THE CLASSIFICATION OF THE STUDENTS SUCCESS VIA THE INFORMATIONS EXISTING IN E-SCHOOL SYSTEM

Uğur ERKAN[1], Levent GÖKREM[2]

[1]Erbaa Vocational School, Gaziosmanpasa University, 60500 Erbaa, Tokat, Turkey
[2]Faculty of Natural Sciences and Engineering, Gaziosmanpasa University, 60100 Tokat, Turkey

Corresponding author, e-mail: leventgokrem@gmail.com

ABSTRACT

This study focuses on the e-school system which is located in the e-government system. E-school system is one of the biggest project of Turkish Republic in today's world. In this system, the students' social and economic information are published. By this way, the students' potential success could be overviewed. The study reaches 756 high school students. In e-school system, 22 information which are related with the students' social and economic situations exist. Some of them are about the students' parent job, education or the houses they live. Some classification methods in Matlab programme such as Naïve Bayes, Knn, Euclidean, Mahalanobis are applied and by this way, their potential success have been tried to be guessed. % 75 of the students' success have been identified. 2/3 of the data (503 students) is used as education data, and 1/3 of them (253 students) is used as test data. The results show that the students' successes are directly related with their social and economic conditions.

Key words: Education, success, classification of student achievement, social conditions

1. INTRODUCTION

As it is searched before in the literature, it is proven that students' success are related with some factors such as social positions, economical factors. In this searches, using questionnaires is the most popular research technique. However; the questionnaires are not valid because the students don't answer the questions carefully. Moreover; they may not give the correct information in questionnaires. In Turkey, e-school system is used actively in every areas. This system provides the students' social and economic situations. These informations are used when necessary in problematic situations. In data mining, there are 3 basic issues: classification, clustering and regression. Classification is a perfect teaching approach when the students are divided into certain groups (1). Classification rules consist of two parts: education and test (2). The effectiveness of the classification could be tested with the rules on the test data. On the other hand, condensation is based on unattended learning because of not knowing the classes (3). In condensation, data mining expert view is needed for applying of this approach. Lastly, the regression could be used on the interpretation of the last variables in data mining (2). In literature, lots of studies have researched this topic previously. Automatic Interaction Sensor (CHAID) is developed for testing the secondary school students' performances (4). This study claims that some factors such as school type, settlement, parents, education, and organization play an important role in the students' in-class success. In the classification of the students' data, Cortez could be used, which enable Bayes method

(5). Binary data classification is applied to divide the data into 5 class. Moreover, regression is based on testing the algorithm's performances (5). The method classifies the students' performances via different data. B interest warehouse and data mining also include the different kind of data resources, deleting, data transformation, and periodical updating of the data. (6). Data mining is used in different areas of education. El Halees prefers the data mining in advanced studies on the evaluation of the students' performances (7). He aims to show how to increase the students' performances with usage of data mining. Moreover; Al- Radaideh adds that students' performances in university education could improve with the help of data mining techniques. On the other hand, Baradwajand Pal focuses on the evaluation of the students' performances by using decision making trees which is one of powerful techniques data mining (8). In education, data mining's different techniques such as decision making trees, rule making, Knn, and Naivela are actively used. These techniques help to classification, clustering, rule making (9-10). This study searches e-school system's effect on the students' success via classification of the data, which is a branch of data mining.

2. METHOD

In this study, the data is gathered from 756 students in Erbaa, Tokat. They are students in secondary education institutions. Table 1 shows the schools and the number of the students who attend the study.

Table 1. School and Student of Total

Schools	Student of Total
Science High School of Erbaa	281
Anatolian Teacher High School of Erbaa	222
Central Anatolian High School of Erbaa	253
Total	**756**

Table 1: Schools and the number of students As it is understood in the table, 756 students from the 3 different high schools participate to the study.

Table 2 shows the list of information gathered from the students and their parents.

Table 2. Note Classification by Average

Note	Group
X>=80	1
80<X>=60	2
X<60	3

Table 2: Classification according to their cumulative grades. Table 2 shows the all information about the students except their ID in e-school system. Some information about the students don't have access fully because of some hitches in the application. Moreover; the parents or students may not give the correct and full information about themselves. Being a martyrs or not has been deleted from the classification method because it is not completed by the students or the administration. While gathering information in this study, 10-11-12 class students' 9. Class cumulative grades have been used because 9. Class students have not finished the class yet. In Table 2, students are classified as to their year-end grades. When examining the students' cumulative grades, it is clear that the students don't take the grades under 40. So, the classification in the system is used in Table 3 instead of the data in Table 2. (Cumulative grade is equal to X).

In Table 2, 1 means that they are successful, 2 are equal to medium level and 3 are the ones who are unsuccessful. The methods in the classification go step by step.

Step 1: Gathering the data.

Step 2: The classification of the data according to the cumulative grades in Table 2.

Step 3: The writing of the students' classes, cumulative grades, and lack of the students' information in Table 3.

Step 4: Min-max. Normalization is used for the data.

$$Y_n = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Yn=Normalized data

X= To be normalized data

X min= The smallest data belonging to X.

Xmax=The biggest data belonging to X.

Table 3. Student Information

Scholarships	Yes-1 No-2	Education status of Mother	Primary School -1 Middle School-2 High School-3 Associate-4 Lisans-5 Master-6 Doktorate-7 Assistant Professor-8 Associate Professor-9 Professor-10
To go student by bus	Yes-1 No -2		
Regarding Child Protection	Yes-1 No -2		
Not relative by Martyr	Yes -1 No -2		
To stay of hostel	Yes -1 No -2		
Disabled status of father	Yes -1 No -2	Job of Mother	Tradesman-1 Private Sector-2 Government Official-3 Teacher-4 Self-employment-5 academician-6 Housewife-7 Retired-8
Education status of father	Primary School-1 Middle School-2 High School-3 Associate-4 Lisans-5 Master-6 Doktorate-7 Assistant Professor-8 Associate Professor-9 Professor-10		
		Mother is live or dead	Living-1 Dead-2
		Are mother and family stay together?	Yes -1 No -2
Job of father	Tradesman-1 Private Sector-2 Government Official-3 Teacher-4 Self-employment-5 academician-6 Soldier-7 Retired 8	Number of siblings	
		Status of home	Own-1 Rent-2 Private-3
		House warning	Natural gas-1 Central Heating-2 Stove-3
		Income status of family	High-1 Middle-2 Low-3
Father is living or dead	Living-1 Dead-2		
Are father and family stay together	Yes -1 No -2	Have you room different	Yes -1 No -2
Disabled status of Mother	Yes -1 No -2	Nonattendance	
		High School 1.Class(9. Class) CGPA	
Which class			

Value: to bring the interval [-1 1] To bring the interval (n>m) to the normalized data (mn) Ya=Yn (n-m) +m is applied. Ya is normalized data in [mn] interval.

$$Y_a = Y_n(n - m) + m$$

Step 5: The separation of the education and test data.

Step 6: The classification of the data with chosen restrictive. Matlab is used for classification process. Moreover; Knn classifier in Table 4 is used for classification. Table 4: selection of classifier. K value for Km classifier in Table 4 is chosen between 1-100.

Table 4. Classifier Selection

k-nearestneighbors (knn) classifier
NaiveBayes classifier
Back propagation
Euclidean Distance classifier
Mahalanobis Distance classifier

3. RESULTS

The data which is gathered with the study consist of 503 education and 253 test data. Education data are shown in Table 5 and test data in the Table 6.

Table 5. Data set of Education

	Data set of Education									
	School Classes			School Classes			School Classes			Total
	10	11	12	10	11	12	10	11	12	
	Class of			Class of			Class of			
Schools	3	2	1	3	2	1	3	2	1	
Anatolian Teacher H. S. Of Erbaa	-	3	1	40	35	29	20	5	15	148
Science High School of Erbaa	2	1	10	16	45	39	46	15	13	187
Central Anatolian H.S. of Erbaa	2	25	11	45	49	29	5	2	-	168
									G. Total	**503**

Table 6. Data set of Testing

	Data set of Testing									
	School Classes			School Classes			School Classes			Total
	10	11	12	10	11	12	10	11	12	
	Class of			Class of			Class of			
Schools	3	2	1	3	2	1	3	2	1	
Anatolian Teacher H. S. Of Erbaa	-	1	1	21	17	15	10	2	7	74
Science High School of Erbaa	1	1	5	8	22	20	23	8	6	94
Central Anatolian H.S. of Erbaa	1	12	6	24	24	14	3	1	-	85
									G. Total	**253**

When analyzing Table 5 and Table 6, 10, 11, 12 class students in different schools use it with a % 66 rate. The rest is test data. While making this grouping, it has been given importance to the equal distribution of the schools, classrooms and students' numbers. Therefore; it could be said that education and test data have a homogeneous form. Moreover; education and test data are recognized

with their all features. Therefore; more reliable results could be obtained. Figure 1 shows K' different rates which is one of Knn classification method.
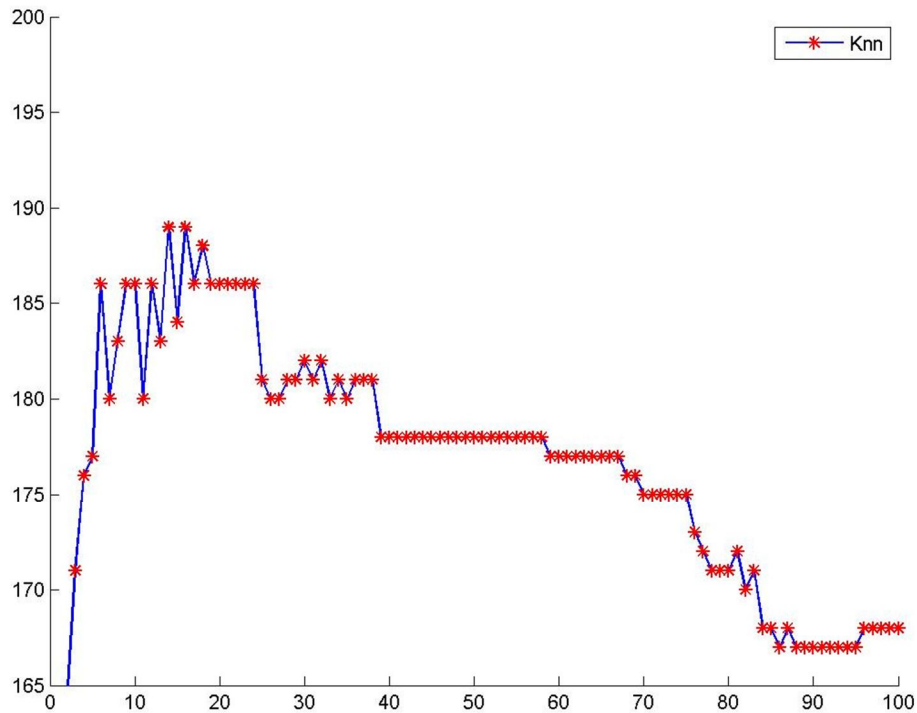

Figure 1. Located value for KNN classifier

When examining Figure 1, the highest rate in Knn classifier are K= 14 and K=16 , T=189. Success rate of the classifiers are shown in Table 7.

Table 7. According to the statement of success classifier

Sorter	TRUE	FALSE	SUCCESS
Knn, for k=15 ve k=16	189	64	75 %
NaiveBayes classifier	186	67	74 %
Classification based on Euclidean distance	177	76	70 %
Classification based on the Mahalanobis distance	176	77	70 %
Back Propagation (Artificial Neural Networks)	185	68	73 %

Table 7: Success table in terms of classifiers. In Table 7, the classifiers' performances are shown. The best classification is made by Knn classifier. However; Naive Bayes classifier looks like a successful result with 186 rates. Naïve Bayes classifier works better when chosen a more extensive interval, not in (-1, 3) interval. Figure 2 shows the classifiers' performances with bar graphics.
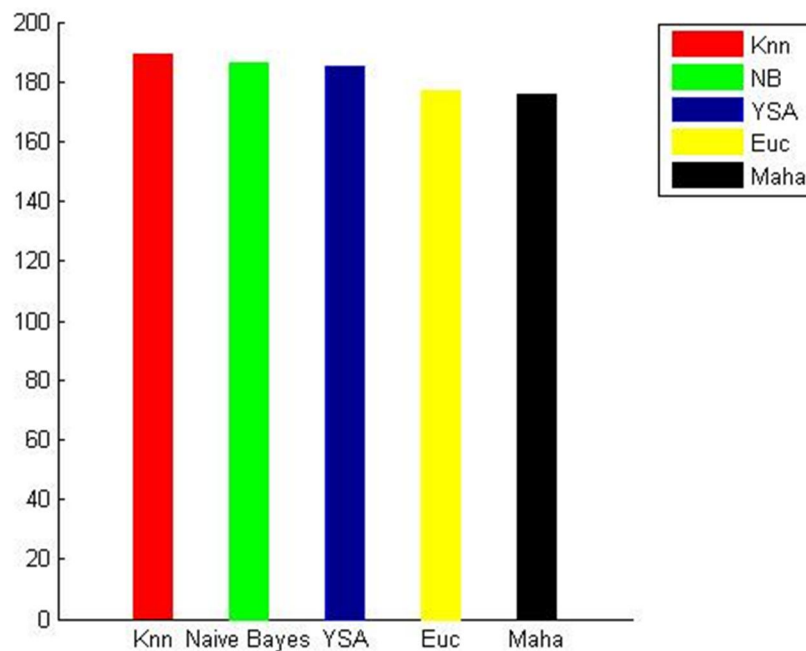
Figure 2. Classifiers Performance

When viewing Figure 2, the performances are line up as Knn, BN, YSA, Eve and Maha. Moreover, Knn rates are taken into account in this study.

4. RESULTS AND DISCUSSION

As it is seen in the research, there is a strong relation between the students' social and economic information in the e-school system and the students' success. The e-school system which is actively applied in Turkey enables to suppose the students' success. The e-school system gain importance in Turkey day by day. However; it could be supported with more advanced methods. As an example, more social areas which are identified to students' success with correlations could be added to the e-school system. Moreover; the existing areas may be developed and by this way, classifiers could work better. To sum up, the information in the e- school system are important for predicting the students' success in todays' world. The e- school system enables the educators to be more aware of their students' proficiency if they use it before. By this way, they prepare their classes in terms of their academicals success. Moreover; e-school system could enable more qualified education processes with the help of supporting it with different activities.

5. SUGGESTIONS

As it is studied in the research, e-school system has lots of advantages for both the students and the educators. However; some suggestions could be made for it. For example; the students' information don't exist fully in the e-school system. So, it may cause some hitches on the classifiers. If more information are located in the system, the application's success become more than %75. Moreover, classification methods could be applied with additional teaching programmer and different supportive methods. A similar study may be applied for the other classifiers. Lastly, the validity of the findings could be searched.

6. REFERENCES

1- A. Merceronand K. Yacef, "Educational Data Mining: a Case Study," InC. Looi; G. McCalla; B. Bredeweg; J. Breuker, editor, Proceedings of the 12th International Conference on Artificial Intelligence in Education AIED, pp. 467–474. Amsterdam, IOS Press, 2005.

2- I. H. Wittenand E. Frank, "Data mining: Practical Machine Learning Tools and Techniques," 2nd ed, Morgan-Kaufman Series of Data Management Systems, Elsevier, 2005.

3. J. Han and M. Kamber, "Data Mining: ConceptsandTechniques," (Morgan-Kaufman Series of Data Management Systems). San Diego: Academic Press, 2001.

4. M. Ramaswamiet al., "A CHAID Based Performance Prediction Model in Educational Data Mining," IJCSI International Journal of Com. Sci., vol. 7, Issue 1, no. 1, January 2010.

5. P. Cortezand A. Silva, "Using Data Mining To Predict Secondary School Student Performance," In *EUROSIS*, A. Britoand J. Teixeira (Eds.), pp. 5-12, 2008.

6. Q. A. AI-Radaideh, E. M. AI-Shawakfa, and M. I. AI-Najjar, "Mining Student Data using Decision Trees,," International Arab Conference on Information Technology (ACIT 2006), Yarmouk University, Jordan, 2006.

7. El-Halees, A., "Mining Students Data to Analyze Learning Behavior: A Case Study", International Arab Conference of Information Technology (ACIT2008), Conference Proceedings, University of Sfax, Tunisia, Dec 15- 18, 2008.

8. Baradwaj, B. and Pal, S., "Mining Educational Data to Analyze Students Performance", International Journal of Advanced Computer Science and Applications, vol. 2, no. 6, pp. 63-69, 2011.

9. Romero, C. and Ventura, S., "Educational data Mining: A Survey from 1995 to 2005", Expert Systems with Applications (33), pp. 135-146, 2007.

10. Romero, C. , Ventura, S. And Garcia, E., "Data mining in course management systems: Moodle case study and tutorial", Computers&Education, vol. 51, no. 1, pp. 368-384, 2008.