Cumhuriyet Science Journal

e-ISSN: 2587-246X
ISSN: 2587-2680

Cumhuriyet Sci. J., 41(2) (2020) 467-471
http://dx.doi.org/10.17776/csj.410709

A matching model to measure compliance between department and student

***Hidayet TAKCI*[1] *Kali GÜRKAHRAMAN*[1] *Emre ÜNSAL*[2,*] *Ahmet Fırat YELKUVAN*[1]**

[1]*Department of Computer Engineering, Faculty of Engineering, Sivas Cumhuriyet University, 58140, Sivas/ TURKEY*

[2]*Department of Software Engineering, Faculty of Technology, Sivas Cumhuriyet University, 58140, Sivas/ TURKEY*

Abstract

The aim of all education systems is to train students who are equipped with knowledge. In that case, that student is able to determine the most suitable profession for him/her success in education and career that are related to this profession will be higher. Studies done up to this day have been focused on finding out the factors affecting the career choice of the student, but they have not suggested any method for determining the most suitable procession. It is not possible to obtain satisfying results from a system that does not lead students to appropriate higher education departments. In this context, a student- department matching system is proposed which aims to increase the success of the education systems in our study. The department of computer engineering was dealt with as a sample department and the proposed study was examined to determine whether a student was suitable for computer engineering or. The required data was obtained with the help of the questionnaire, and then a model of successful and unsuccessful students was created. Data mining algorithms such as C4.5, C-SVC, MLP, and Naïve Bayes are used during the test of the generated model. The best result was obtained by the C-SVC algorithm and the second best result by Naive Bayes. The lowest error rate achieved was 0.2700 and the highest accurate recognition rate was 73.00%.

Article info

History:
Received:29.03.2018
Accepted:22.04.2020
Keywords:
Matching system,
Data mining,
Predictive Models.

1. Introduction

Individuals begin to receive education from the moment that they born. The education, which starts at the family passes through the various steps and moves towards the final destination. At these steps, the transition from a lower to a higher education institution often takes place based on the test score rankings. However, evaluation and transposition based on test score ranking are far from being the right approach. Although the test score is an important factor, it cannot provide the complete information on which educational institution is better individually.

Nowadays, it is a necessity to take into account additional factors, because the test scoring based approaches only evaluate cognitive skills. For instance, it is possible that a student will be unhappy first and then unsuccessfully when he or she goes to a school with only his test score. Although the student completes the school with the family enforcement or even have a business owner in that area, the unhappiness continues. The schools that have gone incorrect and the profession obtained will cause low workforce in the future. This situation may be one of the biggest but most overlooked problems today. Even if it is not possible to overcome this problem, it may be a better solution to take the other factors apart from the test scores.

In this study, a matching model is presented for students to choose the right academic education or profession. The aim of the model is to fulfill the match between the students and the department based on the ability, expectations, and interests. The reason for working with these categories is that a student cannot succeed if he or she is not interested or talented in a profession. In addition, it is also an important factor for a department/profession to meet student expectations.

2. Previous Studies

The proposal confessed in this study is novel; therefore, there are not any previous studies that will overlap with the study. For this reason, similar studies will be considered as previous studies and the differences will be revealed.

The studies have been dealt with the title of factors that affect the selection of professions until now. In 2014, Çelik and Üzmez [1] presented a comprehensive study on this subject. In the related study, the factors such as

*Corresponding author. *Email address: eunsal@cumhuriyet.edu.tr* teaching, medicine, nursing, accounting, information technology, textile ready-to-wear program, tourism management and many fields were studied. As a result of this study, the factors affecting the selection of profession include family approval, social benefit, social expectations, career opportunity, salary, job security and interest. In addition, it would not be misleading to say that one of the factors influencing students' career choice is guidance units. One of the valuable studies about this subject is the study related to random departments by Sarıkaya and Korshid [2] in 2009. In this study, it was revealed that factors such as interest, helplessness, occupational advantage, recommendations, family factors, grades and personal characteristics had an effect on career choice. The last study about the factors influencing the choice of profession was presented in 2011 by Sathapornvajana and Watanapa [3]. This study was conducted on the students who selected areas related to information technologies. The key factors who choose this area are self-sufficiency, self-criticism, self-consciousness, social outlook, career, reputation, ease of professionalism and innovation.

Although there are many factors which affecting the choice of profession are handled one by one; it can be categorized as skills, expectations, and interests. Therefore, the sub-elements of these categories will be revealed, and data collection and data analysis with the correct tools will be contributed solving the problem. Surveys record what people say, and sometimes the questionnaire participant can give misleading or incorrect information. Moreover, the use of psychometric tests will be valuable. Psychometry is a science that evaluates the characteristics and qualities of people. Psychometric tests measure the relevance of person's abilities and skills, abilities he or she needs to perform required by the job [4]. The goal is to guide people to the profession in accordance with their abilities and qualifications. Psychometric tests have been used in the recruitment of candidates, evaluation of working personnel, the decision of promotion, analysis studies and career counseling.

In this work, data mining techniques will be used for data analysis. Data mining is an interdisciplinary way of discovering new and potentially useful information from huge amounts of data [5] which is applied in many sectors such as banking, finance, marketing, insurance, and medicine, also includes appropriate techniques for the education field. For example, students can be divided into categories by using the classifier [6] and prediction models can show whether students will pass a course or fail [7] by using prediction models. Moreover, students could be classified into similar groups with the help of clustering analysis [8]. Additionally, in order to increase student participation, sample works have been done, recently such as merging educational software and student information [9].

3. Materials and Methods

In this study, a machine learning based system design will be proposed, in order to place the students in the correct educational departments. One side of the system will contain student information, the other side will contain section information, and the black box of the system will measure the compatibility between the two sides. The output of this compatibility analysis can be labeled as "Appropriate" or "Not Appropriate", so that if the compatibility between the student and the department is positive, then the output of the system labeled as “Appropriate”, otherwise it labeled as “Not Appropriate”. The proposed matching system design is demonstrated in Figure 1.

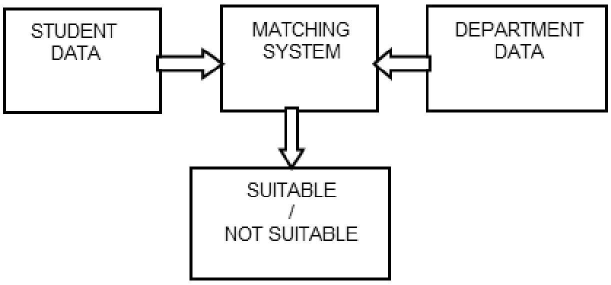

Figure 1. Proposed matching system diagram.

The proposed system will be a system based on predictive models. Therefore, first, the system will be trained with training data, and then the trained system will predict the results for new examples. In this context, some studies have been carried out in accordance with the stages of data mining, and the parameters of the predictive model have been revealed. The data source of the model to be created will be the data obtained from the survey applied to the students who study at the department of computer engineering.

The matches between student information and departmental information will be performed with the help of a predictive model. Before processing the data obtained from the survey, a pre-process is required. In this context, in the first step, the missing, inconsistent, meaningless data in the dataset should be deleted. Subsequently, the survey data will be digitized and transformed, so that data mining methods can be performed. The numerical data which is obtained from the survey will be presented in the vector space model [10], and the data will be presented as a vector for each student. In the data mining phase, a part of the data obtained from the survey will be defined as training data and the other part used as test data. Models will be built for Successful and Unsuccessful students, through the part of the training data. In the test phase, students whose class is unknown will be tested whether successful or unsuccessful.

If we look at the problem regardless of the data mining process, the performed operations can be expressed simply as follows. The data of the students, that currently studying in the computer engineering department, are collected from the survey. The class label of the students is given based on the transcript information so that the model can be built for the successful student as well as for the unsuccessful student. After that, the survey is applied to other students and it is measured using the previously collected data that the new students will or will not be successful in the computer engineering department. As a result, if the survey profile of a new student is similar to the profiles of the students who are now successfully studying in the department, it can be said that the new student is appropriate, otherwise, the new student is not appropriate.

4. Experimental Work

This section includes studies that we have undertaken within the scope of the proposed system model for students to be placed in the appropriate department.

4.1. Data set

The proposed model test was performed on the data collected by the researchers. The survey was chosen as a measurement tool for data collection, and 210 students participated. A part of this survey was taken from the study of the academic self-concept scale presented by Kuzgun [11]. There are four main categories and several sub-categories of the survey study. The details about the survey are given in Table 1.

Table 1. Distribution of The Applied Survey Questions

Main Category	Sub Categories	Question Count
Ability	Literature, science, shape space and hand-eye coordination	16
Expectations	Professional expectations	14
Demographic data	Demographic and registration data	10
Interest	Science, mechanic, hobbies, social, welfare, foreign language etc.	75

We have used the likert scale corresponding to five different values which are between 1 and 5. These values are strongly disagree, disagree, neutral, agree, strongly agree. Besides, the answers of some questions are in the case of 1 and 0.

The developed survey was applied to the students who are studying in Cumhuriyet University, Computer Engineering Department in two stages as pre-application and the main application. Sub-dimensions of the scales were determined by applying Exploratory Factor Analysis (EFA) to the data obtained from the pre-application. In the main application, Confirmatory Factor Analysis (CFA) was used to verify both the outputs of the preliminary application and the suitability of the data to the sample group. Students who participated in the survey were selected from 2, 3 and 4 classes. The distribution of the students according to their classes, and label distributions are given in Table 2.

Table 2. Data Characteristics

Data Source	Students of Computer Engineering Department of Cumhuriyet University
Sample Count	210
Record Distribution	76 (4. Class), 76 (3. Class) and 58 (2. Class)
Class Distribution	Successful (141), Unsuccessful (69)

At the end of the main application, 4 of the 210 students survey records were eliminated because of a determined problem in the survey records. Therefore, the experimental study can be done only with 206 survey records of the students.

4.2. Experimental design

For the accordance of predictive modeling, two different labeling studies have been carried out on collected data. In the first case, transcript grades of the students are directly taken into account and, the grade of the students 2.0 and above are labeled Successful the rest of them labeled Unsuccessful. In the second case, not only the transcript grades but also the opinions of the instructors are also taken into account. Successful and unsuccessful labels were obtained from the opinions of the 5 lecturers who entered the course of the students who participated in the survey and the labeling was made accordingly. After that, decision trees (C4.5), support vector machines (C-SVC), artificial neural networks (MLP) and naive Bayes continuous classification algorithms, which are frequently used in past studies, were used to test our recommendation. The success of the algorithms was tested using the Tanagra [12] named machine learning software for the model tests.

Each algorithm was evaluated with optimum parameters and 10-fold cross-validation method was used for model evaluation.

4.3. Experiment results

In the classification experiments to be performed on the data, the results are presented as error rates. In the first step, the obtained results on the labeled data based on the transcript grades represented in terms of all, ability Test, expectancy test, and interests test rates. The test results are given in Table 3.

Table 3. Classification errors from labeling according to transcript grades

Algorithm	All	Ability Test	Expectancy test	Interests Test
C4.5	0.5450	0.4750	0.4550	0.5700
C-SVC	0.4900	0.4300	0.5200	0.4500
MLP	0.4800	0.5050	0.5350	0.5150
Naive Bayes	0.5050	0.5200	0.5750	0.5050

However, in the second stage, the results of the classification errors are obtained from the opinions of the instructors are examined as given in Table 4.

Table 4. Classification errors obtained from the opinions of the instructors

Algorithm	All	Ability Test	Expectancy test	Interests Test
C4.5	0.4550	0.4000	0.4050	0.4400
C-SVC	0.4050	0.2700	0.3350	0.4400
MLP	0.4250	0.3000	0.4300	0.4650
Naive Bayes	0.3650	0.3100	0.3800	0.4200

When the C4.5 algorithm gave an error rate of 0.5450 in the experiments with the questions of the survey completely, the ability test obtained with 16 questions gave 0.4750 error rate, and a better result was obtained. Similarly, the values dropped from 0.4900 to 0.4300 for the C-SVC algorithm. Unlike the other two methods, the error rate of MLP and Naive Bayes algorithms is increased. The error rate of 0.4800 for MLP has increased to 0.5050, and, it rose from 0.5050 to 0.5200 for Naive Bayes.

The second subcategory is the category of occupational expectations which is measured by the Expectancy test. In this category, the C4.5 algorithm yielded better results than both of the previous test results. However, the C-SVC algorithm, in contrast to the other two experiments, gave the most unsuccessful result in the category of occupational expectations. Likewise, the MLP and Naive Bayes algorithm also yielded unsuccessful results in this category. As a result, occupational expectations carry valuable information for the decision tree classifier but give lower results for others.

Finally, experiments based on areas of interest were carried out and the results are obtained. The results of the Interest test gave values close to the error results as in the Expectancy test. In addition, another remarkable detail is that the C-SVC algorithm yields similar results to the ability test. This situation can be interpreted as a relationship between interests and abilities.

Generally, it can be said that the data set of the second experiment gave the better test results when compared with the first one.

5. Discussion

The transition of the current education system to a higher education institution by exam makes the exam scores significant. While exam scores provide information about the students relatively, in fact, more information can be obtained by assessing the skills. A student who is more intelligent but has less financial facility may not win good department, in contrast, less intelligent, have better financial status, and the hardworking student may win that department. This is an anomaly and does not fit into equal opportunity. In that vein, there are many departments that can be reached with an exam score when the student makes a preference with an exam score alone. For example, a student may prefer computer engineering or mine engineering in terms of test scores. However, there are very different professions than each other, although both departments are found in engineering faculty. As it can be seen from the example, evaluation of the students by exam score alone is not enough; in addition to that, there should be a selection process with more factors. So that, there is a need to collect additional student information and analyze this data correctly for a better evaluation.

For this aim, 210 students from Computer Engineering Department of Cumhuriyet University is selected and asked to fill a survey consist of 115 questions, in order to create a dataset in this study. A matching system based on classification algorithms was established and experiments were carried out with the collected dataset. The class label for student data is given first by reference to the transcript grade. Students with transcript grade 2.00 and above were tagged successfully the rest of them tagged unsuccessfully, and some results were obtained. After that, a labeling was made according to the opinions of the instructors for the students who participated in the survey, and the experiments were repeated with this data. As a result of the repeated experiments, it was observed that the second data set gave better results. This is because the transcript grades are not enough alone for extracting a good student profile.

The success rate being around 75% is not a value that falls behind the literature. Since the problem we are trying to solve is based on many parameters and the training data obtained is not so much, the average of success are behind the literature. Another reason for the success value to remain relatively low is that it was obtained without pretreatment on the data.

6. Conclusion

The existing education system evaluates the students by only sorting with their exam scores, and ignores the personals skills of the students. This issue causes an anomaly and does not fit into equal opportunity. The university student selection system should be needed to revise with more factors. In this study, a student-department matching system is proposed which aims to increase the success of the existing education systems. For this aim, computer engineering department was selected as the pilot department and the data was collected with a measuring instrument composed of 115 questions on 210 students who were studying in the Department of Computer Engineering of Cumhuriyet University. The proposed model based on predictive models. Two datasets are created from the data obtained from the survey. The first experiment includes only the transcript grades of the students. The second data set consist of both grades and the opinions of the instructors and the experiments were repeated with this data. As a result of the repeated experiments, it was observed that the second data set gave better results. The best result was obtained by the C-SVC algorithm and the second best result by Naive Bayes. The lowest error rate achieved was 0.2700 and the highest accurate recognition rate was 73.00%.

Future studies may investigate the success of the students by adding more factors into the experiment and test with different data mining algorithms.

Acknowledgment

This study was supported by funding from the Scientific and Technological Research Council of Turkey (TÜBİTAK). (Project Number: 115E837).

Conflicts of interest

There is no conflict of interest.

References

[1] Çelik N, Üzmez U., Evaluation of university students' affecting factors choice of profession. *Electronic Journal of Occupational Improvement and Research (EJOIR)*, 2(1) (2014) 94-105.

[2] Sarıkaya T., Khorshid L., Üniversite öğrencilerinin meslek seçimini etkileyen etmenlerin incelenmesi: üniversite öğrencilerinin meslek seçimi. *Journal of Turkish Educational Sciences*, 7(2) (2009) 393-423.

[3] Sathapornvajana S., Watanapa B., Factors affecting student's intention to choose IT program. *Procedia Computer Science*, 13, (2012), 60-67.

[4] Çoban, A., Psikometrik testler: Available at: http://www.adnancoban.com.tr/psikometrik_testler.html. Retrieved: July 2015.

[5] Witten I. H., Frank E., Data mining: practical machine learning tools and techniques with Java implementations. 3rd ed. San Francisco CA: Morgan Kaufmann (2002) 76-77.

[6] Cha H., Kim Y. S., Park. S. H., Yoon T., Jung Y., Lee J. H., Learning styles diagnosis based on user interface behaviors for the customization of learning interfaces in an intelligent tutoring system. *In Proceedings of the 8th International Conference on Intelligent Tutoring Systems*, (2006) 513-524.

[7] Hämäläinen W., Vinni M., Comparison of machine learning methods for intelligent tutoring systems. *In International Conference on Intelligent Tutoring Systems*, Springer, Berlin (2006) 525-534.

[8] Perera D., Kay J., Koprinska I., Yacef K., Zaiane O. R., Clustering and sequential pattern mining of online collaborative learning data. *IEEE Transactions on Knowledge and Data Engineering*, 21(6) (2009) 759-772.

[9] Pardos Z. A., Gowda S. M., Baker S.J.d R., Heffernan N. T., The sum is greater than the parts: ensembling models of student knowledge in educational software. *ACM SIGKDD Explorations Newsletter*, 13(2) (2011) 37-44.

[10] Salton G., Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer. Addison-Wesley, (1989).

[11] Kuzgun Y., Mesleki ve Teknik Öğretim Kurumları ve Meslekler Rehberi. National Education Press, Istanbul, (2006).

[12] Rakotomalala R., TANAGRA: a free software for research and academic purposes. *Proceedings of EGC'2005, RNTI-E-3*, 2 (2005) 697-702.