International Journal of Assessment Tools in Education
2021, Vol. 8, No. 4, 928–947
https://doi.org/10.21449/ijate.880914
Published at https://ijate.net/ https://dergipark.org.tr/en/pub/ijate *Research Article*

Classification of Scale Items with Exploratory Graph Analysis and Machine Learning Methods

Ilhan Koyuncu [1,*], **Abdullah Faruk Kilic** [2]

[1]Adıyaman University, Faculty of Education, Department of Educational Sciences, Measurement and Evaluation in Education
[2]Adıyaman University, Faculty of Education, Department of Educational Sciences, Measurement and Evaluation in Education

ARTICLE HISTORY
Received: Feb. 15, 2021
Revised: Oct. 22, 2021
Accepted: Nov. 05, 2021

Keywords:
PISA,
Machine learning,
Exploratory factor analysis,
Exploratory graph analysis,
Monte Carlo simulation,
Scale development.

Abstract: In exploratory factor analysis, although the researchers decide which items belong to which factors by considering statistical results, the decisions taken sometimes can be subjective in case of having items with similar factor loadings and complex factor structures. The aim of this study was to examine the validity of classifying items into dimensions with exploratory graph analysis (EGA), which has been used in determining the number of dimensions in recent years and machine learning methods. A Monte Carlo simulation was performed with a total number of 96 simulation conditions including average factor loadings, sample size, number of items per dimension, number of dimensions, and distribution of data. Percent correct and Kappa concordance values were used in the evaluation of the methods. When the findings obtained for different conditions were evaluated together, it was seen that the machine learning methods gave results comparable to those of EGA. Machine learning methods showed high performance in terms of percent correct values, especially in small and medium-sized samples. In all conditions where the average factor loading was .70, BayesNet, Naive Bayes, RandomForest, and RseslibKnn methods showed accurate classification performances above 80% like EGA method. BayesNet, Simple Logistic and RBFNetwork methods also demonstrated acceptable or high performance under many conditions. In general, Kappa concordance values also supported these results. The results revealed that machine learning methods can be used for similar conditions to examine whether the distribution of items across factors is done accurately or not.

1. INTRODUCTION

Exploratory factor analysis (EFA) is frequently used in scale development or adaptation studies (Fabrigar et al., 1999; Floyd & Widaman, 1995; Kline, 1994). There is a wide acceptance in the literature that EFA can be used in the process of searching evidence for construct validity (Nunnally & Bernstein, 1994). For this reason, the correct use of this frequently used method becomes important in terms of the correctness of decisions made by the researchers (Kılıç & Koyuncu, 2017; Koyuncu & Kılıç, 2019).

*CONTACT: Ilhan Koyuncu ilhankync@gmail.com Adıyaman University, Faculty of Education, Department of Educational Sciences, Measurement and Evaluation in Education

e-ISSN: 2148-7456 /

While performing EFA, it should be examined whether the data set meets the assumptions or not. The assumptions of EFA are the fact that the variables need to have a multivariate normal distribution, the sample size should be sufficient, there must be a linear relationship between the variables, and there need to be no extreme values in the data set, and no multicollinearity and singularity among the variables (Tabachnik & Fidell, 2012). After the data sets are analyzed in terms of assumptions, some methods are used to decide the number of factors. These methods include parallel analysis (Horn, 1965; Timmerman & Lorenzo-Seva, 2011), scree plot (Cattell, 1966), Minimum Average Partial (MAP) analysis (Guadagnoli & Velicer, 1988), and K1 (Kaiser, 1960) rule. After deciding on the number of factors by using several of these methods, the distribution of the items to the factors is examined. It is suggested that the factor loadings of the items should be above .30 (Costello & Osborne, 2005), .32 (Tabachnik & Fidell, 2012), or .40 (Howard, 2016). In this case, using different rotation methods in multi-dimensional structures, the items are to be placed in the dimensions in a meaningful way.

There are both vertical and oblique rotation methods used in multidimensional structures in placing the items to the dimensions. Oblique rotation methods are used if there is a relationship between factors, and vertical rotation methods are used if there is no relationship (Osborne, 2015). However, because there are many rotation methods and different rotation methods give different results, researchers may have difficulty in interpreting factor structures. At this stage, whether the revealed factor structure is compatible with the relevant literature or not is evaluated. On the other hand, it becomes difficult to decide which item will be included in which dimension, especially in cases with overlapping factor loadings. At this point, establishing a mainstay in placing items into dimensions will make it easier for researchers to have accurate decisions. Machine learning methods do not prevent cross loadings; however, they can be used to give researchers an idea about placing the items having cross loadings into the accurate dimension. Therefore, the primary purpose of this study is to use machine learning methods to classify items. However, in the related literature (e.g. Belvederi Murri et al., 2020; Fischer & Alfons Karl, 2020; Kjellström & Golino, 2019; Panayiotou et al., 2020), it is seen that EGA (Golino & Epskamp, 2017) is also used to reveal the relationships between the items. To this end, this study aimed to compare the results of machine learning methods, whose purpose is to make classification, with the EGA, which was developed to explain the relationships between items. Therefore, whether machine learning and EGA would give valid results in the classification of items to the dimensions was examined in this particular study.

EGA is a technique of estimating the number of dimensions and classification of items based on network psychometrics. Network psychometrics is a field that was developed to model networks in psychological data and at the same time, it has undergone advances that allow examining relationships between items (Golino & Epskamp, 2017). EGA makes estimates using the Gaussian graphic model. The Gaussian graphical model predicts the common distribution of variables using the inverse of the variance-covariance matrix. As a result of the estimation, nodes and edges connecting these nodes are obtained. In factor analysis, nodes correspond to items, while edges correspond to factor loadings (Golino & Epskamp, 2017; Golino et al., 2020). As a result of EGA, both the number of factors and those items grouped together are obtained. Golino and Epskamp (2017) and Golino et al. (2020) compared EGA with the methods of determining the number of dimensions (such as parallel analysis, K1 rule, and MAP analysis) and reported EGA as the method that gave the most accurate results when sample size was 5000, the factor structure was four-dimensional, and the correlation between dimensions was .70. Moreover, there are also many researchers who use the EGA in their studies in terms of examining individual differences (Fischer & Alfons Karl, 2020), relationship between observed and latent variables (Belvederi Murri et al., 2020), estimating the number of latent variables (Kjellström & Golino, 2019), and exploring the dimensionality of the social skills (Panayiotou et al., 2020).

Since the focus of this study was the classification of scale items, the performance of machine learning methods seeking answers to classification problems was evaluated. Machine learning methods mostly focus on classification, estimation, and clustering problems. Machine learning which is used to analyze a variety of data structures is one of the fastest developing technical areas of today. This technique, located in the center of artificial intelligence and data science, is at the intersection of computer science and statistical machine learning methods are used in many fields (Jordan & Mitchell, 2015). In this study, whether scale items were correctly classified into the dimensions or not was investigated. For this purpose, some frequently used machine learning methods (Pu et al., 2020) given under the titles Bayes, functions, lazy and trees in the Waikato Environment for Knowledge Analysis (WEKA) software (Hall et al., 2009) were compared. The reason for selecting many methods that are frequently used in the machine learning is based on the fact that different methods are effective for different data structures in the related literature (see Barker et al., 2004; Romero et al., 2013). Summary information about the algorithms used is given in Table 1.

Table 1. *Machine learning algorithms used in the study.*

Title	Algorithm	Explanations
Bayes	BayesNet	Classifies with the method of Bayesian networks. Outputs for network structure, conditional probability distributions, and Bayesian networks are obtained. Various search algorithms and quality measures are used (Hall et al., 2009).
	NaiveBayes	The main purpose of this algorithm that is used in supervised learning is to predict classification probabilities based on estimated class probabilities (John & Langley, 1995).
functions	RBFNetwork	Uses the normalized Gaussian Radial Basis Function network. Its main function is the k-mean clustering method, while training is performed by logistic or linear regression. Standardizes all numerical variables to 0 mean and unit variance (Hall et al., 2009).
	SimpleLogistic	It is a classifier that generates linear logistic regression models. LogitBoost, which uses simple logistic regression functions, is used to fit logistic models (see Landwehr et al., 2006; Sumner et al., 2005).
lazy	KStar	It differs from other instance-based algorithms in terms of being entropy-based. This method enables the classification of the tested objects according to their proximity to similar objects in the learning data, based on some proximity functions (Hall et al., 2009). Detailed information on the technical structure and usefulness of the method was provided by Cleary and Trigg (1995).
	RseslibKnn	This *k*-closest neighborhood classifier with many distance criteria finds fast neighborhoods in large samples and can be applied to numerical and categorical data (see Wojna & Latkowski, 2018; Wojna et al., 2019).
trees	J48Consolidated	With or without pruning, C4.5 creates a consolidated decision tree. Consolidated Tree Construction (CTC) creates a single decision tree based on subsets (see Pérez et al., 2007). A new method has been added to this algorithm to determine the number of clusters to be used in the consolidation process (see Ibarguren et al., 2015).
	RandomForest	It is a classifier based on the generation of random decision trees (see Breiman, 2001).

BayesNet and NaiveBayes given in Table 1 are algorithms based on Bayes theorem. Bayesian methods, which make inferences based on probabilistic estimates, have been an important alternative to usual methods in machine learning such as decision trees and artificial neural networks (John & Langley, 1995). Naive Bayes algorithm, which is frequently used in machine learning field as well as decision trees and neural networks, can perform in estimation and

predictions as well. The fact that the method has conditional independence assumption caused it to be described as 'naive' (Han et al., 2011). Similarly, BayesNet makes a graphical classification process, which makes estimations according to network structures obtained based on conditional probability distributions (Alpaydin, 2010; Bouckaert, 2008). Generally, a Bayes classifier assigns an instance with the highest value to the class after selecting that class with the highest probability in the model having the least error according to the Bayes rule (Alpaydin, 2010; John & Langley, 1995).

In logistic regression models, the probability of a data set belonging to the last class is estimated by subtracting the sum of the probabilities of belonging to each class from the value 1 (Landwehr et al., 2006; Sumner et al., 2005). Radial basis functions (RBF), one of the artificial neural network models, work similarly to perceptron models, but use the gauss function as the threshold function (Akpınar, 2014). RBF network (Hall et al., 2009), whose basic function is obtained with the k-mean clustering method and training with logistics or linear regression (Hall et al., 2009), is generally used in classification problems, modeling and system control fields, and time series analysis. The nearest neighborhood methods that belong to the family of instance-based classification algorithms perform analysis based on distance measures and have many types (Aha et al., 1991). In this mathematics-based method, the instances in the test data are classified according to their positions in the training data in a multidimensional space (Larose & Larose, 2014). KStar algorithm differs from other object-based algorithms in terms of using entropy-based functions (Cleary & Trigg, 1995). RseslibKnn algorithm, which can find fast neighborhoods in large samples, is a method that includes different distance metrics for different types of attributes (see Wojna & Latkowski, 2018; Wojna et al., 2019).

While the classification algorithms based on decision trees are very diverse, J48 and random forest methods are among the most frequently used machine learning methods (Pu et al., 2020). With the addition of new options to the J48 algorithm, the J48 Consolidated algorithm, which creates a single decision tree based on subsets, has been developed as a robust method for classification problems with its high performance (Ibarguren et al., 2015; Pérez et al., 2007). This algorithm generates a consolidated C4.5 decision tree (Quinlan, 1993) with or without pruning (Hall et al., 2009). The random forest classifier (Breiman, 2001) has become one of the most popular machine learning techniques used in such fields as mining, archeology, engineering and wine (Li et al., 2019) in recent years, due to its highly reliable and interpretable results in complex data and its performance comparable with other frequently used machine learning techniques (Zhang & Yang, 2020). In addition, random forests have many advantages such as high classification performance in many data types, handling dimensionality, being capable of variable importance analysis, highy adaptability and time efficiency (Li et al., 2019). The random forest method is a mixture of tree estimators in which each tree has the same distribution for every other tree in the forest and each tree is autonomously dependent on the values of the random vector sets (Breiman, 2001).

Although there are many studies on the effectiveness a wide variety of machine learning techniques on different data types in different fields such as education (e.g. Baker, 2010; Berens et al., 2019; Bulut & Yavuz, 2019; Güre et al., 2020; Hamalainen & Vinni, 2006; Koyuncu , & Gelbal, 2020; Romero & Ventura, 2013), health sciences (e.g. Beleites et al., 2013; Chu et al., 2012; Figueroa et al., 2012; Shao et al., 2013), engineering sciences (e.g. Brain & Webb, 1999; Hegde & Rokseth, 2020; Reich & Barai, 1999), economics (e.g. Azqueta-Gavaldón, 2017; Mele & Magazzino , 2020; Mullainathan & Spiess, 2017), politics (Grimmer, 2015; Guess et al., 2019), environmental sciences (e.g. Heydari & Mountrakis, 2018; Zhang, & Yang, 2020; Mele & Magazzino, 2020). However, in the relevant literature, studies on how machine learning methods can bring solutions to problems in the field of scale development are limited (e.g. Auerswald & Moshagen, 2019: Baldi & Hornik, 1989; Chattopadhyay et al., 2011; Goretzko &

Bühner, 2020; Tezbaşaran, & Gelbal, 2018). Therefore, there is a need to examine how the use of machine learning methods in scale development studies will bring solutions to existing problems. This study, in line with this need, has examined whether machine learning methods and EGA can be a solution to the problems encountered in placing the items in the dimensions.

When studies on exploratory factor analysis using machine learning methods are examined, it can be seen that such studies generally focus on factor retention (e.g. Goretzko & Bühner, 2020; Iantovics et al., 2019). As a result of these studies, it has been reported that machine learning methods can generally be used with traditional methods. In the study conducted by Goretzko and Bühner (2020), it was stated that the ranger and xgboost algorithm were the most accurate methods for 3204 conditions in determining the number of factors. However, these studies do not seek answers to the research problem of correctly classifying the scale items into the factors that the current study deals with. Therefore, it is important to examine whether machine learning methods, which provide solutions to classification problems, can be used in scale development and adaptation studies. In addition, researchers can evaluate the accuracy of their decisions by using these methods in cases where their correct classification percentages are high. For example, such methods let the researchers place the items on a two-dimensional scale as a result of their EFA. In this case, according to the characteristics of the data set, it can be checked whether the items are correctly classified to the dimensions by machine learning methods or EGA. Hence, an evidence related to decision validity can be obtained. For this reason, this study is important in terms of its contribution to the relevant literature and the convenience it will provide to researchers. This study is also important in terms of allowing practitioners to test the correct classification of the items into their dimensions by using machine learning methods. Therefore, this study seeks answers to the following research problems:

Under different simulation conditions for EGA and machine learning methods:

1) What are the correct classification percentage values?
2) How are Kappa concordance values for confusion matrices?

2. METHOD

This study is a Monte Carlo simulation since it was carried out to compare the classification performances of machine learning methods in different factor structures. In Monte Carlo simulation studies, sample data are generated in accordance with the desired distribution characteristics (Bandalos & Leite, 2013). In this study, the data sets were generated as 5-point likert type scale. The skewness of data was adjusted as left-skewed, normal, and right-skewed.

2.1. Simulation Conditions

In the present study, in order to examine the performance of different methods, a set of simulation conditions were determined. These conditions included average factor loadings (.40 and .70), sample siz (100, 200, 500 and 1000), number of items per dimension (5 and 10), and number of dimensions (2 and 3). In addition, distribution of data (left-skewed, normal, and right-skewed) conditions were investigated. In the study, a total of 2 x 4 x 2 x 2 x 3 = 96 simulation conditions were studied and 100 replications were made.

The conditions for the average factor loadings were manipulated to be .40 and .70. In addition to the researchers who state that the factor loadings of the items in the scales should be at least .30 (Costello & Osborne, 2005), there are also researchers who state that it should be at least .32 (Tabachnik & Fidell, 2012). Besides, Howard (2016) states that this value should be at least .40. For this reason, in this study, data sets were produced with an average factor loading of .40 by taking the average factor loadings. On the other hand, .70 was added as another factor loading condition in order to examine how the increase in the average factor loadings affects the performance of the methods.

The conditions for sample size were manipulated to be 100, 200, 500, and 1000. The sample size is frequently selected as 200, 500, and 1000 in factor analysis studies and is defined as small, medium, and large (Beauducel & Herzberg, 2006; Li, 2016b; West et al., 1995). In addition, Gorsuch (1974) suggested that the sample size should be at least 200. On the other hand, since this study investigated the classification performance of machine learning methods, samples with 100 instances were also added to the sample size conditions in order to examine the classification performance in smaller samples. For example, in educational data, it is possible to have data for 50 or even fewer students. Therefore, in this study, small sample sizes were preferred in order to examine the performances of methods at the same time.

The conditions for the number of items per dimension were manipulated to be 5 and 10. In classification methods, imbalanced or balanced distribution of class variable can cause different results (Sun et al., 2006). For this reason, only a balanced distribution (the same number of items per dimension) was examined in this study. Although it is suggested that a dimension should be defined with at least 3 items, it is stated that more items would increase the reliability of the dimension (Brown, 2015). For this reason, 5-item conditions for one dimension were added to the study. In addition, 10-item condition was also added to the study to examine the effect of increasing the number of items on the performance of the methods.

The conditions for the number of dimensions were manipulated to be 2 and 3. 2-dimension condition was investigated because there had to be a dependent variable with at least two categories to make the classification. In addition, 3-dimensional condition was also included to examine how the increase in the number of dimensions would affect the performance of the methods. Since the interfactor correlations in the real data sets were mostly between .20 and .40 (Li, 2016a), it was fixed to .30, the value in the middle of this interval in the present study.

The conditions for the distribution of the data were manipulated to be left-skewed, normal, and right-skewed. This condition was added to the study in order to examine how the change in the distribution of data would affect the performance of methods. Since it was stated that the skewness coefficient can be considered normal for the interval [-2, 2] (Chou & Bentler, 1995; Curran et al., 1996; Finney & DiStefano, 2013), data was categorized in such a way that the coefficient of skewness was 2.5 for a right-skewed distribution and -2.5 for a left-skewed distribution. The data was first generated to show a continuous normal distribution and then it was categorized according to threshold values.

2.2. Data analysis

The lavaan (Rosseel, 2012) package included in the R software (R Core Team, 2020) was used to generate the data. EGAnet (Golino & Christensen, 2020) package was used for exploratory graphic analysis. There are two different methods when performing EGA. These are the graphical least absolute shrinkage and selection operator (GLASSO), and triangulated maximally filtered graph approach (TMFG). In this study, the TMFG method, which was found to give more accurate results (Golino et al., 2020) in many conditions, was used. While the codes written by the researchers were used to calculate the percent correct values from the EGA results, the Kappa values were obtained with the caret (Kuhn, 2020) package.

Analyzes for machine learning methods were performed in the Experimenter module of WEKA (Hall et al., 2009, Bouckaert et al., 2020) with 10-fold cross-validation (Lachenbruch & Mickey, 1968). Cross-validation which was performed by dividing data into a number of folds (usually 10 folds) is a method used when the data is not large enough to divide it into training and test data (Witten et al., 2017). Since the scale items were classified instead of subjects in this study, the data set was transposed, and hence the number of instances was limited to the number of items. Therefore, 10-cross-validation method was used in this study. Boostrapping (Efron, 1983) method is used when the data sets are medium (approximately 1000) or larger

(more than 1000); otherwise, holdout methods are used for small (less than 1000 subjects) sample sizes in machine learning.

2.3. Model Evaluation Criteria

Percent correct values were used to compare the performance of EGA and machine learning methods in the study. The percentage of correctly classified items into the dimensions for 100 replications was calculated. For this purpose, first, it was checked whether the number of recommended factors was estimated correctly. If it was correct, then it was examined whether the items were correctly classified into the dimensions. The percent correct values were obtained by dividing the number of replications in which the items were in the correct factors by the number of replications (100). Since it was stated that percent correct values should be above 80% (Hartmann, 1977), it was used as cut off value for percent correct.

There are many criteria to evaluate the classification performance of machine learning methods. The most used criteria are accuracy (percent correct), error rate, precision, recall, sensitivity, specificity, receiver operating characteristic (ROC) curve, F criterion, and Kappa statistics. These values are calculated by creating a confusion matrix via the classification results. In classification, there are frequencies belonging to the instances classified into the cells of confusion (error) matrix. In an error matrix consisting of 2x2 classes a and b, there are frequencies belonging to instances classified correctly into classes a and b (True positive [TP] and True Negative [TN]). Also, there are instances classified into class b while it should be in class a (False positive [FP]), and in class b while it should be in class a (False Negative [FN]). Based on these frequencies, the accuracy rate is obtained by dividing the number of correctly classified instances to the total number of instances. Error rate is obtained by subtracting the accuracy rate from 1. Precision is calculated by dividing the value of TP by the sum of TP and FP. Sensitivity and Recall (True positive rate) measures are calculated by dividing the TP value to the sum of TP and FN. Specificity (True negative rate) value is calculated by dividing the TN value by the sum of TN and FP.

F value (Rijsbergen, 1979), another measure used in the evaluation of models, is an equally weighted function of precision and recall values, while Fβ value is not a function of equal weights (Han et al., 2011). The Kappa statistic (Cohen, 1960; Fleiss, 1971), which evaluates the concordance in the confusion matrix, is evaluated as low if it is between 0-.20, acceptable if it is between .21-.40, medium if it is between .41-.60, very good if it is between .61-.80, and perfect if it is between .81-1.00 (Landis & Koch, 1977). The ROC curve (Egan, 1975) is a measure which is frequently used in binary classification. It is a graphical representation of TP value on the vertical axis and FP values on the horizontal axis regardless of class memberships or error cost (Witten et al., 2017). If the area value under this curve is around .50, it indicates that the model performance is low, and when it is approximately 1, the performance is high (Han et al., 2011).

There are many model evaluation criteria, but it is important to determine which one is suitable for the data set. In this sense, whether the dependent variable is binary or multinominal is an important case in choosing the evaluation criterion to be used. Since the ROC curve, precision, recall, and specificity measures are used when dependent variable is binary, these evaluation criteria were not used for the multinominal form of dependent variables in this study. Similarly, the F criterion is a measure that can be used if the number of observations in the class variable is unbalanced (Branco et al., 2015). For these reasons, percent correct values and Kappa concordance statistics were used as model evaluation criteria in the present study.

3. RESULT / FINDINGS

Findings of the study are presented in this section according to the order of the research problems given in the introduction section.

3.1. Comparison of Percent Correct Values

Percent correct (PC) values obtained from EGA and machine learning methods are presented in Figure 1. In addition, PC values are given in Appendix A for the ones who want to examine these values in detail. The findings obtained in this section were examined for percent correct values of each method.

The increase in the average factor loading and the sample size increased the PC value in general. The factor loading was primarily effective on the classification performance of EGA. When the factor loading was .70, EGA had sufficient PC performance (>80%) even if sample size was small for normally distributed data. As the sample size increased, EGA had sufficient PC performance under the conditions where the average factor loading was .40. EGA had sufficient PC performance in 52.08% of all conditions.

BayesNet had sufficient PC performance in approximately 98% of the conditions where the average factor loading was .70. When the conditions in which average factor loading was .40 were examined, the increase in the sample size increased the PC performance of the methods. BayesNet had sufficient PC performance in 52.08% of all conditions.

J48Consolidated had sufficient PC performance in all conditions where the average factor loading was .70, the number of items was 10, the number of dimensions was 2, and sample sizes were 200 and 500 regardless of the distribution of the data. However, the PC performance was below 80% under the conditions where sample sizes were 500 and 1000 and average factor loading was .40 for the normally distributed data. TJ48Consolidated had sufficient PC performance in 10.41% of all conditions.

KStar had sufficient PC performance in all conditions where the sample sizes were 100 and 200 and the average factor loading was .70. With the increase of the sample size to 500, it did not have sufficient PC performance for normally distributed data sets. In all conditions where the sample size was 1000, the PC value was below 80%. When average factor loading was .40, the number of dimensions was 2, the distribution of data was normal, and KStar had sufficient PC performance. Generally, KStar had sufficient PC performance in 40.63% of all conditions.

NaiveBayes had sufficient PC performance in all conditions where the average factor loading was .70. The conditions where average factor loading was .40 and the number of items was 10 positively affected the PC performance of the method. The sample size being 500 and above made the PC performance of the method independent from the distribution. However, the method had a better performance in conditions where data were normally distributed and the sample sizes were 100 and 200. NaiveBayes had sufficient PC performance in 67.71% of all conditions.

RandomForest had sufficient PC performance in all conditions where the average factor loading was .70. However, it had not sufficient PC performance under any conditions where the average factor loading was .40 and the number of items was 5. Increasing the number of items and sample size increased the PC performance of the method. RandomForest had sufficient PC performance in 67.71% of conditions.

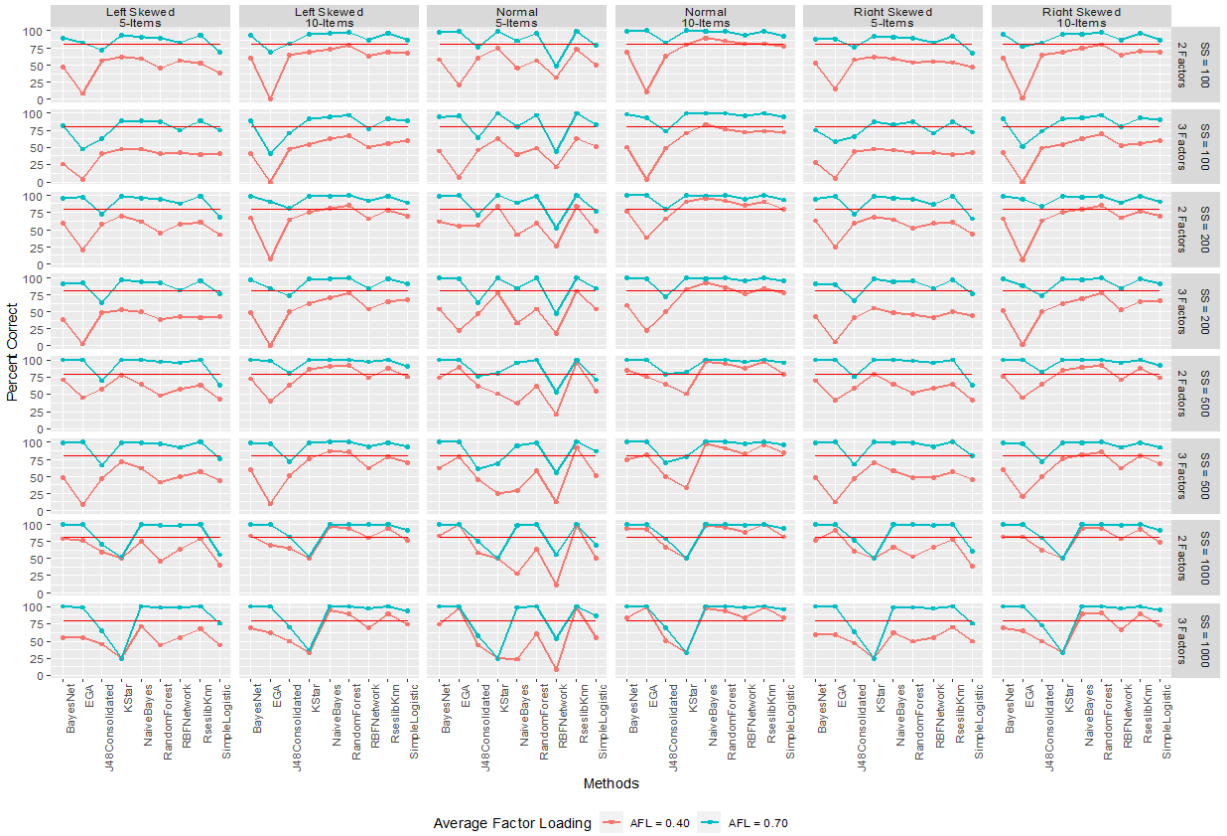
RBFNetwork had generally sufficient performance under conditions where the average factor loading was .70. However, it had not sufficient PC performance under any conditions where data were normally distributed and the number of items was 5. As the sample and the number of items increased, the PC values of the method also increased. RBFNetwork had sufficient PC performance in 45.83% of all conditions.

RseslibKnn had sufficient performance in all conditions where the average factor loading was .70. The PC value of the method was bigger than 80% in the conditions where sample sizes were 500 and 1000 and the number of items was 10. The number of dimensions did not have any effect on the PC performance of the method. It was especially noteworthy that it had 100%

PC value under all conditions where sample size was 1000 and average factor loading was 0.70. RseslibKnn had sufficient PC performance in 69.69% of all conditions.

SimpleLogistic had not sufficient performance under any conditions where the number of items was 5 in skewed data. The method had more than 80% PC values in 3 dimensional structures compared to 2 dimensional ones. However, the method had not sufficient PC performance under any conditions where the average factor loading was .40, and sample sizes were 100 and 200. Its PC values were higher than 80% in only one condition where the sample size was 500. SimpleLogistic had sufficient PC performance in 32.29% of all conditions.

Figure 1. *Comparison of percent correct values of the methods.*



3.2. Comparison of Kappa Concordance Values

Kappa values obtained from EGA and machine learning methods are presented in Figure 2. In addition, Kappa values are given in Appendix B for researchers who would like to examine the details.

EGA's Kappa values varied between .69 and 1.00 for all simulation conditions. Accordingly, EGA had a very good matrix concordance in all conditions. However, it should be kept in mind that kappa values were calculated only with replications where the number of dimensions was estimated correctly. In other words, Kappa values should be evaluated together with percent correct values. According to these results, it can be said that EGA could classify items at a fairly good level in cases where the number of dimensions was estimated correctly.

BayesNet had good Kappa values above .60 in all conditions where the average factor loading was .70. When the conditions with an average factor loading of 0.40 were examined, it had more acceptable Kappa values that were obtained in large samples compared to small ones, in 2 dimensions compared to 3 dimensions, and in normal distribution compared to skewed distributions. In about 60% of the conditions where the average factor loading was .40,

moderate concordance was observed. BayesNet had good Kappa values in 66.66% of all conditions.

J48Consolidated had good Kappa values in 50% of the conditions where the average factor loading was .70, and the other 50% of the conditions had medium or acceptable Kappa values. In 25% of the conditions where the average factor loading was .40, medium and above concordance was observed, while acceptable concordance was obtained in other conditions. In this method, in general, higher Kappa values were obtained in 2 dimensions compared to 3 dimensions, for 5 items compared to 10 items per dimension. Changing the skewness and sample size did not cause a northwothy change in the Kappa values. J48Consolidated had good Kappa values in 30.21% of all conditions.

KStar had perfect concordance in all conditions where sample sizes were 100, 200 and average factor loading was .70. Kappa values were slightly lower under conditions where sample size was 500 and there were normally distributed data than the skewed ones. However, Kstar had insufficient concordance under conditions where sample size was 1000 and the number of dimensions was 3. When the conditions with an average factor loading of .40 were examined, it tended to show a higher and better level of concordance in conditions where the number of dimensions was low, the number of items was 5, and the distribution of variables was normal. KStar had good Kappa values in 56.25% of all conditions.

NaiveBayes had good concordance in all conditions where the average factor loading was .70. Under conditions where the average factor loading was .40, it had much better Kappa values obtained in large samples compared to small ones and 3-dimensional structure compared to 2-dimensional ones. However, a better concordance was observed under conditions where data was skewed, and the number of items was 5 when compared to 10 items. The opposite of this case was true when there were 10 items per dimension. NaiveBayes had good Kappa values in 72.92% of all conditions.

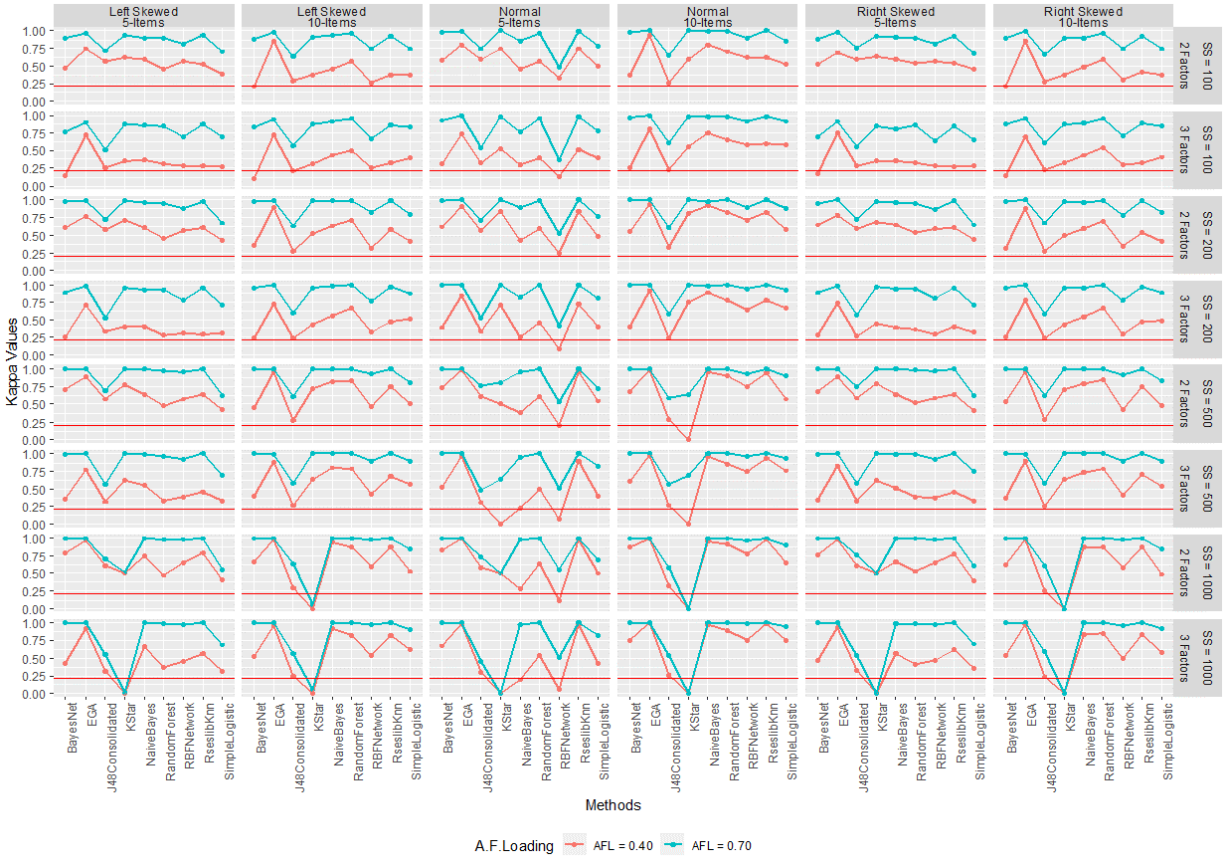
RandomForest had perfect Kappa values in all conditions where the average factor loading was .70. Under conditions where the average factor loading was .40, the increase in the number of items per dimension, sample size and the decrease in skewness increased the performance of the method. Generally, acceptable Kappa values were obtained. NaiveBayes had good Kappa values in 75% of all conditions.

RBFNetwork generally had perfect concordance in all conditions where the average factor loading was .70, except for conditions that the data were normally distributed, and the number of items was 5. In the conditions where the average factor loading was .40, the concordance was generally above the acceptable level, except for the conditions with a normal distribution and 5 items per factor. Overall, as the sample size got larger, the concordance increased. While the increase in the number of items per dimension decreased the performance in skewed data, it had the opposite in normal distributions. RBFNetwork had good Kappa values in 50% of all conditions.

RseslibKnn had perfect concordance in all conditions where the average factor loading was .70. There was an acceptable concordance when the average factor loading was .40 and the sample size was small. Under conditions where the sample size was over 200, a fairly good concordance was observed. In general, getting a larger sample size increased the concordance, especially in skewed data. The change in the number of dimensions did not have a northwhorty effect on the overall concordance for RseslibKnn. Especially in all conditions where the sample sizes were 500 and 1000 and the average factor loading was .70, it is noteworthy that the agreement was 1. The decrease in the number of items per dimensions and skewness of data increased the concordance. RseslibKnn had good Kappa values in 81.25% of all conditions.

SimpleLogistic had a very good or perfect concordance in all conditions where the average factor loading was .70. In all conditions where the average factor loading was .40, SimpleLogistic had an acceptable or above concordance level. In general, the increase in the number of items and the sample size increased concordance. Increasing the number of dimensions and skewness of data decreased the concordance. SimpleLogistic had good Kappa values in 54.17% of all conditions.

Figure 2. *Comparison of Kappa concordance values of the methods.*



4. DISCUSSION and CONCLUSION

In this study, the usability of exploratory graph analysis (EGA) and machine learning methods in deciding which item should be included in which dimension in the exploratory factor analysis was examined. The results obtained for different conditions were successively discussed for the performance of methods with regard to different sample sizes, average factor loadings, the number of items per dimensions, the number of dimensions, and the distribution of data.

When the findings obtained for different conditions were evaluated together, it was seen that machine learning methods gave comparable results to EGA. Machine learning methods showed high performance, especially in small and medium sample sizes. For example, in all conditions where the average factor loading was .70, BayesNet, Naive Bayes, RandomForest, and RseslibKnn methods had bigger values than 80% PC values similar to the values of EGA. BayesNet, Simple Logistic and RBFNetwork methods had also an acceptable or high PC performance under many conditions such as different sample sizes, factor loadings, and the number of items. These methods had better classification performance than that of EGA when factor loading was .40. Kappa concordance values also support these results. In general, higher percent correct and Kappa values were obtained in conditions where the average factor loading was .70 compared to the average factor loading .40.

Under conditions where the average factor loadings and the number of items per dimension were low, percent correct values below 80% were obtained regardless of the number of factors, skewness of data, and sample size. However, in conditions where the average factor loading was low and the number of items per dimension was high, sample size was small and data were normally distributed and PC values of Naive Bayes, RandomForest, RBFNetwork, RseslibKnn, and SimpleLogistic methods were close to 80% or above. These methods showed the same performance even if there was skewness in large sample sizes. Kappa values also greatly supported such a result. Especially when the number of items per dimension was more than 5, it was seen that these methods performed well even if the average factor loading was low. The fact that the methods were Bayesian, decision trees, artificial neural networks and instance-based showed that classification decisions can be made with different statistical and mathematical based methods. In addition, it was observed that the performance of some methods such as RBFNetwork and Kstar decreased in the conditions having 5 items and large sample sizes. This interesting result was considered to be obtained due to the mathematical structure of those methods.

Machine learning methods generally do not require any assumptions (except the conditional independence assumption for the Naïve Bayes). The results of this study showed that the number of categories, skewness of data, and sample size had an effect on the classification performance of these methods. Although they were not based on factor analysis, the results of other studies revealed that sample size (Beleites et al., 2013; Brain & Webb, 1999; Chu et al., 2012; Figueroa et al., 2012; Heydari & Mountrakis, 2018; Hamalainen & Vinni, 2006; Shao et al., 2013), feature selection (Chu et al., 2012), and the number of nominal classes (Minaei-Bidgoli et al., 2003; Nghe et al., 2007) had effects on the performance of machine learning methods. On the other hand, studies on factor analysis using machine learning generally focused on factor retention (e.g. Goretzko & Bühner, 2020; Iantovics et al., 2019). Therefore, the results of the present study provide researchers with a reference point in using and selecting the most suitable machine learning method for their data structure to decide on which items will be included in which factors. For example, assume that when a researcher cannot decide on which item belongs to which dimension after EFA analysis because an item can load more than one dimension at the same time (cross loading), the researcher in such a situation can try different methods given in the present study and place the item into the appropriate dimension by taking into account the conditions similar to her/his own study. In addition, in cases where it is necessary to perform item parceling, items can be grouped by using methods that give accurate results in the current study.

Due to many simulation conditions handled, the discussions were formed from generalized results for different conditions in the present study. Researchers who perform exploratory factor analysis can choose machine learning methods and classify scale items according to the characteristics of their data sets (sample size, average factor loading, and skewness of the data). In this case, they can compare the percent correct and Kappa values obtained from their study with the results of this study. For example, let us consider a method where PC value was obtained as 100% in current study. If the researcher obtains a very low value when he/she uses this method in his/her own data set, he/she may consider re-classifying the items. Thus, it will be possible to examine whether the items are in the right dimension or not. In addition, assume that researchers have been given a basis for decision-making. However, it should also be taken into account that this study does not cover all of the real situations that may actually occur. The level of similarity of the characteristics of the real data set with the conditions examined in the current study should also be taken into consideration.

In this study, eight machine learning methods based on different statistical and mathematical basis included in the WEKA software were examined. In future studies, the performance of

other methods such as Bayesian, artificial neural networks, instance based, rule based, decision tree, and support vector machine can also be examined. In addition, the number of conditions used in this research can be increased or the performance of EGA and machine learning methods used in the current study can be compared for different conditions such as inter-factor correlations. Since this study was carried out with simulated data sets, the performance of the EGA and machine learning methods can be examined over real data sets in similar conditions.

Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the authors.

Authorship Contribution Statement

Ilhan Koyuncu: Investigation, Resources, Visualization, Software, Formal Analysis, Writing original draft, Methodology, Supervision, and Validation. **Abdullah Faruk Kilic**: Investigation, Resources, Software, Formal Analysis, Writing original draft, Methodology, Supervision, and Validation.

ORCID

Ilhan Koyuncu https://orcid.org/0000-0002-0009-5279
Abdullah Faruk Kilic https://orcid.org/0000-0003-3129-1763

5. REFERENCES

Aha, D. W., Kibler, D., & Albert, M. K. (1991). Instance-based learning algorithms. *Machine Learning 6*, 37-66.

Akpınar, H. (2014). *Veri madenciliği veri analizi [Data mining data analysis]*. Papatya Yayınları.

Alpaydin, E. (2010). *Introduction to machine learning: Adaptive computation and machine learning series*. MIT Press.

Auerswald, M., & Moshagen, M. (2019). How to determine the number of factors to retain in exploratory factor analysis: A comparison of extraction methods under realistic conditions. *Psychological Methods, 24*(4), 468–491. https://doi:10.1037/met0000200

Azqueta-Gavaldón, A. (2017). Developing news-based economic policy uncertainty index with unsupervised machine learning. *Economics Letters*, *158*, 47-50.

Baker, R. S. J. (2010). Machine learning for education. *International Encyclopedia of Education, 7*(3), 112-118.

Baldi, P., & Hornik, K. (1989). Neural networks and principal component analysis: Learning from examples without local minima. *Neural Networks, 2*, 53-58.

Bandalos, D. L., & Leite, W. (2013). Use of Monte Carlo studies in structural equation modeling research. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed.). Information Age.

Barker, K., Trafalis, T., & Rhoads, T. R. (2004). Learning from student data. *In Proceedings of the 2004 Systems and Information Engineering Design Symposium* (pp. 79-86). IEEE.

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(2), 186-203. https://doi.org/10.1207/s15328007sem1302_2

Beleites, C., Neugebauer, U., Bocklitz, T., Krafft, C., & Popp, J. (2013). Sample size planning for classification models. *Analytica Chimica Acta, 760*, 25-33.

Belvederi Murri, M., Caruso, R., Ounalli, H., Zerbinati, L., Berretti, E., Costa, S., … Grassi, L. (2020). The relationship between demoralization and depressive symptoms among patients from the general hospital: network and exploratory graph analysis: Demoralization and depression symptom network. *Journal of Affective Disorders*, *276*(June), 137–146. https://doi.org/10.1016/j.jad.2020.06.074

Berens, J., Schneider, K., Gortz, S., Oster, S., & Burghoff, J. (2019). Early detection of students at risk - predicting student dropouts using administrative student data from German universities and machine learning methods. *Journal of Educational Machine learning, 11*(3), 1-41. https://doi.org/10.5281/zenodo.3594771

Bouckaert, R. R. (2008). Bayesian network classifiers in Weka for Version 3-5-7. *Artificial Intelligence Tools*, *11*(3), 369-387.

Bouckaert, R. R., Frank, E., Hall, M., Kirkby, R., Reutemann, P., Seewald, A., & Scuse, D. (2020). *WEKA manual for version 3-9-5*. University of Waikato.

Brain, D., & Webb, G. (1999). On the effect of data set size on bias and variance in classification learning. In *Proceedings of the Fourth Australian Knowledge Acquisition Workshop*, University of New South Wales (pp. 117-128), December 5-6, Sydney, Australia.

Branco, P., Torgo, L., & Ribeiro, R. (2015). A survey of predictive modelling under imbalanced distributions. *arXiv preprint arXiv:1505.01658*.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). The Guilford.

Bulut, O., & Yavuz, H. C. (2019). Educational machine learning: A tutorial for the" Rattle" package in R. *International Journal of Assessment Tools in Education, 6*(5), 20-36.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*(2), 245–276. https://doi.org/10.1207/s15327906mbr0102_10

Chattopadhyay, M., Dan, P. K., & Mazumdar, S. (2011). Principal component analysis and self-organizing map for visual clustering of machine-part cell formation in cellular manufacturing system. In *Systems Research Forum* (Vol. 5, No. 01, pp. 25-51). World Scientific Publishing Company.

Chou, C. P., & Bentler, P. M. (1995). Estimates and tests in structural equation modeling. In Rich H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications*. Sage.

Chu, C., Hsu, A. L., Chou, K. H., Bandettini, P., Lin, C., & Alzheimer's Disease Neuroimaging Initiative (2012). Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. *Neuroimage*, *60*(1), 59-70.

Cleary, J. G., & Trigg, L. E. (1995). K*: An instance-based learner using an entropic distance measure. In *Machine Learning Proceedings 1995* (pp. 108-114). Morgan Kaufmann.

Cohen, J. (1960). A coefficient of agreement for nominal scales, *Educational and Psychological Measurement*, *20*(1), 37-46.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, *10*(7), 27–29. https://doi.org/10.1.1.110.9154

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*(1), 16-29. https://doi.org/10.1037/1082-989X.1.1.16

Efron, B. (1983). Estimating the error rate of a prediction rule: Improvements on crossvalidation. *J. Amer. Stat. Ass.*, *78*, 316–331.

Egan, J. P. (1975). *Signal detection theory and ROC analysis*. Academic Press.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272–299. https://doi.org/10.1037/1082-989X.4.3.272

Figueroa, R. L., Zeng-Treitler, Q., Kandula, S., & Ngo, L. H. (2012). Predicting sample size required for classification performance. *BMC Medical Informatics and Decision Making*, *12*(1), 8.

Finney, S. J., & DiStefano, C. (2013). Nonnormal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed., pp. 439–492). Charlotte, NC: IAP.

Fischer, R., & Alfons Karl, J. (2020). The network architecture of individual differences: Personality, reward-sensitivity, and values. *Personality and Individual Differences*, *160*(February), 109922. https://doi.org/10.1016/j.paid.2020.109922

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, *76*(5), 378-382.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, *7*(3), 286-299. https://doi.org/10.1037/1040-3590.7.3.286

Golino, H. F., & Christensen, A. P. (2020). *EGAnet: Exploratory Graph Analysis -- A framework for estimating the number of dimensions in multivariate data using network psychometrics*. Retrieved from https://CRAN.R-project.org/package=EGAnet

Golino, H. F., & Epskamp, S. (2017). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PLOS ONE*, *12*(6), 1-26. https://doi.org/10.1371/journal.pone.0174035

Golino, H. F., Moulder, R., Shi, D., Christensen, A. P., Garrido, L. E., Nieto, M. D., … Boker, S. M. (2020). Entropy fit indices: New fit measures for assessing the structure and dimensionality of multiple latent variables. *Multivariate Behavioral Research*, 1–29. https://doi.org/10.1080/00273171.2020.1779642

Golino, H. F., Shi, D., Christensen, A. P., Garrido, L. E., Nieto, M. D., Sadana, R., … Martinez-Molina, A. (2020). Investigating the performance of exploratory graph analysis and traditional techniques to identify the number of latent factors: A simulation and tutorial. *Psychological Methods*, *25*(3), 292–320. https://doi.org/10.1037/met0000255

Goretzko, D., & Bühner, M. (2020). One model to rule them all? Using machine learning algorithms to determine the number of factors in exploratory factor analysis. *Psychological Methods*, *25*(6), 776–786. https://doi.org/10.1037/met0000262

Gorsuch, R. L. (1974). *Factor analysis*. W. B. Saunders.

Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin*, *103*(2), 265–275.

Guess, A., Munger, K., Nagler, J., & Tucker, J. (2019). How accurate are survey responses on social media and politics?. *Political Communication*, *36*(2), 241-258.

Güre, Ö. B., Kayri, M., & Erdoğan, F. (2020). Analysis of factors effecting PISA 2015 mathematics literacy via educational machine learning. *Education and Science*, *45*(202), 393-415.

Grimmer, J. (2015). We are all social scientists now: How big data, machine learning, and causal inference work together. *PS, Political Science & Politics*, *48*(1), 80.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Peter, R., & Witten, I. H. (2009). The WEKA machine learning software: An update. *SIGKDD Explorations*, *11*(1), 10-18.

Hamalainen, W., & Vinni, M. (2006). Comparison of machine learning methods for intelligent tutoring systems. In *Proceedings of International Conference on Intelligent Tutoring Systems* (pp. 525-534). Springer Berlin/Heidelberg.

Han, J., J. Pei, & Kamber, M. (2011). *Machine learning: Concepts and techniques*. Elsevier.

Hartmann, D. P. (1977). Considerations in the choice of interobserver reliability estimates. *Journal of Applied Behavior Analysis*, *10*(1), 1311156. https://doi.org/10.1901/jaba.1977.10-103

Hegde, J., & Rokseth, B. (2020). Applications of machine learning methods for engineering risk assessment–A review. *Safety Science*, *122*, 104492.

Heydari, S. S., & Mountrakis, G. (2018). Effect of classifier selection, reference sample size, reference class distribution and scene heterogeneity in per-pixel classification accuracy using 26 Landsat sites. *Remote Sensing of Environment*, *204*, 648-658.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. https://doi.org/10.1007/BF02289447

Howard, M. C. (2016). A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve? *International Journal of Human-Computer Interaction*, *32*(1), 51-62. https://doi.org/10.1080/10447318.2015.1087664

Iantovics, L. B., Rotar, C., & Morar, F. (2019). Survey on establishing the optimal number of factors in exploratory factor analysis applied to machine learning. *Wiley Interdisciplinary Reviews: Machine learning and Knowledge Discovery*, *9*(2), 1-20. https://doi.org/10.1002/widm.1294

Ibarguren, I., Pérez, J. M., Muguerza, J., Gurrutxaga, I., & Arbelaitz, O. (2015). Coverage-based resampling: Building robust consolidated decision trees. *Knowledge-Based Systems*, *79*, 51-67. https://doi.org/10.1016/j.knosys.2014.12.023

John, G. H., & Langley P. (1995). Estimating continuous distributions in Bayesian classifiers. In P. Besnard & S. Hanks (Eds.), *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence* (pp. 338–345). San Francisco, Morgan Kaufmann.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: trends, perspectives, and prospects. *Science*, *349*(6245), 255-260, https://doi.org/10.1126/science.aaa8415

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*(1), 141-151. https://doi.org/10.1177/001316446002000116

Kılıç, A. F., & Koyuncu, İ. (2017). Ölçek uyarlama çalışmalarının yapı geçerliği açısından incelenmesi [Examination of scale adaptation studies in terms of construct validity]. In Ö. Demirel & S. Dinçer (Eds.), *Küreselleşen dünyada eğitim [Education in a globalizing world]* (pp. 1202–1205). Pegem Akademi.

Kjellström, S., & Golino, H. (2019). Mining concepts of health responsibility using text mining and exploratory graph analysis. *Scandinavian Journal of Occupational Therapy*, *26*(6), 395–410. https://doi.org/10.1080/11038128.2018.1455896

Kline, P. (1994). *An easy guide to factor analysis*. Routledge.

Koyuncu, İ., & Gelbal, S. (2020). Comparison of machine learning classification algorithms on educational data under different conditions. *Journal of Measurement and Evaluation in Education and Psychology*, *11*(4), 325-345.

Koyuncu, İ., & Kılıç, A. F. (2019). The use of exploratory and confirmatory factor analyses: A document analysis. *Education and Science*, *44*(198), 361-388. https://doi.org/10.15390/EB.2019.7665

Kuhn, M. (2020). *caret: Classification and Regression Training*. Retrieved from https://cran.r-project.org/package=caret

Lachenbruch, P. A., & Mickey, M. R. (1968). Estimation of error rates in discriminant analysis. *Technometrics*, *10*(1), 1-11.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159-174.

Landwehr, N., Hall, M., & Frank, E. (2006). *Logistic model trees*. Kluwer Academic Publishers.

Larose, D. T., & Larose, C.D. (2014). *Discovering knowledge in data: An introduction to machine learning.* John Wiley and Sons.

Li, C.-H. (2016a). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, *48*(3), 936–949. https://doi.org/10.3758/s13428-015-0619-7

Li, C.-H. (2016b). The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. *Psychological Methods*, *21*(3), 369-387. https://doi.org/10.1037/met0000093

Li, N., Qi, J., Wang, P., Zhang, X., Zhang, T., & Li, H. (2019). Quantitative structure–activity relationship (QSAR) study of carcinogenicity of polycyclic aromatic hydrocarbons (PAHs) in atmospheric particulate matter by random forest (RF). *Analytical Methods*, *11*(13), 1816-1821.

Mele, M., & Magazzino, C. (2020). A machine learning analysis of the relationship among iron and steel industries, air pollution, and economic growth in China. *Journal of Cleaner Production*, *277*, 123293.

Minaei-Bidgoli, B., D.A. Kashy, G. Kortemeyer, & W. Punch (2003). Predicting student performance: An application of machine learning methods with an educational web-based system. In Proceedings of *33rd Frontiers in Education Conference*, (pp. 13-18). Westminster, CO.

Mullainathan, S., & Spiess, J. (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, *31*(2), 87-106.

Nghe, N. T., Janecek, P., & Haddawy, P. (2007). A comparative analysis of techniques for predicting academic performance. In *Frontiers in Education Conference-Global Engineering: Knowledge Without Borders, Opportunities Without Passports*, (pp. T2G-7). IEEE.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd. ed.). McGraw-Hill.

Osborne, J. W. (2015). What is rotating in exploratory factor analysis? *Practical Assessment Research & Evaluation*, *20*(2), 1–7.

Panayiotou, M., Santos, J., Black, L., & Humphrey, N. (2020). Exploring the dimensionality of the social skills improvement system using exploratory graph analysis and bifactor-(S−1) modeling. *Assessment*, 1-15. https://doi.org/10.1177/1073191120971351

Pérez, J. M., Muguerza, J., Arbelaitz, O., Gurrutxaga, I., & Martín, J. I. (2007). Combining multiple class distribution modified subsamples in a single tree. *Pattern Recognition Letters*, *28*(4), 414-422. https://doi.org/10.1016/j.patrec.2006.08.013

Pu, Y., Apel, D. B., & Hall, R. (2020). Using machine learning approach for microseismic events recognition in underground excavations: Comparison of ten frequently-used models. *Engineering Geology*, *268*, 105519.

Quinlan, J. R. (1993). *C4.5: programs for machine learning*. Morgan Kaufmann Publishers, Inc.

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. Retrieved from https://www.r-project.org/

Reich, Y., & Barai, S. V. (1999). Evaluating machine learning models for engineering problems. *Artificial Intelligence in Engineering*, *13*(3), 257-272.

Rijsbergen CV. (1979). *Information retrieval* (2nd ed.). Butterworth.

Romero, C., Espejo, P. G., Zafra, A., Romero, J. R., & Ventura, S. (2013). Web usage mining for predicting final marks of students that use Moodle courses. *Computer Applications in Engineering Education*, *21*(1), 135- 146.

Romero, C., & Ventura, S. (2013). Machine learning in education. *WIREs Machine learning Knowledge Discovery 3*(1), 12- 27.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2), 1–36.

Shao, L., Fan, X., Cheng, N., Wu, L., & Cheng, Y. (2013). Determination of minimum training sample size for microarray-based cancer outcome prediction–an empirical assessment. *PloS one*, *8*(7), e68579. https://doi.org/10.1371/journal.pone.0068579

Sumner, M., Frank, E., & Hall, M. (2005, October). Speeding up logistic model tree induction. In *European conference on principles of machine learning and knowledge discovery* (pp. 675-683). Springer, Berlin, Heidelberg.

Sun, Y., Kamel, M. S., & Wang, Y. (2006). Boosting for learning multiple classes with imbalanced class distribution. In *Sixth international conference on data mining (ICDM'06)* (pp. 592-602). IEEE.

Tabachnik, B. G., & Fidell, L. S. (2012). *Using multivariate statistics* (6th ed.). Pearson.

Tezbaşaran, E., & Gelbal, S. (2018). Temel bileşenler analizi ve yapay sinir ağı modellerinin ölçek geliştirme sürecinde kullanılabilirliğinin incelenmesi [An investigation on usability of principal component analysis and artificial neural network models in the process of scale development]. *Mersin University Journal of the Faculty of Education*, *14*(1), 225-252.

Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods*, *16*(2), 209-220. https://doi.org/10.1037/a0023353

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications* (pp. 56-75). Sage.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2017). *Machine learning: Practical machine learning tools and techniques* (4th Edition). Morgan Kaufmann.

Wojna, A., Latkowski, R. (2018): Rseslib 3: Open source library of rough set and machine learning methods. In: Proceedings of *the International Joint Conference on Rough Set (LNCS*, vol. 11103, pp. 162-176). Springer.

Wojna, A., Latkowski, R., Kowalski, (2019). *RSESLIB: User guide*. Retrieved from http://rseslib.mimuw.edu.pl/rseslib.pdf

Zhang, F., & Yang, X. (2020). Improving land cover classification in an urbanized coastal area by random forests: The role of variable selection. *Remote Sensing of Environment*, *251*, 112105. https://doi.org/10.1016/j.rse.2020.112105

6. APPENDICES

APPENDIX A

Percent correct values of the methods.

Skewness	Items per Factor	Methods	Sample Size 100				200				500				1000			
			Average Factor Loadings 0.40		0.70		0.40		0.70		0.40		0.70		0.40		0.70	
			Number of Factors 2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3
Left Skewed	5 Items	BayesNet	46.70	25.25	89.30	81.30	59.60	39.05	96.50	91.05	71.10	49.00	99.60	98.90	78.60	54.85	100	99.90
		EGA	9.00	3.00	83.00	47.00	21.00	4.00	98.00	93.00	45.00	8.00	100	100	77.00	55.00	100	99.00
		J48Consolidated	56.80	40.50	71.80	62.05	58.10	48.65	72.90	63.25	57.80	46.60	70.20	66.20	60.60	46.05	70.50	64.25
		KStar	61.90	47.50	94.10	89.30	70.80	52.55	98.70	97.00	78.50	71.55	100	99.70	50.00	25.00	51.60	25.45
		NaiveBayes	59.20	47.60	90.40	88.55	61.40	50.10	96.00	94.40	65.00	62.00	99.80	99.10	74.70	70.90	100	100
		RandomForest	45.20	40.15	89.20	87.30	44.80	38.45	94.80	93.00	47.60	42.00	97.70	97.45	46.80	44.40	98.90	98.75
		RBFNetwork	56.80	42.05	82.30	75.45	57.10	42.65	87.90	81.90	57.20	49.90	96.10	93.05	64.60	55.05	98.80	98.55
		RseslibKnn	52.60	39.75	93.60	89.45	60.10	41.80	98.50	95.80	63.50	56.50	100	99.95	78.70	66.70	100	100
		SimpleLogistic	38.90	40.40	69.50	75.35	42.80	43.75	68.20	76.75	42.40	44.65	62.90	75.45	40.50	43.45	56.40	75.50
	10 Items	BayesNet	60.35	40.93	93.75	88.43	68.00	49.27	99.00	97.37	72.65	60.07	100	99.90	83.60	68.30	100	100
		EGA	1.00	0.00	69.00	41.00	7.00	1.00	91.00	85.00	40.00	10.00	99.00	98.00	70.00	62.00	100	100
		J48Consolidated	64.70	47.60	82.00	71.47	64.15	49.47	81.40	73.07	63.55	51.60	81.10	71.70	64.90	49.83	81.80	70.43
		KStar	68.45	53.80	95.70	91.00	75.80	62.10	99.30	97.60	86.00	76.13	99.95	99.80	50.00	33.33	53.00	36.33
		NaiveBayes	72.75	62.70	97.10	94.33	81.55	70.60	99.40	98.70	91.10	86.37	99.95	99.97	97.10	94.73	100	100
		RandomForest	78.60	66.57	97.95	96.40	85.65	78.13	99.70	99.37	91.40	86.17	99.90	99.93	93.65	88.67	100	100
		RBFNetwork	63.05	50.93	87.25	78.33	65.95	54.33	91.65	84.27	73.40	61.77	96.70	93.13	80.15	68.83	99.45	97.67
		RseslibKnn	69.15	55.30	96.05	91.00	78.95	65.17	99.35	98.00	87.40	78.30	100	99.87	93.75	88.70	100	100
		SimpleLogistic	68.30	59.97	87.10	88.60	70.65	67.67	89.85	91.10	75.25	71.00	90.65	93.40	76.60	74.70	91.95	93.83
Normal	5 Items	BayesNet	57.90	44.30	97.50	94.75	61.80	53.70	99.50	99.00	73.60	62.20	100	100	83.40	74.85	100	100
		EGA	21.00	6.00	99.00	95.00	55.00	23.00	100	98.00	89.00	79.00	100	100	100	97.00	100	100
		J48Consolidated	60.20	46.20	75.50	64.55	56.90	47.95	71.90	63.15	61.60	45.15	76.70	60.50	58.40	44.65	74.50	57.60
		KStar	74.50	63.25	99.50	99.60	83.60	77.90	100	100	50.00	25.00	81.40	69.20	50.00	25.00	50.00	25.00
		NaiveBayes	45.60	39.00	85.20	80.75	43.10	33.80	89.30	83.85	37.50	28.75	96.20	95.20	28.90	23.95	98.20	98.40
		RandomForest	56.80	48.70	96.50	97.15	59.40	54.15	99.30	99.00	61.40	57.50	99.80	99.55	64.10	60.30	99.80	100
		RBFNetwork	32.60	22.15	48.20	43.80	25.50	18.20	52.10	47.35	21.00	12.20	53.00	54.75	12.30	8.10	56.30	53.50
		RseslibKnn	73.80	62.45	99.00	99.45	84.50	79.65	100	100	96.10	93.00	100	100	98.70	98.60	100	100
		SimpleLogistic	50.00	52.15	78.10	82.80	47.90	54.30	77.10	84.45	55.40	51.75	71.80	87.25	50.50	55.20	68.90	86.90
	10 Items	BayesNet	68.55	49.83	99.20	98.20	77.50	59.73	99.80	99.93	84.35	73.87	100	100	94.10	84.07	100	100
		EGA	12.00	3.00	100	93.00	38.00	23.00	100	98.00	76.00	81.00	100	100	93.00	99.00	100	100
		J48Consolidated	62.90	48.50	82.45	74.10	66.55	49.47	80.30	72.27	63.95	50.43	79.50	71.03	66.55	50.40	79.10	69.00
		KStar	80.00	70.57	100	99.50	90.30	83.13	100	99.97	50.00	33.33	82.60	79.23	50.00	33.33	50.00	33.33
		NaiveBayes	89.80	83.63	99.45	99.27	95.85	92.57	99.20	98.70	97.85	97.67	100	100	98.10	97.97	100	100
		RandomForest	85.60	77.10	99.70	99.40	91.70	85.33	99.85	100	95.00	90.53	100	99.97	96.15	92.90	100	100
		RBFNetwork	80.85	71.87	94.25	95.00	85.40	75.97	94.40	95.80	88.25	83.00	96.65	98.07	89.10	84.00	98.65	99.33
		RseslibKnn	80.85	73.57	99.85	99.53	91.00	84.57	99.95	99.97	97.65	96.33	100	100	99.75	99.57	100	100
		SimpleLogistic	76.70	72.73	92.30	94.57	79.15	78.13	94.15	95.13	79.15	84.07	95.45	96.30	82.30	83.57	94.75	96.87
Right Skewed	5 Items	BayesNet	52.80	28.40	87.80	75.40	63.80	42.60	95.40	91.60	69.20	48.10	99.70	98.85	75.90	58.90	100	100
		EGA	16.00	5.00	88.00	59.00	25.00	6.00	99.00	90.00	41.00	13.00	100	100	92.00	59.00	100	100
		J48Consolidated	58.60	43.85	76.50	65.40	59.30	42.15	72.60	66.55	58.80	47.05	75.40	67.00	62.00	47.45	75.90	63.50
		KStar	62.50	48.00	92.40	87.70	69.20	55.25	98.30	97.80	78.90	70.55	100	100	50.00	25.00	50.20	25.00
		NaiveBayes	58.80	46.20	91.40	83.80	64.30	49.25	96.40	94.65	63.70	58.25	99.80	99.40	67.20	62.35	99.60	99.60
		RandomForest	53.60	42.15	90.00	87.95	52.80	46.30	94.90	94.95	52.00	47.75	98.70	99.30	52.60	49.55	99.60	99.40
		RBFNetwork	55.90	41.70	82.40	70.55	59.30	42.20	86.10	83.85	58.90	48.75	96.50	93.50	66.20	55.35	98.90	98.00
		RseslibKnn	54.10	39.60	92.90	87.15	60.70	50.50	98.70	97.00	63.90	56.75	100	100	77.60	70.35	100	100
		SimpleLogistic	46.40	41.45	67.60	72.75	43.80	44.50	65.40	76.25	40.60	45.55	63.40	79.50	39.10	48.85	61.70	76.15
	10 Items	BayesNet	60.55	42.57	95.25	91.07	65.70	51.00	98.30	97.63	77.10	58.80	99.90	99.80	81.35	69.30	100	100
		EGA	2.00	0.00	77.00	52.00	5.00	2.00	95.00	88.00	45.00	21.00	100	98.00	82.00	64.00	100	100
		J48Consolidated	64.25	48.43	82.80	73.70	63.65	49.43	83.80	72.90	64.50	50.00	81.70	72.20	63.15	48.87	80.55	72.60
		KStar	68.40	54.70	95.10	91.50	75.20	62.00	99.10	97.63	85.55	75.53	99.95	99.87	50.00	33.33	50.05	33.40
		NaiveBayes	74.20	62.93	95.20	92.73	79.40	69.30	98.05	96.93	89.70	82.00	100	99.77	93.80	89.43	100	99.93
		RandomForest	79.80	70.00	98.40	97.07	84.90	77.47	99.65	99.27	92.50	85.57	100	99.93	94.05	90.10	100	100
		RBFNetwork	65.25	53.63	87.50	80.93	67.45	53.40	89.25	84.77	71.10	61.50	95.90	92.40	79.60	66.43	98.95	97.57
		RseslibKnn	70.70	55.30	96.00	92.97	76.65	65.50	99.40	98.30	87.35	80.33	100	99.87	93.40	89.53	100	100
		SimpleLogistic	69.15	60.40	87.55	89.67	70.30	65.90	90.90	91.73	74.15	69.33	91.40	92.80	74.30	72.20	91.90	94.80

APPENDIX B

Kappa concordance values of the methods

Skewness	Items per Factor	Methods	Sample Size: 100				200				500				1000			
			Average Factor Loadings: 0.40		0.70		0.40		0.70		0.40		0.70		0.40		0.70	
			Number of Factors: 2	3	2	3	2	3	2	3	2	3	2	3	2	3	2	3
Left Skewed	5 Items	BayesNet	0.47	0.15	0.89	0.76	0.60	0.25	0.97	0.88	0.71	0.36	1.00	0.99	0.79	0.43	1.00	1.00
		EGA	0.74	0.72	0.97	0.91	0.77	0.70	1.00	0.99	0.89	0.78	1.00	1.00	0.97	0.92	1.00	1.00
		J48Consolidated	0.57	0.25	0.72	0.52	0.58	0.33	0.73	0.53	0.58	0.32	0.70	0.56	0.61	0.31	0.71	0.55
		KStar	0.62	0.35	0.94	0.87	0.71	0.40	0.99	0.96	0.78	0.62	1.00	1.00	0.50	0.00	0.52	0.01
		NaiveBayes	0.59	0.36	0.90	0.86	0.61	0.40	0.96	0.93	0.65	0.55	1.00	0.99	0.75	0.66	1.00	1.00
		RandomForest	0.45	0.31	0.89	0.85	0.45	0.28	0.95	0.92	0.48	0.33	0.98	0.97	0.47	0.37	0.99	0.99
		RBFNetwork	0.57	0.29	0.82	0.70	0.57	0.31	0.88	0.77	0.57	0.39	0.96	0.92	0.65	0.46	0.99	0.98
		RseslibKnn	0.53	0.28	0.94	0.87	0.60	0.29	0.98	0.95	0.64	0.46	1.00	1.00	0.79	0.57	1.00	1.00
		SimpleLogistic	0.39	0.27	0.70	0.69	0.43	0.31	0.68	0.71	0.42	0.33	0.63	0.69	0.41	0.32	0.56	0.69
	10 Items	BayesNet	0.21	0.11	0.88	0.83	0.36	0.24	0.98	0.96	0.45	0.40	1.00	1.00	0.67	0.52	1.00	1.00
		EGA	0.85	0.72	0.98	0.94	0.89	0.73	0.99	0.99	0.96	0.88	1.00	1.00	0.98	0.97	1.00	1.00
		J48Consolidated	0.29	0.21	0.64	0.57	0.28	0.24	0.63	0.60	0.27	0.27	0.62	0.58	0.30	0.25	0.64	0.56
		KStar	0.37	0.31	0.91	0.87	0.52	0.43	0.99	0.96	0.72	0.64	1.00	1.00	0.00	0.00	0.06	0.05
		NaiveBayes	0.46	0.44	0.94	0.91	0.63	0.56	0.99	0.98	0.82	0.80	1.00	1.00	0.94	0.92	1.00	1.00
		RandomForest	0.57	0.50	0.96	0.95	0.71	0.67	0.99	0.99	0.83	0.79	1.00	1.00	0.87	0.83	1.00	1.00
		RBFNetwork	0.26	0.26	0.74	0.67	0.32	0.32	0.83	0.76	0.47	0.43	0.93	0.90	0.60	0.53	0.99	0.97
		RseslibKnn	0.38	0.33	0.92	0.86	0.58	0.48	0.99	0.97	0.75	0.67	1.00	1.00	0.88	0.83	1.00	1.00
		SimpleLogistic	0.37	0.40	0.74	0.83	0.41	0.52	0.80	0.87	0.50	0.56	0.81	0.90	0.53	0.62	0.84	0.91
Normal	5 Items	BayesNet	0.58	0.31	0.98	0.93	0.62	0.39	0.99	0.99	0.74	0.52	1.00	1.00	0.83	0.68	1.00	1.00
		EGA	0.80	0.74	1.00	0.99	0.91	0.84	1.00	1.00	0.98	0.97	1.00	1.00	1.00	1.00	1.00	1.00
		J48Consolidated	0.60	0.32	0.75	0.54	0.57	0.34	0.72	0.53	0.62	0.30	0.77	0.49	0.58	0.30	0.74	0.46
		KStar	0.74	0.53	1.00	0.99	0.84	0.70	1.00	1.00	0.50	0.00	0.81	0.63	0.50	0.00	0.50	0.00
		NaiveBayes	0.46	0.30	0.85	0.77	0.43	0.25	0.89	0.82	0.38	0.22	0.96	0.95	0.29	0.19	0.98	0.98
		RandomForest	0.57	0.39	0.97	0.96	0.59	0.46	0.99	0.99	0.61	0.50	1.00	1.00	0.64	0.53	1.00	1.00
		RBFNetwork	0.33	0.13	0.48	0.37	0.25	0.09	0.52	0.42	0.21	0.07	0.53	0.51	0.12	0.05	0.56	0.51
		RseslibKnn	0.74	0.52	0.99	0.99	0.84	0.72	1.00	1.00	0.96	0.90	1.00	1.00	0.99	0.98	1.00	1.00
		SimpleLogistic	0.50	0.39	0.78	0.78	0.48	0.41	0.77	0.80	0.55	0.40	0.72	0.83	0.51	0.43	0.69	0.83
	10 Items	BayesNet	0.37	0.25	0.98	0.97	0.55	0.40	1.00	1.00	0.69	0.61	1.00	1.00	0.88	0.76	1.00	1.00
		EGA	0.93	0.80	1.00	0.99	0.94	0.91	1.00	1.00	0.99	0.98	1.00	1.00	1.00	1.00	1.00	1.00
		J48Consolidated	0.26	0.23	0.65	0.61	0.33	0.24	0.61	0.58	0.28	0.26	0.59	0.57	0.33	0.26	0.58	0.53
		KStar	0.60	0.56	1.00	0.99	0.81	0.75	1.00	1.00	0.00	0.00	0.65	0.69	0.00	0.00	0.00	0.00
		NaiveBayes	0.80	0.75	0.99	0.99	0.92	0.89	0.98	0.98	0.96	0.97	1.00	1.00	0.96	0.97	1.00	1.00
		RandomForest	0.71	0.66	0.99	0.99	0.83	0.78	1.00	1.00	0.90	0.86	1.00	1.00	0.92	0.89	1.00	1.00
		RBFNetwork	0.62	0.58	0.89	0.92	0.71	0.64	0.89	0.94	0.76	0.75	0.93	0.97	0.78	0.76	0.97	0.99
		RseslibKnn	0.62	0.60	1.00	0.99	0.82	0.77	1.00	1.00	0.95	0.94	1.00	1.00	0.99	0.99	1.00	1.00
		SimpleLogistic	0.53	0.59	0.85	0.92	0.58	0.67	0.88	0.93	0.58	0.76	0.91	0.94	0.65	0.75	0.90	0.95
Right Skewed	5 Items	BayesNet	0.53	0.17	0.88	0.69	0.64	0.28	0.95	0.89	0.69	0.35	1.00	0.99	0.76	0.47	1.00	1.00
		EGA	0.69	0.76	0.98	0.92	0.79	0.73	1.00	0.99	0.89	0.83	1.00	1.00	0.99	0.93	1.00	1.00
		J48Consolidated	0.59	0.29	0.76	0.56	0.59	0.27	0.73	0.57	0.59	0.34	0.75	0.58	0.62	0.33	0.76	0.53
		KStar	0.63	0.35	0.92	0.84	0.69	0.44	0.98	0.97	0.79	0.62	1.00	1.00	0.50	0.00	0.50	0.00
		NaiveBayes	0.59	0.35	0.91	0.80	0.64	0.39	0.96	0.94	0.64	0.51	1.00	0.99	0.67	0.56	1.00	0.99
		RandomForest	0.54	0.32	0.90	0.86	0.53	0.36	0.95	0.94	0.52	0.39	0.99	0.99	0.53	0.41	1.00	0.99
		RBFNetwork	0.56	0.29	0.82	0.64	0.59	0.30	0.86	0.80	0.59	0.38	0.97	0.92	0.66	0.47	0.99	0.97
		RseslibKnn	0.54	0.27	0.93	0.84	0.61	0.40	0.99	0.96	0.64	0.46	1.00	1.00	0.78	0.62	1.00	1.00
		SimpleLogistic	0.46	0.29	0.68	0.66	0.44	0.32	0.65	0.70	0.41	0.33	0.63	0.74	0.39	0.36	0.62	0.70
	10 Items	BayesNet	0.21	0.14	0.90	0.87	0.31	0.26	0.97	0.96	0.54	0.38	1.00	1.00	0.63	0.54	1.00	1.00
		EGA	0.85	0.70	0.99	0.96	0.88	0.78	1.00	0.99	0.96	0.90	1.00	1.00	0.99	0.97	1.00	1.00
		J48Consolidated	0.28	0.23	0.66	0.61	0.27	0.24	0.68	0.59	0.29	0.25	0.63	0.58	0.26	0.23	0.61	0.59
		KStar	0.37	0.32	0.90	0.87	0.50	0.43	0.98	0.96	0.71	0.63	1.00	1.00	0.00	0.00	0.00	0.00
		NaiveBayes	0.48	0.44	0.90	0.89	0.59	0.54	0.96	0.95	0.79	0.73	1.00	1.00	0.88	0.84	1.00	1.00
		RandomForest	0.60	0.55	0.97	0.96	0.70	0.66	0.99	0.99	0.85	0.78	1.00	1.00	0.88	0.85	1.00	1.00
		RBFNetwork	0.30	0.30	0.75	0.71	0.35	0.30	0.79	0.77	0.42	0.42	0.92	0.89	0.59	0.50	0.98	0.96
		RseslibKnn	0.41	0.33	0.92	0.89	0.53	0.48	0.99	0.97	0.75	0.70	1.00	1.00	0.87	0.84	1.00	1.00
		SimpleLogistic	0.38	0.41	0.75	0.84	0.41	0.49	0.82	0.88	0.48	0.54	0.83	0.89	0.49	0.58	0.84	0.92