to get better. One method that provides an advantage to overcome their difficulties is data mining, with data being "the new oil of the digital economy" (Wired, 2014).

With the 84.2% success rate in classifying students according to *Health-Related Physical Fitness Report* measurements and the 91.1% correct classification with the artificial neural network analysis in this study, the artificial neural network analysis, compared to the CHAID analysis, is clearly better at predicting BMI category of students. It is thought that the primary factor in this success of artificial neural networks is to carry out a training process using data and to try to reduce the error rate by constantly renewing itself after this process. Other studies show similar results. The Tosun (2007) study reveals artificial neural network analysis better classifies student achievement than does a decision trees analysis. The Toprak (2017), which classifies students according to their PISA test mathematics achievements, reveals artificial neural networks are better than decision trees and discrimination analysis methods. These studies and other studies in the literature show that artificial neural networks can be used in classification problems in the field of education (Algarni, 2016; Altun et al., 2019; Aybek, 2016; J. Chen et al., 2014; Çırak & Çokluk, 2013; Demiralay et al., 2017; Romero & Ventura, 2012). However, the CHAID analysis also classifies at a high level of accuracy. With the advantage of CHAID analysis, students with similar physical characteristics can organize activities to be included in the same group. Thus, it is possible for the students in the obese category to be included in the normal weight category more systematically and quickly.

The two most important variables that predict *Health-Related Physical Fitness Report* classification among the independent variables in the study are body mass index and body weight. This is due to the World Health Organization's classification of BMI with the Body Weight / (Height*Height) formula. In addition, it is an expected result that the BMI variable calculated from the ratio of the two is among the important variables. It is known that as the number of sit-up and push-up repetitions increases, individuals are more likely to be classified in the normal weight category, and that as the number of repetitions decreases, they are more likely to be classified in a category other than the normal weight category (X. Chen et al., 2020). In the literature, many studies emphasize that individuals in the obese category have significantly lower number of sit-up and push-up repetitions than do their healthy peers (X. Chen et al., 2020; Deforche et al., 2003; İskenderoğlu, 2020; Kamuk, 2019; Kim et al., 2005; Mak et al., 2010; Orjan et al., 2005; Tulum, 2002). In this context, our study shows consistent results with studies on similar age groups in the literature.

Considering these results, it is thought that the possibility of classification of students as normal weight in the *Health-Related Physical Fitness Report* classification increases with the fact that physical education and sports teachers have more room for sit-up and push-up exercises in the classroom and extracurricular activities. In addition, there are studies in the literature on the variables affecting classification by artificial neural network analysis. For example, the Ivankovic et al. (2010) study using artificial neural network analysis to determine the levels at which certain variables affect the wins of the teams in the Serbian Basketball League points to some variables (i.e., numbers from the painted area) bringing teams closer to winning. In this sense, it can be said that artificial neural network analysis is a successful method in determining the degree to which variables affect classification.

The multilayer perceptron model created with artificial neural networks makes the most accurate classification in BMI classification in the overweight and normal weight categories. The normal weight category with the most agglomeration is expected to be highly accurate. However, the highest estimation of the overweight category is the most striking result of artificial neural network analysis. This shows that the classification performance is higher for students in the overweight category than in the other categories. For this reason, the model established with artificial neural network analysis chooses the students in the overweight category better and gives better results in its classification. It is, however, possible to say that the CHAID analysis gives better results, especially in the classification of students in the normal weight category. The artificial neural network analysis fails to correctly classify Class 1, 2 and 3 obesity categories, whereas the CHAID analysis fails to do so for Class 2 and 3 obesity categories. This may happen because these categories contain fewer samples.

In light of these results, imagine a student who has a body weight of 73 kg, height of 135 cm, 3 push-ups, 2 sit-ups, 18 cm sit-and-reach flexibility (right) test, and 17 cm sit-and-reach flexibility (left) test. Artificial neural networks is used in predicting which BMI classification this student will belong to in the future. As a result of the analysis made with variable values added to the data set, this student is assigned to the "overweight" class. The probability of this estimate being correct is 91%. Of the parameters that will allow the obesity level to be brought to the normal weight level, shuttle and push-ups can have an important effect in lowering this level normally for this student, who is predicted to be in the overweight category.

For the generalization of the findings within the scope of this research, further studies with larger samples are recommended. However, this study is thought to be an example for the use of data mining methods in the process of physical education and sports education research. On the other hand, in terms of parameters that influence students' *Health-Related Physical Fitness Report*, researchers must consider BMI classification. With more importance given to these parameters in the classroom and extracurricular activities, more students can be included in the normal weight category in BMI classification.

References

Algarni, A. (2016). Data mining in education. *International Journal of Advanced Computer Science and Applications*, *7*(6). https://doi.org/10.14569/IJACSA.2016.070659

Altun, M., Kayıkçı, K., & Irmak, S. (2019). Sınıf öğretmenliği öğrencilerinin mezuniyet notlarının regresyon analizi ve yapay sinir ağları yöntemleriyle tahmini [Estimation of graduation grades of primary education students by using regression analysis and artificial neural networks]. *E-Uluslararası Eğitim Araştırmaları Dergisi*, *10*(3), 29–43. https://doi.org/10.19160/ijer.624839

Atasoy, M., Dalkılıç, M., & Uğraş, S. (2017). Yapay sinir ağları ile dövüş sporları alanında lisanslı sporcu sayılarının tahmini [Estimation of licensed sportsman-woman in area of martial sports by artificial neural networks]. *Üniversitesi Beden Eğitimi ve Spor Bilimleri Dergisi*, *1*(1), 33–37.

Avşar, A. Ş., & Yalçın, S. (2015). Öğrencilerin okuma başarılarını açıklayan ailesel değişkenlerin chaid analizi ile belirlenmesi [Determining the parental variables that explain students' reading success by using CHAID analysis]. *Eğitim ve Bilim Dergisi*, *40*(179), 1–9. http://egitimvebilim.ted.org.tr/index.php/EB/article/view/2890/1023

Aybek, H. S. (2016). *Öğrenci başarısının yapay sinir ağları ile kestirilmesi: Anadolu Üniversitesi Açıköğretim sistemi örneği* [Predicting achievement with artificial neural networks: The case of Anadolu University open education system; unpublished master's thesis]. Anadolu Üniversitesi, Eskişehir.

Chen, J. F., Hsieh, H. N., & Do, Q. H. (2014). Predicting student academic performance: A comparison of two meta-heuristic algorithms ınspired by cuckoo birds for training neural networks. *Algorithms*, *7*(4), 538–553. https://doi.org/10.3390/a7040538

Chen, X., Cui, J., Zhang, Y., & Peng, W. (2020). The assosciation between BMI and health related physical fitness among Chinese college students: A cross-sectional study. *BMC Public Health*, *20*, Article 444. https://doi.org/10.1186/s12889-020-08517-8

Chinn, S., & Rona, R. J. (2001). Prevalence and trends in overweight and obesity in three cross sectional studies of British children. *British Medical Journal*, *322*, 24–26. https://doi.org/10.1136/bmj.322.7277.24

Çırak, G., & Çokluk, Ö. (2013). Yükseköğretimde öğrenci başarılarının sınıflandırılmasında yapay sinir ağları ve lojistik regresyon yöntemlerinin kullanılması [The usage of artifical neural network and logistic regresssion methods in the classification of student achievement in higher education].

Mediterranean Journal of Humanities, *3*(2), 71–79. https://doi.org/10.13114/MJH/201322471
Corbin, C., & Pangrazi, R. (1993). The health benefits of physical activity. *Physical Activity and Fitness Research Digest*, *1*(1), 1–7.
Deforche, B., Lefevre, J., De Bourdeaudhuij, I., Duquet, W., & Bouckaert, J. (2003). Physical fitness and physical activity in obese and nonobese Flemish youth. *Obesity Research*, *11*(3), 434–441. https://doi.org/10.1038/oby.2003.59
Demiralay, R., Akdenizli, İ., & Halime, B. (2017). Estimating of student success with artificial neural networks. *New Trends and Issues Proceedings on Humanities and Social Sciences*, *3*(7), 21–27. https://doi.org/10.18844/prosoc.v2i7.1980
Doinne, L., Aimeras, N., & Bouchard, C. (2000). The association between vigorous physical activities and fat deposition in male adolescents. *Medicine & Science in Sports & Exercise*, *32*(2), 392–397. https://doi.org/10.1097/00005768-200002000-00020
Fan, Y., Liu, Y., Chen, H., & Ma, J. (2019). Data mining-based design and implementation of college physical education performance management and analysis system. *International Journal of Emerging Technologies in Learning*, *14*(06), 87–96. https://doi.org/10.3991/ijet.v14i06.10159
Haykin, S. (2009). *Neural networks and learning machines* (3rd ed.). Pearson.
İskenderoğlu, C. (2020). *Obez ve sağlıklı adölesanlarda günlük yaşam aktiviteleri, fiziksel uygunluk ve yaşam kalitesinin karşılaştırılması* [A comparison activities of daily living, physical fitness, and quality of life in obese and healthy adolescents; unpublished master's thesis]. Hacettepe Üniversitesi, Ankara.
Ivankovic, Z., Rackovic, M., Markoski, B., Radosav, D., & Ivkovic, M. (2010). Analysis of basketball games using neural networks. In *11th IEEE International Symposium on Computational Intelligence and Informatics: Proceeding*s (pp. 251–256). IEEE. https://doi.org/10.1109/CINTI.2010.5672237
Jiang, B., Sun, H., Bai, W., Li, H., Wang, Y., Xiong, H., & Wang, N. (2018). Data analysis of soccer athletes' physical fitness test based on multi-view clustering. *Journal of Physics Conference Series*, *1060*, Article 012024. https://doi.org/10.1088/1742-6596/1060/1/012024
Kamuk, Y. U. (2019). *Türk Silahlı Kuvvetleri'nde fiziksel uygunluğun değerlendirilmesi* [Evaluation of physical fitness in the Turkish Armed Forces]. Nobel.

Kantardzic, M. (2011). *Data mining: Concepts, models, methods, and algorithms*. Wiley. https://doi.org/10.1002/9781118029145

Kim, J., Must, A., Fitzmaurice, G., Gillman M., Chomitz, V., Kramer, E., Mcgowan, R., & Peterson, K. (2005). Relationship of physical fitness to prevalence and incidence of overweight among schoolchildren. *Obesity Research*, *13*(7), 1246–1254. https://doi.org/10.1038/oby.2005.148

Kuyucu, Y. E. (2012). *Lojistik regresyon analizi (lra), yapay sinir ağları (ysa) ve sınıflandırma ve regresyon ağaçları (cart) yöntemlerinin karşılaştırılması ve tıp alanında bir uygulama* [Comparison of logistic regression analysis (LRA), artificial neural networks (ANN), and classfication and regression trees (C&RT) methods and an application in medicine; unpublished master's thesis]. Gaziosmanpaşa Üniversitesi, Tokat.

Mak, K., Ho, S., Lo, W., Thomas, N., Mcmanus, A., Day, J., & Lam, T. (2010). Health related physical fitness and weight status in Hong Kong adolescents. *BMC Public Health*, *10*, Article 88. https://doi.org/10.1186/1471-2458-10-88

Mattison, R. (1997). *Data warehousing and mining for telecommunications*. Artech House.

McGinnis, M., Kanner, L., & Degraw, C. (1991). Physical education's role in achieving national health objectives. *Research Quarterly for Exercise and Sport*, *62*(2), 138–142. https://doi.org/10.1080/02701367.1991.10608702

Orjan, E., Kristjan, O., & Bjorn, E. (2005). Physical performance and body mass index in Swedish children and adolescents. *Scandivavian Journal of Nutrition*, *49*(4),172–179. https://doi.org/10.1080/11026480500441275

Öztemel, E. (2012). *Yapay Sinir Ağları* [Artificial neural network]. Papatya Yayıncılık.

Pan, L. (2019). A big data-based data mining tool for physical education and technical and tactical analysis. *International Journal of Emerging Technologies in Learning*, *14*(22), 220–231. https://doi.org/10.3991/ijet.v14i22.11345

Rani, B., Kaur, P., & Bansal, A. (2012). Analysis of student physical fitness data using data mining algorithm. *International Journal of Applied Engineering Research*, *7*(11), 1659–1662.

Republic of Turkey Ministry of Health. (2018). *Sağlıkla İlgili Fiziksel Uygunluk Karnesi Raporu* [Health-related physical fitness report]. https://www.saglik.gov.tr/TR,50173/fiziksel-uygunluk-karnesi-ilk-raporu-hazirlandi-07112018.html

Republic of Turkiye Ministry of National Education. (2017). *Fiziksel Uygunluk Karnesi* [Health-related physical fitness report]. http://okulsagligi.meb.gov.tr/www/beden-egitimi-ve-spor-dersi-ogretmenlerine-verilecek-quotfiziksel-aktivite-uygunluk-karnesiquot-egitimi/icerik/47

Republic of Turkiye Ministry of National Education. (2018). *Ortaöğretim Beden Eğitimi ve Spor Dersi Öğretim Programı* [Secondary education physical education and sports lesson curriculum]. https://bit. ly/2X4zpWf

Romero, C., & Ventura, S. (2012). Data mining in education. *WIREs Data Mining and Knowledge Discovery*, *3*(1), 12–27. https://doi.org/10.1002/widm.1075

Sever, H., & Oğuz, B. (2002). Veri tabanlarında bilgi keşfine formel bir yaklaşım, kısım 1: eşleştirme sorguları ve algoritmalar [A Formal approach to discovery in very large databases: Part 1 - Association queries and algorithms]. *Journal of Information World*, *3*(2), 173–204. https://doi.org/10.15612/BD.2002.517

Silahtaroğlu, G. (2013). *Veri madenciliği kavram ve algoritmaları* [Data mining concepts and algorithms]. Papatya.

Toprak, E. (2017). *Yapay sinir ağı, karar ağaçları ve ayırma analizi yöntemleri ile PISA 2012 matematik başarılarının sınıflandırılma performanslarının karşılaştırılması* [Comparison of classification performances of mathematics achievement at Pisa 2012 with artificial neural network, decision trees, and discriminant analysis; unpublished doctoral thesis]. Hacettepe Üniversitesi, Ankara.

Tosun, S. (2007). *Sınıflandırmada yapay sinir ağları ve karar ağaçları karşılaştırması: Öğrenci başarıları üzerine bir uygulama* [Artificial neural networks and decision tree comparison in classification analysis: An application on students' success; unpublished master's thesis]. İstanbul Teknik Üniversitesi.

Tulum, Z. (2002). *7–11 yaş grubu obez çocuklarda fiziksel uygunluk* [Physical fitness of obese and nonobese children aged 7–11 years; unpublished master's thesis]. Dokuz Eylül Üniversitesi, İzmir.

Wang, C. (2020). Analysis method of college student physical education quality based on big data analysis. In Z. Xu, R. M. Parizi, M. Hammoudeh, & O. Loyola-González (Eds.), *Cyber security intelligence and analytics: Proceedings of the 2020 International Conference on Cyber Security Intelligence and Analytics (CSIA 2020)* (pp. 576–581). https://doi.org/10.1007/978-3-030-43306-2_81

Wired. (2014). *Data is the new oil of the digital economy*. https://www.wired.com/insights/2014/07/data-new-oil-digital-economy/
YoussefAgha, A., Lohrmann, D., & Jayawardene, W. (2013). Use of data mining to reveal body mass index (BMI): Patterns among Pennsylvania schoolchildren, pre-k to grade 12. *Journal of School Health*, *83*(2), 85–92. https://doi.org/10.1111/josh.12002
Zhu, S. (2019). Research on data mining of education technical ability training for physical education students based on apriori algorithm. *Cluster Computing*, *22*(Suppl. 6), 14811–14818. https://doi.org/10.1007/s10586-018-2420-8