Türkiye Klinikleri Biyoistatistik Dergisi
Turkiye Klinikleri Journal of Biostatistics Turkiye Klinikleri J Biostat. 2022;14(2):55-62

ORİJİNAL ARAŞTIRMA ORIGINAL RESEARCH DOI: 10.5336/biostatic.2022-88297

Comparison of Prediction Performance of Computer Adaptive Testing with Machine Learning Methods in Response Patterns Suitable for Rasch Model with Simulation Study: A Methodological Research

Rasch Modeline Uygun Yanıt Desenlerinde Bilgisayar Uyarlamalı Test Yöntemi ile Makine Öğrenmesi Yöntemlerinin Tahmin Performanslarının Benzetim Çalışması ile Karşılaştırılması: Metodolojik Araştırma

Emrah Gökay ÖZGÜR[a], Beyza DOĞANAY ERDOĞAN[b]

[a]Department of Biostatistics and Medical Informatics, Kocaeli University Faculty of Medicine, Kocaeli, Türkiye
[b]Department of Biostatistics, Ankara University Faculty of Medicine, Ankara, Türkiye

The present study has been produced from the doctorate thesis of Emrah Gökay özgür entitled "Comparison of Prediction Performance of Computer Adaptive Testing with Machine Learning Methods in Response Patterns Suitable for Rasch Model with Simulation Study" (Ankara, Ankara University, 2020).

ABSTRACT Objective: The scales measuring latent variables are used to gain information about the characteristic (θ) levels of individuals. Scales can be investigated with classical methods as well as the Computer Adaptive Testing (CAT) method. In this study, the performance of machine learning algorithms, including Classification and Regression Tree (CART), Random Forest (RF), Gradient Boosting Machines (GBM) and Extreme Gradient Boosting Machines (XGBoost), were tested as a new approach to a CAT application with the algorithm routinely used in CAT in simulation data derived from different scenarios from the Rasch model. **Material and Methods:** In the CAT application, the Rasch model was used as the probabilistic model and the question selection based on the information criterion was used as the question selection criterion. The performances of the methods were compared, based on the average number of items, the square root of the mean squared error (RMSE), the intraclass correlation coefficient (ICC) and the average prediction value obtained from these predictions. **Results:** Different methods become superior to others as category numbers increased. When the number of items was considered, the CART method made good predictions with the least number of items. When RMSE values were analysed, both GBM and XGBoost methods had low RMSE values. The methods compared have a good ICC value in estimating total scores. **Conclusion:** As a result of general comparisons, we recommend that a person planning a new study use machine learning methods which are frequently used in different fields recently as an alternative to the CAT method in accordance with its purpose.

Keywords: Computerized adaptive testing; machine learning; Rasch; scale

ÖZET Amaç: Örtük (latent) değişkenlerin değerlendirilmesinde kullanılan ölçekler, bireylerin (θ) düzeyleri hakkında bilgi sahibi olmak için kullanılır. Ölçekler, klasik kalem-kâğıt uygulaması yanı sıra Bilgisayar Uyarlamalı Test (BUT) yöntemi ile de uygulanabilir. Bu çalışmada, BUT uygulamasında rutin olarak kullanılan algoritma ile BUT uygulamasına yeni bir yaklaşım olarak makine öğrenmesi algoritmalarının Sınıflandırma ve Regresyon Ağacı [Classification and Regression Tree (CART)], Rastgele Orman [Random Forest (RF)], Gradyan Artırma Makineleri [Gradient Boosting Machines (GBM)] ve Ekstrem Gradyan Artırma [Extreme Gradient Boosting Machines (XGBoost)] Rasch modelinden farklı senaryolardan türetilen bir simülasyon verisinde performansları incelenmiştir. **Gereç ve Yöntemler:** BUT algoritmasında olasılıksal model olarak Rasch modeli ve soru seçim kriteri olarak da bilgi kriterine dayalı soru seçimi kullanılmıştır. Bununla birlikte yöntemlerin performansları, ortalama madde sayısı, hata kareler ortalamasının karekökü [root mean square error (RMSE)], sınıf içi korelasyon katsayısı (SKK) ve bu tahminlerden elde edilen ortalama tahmin değerine göre karşılaştırılmıştır. **Bulgular:** Her parametre için farklı yöntemler diğerlerinden üstün sonuçlar ortaya çıkarmıştır. Madde sayısı dikkate alındığında, CART yönteminin en az madde ile tahminde bulunduğu ortaya çıkmıştır. RMSE değerleri incelendiğinde, GBM ve XGBoost yöntemlerinden elde edilen değerlerin düşük olduğu görülmektedir. Karşılaştırılan yöntemler, toplam puanları tahmin etmede iyi bir SKK değerine sahiptir. **Sonuç:** Günümüzde farklı alanlarda sıklıkla kullanılan makine öğrenme yöntemlerinin genel karşılaştırmalar sonucunda yeni bir çalışma planlayan kişinin amacına uygun olarak BUT yöntemine alternatif olarak kullanılabileceği görülmekte ve önerilmektedir.

Anahtar kelimeler: Bilgisayar uyarlamalı test; makine öğrenmesi; Rasch; ölçek

Correspondence: Emrah Gökay ÖZGÜR
Department of Biostatistics and Medical Informatics, Kocaeli University Faculty of Medicine, Kocaeli, Türkiye
E-mail: emrahgokayozgur@gmail.com

Peer review under responsibility of Turkiye Klinikleri Journal of Biostatistics.

Received: 20 Jan 2022 ***Received in revised form:*** 14 Mar 2022 ***Accepted:*** 30 Mar 2022 ***Available online:*** 25 Apr 2022

The center of any health system can be regarded as “the patient”. In recent years, patient-centered health systems have become increasingly important. It is anticipated that the patient-reported outcomes of any clinical intervention that the patient has undergone will be more important in the future than other results, such as clinical, physiological or caregiver reports. Measurements of physical function, symptoms, global health assessment, psychological well-being, social well-being, cognitive function, and health-related quality of life (QoL) can be made using patient reported results.

Especially in diseases such as cancer, there may be situations where the QoL depends on the progression of the disease. Factors affecting the QoL are often multiple due to the progression of cancer, the economic burden of the disease, negative effects on home life and negative effects on patient psychology. In such cases, patient reports are helpful in determining the QoL in cancer patients.

In the field of healthcare, scales are often used to measure latent variables that cannot be measured directly. Scales are measurement tools that consist of appropriate items, and should be reliable, valid and sensitive to change. The results obtained from the scales are used to make value judgments about the examined property levels of individuals (θ). Scales can be completed directly by the patient with pen and paper, or they can be applied using a Computer Adaptive Testing (CAT) application. CAT is a form of test that is applied with the help of a computer and adapted to the selection of questions that will measure the level of (θ) of the person to whom the test is applied in the most accurate way. Participants do not need to look at the entire test or scale. CAT models hold information about success and failure in problem solving from previous steps.[2] Based on previous responses and appropriate models, we can precisely estimate a reliable assessment of the level of (θ) of the person. The important thing here is to obtain the level of (θ) of the person with minimum error without looking at the whole test or scale.[1]

Machine learning has been used widely and often in many fields in recent years. There are studies conducted for both prediction and classification purposes. In addition, there are also studies in which scales are used as a size reduction method to reduce the number of items. In light of this, the idea has emerged that machine learning methods can be used to predict the total scores of individuals in scales.

In this study, the total scores of individuals on datasets with different numbers of categories and different numbers of individuals, derived from the Rasch model will be predicted with the algorithm commonly used in CAT application and the addition of machine learning algorithms used as a new approach to CAT application. The machine learning algorithms used in this study were Classification and Regression Tree (CART), Random Forest (RF), Gradient Boosting Machines (GBM) and Extreme Gradient Boosting Machines (XGBoost). Furthermore, the prediction performances of the different machine learning algorithms will be compared according to certain study criteria.

MATERIAL AND METHODS

COMPUTERIZED ADAPTIVE TESTING

In the traditional or fixed form test, the same items are applied to individuals using a paper-pencil form or on a computer screen. An adaptive test sets the difficulty for different people by going beyond the concept of the same item, and items of different difficulty are applied to evaluate the performance of each person. This is the most important feature that distinguishes CAT from other traditional approaches.

One of the fundamentals of an adaptive test is that the test experience is dynamic and sensitive to the test participant's performance.[2] The idea of adaptive testing is not new and it was first applied by Binet in 1905 with the intelligence test.[3] In this test, Binet selected each successive question according to their performance in the previous questions and ensured that the IQ test was both effective and suitable for the respondent's skill. The aim here is to estimate the level of individuals (θ) by choosing separate questions according to their performance for each individual rather than the group.[3]

Using a calibrated question bank, the test starts with the starting question selected based on a certain rule for each person to whom the test will be applied and the person's (θ) level is estimated. Afterwards, a new question is selected according to the determined item selection rule and again (θ) estimation is made. The question selection and (θ) estimation steps continue iteratively until the specified stopping criterion is met. When the stopping rule is met, the test ends and the last estimate is taken as the person's (θ) level estimate. All CAT algorithms are based on this basic format. However, some details such as question selection rules vary.

CART

CART is a versatile machine learning algorithm that can perform both classification and regression tasks.[4] Tree-based learning algorithms are considered to be one of the best and most used supervised learning methods.[5] Tree-based methods strengthen models used for prediction with high accuracy, stability and ease of interpretation. Unlike linear models, tree-based models also show nonlinear relationships quite well. They can be adapted to solve any problem (classification or regression) available.[6]

CART is divided into two categories, categorical or quantitative, according to the type of dependent variable. The CART to be used when the dependent variable is categorical is called the classification tree.

RF

RF, Bagging and Random Subspace (RS) methods, which are the types of community learning methods. In the Bagging method, only sample selection is made randomly, while in the RF method, there is randomness in both sample and variable selection.[7] In the RF method, sample selection is made with bootstrap and variable selection with RS.

In the RF method, a series of decision trees are created on training samples. However, when constructing these decision trees, considering each split in a tree, instead of all p estimators, a random subset of these estimators of size m is chosen as candidate estimators to be used in the cleavage. Division allows only one of these m estimators to be used. A new sample of the estimator m is taken at each division, and typically $m \approx \sqrt{p}$ for classification problems and m=p/3 for regression problems.[4] That is, the number of estimators taken into account in each division is approximately equal to the square root of the total number of estimators in classification problems, and one third in regression problems.

GBM

In the Boosting method, an attempt is made to derive strong learners by forming them from weak learners. This process is done iteratively. New models are established to obtain a more precise estimation of the response variable.[8] The difference between boosting algorithms usually emerges in the way weak learners define their deficiencies.[9]

The first leaf is created in GBM. Then, new trees are created by considering the prediction errors. This situation continues until the number of decided trees is reached or no further development can be made from the model.[8]

EXTREME GRADIENT BOOSTING

XGBoost is an optimized, high performance version of the GBM algorithm. It was introduced in the article "XGBoost: A Scalable Tree Boosting System" published by Tianqi Chen and Carlos Guestrin in 2016. The most important features of the algorithm are that it can quickly obtain high predictive power, prevent over-learning and manage missing data. According to Chen and Guestrin, XGBoost works 10 times faster than other popular algorithms. Software and hardware optimization techniques have been applied to achieve superior results using less resource. It is cited as the best of decision tree based algorithm.[10]

SIMULATION

The purpose of this study was to examine the performance of machine learning methods as a new approach to CAT application with the algorithm commonly used in CAT application on data derived from the Rasch model, and to examine whether this approach could be an alternative to the CAT algorithm. In the scope of the study, deriving data from the Rasch model, the number of items was conducted in different scenarios according to the level of the individuals, item difficulties and the number of categories of item response options.

For the derived data set, the number of items were determined as 50, 150 and 250, the number of people was 100. The number of response categories for each item was determined as 2, 3, and 5. Within the scope of the study, 1000 Monte Carlo repetitions were made. For the derived data set, the number of items were determined as 50, 150 and 250, the number of people was 100. The number of response categories for each item was determined as 2, 3, and 5. Within the scope of the study, 1000 Monte Carlo repetitions were made. All scenarios are given in Table 1.

DETERMINING THE (Θ) LEVEL OF PERSONS

The vector containing the individual parameters was created in increments of $(\theta_{maximum} - \theta_{minimum})/(number\ of\ person - 1)$ units, with a minimum of -4 logit and a maximum of 4 logit.

DETERMINATION OF ITEM PARAMETERS (β)

Item difficulties, with a minimum -4 logit, maximum 4 logit were used and the number of items was created in increments of $(\beta_{maximum} - \beta_{minimum})/(number\ of\ person - 1)$ units to be 50.150 and 250. The average of the threshold difficulties for the items was equal to the item difficulty. Threshold difficulties (β_{jk}) are obtained by adding the threshold values (τ_{jk}) to the item difficulty (β_i). Therefore, the sums of threshold values must be zero. Threshold difficulties were obtained by deriving random numbers equal to the threshold number from the Uniform (0.2) distribution for three categories and Uniform (0.4) for five categories, thus according to the number of categories for each item, and adding and subtracting this random number from the item difficulties.

Within the scope of this study, the feature estimation examined for the CAT algorithm was carried out with the expected value of the Expected a Posteriori and the question selection was performed using the Maximum Posterior Weighted Information method. The standard error of θ level estimation was determined as 0.33 as the stopping criterion in the algorithm.

TABLE 1: Information of scenarios S1-S9.

	Number of items	Number of categories	θ	β
S_1	50	2.3.5	-4.5;4.5	-4.5;4.5
S_2	150	2.3.5	-4.5;4.5	-4.5;4.5
S_3	250	2.3.5	-4.5;4.5	-4.5;4.5
S_4	50	2.3.5	-4.5;2.5	-2.5;4.5
S_5	150	2.3.5	-4.5;2.5	-2.5;4.5
S_6	250	2.3.5	-4.5;2.5	-2.5;4.5
S_7	50	2.3.5	-2.5;4.5	-4.5;2.5
S_8	150	2.3.5	-2.5;4.5	-4.5;2.5
S_9	250	2.3.5	-2.5;4.5	-4.5;2.5

STATISTICAL ANALYSIS

The CAT algorithm was implemented in the SmartCat program (Ankara University Faculty of Medicine, Department of Biostatistics, Psychometric Lab.) and R v.3.6.3 program.

Analysis of machine learning methods was performed with the rpart, rpart.plot, caret, randomForest, gbm and xgboost libraries in R v3.6.3 program. In addition, the irr library was used for the intraclass correlation coefficient (ICC) between the predictions and the total score.

The simulated data set was divided into 80% training data and 20% test data. The results of the test data were obtained on the models obtained from the training set. The CAT method was also applied for test data in terms of integrity in the comparison of CAT method and machine learning methods.

RESULTS

When making comparisons on the basis of category and scenario, low root mean square error (RMSE), low item number and high ICC are desired. Looking at scenario 1 (S1) in Table 2, when the number of categories was 2, the lowest RMSE value was obtained from the CAT method. As the number of categories increased, different methods come to the fore. When there were 3 and 5 categories, the lowest RMSE value was obtained from the XGBoost method. Considering the number of items, it is evident that the CART method is capable of performing the least item estimate. As the number of categories increased in ICC values, the XGBoost method became the best performer.

TABLE 2: Results of scenarios S_1. S_2 and S_3.

		2 Categories			3 Categories			5 Categories		
		RMSE	**Item**	**ICC**	**RMSE**	**Item**	**ICC**	**RMSE**	**Item**	**ICC**
S_1	CAT	**1.505**	49.997	**0.984**	4.845	19.193	0.993	9.580	8.898	0.993
	CART	4.473	**7.241**	0.743	10.478	**7.701**	0.954	16.800	**7.612**	0.972
	RF	2.507	16	0.922	3.382	16	0.995	4.986	16	0.997
	GBM	1.965	18	0.962	3.626	42	0.995	6.673	43	0.995
	XGBoost	2.527	48	0.917	**1.978**	50	**0.998**	**2.141**	50	**0.999**
S_2	CAT	4.700	49.997	0.864	13.912	19.193	0.993	28.478	9.150	0.912
	CART	5.276	**7.419**	0.725	10.332	**7.293**	0.930	16.434	**7.469**	0.959
	RF	2.889	16	0.922	3.850	16	0.990	5.647	16	0.995
	GBM	**2.087**	29	**0.969**	3.498	30	0.995	6.204	39	0.994
	XGBoost	2.983	50	0.913	**2.211**	49	**0.997**	**1.966**	50	**0.999**
S_3	CAT	4.426	49.997	0.864	11.413	17.002	0.940	28.779	9.150	0.912
	CART	5.271	**7.418**	0.725	10.353	**7.778**	0.930	16.363	**7.432**	0.959
	RF	2.864	16	0.924	3.834	16	0.991	5.559	16	0.995
	GBM	**2.069**	28	**0.969**	3.570	32	0.992	3.238	38	0.994
	XGBoost	2.494	50	0.944	**1.287**	50	**0.999**	**1.431**	50	**0.999**

CAT: Computer Adaptive Testing; CART: Classification and Regression Tree; RF: Random Forest; GBM: Gradient Boosting Machines; XGBoost: Extreme Gradient Boosting Machines.

While the number of categories was 2 in S2 and S3, the lowest RMSE value was obtained from the GBM method. In other cases, a lower RMSE value was obtained with the XGBoost method. Considering the number of items, again prediction was made with the least number of items by the CART method. Considering the agreement between the predicted value and the actual value, when the number of categories was 2, GBM produced the lowest ICC values and as the number of categories increased, higher ICC values were obtained from the XGBoost method.

Looking at the other scenarios (S_4, S_5,…., S_9), the lowest RMSE and highest ICC values were obtained from the XGBoost method in all category numbers, while the lowest item number was obtained from the CART method. Looking at the other scenarios (S4, S5,…., S9) in Table 3 and Table 4, the lowest RMSE and highest ICC values were obtained from the XGBoost method in all category numbers, while the lowest item number was obtained from the CART method.

TABLE 3: Results of scenarios S_4. S_5 and S_6.

		2 Categories			3 Categories			5 Categories		
		RMSE	**Item**	**ICC**	**RMSE**	**Item**	**ICC**	**RMSE**	**Item**	**ICC**
S_4	CAT	5.372	38.242	0.976	16.165	16.965	0.991	29.979	8.855	0.992
	CART	13.400	**7.721**	0.708	32.393	**7.869**	0.952	51.616	**7.642**	0.971
	RF	7.539	50	0.909	10.272	50	0.995	14.464	50	0.997
	GBM	5.226	56	0.976	11.356	93	0.994	20.367	101	0.995
	XGBoost	**3.969**	150	**0.979**	**3.185**	150	**0.999**	**7.347**	150	**0.999**
S_5	CAT	20.875	37.106	0.687	44.017	16.913	0.906	74.630	8.219	0.928
	CART	15.818	**7.763**	0.710	31.259	**7.735**	0.930	50.290	**7.784**	0.957
	RF	8.652	50	0.915	11.620	50	0.990	16.658	50	0.995
	GBM	5.992	62	0.969	11.195	81	0.991	19.418	87	0.994
	XGBoost	**4.797**	149	**0.979**	**3.511**	148	**0.999**	**4.612**	150	**0.999**
S_6	CAT	20.114	37.423	0.704	42.624	17.864	0.910	87.710	9.236	0.906
	CART	15.894	**7.763**	0.697	32.132	**7.649**	0.925	50.566	**7.818**	0.956
	RF	8.580	50	0.915	11.542	50	0.990	16.479	50	0.995
	GBM	5.949	74	0.968	11.165	83	0.991	18.918	82	0.994
	XGBoost	**3.640**	150	**0.988**	**2.155**	150	**0.999**	**7.782**	150	**0.999**

CAT: Computer Adaptive Testing; CART: Classification and Regression Tree; RF: Random Forest; GBM: Gradient Boosting Machines; XGBoost: Extreme Gradient Boosting Machines.

TABLE 4: Results of scenarios S_7. S_8 and S_9.

		2 Categories			3 Categories			5 Categories		
		RMSE	**Item**	**ICC**	**RMSE**	**Item**	**ICC**	**RMSE**	**Item**	**ICC**
S_7	CAT	10.063	36.618	0.969	27.597	16.927	0.991	50.285	8.854	0.992
	CART	22.489	**7.830**	0.702	54.192	**7.843**	0.951	85.709	**7.647**	0.971
	RF	12.634	83	0.907	16.936	83	0.995	23.582	83	0.997
	GBM	8.786	74	0.966	19.330	126	0.994	34.058	122	0.995
	XGBoost	**5.014**	248	**0.989**	**5.408**	250	**0.999**	**14.703**	250	**0.999**
S_8	CAT	33.896	36.153	0.687	72.553	16.530	0.906	146.092	8.143	0.906
	CART	26.834	**7.809**	0.981	53.547	**7.776**	0.924	84.321	**7.872**	0.957
	RF	14.330	83	0.912	19.308	83	0.990	27.777	83	0.995
	GBM	10.044	89	0.967	18.654	120	0.991	31.888	122	0.994
	XGBoost	**5.966**	245	**0.988**	**4.519**	249	**0.999**	**9.662**	249	**0.999**
S_9	CAT	30.566	36.550	0.747	66.878	17.567	0.919	151.493	9.178	0.900
	CART	26.636	**7.826**	0.693	53.082	**7.789**	0.927	83.750	**7.875**	0.957
	RF	14.446	83	0.912	19.237	83	0.990	27.066	83	0.995
	GBM	10.128	80	0.966	18.801	107	0.991	31.192	104	0.994
	XGBoost	**4.445**	250	**0.994**	**5.097**	250	**0.999**	**15.565**	250	**0.998**

CAT: Computer Adaptive Testing; CART: Classification and Regression Tree; RF: Random Forest; GBM: Gradient Boosting Machines; XGBoost: Extreme Gradient Boosting Machines.

DISCUSSION

In the field of health, scales are used in the assessment of latent variables that cannot be measured directly in order to gain information about the θ levels of individuals. Scales can be applied using the CAT method, as well as in the traditional fashion with pen and paper. Using the CAT method, means that θ levels of individuals can be obtained in a shorter time and using fewer items. Nowadays, CAT applications are used in most of the studies in this field.

Machine learning methods, which have become increasingly important in recent years, have been used in many areas. Machine learning techniques are gaining importance in the field of health. In this study, the research question was whether machine learning methods can be used as an alternative to the CAT method to obtain θ levels of individuals from scales.

Michel et al. compared the estimation performance of CART and CAT methods.[5] In this study, an agreement was found between the predictive values of the scores. They estimated the total score of individuals as 50.09 with CART and 50.00 with CAT. In addition, when the RMSE values of these two methods were examined, the RMSE values of the models that are the best for both methods are found to be 0.16 for CART and 0.22 for CAT. As a result, these authors suggested that the CART method can be used as an alternative to the CAT method.

In another study, Peute et al. investigated if the CART or CAT method was better in reducing the number of items in the resulting scale.[11] They suggested that the scale size can be used in both methods after finding that the CART method made acceptable estimations with 2.4 items while this was 5.3 items with the CAT method. In addition, the CART method gave better sensitivity and specificity values.

Harrison et al. estimated less items with the decision tree method in a comparative simulation study (mean of 7.32 items in decision tree assessments vs 9.00 items in CAT assessments).[12] However, it was concluded that the RMSE and Mean Absolute Error values of the estimation values obtained by the decision tree method were higher than the CAT method. Harrison et al. concluded that the error values were higher, although predictions were made with fewer items with the decision tree method.[12]

CONCLUSION

In this study, machine learning methods were shown to be an acceptable alternative to the CAT method on the basis of a range of indicators. Depending on the aim of the exercise, different machine learning methods can be used as an alternative to the CAT method for low RMSE, low item count or high ICC. When a low average number of items is desired, the CART method can be used. When low RMSE is concerned, the XGBoost method can be used but since the average number of items used in the XGBoost method is high, RF and GBM methods can also be used in this case. When it comes to ICC, we suggest that all of the machine learning techniques we tested, CART, GBM, XGBoost and RF can be used as an alternative to CAT method in situations with large item numbers.

Source of Finance

During this study, no financial or spiritual support was received neither from any pharmaceutical company that has a direct connection with the research subject, nor from a company that provides or produces medical instruments and materials which may negatively affect the evaluation process of this study.

Conflict of Interest

No conflicts of interest between the authors and/or family members of the scientific and medical committee members or members of the potential conflicts of interest, counseling, expertise, working conditions, share holding and similar situations in any firm.

Authorship Contributions

All authors contributed equally while this study preparing.

Availability of Data and Material

The data set was obtained by simulation.

REFERENCES

1. Merembayev T, Amirgaliyeve S, Kozhaly K. Using item response theory in machine learning algorithms for student response data. IEEE Smart Information Systems and Technologies; 2021 April 28-30; Nur-Sultan, Kazakhstan. [Crossref]
2. Rezaie M, Golshan M. Computer Adaptive Test (CAT): advantages and limitations. International Journal of Educational Investigations. 2015;2(5):128-37. [Link]
3. Oztuna D. Kas-iskelet sistemi sorunlarının özürlülük değerlendiriminde bilgisayar uyarlamalı test yönteminin uygulanması [Doktora tezi]. Ankara: Ankara Üniversitesi; 2008. Erişim Tari-hi:16.06.2022 [Link]
4. Michel P, Baumstarck K, Loundou A, Ghattas B, Auquier P, Boyer L. Computerized adaptive testing with decision regression trees: an alternative to item response theory for quality of life measurement in multiple sclerosis. Patient Prefer Adherence. 2018;12:1043-53. [Crossref] [PubMed] [PMC]
5. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning with Applications in R. 4th ed. New York: Springer; 2014. [Crossref]
6. Loh WY. Classification and regression trees. Wires Data Mining and Knowledge Discovery. 2011;1(1):14-23. [Crossref]
7. Keskin MV. Büyük veride makine öğrenmesi uygulaması [Yüksek lisans tezi]. İstanbul: Yıldız Teknik Üniversitesi; 2018. Erişim Tarihi:16.06.2022 [Link]
8. Natekin A, Knoll A. Gradient boosting machines, a tutorial. Front Neurorobot. 2013;7:21. [Crossref] [PubMed] [PMC]
9. Muratlar ER. Gradient Boosted Regression Tree. 2020. Cited: February 15, 2022. Available from: [Link]
10. Tianqi C, Carlos G. XGBoost: A Scalable Tree Boosting System. KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016; 785-794. [Crossref]
11. Peute L, Scheeve T, Jaspers M. Classification and regression tree and computer adaptive testing in cardiac rehabilitation: instrument validation study. J Med Internet Res. 2020;22(1):e12509. [Crossref] [PubMed] [PMC]
12. Harrison CJ, Sidey-Gibbons CJ, Klassen AF, Wong Riff KWY, Furniss D, Swan MC, et al. Recursive partitioning vs computerized adaptive testing to reduce the burden of health assessments in cleft lip and/or palate: comparative simulation study. J Med Internet Res. 2021;23(7):e26412. [Crossref] [PubMed] [PMC]