AN ACHIEVEMENT ANALYSIS OF SPECIAL EDUCATION SAMPLES ACQUIRED BY MEANS OF BOOTSTRAP METHOD WITH WEKA ALGORITHMS

Tolga Demirhan, Özlem Uçar

Trakya University, Balkans Campus, 22030 Edirne, Turkey

Abstract

In this study, an achievement analysis using weka algorithms was conducted on samples acquired by means of bootstrap method. Bootstrap is one of the resampling methods that allow for the replicability of the results without replicating the same method in a new group, a technique which is used in cases with small sample sizes. In this study, this method is used through the instrument of Resampling Excel-add-in macro developed on Excel 2013. As a result, the large samples acquired were analyzed with Weka and the factors which have an effect on student achievement were investigated.

Key words: *bootstrap, weka, hearing ımpaired, resampling, achievement analysis*

1. INTRODUCTION

The type of education which is carried out through advanced educational programs and methods in environments appropriate to the disabilities and features of the individuals, whose qualities and educational competencies are different than what is expected of them for various reasons, to meet their needs is called 'Special Education'. (MEB,2016)

Hearing impaired individuals receive education in special education schools, too. In our country, the National Disability Database includes 11.000 hearing impaired school-age individuals and approximately 1082 of them receive education in 44 Schools for the Deaf within the body of the Ministry of National Education (Millî Eğitim İstatistikleri, 2014)

As in other research studies within the field of education, studies in the field of special education benefit from questionnaires and tests developed by the researcher or student information systems used by schools in the creation of the data set. However, when the statistical data presented above are considered, it is clear that the size of the data to be collected would be small due to the low number of students, in a study to be conducted upon the selection of a school. Furthermore, according to Sümbüloğlu V, Sümbüloğlu K, the inadequacy of the sample size is a sign of the study's lack of a scientific basis. (Orekkici Temel G., Erdoğan, S., 2011)

In the analysis of research data of a small sample size, it can be stated that the results belong to a special sample without regards to how meticulously the study is conducted, which reduces the value of the study. In that respect, the way to show that the results are replicable is to replicate the study with new sample groups. However, this is not widely preferred in situations where the sample size cannot be increased or it requires an extra amount of time, money and labour. In cases where the sample size is small, resampling methods which are also used in social and health sciences and allows for the replicability of the results without replicating the study with a new sample group are used. (Orekkici Temel G., Erdoğan, S., 2012)

Numerous resampling methods are reported. The fact that there are only a few packages through which one can run these tests prompts researchers to create and use macros on existing software. (Koşkan,Ö., 2007)

Tang et al. work with different resampling methods for bilateral dichotomous data which is frequently used in modern comparative medical studies. They evaluate the performances of Wald-type statistic, Fieller's Theorem, Likelyhood Ratio, Score Statistic and Bootstrap Method and find out that the Bootstrap Method produces as reliable results in small samples as it does in large samples. Tang et al. create 8, 15, 30, 50 and 1000 bootstrap instances and as a result, they argue that the bootstrap method is

a noteworthy one among others in small sample sizes. (Nian-Sheng Tang, Shi-Fang, Man-Lai Tang,Yan-Bo Pei Qiu, 2011)

Takma and Atıl provide information on hypothesis testing, regression analysis and confidence intervals in the Bootstrap Method and indicate that the method is rather simple to employ. In their study, 1000 bootstrap instances are derived out of 25 observations and it is concluded that computations produce similar outcomes in both classical methods and the Bootstrap method, while providing practical evidence to support the superiority of the Bootstrap Method over classical methods. Moreover, the Bootstrap Methods does not have assumptions on the data set at hand. In the widely known classical methods, the results may be deemed unreliable in the case of a doubt in terms of the validity of the assumptions required to conduct certain analyses. The Bootstrap method produces reliable outcomes even in such conditions. Hence, the Bootstrap Method is superior to the analysis of variance in very small data sets or non-normally distributed data.

Bootstrap Method is widely used in applied statistics. The core of this method is to resample the existing data set with the purpose of creating very large data sets. Thus, various statistical analyses can be performed on small data sets. (Takma, Ç., Atil,H., 2006)

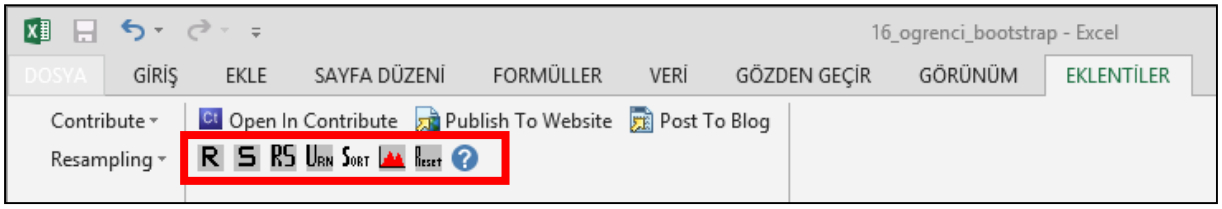
In the present study, the factors which have an effect on student performance, success or failure are revealed through weka, which is a machine learning tool, with a large sample created out of a small sample with Resampling Excel-add-in written on Excel 2013. The data acquired through questionnaires and the achievement test of "Our Planet Earth" chapter of the Science and Technology course taken by 16 fourth-grade students in two different primary schools in Edirne and İstanbul is used for the study.

2. MATERIAL AND METHODOLOGY

The research is completed in three phases: The Creation of the data set, performance analysis and achievement analysis.

2.1. The Creation of the Data Set

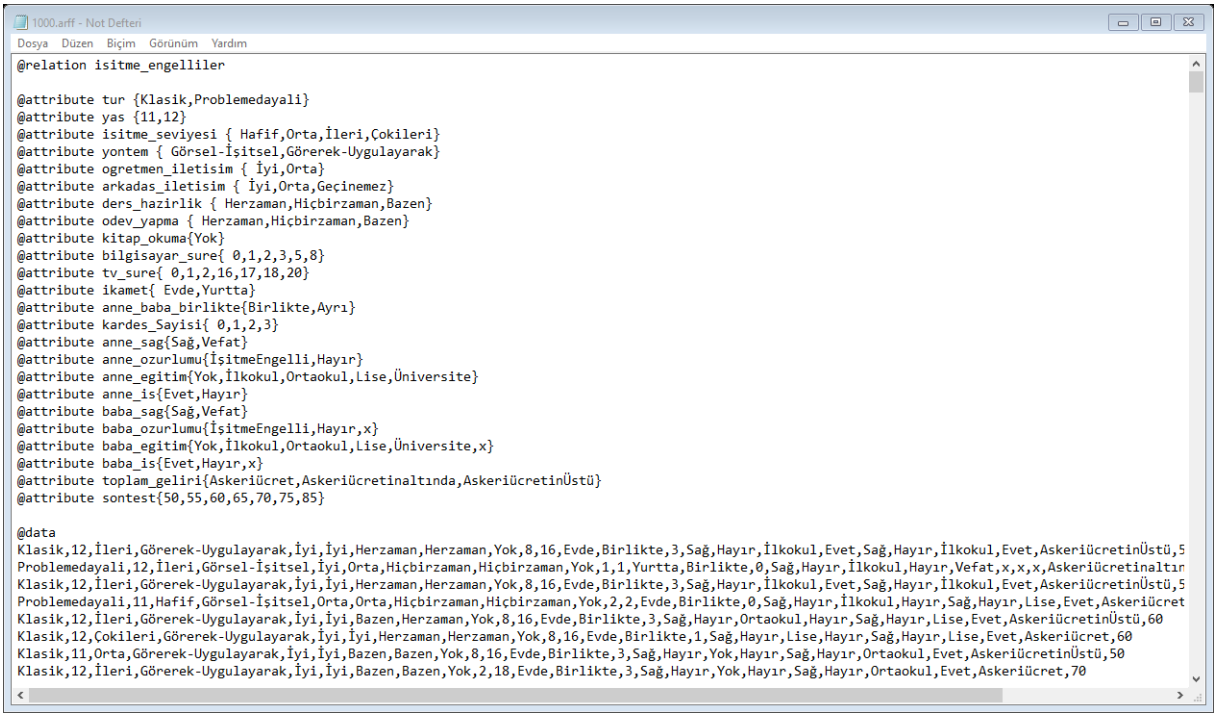
The study was conducted with a sum of 16 students who volunteered to participate in the study and were registered in the fourth grade of two schools for the deaf in Edirne and İstanbul. As a part of the study, the students were given a questionnaire and asked to take the achievement test of "Our Planet Earth" chapter within the Science and Technology course. The data collected through these instruments are resampled as 100, 300, 500, 700, 900 and 1000 instances separately with the Bootstrap Method through Resampling Excel-add-in Macro developed on Excel 2013. For this process, the items in the Plug-ins tab on the ribbon in Excel are utilized.



Picture 1. Resampling Excel-add-in Macro Tab

2.2. Analysis of the Data Set

There are various tools used in the field of data mining and one of the most popular tool among those is Weka. With Weka, it is possible to run numerous algorithms on the desired data set. In the present study, the classification algorithms in Weka 3.7 are benefitted from in the analysis of the data sets in different sizes. For that reason, the data sets are firstly converted into .arff to be read by Weka. The picture below displays a file in .arff format.



Picture 2. Data Set File with 1000 instances in .arff Format

Various criteria can be utilized to measure the performance of a classification algorithm. In this study, the criteria used to measure the performance of the algorithm were prediction accuracy, kappa and f-measure values.

Prediction Accuracy is the simplest and most popular method of measuring the model success. Prediction Accuracy is the value which is utilized by classification algorithms by means of using the observations as both training and test data, thus providing the rate of accurate predictions. (Özhan E.2014)

Kappa Statistic (KS) is a criterion used to quantify the agreement between the predicted and the observed classes in a given data set (Witten et al., 2005; Aydın, F. 2011). KS takes values between -1 and 1. -1 indicates a total disagreement or a negative relationship. 1, on the other hand, signifies a perfect agreement (Kılıçaslan,Y. et al., 2009). If KS has a value of or above 0.4, the level of agreement is considered to be acceptable beyond chance (Landis et al., 1977; Aydın, F. 2011).

F-measure is expressed as the harmonic mean of the precision and recall values. F-measure is an important criterion to increase the performance of the classifier especially in the process of preparing training data. In the performance analysis of classifiers, the relationship between f-measure and data size is put forward (Kılıçaslan, Y. et al., 2009) and through the learning curve plotted between f-measure and data size, a happy graph was obtained (Russell et al., 2003; Aydın, F.2011). Furthermore, f-measure in the application developed in this study was used to decide if the classes were sufficient for diagnosis. The minimum acceptable F-measure is normally 0.5. In this study, this value is taken as 0.5, too. (Aydın, F. 2011)

In the present study, the classification algorithms of Weka were used in default values to avoid affecting model success. Cross-Validation 10 was preferred as a test method. The results of the analyses showed that the LWL classification algorithm was the one which produced the most successful results in the data set with 500 instances. The second position in the ranking was shared by HoeffdingTree and NaiveBayes.

Algorithms	Instances	Correctly Classified Instances %	Kappa	F-measure
NaiveBayes	500	% 95	0,94	0,95
HoeffdingTree	500	% 95	0,94	0,95
LWL	500	% 100	1,00	1,00

Table 1. The Most Successful Algorithms

LWL classification algorithm is similar to other lazy learning methods according to Frank et al. (2003). However, it operates differently while a new instance is being classified. LWL algorithm creates a new Naive Bayes model using the data weight of the learning instances. This outperforms both Naive Bayes and similar methods with many experimentally tested data sets. (Gebeyehu,S., Rao,V.,S., 2014)

2.3. Achievement Analysis

In the acquisition of the posttest attribute, which was the target attribute, with the instances in the study, InfoGainAttributeEval (InfoGainAttributeEval, 2016) and ChiSquaredAttributeEval algorithms in the Select Attributes tab were used to display how much information each attribute which formed each instance in the data set provided.

The data set with 500 instances, on which the most successful results were obtained, was selected to run InfoGainAttributeEval and ChiSquaredAttributeEval algorithms. Cross-Validation 10 was preferred as the selection mode of the algorithms. As a result of the analysis of the algorithms, it was seen that the tv_duration attribute had the strongest effect on the posttest attribute. Moreover, it was also found out that computer_duration and father and mother education level attributes had an effect on posttest achievement.

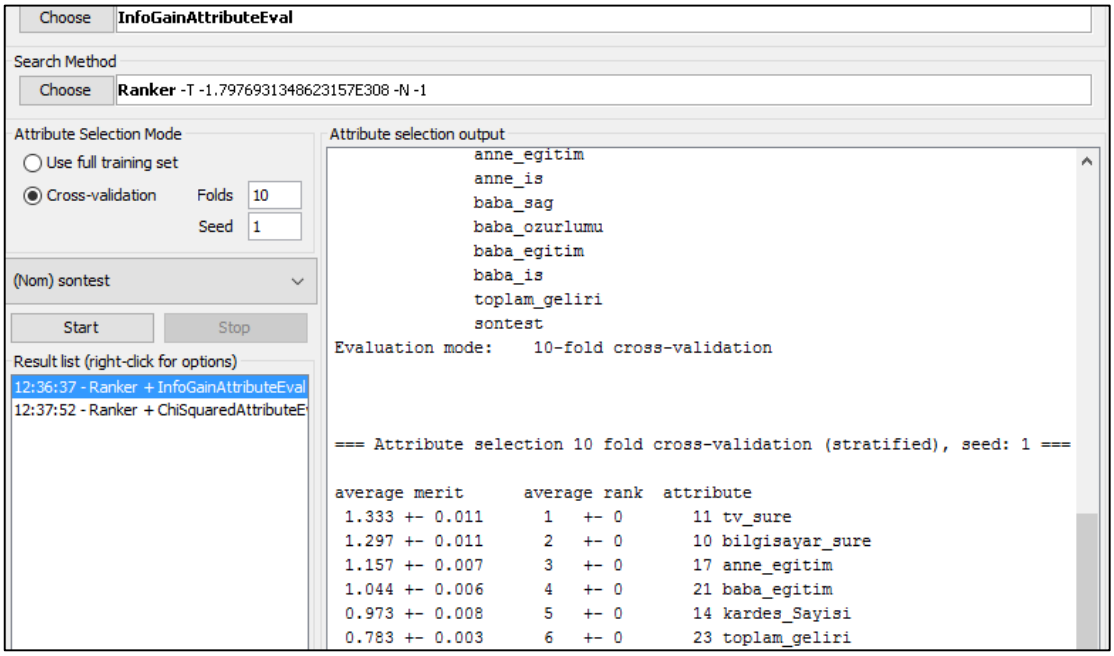
3. FINDINGS

In the study, data sets created by means of bootstrap method in varying sizes were analyzed with the classification algorithms of Weka. Among these algorithms, LWL produced the most succesful result on the data set with 500 instances. The results acquired with this data set are as follows: Prediction accuracy is 100%, Kappa value is 1 and F-measure value is 1.

Sample Size	Correctly Classified Instances %	Kappa	F-measure
16	0	-0,25	0,00
100	92	0,89	0,90
300	90	0,88	0,90
500	100	1,00	1,00
700	92	0,90	0,92
900	98	0,98	0,98
1000	99	0,99	0,99

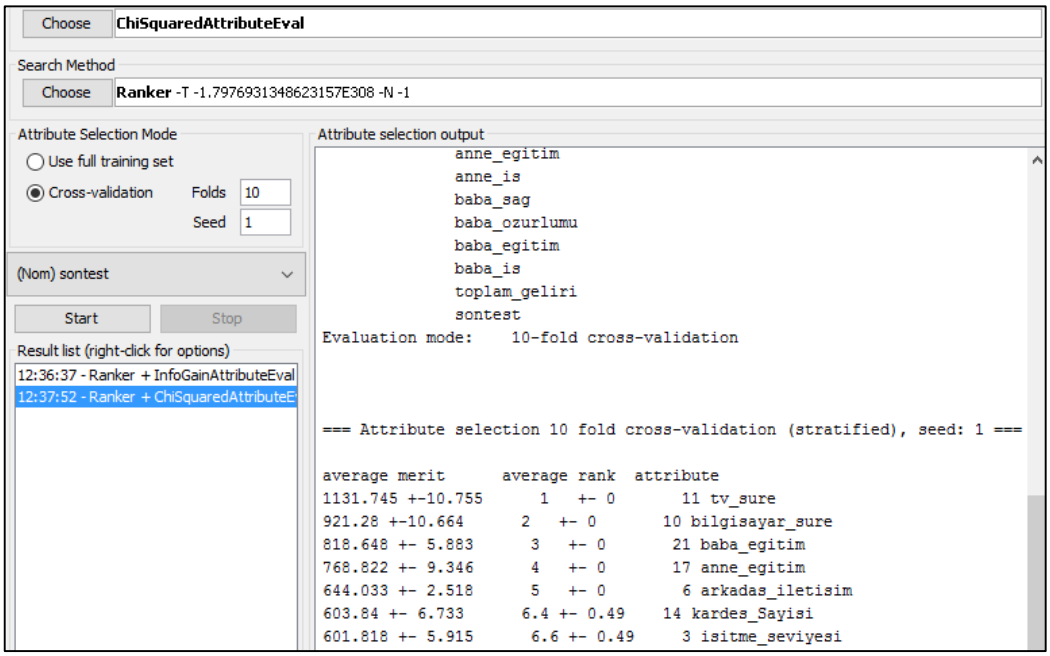
Table 2. LWL Results

When InfoGainAttributeEval algorithm was run on the data set with 500 instances, it was seen that the tv_duration, computer_duration, mother_education and father_education attributes were respectively the four most effective attributes in reaching the posttest class value.



Picture 3. InfoGainAttributeEval Results

When ChiSquaredAttributeEval algorithm was run on the data set with 500 instances, it was seen that the tv_duration, computer_duration, father_education and mother_education attributes were respectively the four most effective attributes in reaching the posttest class value.



Picture 4. ChiSquaredAttributeEval Result

4.CONCLUSION

The present study was carried out with the fourth grade students of two volunteer schools for the deaf in Edirne and İstanbul. The data collected from 16 volunteer students in these schools through questionnaires and tests were converted into data sets with 100, 300, 500, 700, 900 and 1000 instances by means of the Bootstrap Resampling Method. These data sets were separately analyzed with the classification algorithms within Weka. As a result of the analyses, it was seen that the LWL classification algorith had a prection accuracy rate of 100%, a kappa value of 1 and an F-measure value of 1 for the data set with 500 instances. Thanks to these results, it was concluded that the LWL algorithm was more successful than the other algorithms. In addition, the algorithms with a close rate of success to that of the LWL algorithm were found out to have the highest values with the data set which had 500 instances.

Furthermore, it was discovered that the length of the time spent with television or computer had the strongest effect on the level of student achievement, when each attiribute within the valid data set was analyzed by means of ChiSquaredAttributeEval and InfoGaimAttributeEval algorithms to reveal the classifier value, that is, how much information it provides to reach the posttest(sont_test) attribute. Moreover, it was also seen that the level of the participants' mothers and fathers have an effect on the dependent variable.

In the light of the results of the present study, it is believed that similar research applications on larger data sets with different chapter achievement tests in the schools for the deaf will aid in revealing and clarifying the attributes which have an effect on student achievement.

REFERENCES

Aydın F., Kalp Ritim Bozukluğu Olan Hastaların Tedavi Süreçlerini Desteklemek Amaçlı Makine Öğrenmesine Dayalı Bir Sistemin Geliştirilmesi, Yüksek Lisans Tezi,2011;39-41

Gebeyehu,S., Rao,V.,S., A Two Step Data Mining Approach for Amharic Text Classification ,American Journal of Engineering Research (AJER) e-ISSN : 2320-0847 p-ISSN : 2320-0936 Volume-03, Issue-04, 2014; 251-259

InfoGainAttributeEval, viewed 15 February 2016. http://wiki.pentaho.com/display/DATAMINING/InfoGainAttributeEval

Kılıçaslan Y., Güner E. S., Yıldırım S., Learning-based pronoun resolution for Turkish with a comparative evaluation; Computer Speech & Language Volume 23, Issue 3, July 2009;311-331.

Koşkan,Ö., 2007, "Grup Karşılaştırmalarında Yeniden Örnekleme(Resamplıng) Yaklaşımı", Doktora Tezi, Ankara Üniversitesi Fen Bilimleri Enstitüsü, Zootekni Anabilim Dalı, Ankara

MEB, Özel Eğitim Nedir?, viewed 10 February 2016. http://surram.meb.k12.tr/meb_iys_dosyalar/21/18/974637/icerikler/ozel-egitim-nedir_273810.html

Millî Eğitim İstatistikleri 2013/'14, Örgün Eğitim T.C. Millî Eğitim Bakanlığı, Strateji Geliştirme Başkanlığı, Resmi İstatistik Programı Yayınıdır Issn 1300-0993,s:36

Nian-Sheng Tang , Shi-Fang, Man-Lai Tang ,Yan-Bo Pei Qiu, 2011, "Asymptotic Confidence İnterval Construction For Proportion Difference İn Medical Studies With Bilateral Data",Statistical Methods İn Medical Research,; 20: 233–259

Orekkici Temel G., Erdoğan, S., Tanı Koyma Amaçlı Yapılan Tıbbi Çalışmalarda Jacknife, Bootstrap Ve Çapraz Geçerlilik Yöntemlerinin Kullanımı , Düzce Üniversitesi Sağlık Bilimleri Enstitüsü Dergisi 2011;1(3): 45-49 Issn: 2146-443x

Orekkici Temel G., Erdoğan, S., Sınıflama Modelinin Performansını Değerlendirmede Yeniden Örnekleme Yöntemlerinin Kullanımı – Gülahn Orekkici Temel, Semra Erdoğan, Handan Ankaralı – Bilişim Teknolojileri Detgisi, Cilt:5, Sayı:3, Eylül 2012

Özhan E., Yapay Zeka Yöntemleri İle Dünya Depremlerinin Modellene Bilirliği, Isıtes2014 Karabük – Turkey,2014; 1188

Takma, Ç., Atil,H., 2006, "Bootstrap Metodu Ve Uygulanışı Üzerine Bir Çalışma 2. Güven Aralıkları, Hipotez Testi Ve Regresyon Analizinde Bootstrap Metodu", Ege Üniv. Ziraat Fak. Derg., , 43(2):63-72, Issn 1018-8851