The Eurasia Proceedings of
Educational & Social Sciences (EPESS)
ISSN: 2587-1730

The Eurasia Proceedings of Educational & Social Sciences (EPESS), 2017

Volume 8, Pages 9-18

IConSE 2017: International Conference on Science and Education (IConSE)

PERFORMANCE AND ACHIEVEMENT ANALYSIS OF A DATASET OF DISTANCE EDUCATION SAMPLES WITH WEKA

Tolga Demirhan
Trakya University

Ilker Hacioglu
Trakya University

Abstract: Data mining methods can be used to create models that will help in making meaningful deductions or even future predictions by establishing relationships within records which have values that can not be understood alone. In this study, a data set was created through the voluntary participation of Trakya University, Tunca Vocational School (Distance Education) students to a questionnaire. Weka, a data mining application, was used to analyze the survey results. The most successful models on Weka for the relevant data set and the attributes that affect student success were investigated.

Keywords: Weka, achievement analysis, data mining, distance education

Introduction

Since the last few years, many countries have faced the failure of their students and the problem of student dropout. For this reason, governments of many countries have been focusing on determining the factors that cause this situation. Foreseeing students' school failure can be a challenging task because it is a multi-factor issue, and the present data are normally imbalanced. In order to solve these problems, information discovery, data mining algorithms and approaches are used in databases to predict student failure. (Bhawana, A., Bharti, G.,2014)

Data mining is the process of discovering information by analyzing large quantities of data from various perspectives and extracting useful information. (Pandeeswari, L., Rajeswari, K., 2014) Data mining has been successfully used in diverse fields, including the academic field. The data mining area which is concerned with developing methods for discovering the unique data types that come from the educational environment and which contain the results of the students, reveal useful patterns in the database for better understanding of the student and evaluating the learning process of the student is referred to as educational data mining. (Chan, A.Y. K., Chow, K.O., Cheung, K. S., 2008. Chandra, E., Nandhini, K., 2010)

The samples forming the dataset in the educational data mining and their attributes can be generated by school automations where information such as gender, age, and history is taken and recorded, or they can be created traditionally by face-to-face interviews or surveys in the classroom environment. (Araque, F., Roldan, C., Salguero, A.,2009)

One of the most useful data mining techniques for educational data sets is classification. Classification maps the data to a set of predefined classes. In classification, classes are determined prior to examining the data due to the supervised learning approach. It is useful to estimate the performance of the learner with high accuracy. (Guleria, P., Sood, M.,2014)

In the study, Weka software was utilized since it presents preprocessing, student achievement and the algorithms which would be needed in the information discovery experiments together. WEKA is an open source software published under the GNU General Public License. WEKA is used as a tool to run different classification

- Selection and peer-review under responsibility of the Organizing Committee of the conference

*Corresponding author: Tolga Demirhan- E-mail: tolgademirhan@trakya.edu.tr

algorithms. WEKA includes tools for data preprocessing, classification, regression, clustering, association rules and visualization. (Dener, M., Dörterler, M., Orman, A.,2009)

Methods

The real world instances used in the study were formed with a total of 156 distance education students who were studying computer programming and volunteered to participate in the study. Instances within the data set were the students enrolled in the computer programming program in the 2016-2017 academic year. In the study, a questionnaire was given to the students, asking their demographic and academic information. The Google forms survey application was used to obtain the survey results from the students. The responses of the students were analyzed by exporting them to an excel file via the Google forms survey program and potentially problematic areas were corrected. The questions and codes used in the study are presented below.

Items	Code
What is your gender?	S1
How old are you?	S2
Please state what year of studies you are in.	S3
Please state your marital status.	S4
Please state the number of dependents as children you have.	S5
What is your most recently received qualification?	S6
Please state the city you live in.	S7
Are you employed?	S8
Which sector does your job belong to?	S9
For how many years have you been working?	S10
What is your average monthly income?	S11
What is your high school diploma grade?	S12
What is your average grade in Turkish, History and Foreign Language courses?	S13
What is your average grade in the Mathematics course?	S14
What is your average grade in the software based courses? (Intro. To Programming, Visual Programming, Internet Programming, Object Oriented Programming etc...)	S15
What is your average grade in theoretical vocational courses? (Server OS, Computer Architecture, Database Management Systems etc...)	S16
What is the type of company you completed your practical training?	S17
In which province is that company located?	S18
In which professional field does the company operate?	S19
In which unit did you complete your practical training?	S20
What was the weekly duration of work? (5-6-7 days...)	S21
How would you evaluate the duration of practical training?	S22
I believe that the practical training was beneficial in terms of knowledge and it contributed to my receiving information related to my profession.	S23
Were you able to use the knowledge you acquired in your department during your practical training?	S24
I believe that practical training helped me gain experience.	S25
Practical training encouraged me to learn more.	S26
I would recomment the company where I completed my practical training to my friends.	S27
The administration of practical training company were concerned with the trainees.	S28
Practical training contributed to my teamwork skills.	S29
Practical training contributed to the improvement of my written and oral communication skills.	S30
Practical training increased my sense of duty and responsibility.	S31
What is your level of computer use?	S32
How frequent do you use computers and the internet?	S33

What is your frequency of using the LMS?	**S34**
For how many hours in a day do you use the internet for learning purposes?	**S35**
Please select the learning material or materials you prefer while studying.	**S36**
The content of the courses are sufficient theoretically.	**S37**
The content of the courses are sufficient in terms of applications.	**S38**
The videos related to the courses are sufficient.	**S39**
I was able to communicate my requests and suggestions regarding the courses.	**S40**
The language of the courses are clear and understandable.	**S41**
Course content fits the aims of the program.	**S42**
Transport to the examination center and physical facilities are sufficient.	**S43**
Exam results are announced timely.	**S44**
I participate in the virtual class communication hours.	**S45**
I am content with getting in contact with the teaching staff during the virtual class communication hours.	**S46**
I am content with the Learning Management System (LMS) that is used.	**S47**
I know how to use the LMS.	**S48**
I can easily access the content visa LMS and the web site of the university.	**S49**
I receive informative e-mails and SMS regarding my distance education courses and exams.	**S50**
I find the announcements and notifications insufficient.	**S51**

Figure1. Table of questionnaire items – codes and responses

The data was converted into .arff format, which can be understood by Weka, to perform analyses on the records with the software. The .arff file is provided below.

```
@relation Uzaktan_Egitim

@attribute s1 { bay,bayan }
@attribute s2 { 23_28,29_34,17_22,41veüzeri,35_40 }
@attribute s3 { 1sinif,2sinif }
@attribute s4 { bekar,evli }
@attribute s5 { cocugum_yok,evet_1_cocugum_var,evet_2_cocugum_var,evet_3_ve_u
@attribute s6 { lise_mezunu,universite_mezunu,yuksekokul_mezunu,yuksek_lisans
@attribute s7 { istanbul,sakarya,tekirdag,kırklareli,edirne,konya,kocaeli,ank
@attribute s8 { evet_calisiyorum,hayir_calisiyorum }
@attribute s9 { kamu_kurumu,ozel_sektor,* }
@attribute s10 { 4_7_yil,10_yildan_fazla,8_10_yil,0_3_yil,* }
@attribute s11 { 1501_3000_TL,3001_4000_TL,4001_TL_ve_üzeri,0_1500_TL,* }
@attribute s12 { 3_3.99,4_5,2_2.99 }
@attribute s13 { basarili,basarisiz }
@attribute s14 { basarili,basarisiz }
@attribute s15 { basarili,basarisiz }
@attribute s16 { basarili,basarisiz }
@attribute s17 { kamu_kurumu,ozel_kurum,* }
@attribute s18 { istanbul,*,ankara,edirne,kocaeli,tekirdag,antalya }
@attribute s19 { bilgisayar,elektrik,belediye,egitim,yazilim,*,tekstil,kamu,d
@attribute s20 { plan,ofis,cagri_merkezi,bilgi_islem,yazilim,*,arge,satis,oto
@attribute s21 { 5gun,6gun,*,7gun }
@attribute s22 { fazla,yeterli,*,yetersiz }
@attribute s23 { evet,*,hayir }
@attribute s24 { evet,*,hayir }
@attribute s25 { evet,*,hayir }
@attribute s26 { evet,*,hayir }
@attribute s27 { evet,hayir,* }
@attribute s28 { evet,*,hayir }
```

Figure2. Arff data set file

When the data set was examined, it was seen that the instance quantities in the four-class structure, which the attributes containing the achievement data has, posits imbalance.In a data set, having the number of class instances in the target attribute close to one another, that is, having a balanced distribution is a desirable condition for algorithm success. (Bulut, F., 2016) Many algorithms and techniques are proposed to solve the problem of instance imbalance. (Longadge, R., S. Dongre, S., Malik, L., 2013) In this study, the attribute expressing course achievement as failure, mediocre, good and very-good was converted into two classes as success and failure.

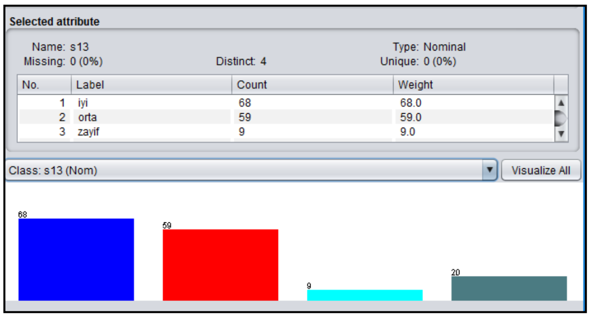

Figure3. Classify value with two classes

Despite many classification models, Naïve Bayes, MLP, SMO, j48, Rep tree, Random tree and Decision table algorithms, which are preferred in data sets with educational instances and nominal type attributes, were selected for the study. (Pandeeswari, L., Rajeswari, K., 2014)

Algorithms

Naive Bayes: The Bayesian classifiers are based on the Naive Bayes theorem and assume that each attribute is independent of other attributes in the sample. The conditional probability of a class label is predicted and assumptions are made on the model to make this probability a product of conditional probabilities. (Mishra, T., Kumar, D., Gupta, S.,2016)

MLP: Almost all the problems encountered in daily life are nonlinear. Multilayer Perceptron (MLP) model has been developed as a result of the studies to solve the XOR problem which shows a non-linear relationship. This model is also referred to as error propagation model or backpropagation network model. The reason why this model is widely used is that numerous training algorithms can be used in training this particular network. (Öztemel, E., 2012)

SMO: Sequential Minimal Optimization is a highly preferred algorithm due to its simplicity and it is used to solve the optimization problems that arise during training. SMO is an algorithm that uses SVM (Support Vector Machine) algorithm. SMO makes choices to solve the smallest possible optimization problems in every single step and produces results. (J.Plat,2000)

J48: In a decision tree all instances start at the root node. The best discriminating attribute is used in the root node and branches to inner nodes based on the partitioning feature. The operation continues as long as all instances in a node belong to the same class, some threshold criteria are met or there are no attributes left. One of the most useful features of decision trees is their easy interpretation in terms of understandability and rules. It is assumed that decision tree has different values in at least one of the attributes of the instances which belong to different classes. J48 is a decision tree algorithm based on the very popular C4.5 algorithm. (Mishra, T., Kumar, D., Gupta, S.,2016)

Rep tree: REPTree is a fast decision tree learner which forms a decision tree by using information acquisition as a criterion to select the attribute to be tested in a node. (Erdogan, S., Timor, M.,2009)

Random tree: Random Tree is a classification algorithm that creates a tree by taking randomly selected attributes at each node in a certain number. No pruning. It also has an option that allows for the prediction of the class probabilities based on the retained data set. (Akçetin, E., Çelik, U.,2015)

Decision Table: The decision table is used for the analysis and representation of complex logical relations. It is ideal for defining situations where various combinations of actions are taken under different condition clusters. In order to define test conditions through decision tables, conditions should be interpreted as input and actions should be interpreted as output. (Noikajana, S., Suwannasart, T.,2008)

InfoGainAttributeEval: The contribution of each attribute in a data set to the target attribute can be found by the Select Attributes algorithms provided by WEKA. In the study, InfoGainAttributeEval was used to reveal how much information is provided by each attribute of each instance in the data set. InfoGainAttributeEval works by evaluating an attribute's value by measuring the informational gain of the class. InfoGainAttributeEval supports Binary class, Missing class values and Numeric values. (Rajpal, R., Kaur, S., Kaur, R., 2016)

Evaluation Metrics

The Confusion Matrix is used to evaluate the performance of the classification models. A two-class confusion matrix is as follows.

Table 1. Confusion matrix

		Predicted Class	
		Class=Success	Class=Failure
Actual Class	Class=Success	TP	FN
	Class=Failure	FP	TN

TP (True Pozitive) - FN (False Negative) - FP (False Pozitive) - TN (True Negative)

TP shows the number of those who are actually successful and who were assigned to the success class by the classifier.

FN shows the number of those who are actually successful but who were assigned to the failure class by the classifier.

FP shows the number of those who actually failed but who were assigned to the success class by the classifier.

TN shows the number of those who actually failed and who were assigned to the failure class by the classifier.

Values such as Accuracy, Precision, Recal, F-Measure, Kappa, ROC and so on, which help interpret the performances of the algorithms, can be obtained using the complexity matrix. In this study, Prediction Accuracy, Kappa and F-measure values were used to measure the performance of the algorithm and to compare the results with those of other classification algorithms.

Prediction accuracy is the most popular criterion that is used for the comparison of model success. It is expressed as the ratio of the sum of the number of correct predictions (TP + TN) made by the classiffication algorithm to the sum of the number of instances (TP + TN + FP + FN) partaken in the classification. (Syahela Hussien, N.,2016)

$$\textbf{Accuracy} = (\sum \textbf{True Positivie} + \sum \textbf{True Negative}) / \sum \textbf{Total Population}$$

Kappa Statistic (KS) is a measure used to quantitatively express the correspondence between predicted and observed classifications in a data set. KS value is between -1 and 1. -1 indicates that there is an incompatibility or a relationship in the opposite direction. 1 indicates perfect correspondence. If KS has a value of 0.4 or above, it can be said that there is an acceptable correspondence beyond chance. (Aydın, F.,2011)

$$\textbf{Kappa} = (\textbf{Observed Accuracy} - \textbf{Expected Accuracy}) / (1 - \textbf{Expected Accuracy})$$

F-Measure is expressed as the harmonic mean of the values of precision and recall. F-Measure is used in particular to have knowledge of the classifier's performance during the preparation of training data and to determine if classes are sufficient for diagnosis. The acceptable F-criterion value is usually taken as a minimum of 0.5. (Aydın, F.,2011)

$$\textbf{F-measure} = 2\textbf{x}\,((\textbf{Precision x Recall}) / (\textbf{Precision + Recall}))$$
$$\text{Precision} = \sum \text{True Positive} / \sum \text{Test Outcome Positive}$$
$$\text{Recall} = \sum \text{True Positive} / \sum \text{Condition Positive}$$

Results and Findings

Experiments were performed using Weka 3.8.1. Naïve Bayes, MLP, SMO, j48, Rep tree, Random tree and Decision table were used as the main classifiers. The results obtained from the data sets are given below. As the classify value, Turkish, History, English S13, Mathematics S14, Applied vocational courses 15 and theoretical vocational courses S16 means were used. The results obtained from experiments in this context are given in the tables below.

While using the Weka classification algorithms, adjustments which would affect the model success were avoided and the algorithms were use in their defaults. In addition, Cross-Validation 10 fold is preferred as the test method. In this method, the data source is divided into 10 equal parts and each piece is once used for testing, and the remaining 9 pieces are used as training sets.

For the means of the Turkish, History and English courses with the classify value of S13, J48 algorithm produced a correct classification rate of 74%, a Kappa value of 0,459 and an F-measure value of 0,735. Regarding these results, it was concluded that the J48 algorithm was more successful than other algorithms.

Table 2. Classify value: (Turkish, history and English mean) S13

Algorithms	*Correctly C.Ins. %*	*F-measure*	*Kappa*
J48	74	0,735	0,459
RepTree	65	0,642	0,270
DecisionTable	71	0,708	0,405
MultiplayerPerceptron	65	0,652	0,291
NaiveBayes	63	0,630	0,254
SMO	71	0,705	0,398
RandomTree	55	0,551	0,088

When the results of the algorithms are examined for the success of the Mathematics course with the classify value of S14, it was concluded that the results obtained from the algorithms were coincidental because the kappa value was below the acceptable limits.

Table 3. Classify value: (Mathematics course) S14

Algorithms	*Correctly C. Ins %*	*F-measure*	*Kappa*
J48	69	0,674	0,228
RepTree	67	0,648	0,166
DecisionTable	73	0,715	0,325
MultiplayerPerceptron	69	0,689	0,278
NaiveBayes	68	0,678	0,256
SMO	69	0,685	0,275
RandomTree	62	0,615	0,117

The correct classification rate obtained by the J48 algorithm for the Applied Vocational Courses mean with the classify value of S15 was found to be 80% with a Kappa value of 0,60, and an F-measure value of 0,80. It was concluded that the J48 algorithm was more successful than other algorithms for the related classify value.

Table 4. Classify value: (Applied vocational courses mean) S15

Algorithms	*Correctly C. Ins. %*	*F-measure*	*Kappa*
J48	80	0,801	0,5954
RepTree	78	0,782	0,5582
DecisionTable	76	0,760	0,5112
MultiplayerPerceptron	71	0,711	0,4144
NaiveBayes	71	0,705	0,4023
SMO	63	0,633	0,2538
RandomTree	59	0,589	0,1659

The correct classification rate of the J48 algorithm for the theoretical vocational courses mean with the classify value of S16 was found to be 79% with a kappa value of 0.57, and F-measure value of 0.79. It was concluded that the J48 algorithm for classify value related to these results was more successful than the other algorithms.

Table 5. Classify value: (Theoretical vocational courses mean) S16

Algorithms	*Correctly C.Ins %*	*F-measure*	*Kappa*
J48	79	0,788	0,5677
RepTree	78	0,782	0,5553
DecisionTable	79	0,787	0,5648
MultiplayerPerceptron	78	0,782	0,5553
NaiveBayes	69	0,687	0,3687
SMO	74	0,743	0,4751
RandomTree	60	0,596	0,1774

The result obtained by running the InfoGainAttributeEval on the data set for the attributes with the classify value of Turkish, History and Foreign Language is given in the figure below. When it is run for the attribute with the code S13 and the classify value of Turkish, History and Foreign Language, it was seen that the 5 most effective attributes in terms of reaching this particular attribute were S16, S6, S14, S19 and S15 respectively.

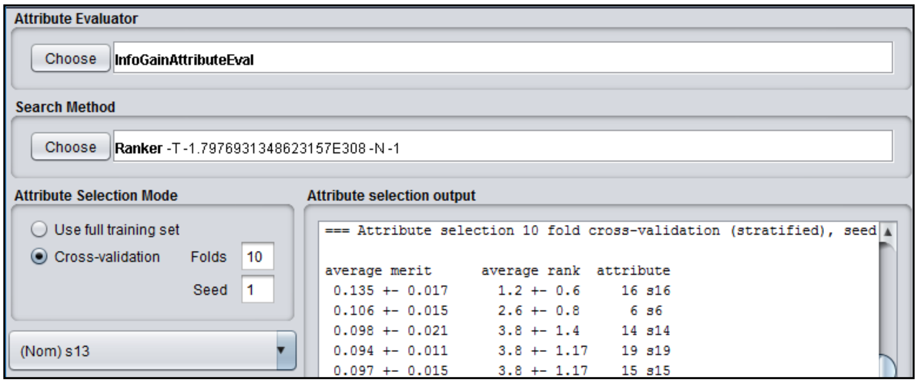

Figure 4. InfoGainAttributeEval Algorithm classify value: Türkçe

The result obtained by running the InfoGainAttributeEval on the data set for the attributes with the classify value of Mathematics is given in the figure below. When it is run for the attribute with the code S14 and the classify value of Mathematics, it was seen that the 5 most effective attributes in terms of reaching this particular attribute were the Education Level S6, S19, S13, S7 and S36 respectively.

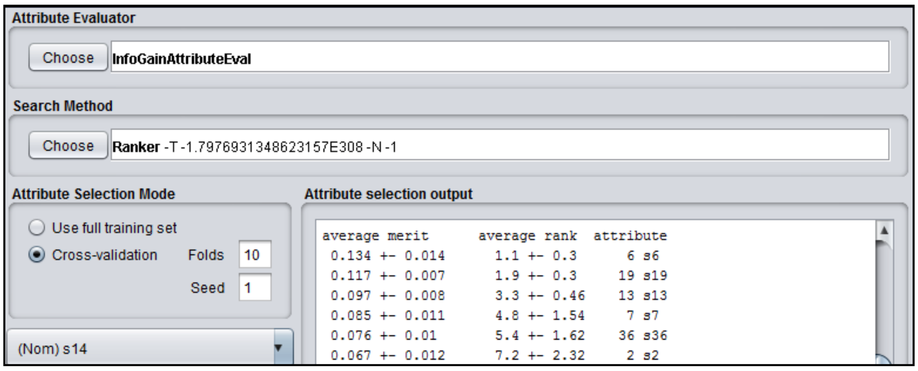

Figure 5. InfoGainAttributeEval algorithm classify value: mathematics

The result obtained by running the InfoGainAttributeEval on the data set for the attributes with the classify value of Applied Vocational Courses is given in the figure below. When it is run for the attribute with the code S15 and the classify value of Applied Vocational Courses, it was seen that the 5 most effective attributes in terms of reaching this particular attribute were the Theoretical Vocational Courses Mean S16, S13, S32, S36 and S51 respectively.

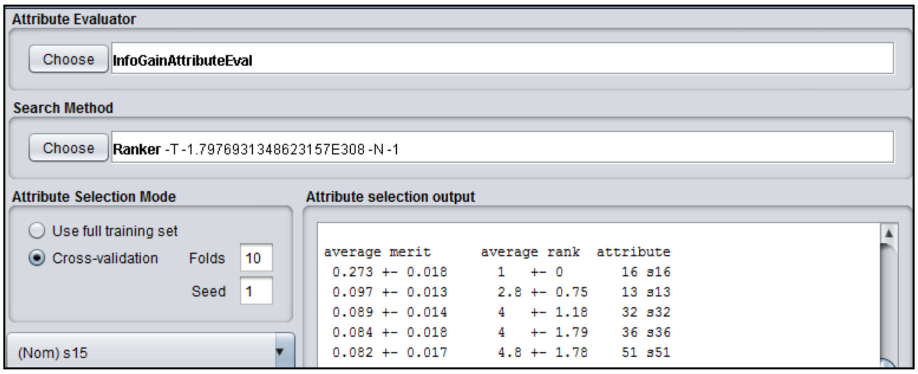

Figure 6. InfoGainAttributeEval algorithm classify value: applied vocational course achievement

The result obtained by running the InfoGainAttributeEval on the data set for the attributes with the classify value of Theoretical Vocational Courses is given in the figure below. When it is run for the attribute with the code S16 and the classify value of Theoretical Vocational Courses, it was seen that the 5 most effective attributes in terms of reaching this particular attribute were the Applied Vocational Courses Mean S15, S32, S13, S19, S7, S3 and S20.

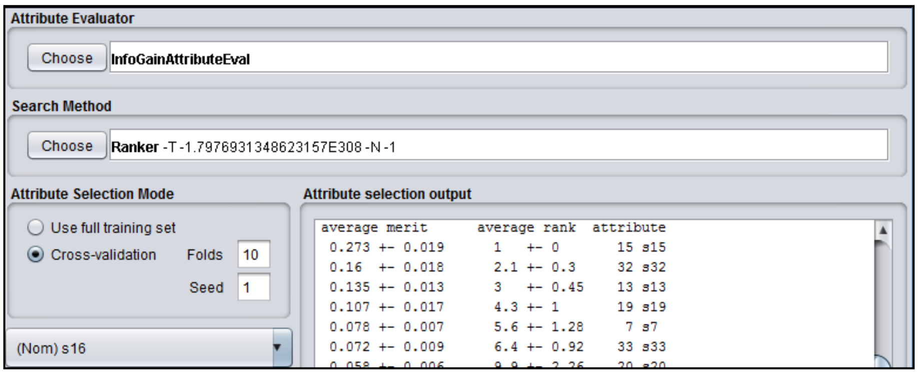

Figure7. InfoGainAttributeEval algorithm classify value: Theoretical vocational course achievement

Confusion Matrix results, tree view and tree structure obtained through J48 classification algorithm for the classify value coded S15 (Applied Vocational Courses Mean), which has the most successful classification performance value, are presented below.

When the tree structure is examined, it is seen that the majority of the instances successfully labeled in the S16 code (Applied Vocational Courses Mean) succeeded for the S15 attribute. When the rules are examined;

For the 88 instances where the S16 = success: success (88.0 / 16.0) line and the value of the S16 code attribute succeeded, it was found that the J48 algorithm successfully predicted the S15 attribute for the 88 records, but 16 of these predictions were incorrect.

For the rows S16=failure, S14=success and S35=2_3hours:success(3.0), the values of the S14 and S35 attributes were failure and success respectively. In 3 records where the value of the S35 attribute was 2_3_hours, J48 algorithm predicted the S15 attribute to be success and these instances are, in fact, successful.

For the rows S16=failure and S14=failure:failure(56.0,/9.0), the S16 attribute value was failure. In 56 records where S14 attribute was failure, J48 algorithm predicted the S15 attribute to be failure and 9 of these values are in fact failure.

```
J48 pruned tree
------------------

s16 = basarili: basarili (88.0/16.0)
s16 = basarisiz
|   s14 = basarili
|   |   s35 = 2_3_saat: basarili (3.0)
|   |   s35 = 4_5_saat: basarili (3.0/1.0)
|   |   s35 = 0_1_saat: basarisiz (5.0)
|   |   s35 = 6_saat_ve_uzeri: basarili (1.0)
|   s14 = basarisiz: basarisiz (56.0/9.0)

Number of Leaves  :     6

Size of the tree :      9
```

Figure8. J48 decision tree

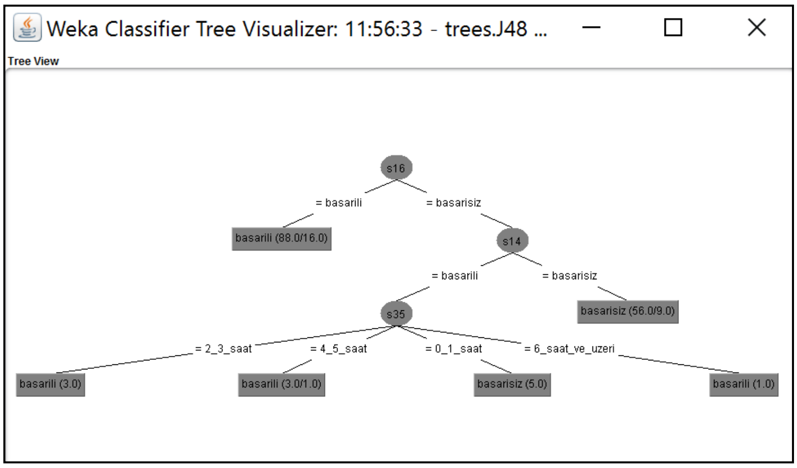

Figure9. J48 tree view

When the results were investigated, it was seen that 74 of 87 instances labelled as success and 52 of 69 instances labeled as failure were accurately predicted.

```
=== Confusion Matrix ===

  a  b   <-- classified as
 73 14 |  a = basarili
 17 52 |  b = basarisiz
```

Figure10. Confusion matrix

Conclusion

The study was conducted with the participation of 156 volunteer students of Trakya University, Tunca Vocational School (Distance Education), Computer Programming Program. The data set created using the questionnaire data was analyzed by classification algorithms provided by the WEKA program.

When the limits for the criteria as a result of the classification tests are considered, it was seen that the best performance for S13, S15 and S16 attributes were observed in the J48 algorithm.

In addition, the amount of information provided by each attribute used in the current data set in the point of reaching S13, S14, S15 and S16 were investigated with InfoGainAttributeEval and it was seen that the most efficient atrribute within the data set with the classify value of S15 was S16.

In the data set with the classify value of S15, with which the most successful results were obtained, S16, S14 and S35 attributes were efficient in classification according to the tree created using the J48 algorithm.

In the next study, firstly, a new data set will be formed by adding the questionnaire results belonging to the freshmen who will register in the 2017-2018 academic year to the current dataset. Secondly, a new data set will be created through resampling the data set which includes 156 instances. Lastly, comparisons of performance and achievement tests will be made on the data sets created.

References

Akçetin, E., & Çelik, U. (2015). İstenmeyen Elektronik Posta (Spam) Tespitinde Karar Ağacı Algoritmalarının Performans Kıyaslaması, *İnternet Uygulamaları ve Yönetimi, (pp.* 43-56). doi: 10.5505/iuyd.2014.43531

Araque, F., & Roldan, C., & Salguero, A. (2009). Factors influencing university drop out rate. *Computer& Education*, 53 (3), (pp. 563-574). https://doi.org/10.1016/j.compedu.2009.03.013

Aydın F. (2011). *Kalp Ritim Bozukluğu Olan Hastaların Tedavi Süreçlerini Desteklemek Amaçlı Makine Öğrenmesine Dayalı Bir Sistemin Geliştirilmesi* (Unpublished master dissertation). Trakya Üniversitesi Fen Bilimleri Enstitüsü, Edirne-Turkey.

Bhawana, A., & Bharti, G. (2014). Review on Data Mining Techniques Used For Educational System, *International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com*, 4 (11), Retrieved from http://www.ijetae.com/files/Volume4Issue11/IJETAE_1114_50.pdf

Bulut, F. (2016). Performance Analysis of Ensemble Methods on Imbalanced Datasets. *Bilişim Teknolojileri Dergisi*, 9 (2), (pp. 153-159), doi: 10.17671/btd.81137

Chan, A.Y.K., & Chow, K.O., & Cheung, K.S. (2008) Online Course Refinement through Association Rule Mining. *Journal of Educational Technology Systems,* 36 (4). (pp. 433 – 44)

Chandra,E., & Nandhini, K. (2010) Knowledge Mining from Student Data. *European Journal of Scientific Research,* ISSN 1450-216X 47 (1), (pp.156-163).

Dener, M., & Dörterler, M., & Orman, A. (2009). Açık kaynak kodlu veri madenciliği programları: WEKA'da örnek uygulama, *Akademik Bilisim'09 - XI. Akademik Bilişim Konferansı*, Şanlıurfa.

Erdoğan, S., & Timor, M. (2005). A Data Mining Application in a Student Database. *Journal of Aeronautics and Space Technologies*, 2(2), (pp.53-57).

Guleria, P., & Sood, M. (2014). Data Mining In Education : A Review On The Knowledge Discovery Perspective. *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, 4(5), (pp.47-60) .

J. Platt., (2000). *Fast training of support vector machines using sequential minimal optimization. Advances in kernel methods: Support vector learning*. Retrieved from https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/smo-book.pdf

Longadge, R., & S. Dongre, & Malik, L., (2013). Class Imbalance Problem in Data Mining: Review *International Journal of Computer Science and Network IJCSN* , 2(1), ISSN (Online) : 2277-5420

Mishra, T., & Kumar, D., & Gupta, S., (2016). Students' Employability Prediction Model through Data Mining. *International Journal of Applied Engineering Research,* 11 (4), ISSN 0973-4562, (pp:2275-2282).

Noikajana, S., & Suwannasart, T., (2008). Web Service Test Case Generation Based on Decision Table, *The Eighth International Conference on Quality Software*, (pp:321-326), doi 10.1109/QSIC.2008.7:

Öztemel, E. (2012). Yapay Sinir Ağları, Papatya Yayıncılık, Papatya & Kelebek Tasarım, ISBN: 978-975-6797-39-6, İstanbul.

Pandeeswari, L., & Rajeswari, K. (2014). Student Academıc Performance Using Data Mining Techniques. *International Journal of Computer Science and Mobile Computing, IJCSMC,* 3 (10), (pp.726–731).

Rajpal, R., & Kaur, S., & Kaur, R. (2016). Improving Detection Rate Using Misuse Detection and Machine Learning. *SAI Computing Conference.* London, UK.

Syahela Hussien, N., & Sulaiman, S., & Mariyam Shamsuddin, S., (2016). Tools in Data Science for Better Processing. *AIP Conference Proceedings*, doi: 10.1063/1.4954530