Discovery of Course Success Using Unsupervised Machine Learning Algorithms

Emre CAM [1], Muhammet Esat OZDAG [2]

http://dx.doi.org/10.17220/mojet.2021.9.1.242

[1] emre.cam@gop.edu.tr,
Tokat Gaziosmanpasa University,
Department of Computer Technology,
Turkey

[2]muhammetesat.ozdag@gop.edu.tr,
Tokat Gaziosmanpasa University,
Department of Computer Technology,
Turkey

ABSTRACT

This study aims at finding out students' course success in vocational courses of computer and instructional technologies department by means of machine learning algorithms. In the scope of the study, a dataset was formed with demographic information and exam scores obtained from the students studying in the Department of Computer Education and Instructional Technology at Gaziosmanpasa University. 127 students, who took the courses of Programming Languages I and Programming Languages II, participated in the study. Model that was suggested in the study was implemented using open source coded Keras library. Students were split into clusters by K-means and Deep Embedded Clustering algorithms which are unsupervised machine learning algorithms. Effect of the attributes that enabled clustering was identified by Kruskal Wallis test. With this study, a model that helps educators and instructional designers build skills for predicting, assures discovering success patterns through data mining and facilitates assisting in the stages of lesson planning was proposed.

Keywords: *Machine Learning, K-means, Deep Embedded Clustering, Educational Data Mining, Course Success*

INTRODUCTION

Researchers carried on artificial intelligence are already used in many fields. They also have been quite prominent in the field of education. Artificial intelligence in the field of education has been creating various discussion topics varying from determining teaching methods and strategies to motivation. The term of artifical intelligence was initally stated by John McCarthy at Dartmouth Conference in 1950. Though it has many definitions it is generally referred as carrying out highly complex cognitive processes like reasoning, inferring, generalizing and learning from past experiences which are consided as the traits that are peculiar to human beings (Kazu & Özdemir, 2009). AI, with the most general definition is defined as the attempt of creating computers or machines which are generally associated with such cognitive tasks as learning and problem solving (Baker & Smith, 2019). When the literature is analyzed, artifical intelligence is seen as a term which is used in defining from machine learning to deep learning or from educational data mining to learning analytics rather than defining a specific technology.

The concept that combines artifical intelligence with educational processes emerge as learning analytics. The concept of learning analytics was firstly coined by Siemens (2010) as using analysis to discover knowledge and social connections and also to make deduction and making use of smart data and the data produced by the learner to offer suggestions. With learning analytics, data belonging to students are analyzed within the framework of internal and external opportunities and restrictions and they are supported by educational theories. There are various tools such as educational theories, algorithms and technologies in learning analytics. Algorithms and technologies used to convert the data in education into information are generally seen within the context of artifical intelligence and educational data mining. Educational data mining is defined as *“is an emerging discipline, concerned with developing methods for exploring the unique and increasingly large-scale data that come from educational settings and using those methods to better understand students, and the settings which they learn in.”* (International Educational Data Mining Society, 2020). Methods that are widely used in the field of educational data mining can be stated as regression, clustering, classifying, association rules (Romero & Ventura, 2010). Educational data mining can be used as an efficient tool for such practices as defining learner characteristics, behaviors and models, and enhance learning processes, offering needs-based services, developing prediction models and accelerating decision support processes.

Educational data mining uses the field of machine learning like artifical intelligence. Machine learning is an artificial intelligence field that allows computers to predict the events in the future and to model using experiences gained through earlier information. It can also be defined as computers making decisions about potential similar events by learning the information and experiences about an event and producing solutions to the problems. There is a direct relation between machine learning and data mining (Figure 1) (Alpaydın, 2004). Application of machine learning techniques in large databases is data mining. Machine learning is in the implementation phase of the data mining process. In this process, a selected machine learning technique is applied on the dataset and results are gathered. Machine learning is not only a technique applied on the data but also it is an artifical intelligence field. Data mining deals with the gathered information and the evaluation of it. On the other hand, machine learning deals with the techniques that allow the extraction of this information and enabling the computers using those techniques to develop themselves. The greatest difference between these methods is as follows, while machine learning deals with how to best extract the predictions and definitions with a high performance, data mining deals with the information gathered (Dalyan, 2006).

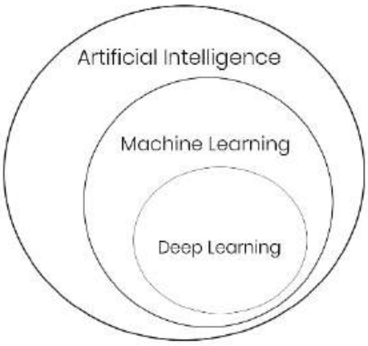


Figure 1. Relationship between Artificial Intelligence, Machine Learning and Deep Learning

Researches carried out recently have focused on deep learning which is a part of machine learning, one of the fields of human and computer interaction. Deep learning that has a multi-layered structure when compared to the method of machine learning has gained a great deal of attraction by being inspired by the functioning of human brain (Koitka & Friedrich, 2016). Deep learning is a subtype of machine learning. The most important feature that separates machine learning from deep learning are the layers in the architecture of Artificial Neural Networks. As learning takes place in deep layers it is identified as deep learning.

Deep learning models in the application fields of artificial intelligence are generally used in many different fields such as image recognition, object reception/identification, voice recognition, natural language processing and genetics. Though use of deep learning is limited in learning environments, it can enable us to make several predictions about learners and teachers when used in educational sense.

Robinson et al. (2016) carried out a poll on learners who would start HarvardX online course consisting of questions such as their motivation for the course, if they had the real intention of finishing it, earlier leaning experiences and demographic information along with an open ended question about their expectations from the online course. Findings obtained from the answers given to the open ended questions analysed by natural language process showed that a machine learning prediction model was successful at predicting which learners would complete the online course.

In a study carried out in England Open University, Hussain, Zhu, Zhang & Abidi (2018) used machine learning algorithms to evaluate the effect of learner participant on performance and thus to identify the learners who performed low participation. They stated that when the model they created was integrated in virtual learning, learners with low participation rate could be detected and special precautions could be taken for the ones who are under risk just before final exams.

In the study named “Evaluation of Students’ Performance and Learning Efficiency Based of ANFIS”, Yusof, Zin, Yassin & Samsuri (2009) focused on the prediction of learners performance in Programming Technique course. Such parameters as learners’ average scores, how much time they spent, if they needed assistance or not were accepted as entry parameters. The study that was previously carried out by just fuzzy logic rules was set as ANFIS model by including artificial neural networks.

In the study that Lykourentzou et al. (2009) carried out in 2009, a learner’s quitting a course or school was predicted by multiple genetic algorithm method through evaluating the results of 3 different methods. In the study test results, project evaluations and demographic data were made use of. In the other study that Vandamme, Meskens & Superby (2007) carried out in 2007, which students would fail the class or quit school was predicted by classifying students into such risk groups as low, medium, high by using the data about students’ demographic information along with their socio-economic and academic background.

Artificial intelligence applications that will directly support the special needs of all the participants taking part in educational processes must be used in education systems. Learners need changes in their learning processes, sources and study patterns in order to enhance their academic performance. Meaningful information obtained by data mining techniques along with the information such as use case of learners, course information, academic information can offer teachers and administrators opportunities to plan and design education system. Thus a whole education system develops with its components. As Yükseltürk, Özekeş & Türel (2014) stated, use of educational data mining techniques in education can offer teachers and researchers opportunities to obtain interesting and useful information on relations among the variables of large datasets. In this study, a model that helps educators and instructional designers build skills for forecasting (predicting), assures discovering success patterns through educational data mining and facilitates assisting in the stages of lesson planning was proposed.

METHODOLOGY

First part of the research discusses the machine learning approaches used for discovering the relation between student success and demographic attributes. In the following parts of the study, K-means clustering, Auto-Encoder (AE) and Neural Network (NN) based Deep Embedded Clustering (DEC) methods were discussed. This part of the study identifies parameter adaptations of housekeeping on dataset, identification of number of clusters and K-means, DEC approaches and ends with processing algorithm pertaining to the method. In the second part, information about the answers given by the students to the questionnaire questions is presented.

K-means

K-means, initially proposed by MacQueen in 1967, is an unsupervised learning algorithm used frequently in data mining and also in clustering large data clusters (Na, Xumin & Yong, 2010). The algorithm is composed of two different stages. First stage is random selection of the centre as much as the number k that was previously identified. Second stage is taking every member data point to the closest centre. Euclidean distance is widely used for measuring distance. When all data points are included into one single data centre, early clusters may come up. The centre point is recalculated and all the points in the dataset are again taken to new centres through a recurring process. This iterative process is repeated until the minimum point is detected.

$$E = \sum_{i=1}^{k} \sum_{x \in C_i} |x - x_i|^2$$

In the equation 3, x_i symbolises the center alignment of C_i cluster, and x symbolises the target object. E is the total for square error of all points. Criteria function, measuring the distance and used for finding the distance of each element of the set and their distance to the center is Euclidean distance. (x_i, y_i) whichis the Euclidean distance between the vectors $x = (x_1, x_2, .. x._i)$ and $y = (y_1, y_2, .. y._i)$is calculated using the equation 4.

$$d(x_i, y_i) = \sqrt{[\sum_{i=1}^{n}(x_i - y_i)^2]} \; (4)$$

Auto-Encoder

AE is an unsupervised learning type and is known as feed forward neural network (Dong, Liao, Liu & Kuang, 2018). It tries to set the output values equal to the width of original entry values. The general model is composed of three basic stages. (1) encoder : a feed forward structure consisting of weight matrix and bias. (2) activation : a nonlinear function that transforms coded coefficients into range of [0-1]. (3) decoder : a structure that performs back propagation. AE aims at excluding significant attributes by decreasing the entry width and thus avoiding the problem of overlifting.

Deep Embedded Clustering (DEC)

Clustering is a substantial research topic for fields of machine learning and data mining. Developments in deep neural networks (DNN) has increased the focus on use of DNNs in clustering problems (Ren et al., 2018). Xie and his friends, in their study in 2016, proposed Deep Embedded Clustering (DEC) as an algorithm that learns representation of properties and clustering simultaneously (Xie, Girshick & Farhadi, 2016).

Let’s discuss a cluster X that consists of n number of points and waiting to be split into k number of clusters. While x_i symbolises each element of Xset, it also symbolises the centroids of the clusters up to u_j, $j, \dots k$. DEC, rather than clustering directly on X space, suggests initially transforming data $f_\theta : X \rightarrow Z$ through nonlinear mapping. At this point, θ is the learnable parameter and Z is covered attribute area.

DEC clusters f_θ'ı , which is a DNN parameter mapping data points of Z on space Z in the centre of cluster k by simultaneously optimizing it. DEC has two phases. (1) : It starts parameters via deep AE. (2) : optimizes the parameters which means that it performs clustering.

Data Preparation

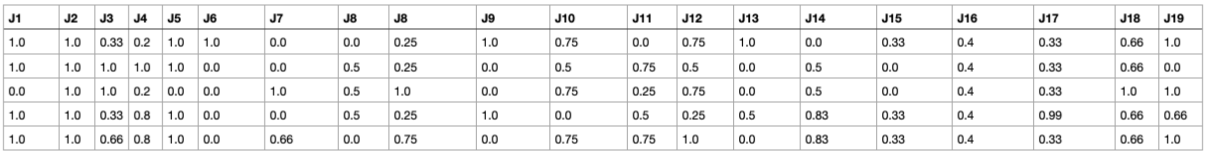
When attribute values in datasets were analysed, values consisting of categorical responses were demoted to numeric values from 1 to 7 as density clustering methods could operate on numeric values. Clustering algorithms such as K-means’ using Euclidean distance in clustering processes makes the criteria for attribute values’ clustering success important. Therefore, all attribute values ranging from [0,1] were rescaled. Scaling formula is given in Equation 1.

$$x_{i,j} = \frac{x_{i,j} - x_j^{min}}{x_j^{max} - x_j^{min}} \quad (1)$$

In Equation 1, j symbolizes attributes and i symbolizes measurement index. In Table 1, first 5 measurement values of dataset scaled in the range of [0-1] are showed. For lost and corrupted measurement values in the study, totality of data was ensured by calculating the average of total values. In Equation 2, j symbolizes attribute and , $x_{i,j}$ symbolizes measurement value. Of the attributes, area of Id was removed as it made measurements unique and affected clustering results.

$$x^{missing} = \frac{\sum_{i=1}^{127} x_{i,j}}{127} \quad (2)$$

Table 1. Normalised first 5 measurement values



J1	J2	J3	J4	J5	J6	J7	J8	J8	J9	J10	J11	J12	J13	J14	J15	J16	J17	J18	J19
1.0	1.0	0.33	0.2	1.0	1.0	0.0	0.0	0.25	1.0	0.75	0.0	0.75	1.0	0.0	0.33	0.4	0.33	0.66	1.0
1.0	1.0	1.0	1.0	1.0	0.0	0.0	0.5	0.25	0.0	0.5	0.75	0.5	0.0	0.5	0.0	0.4	0.33	0.66	0.0
0.0	1.0	1.0	0.2	0.0	0.0	1.0	0.5	1.0	0.0	0.75	0.25	0.75	0.0	0.5	0.0	0.4	0.33	1.0	1.0
1.0	1.0	0.33	0.8	1.0	0.0	0.0	0.5	0.25	1.0	0.0	0.5	0.25	0.5	0.83	0.33	0.4	0.99	0.66	0.66
1.0	1.0	0.66	0.8	1.0	0.0	0.66	0.0	0.75	0.0	0.75	0.75	1.0	0.0	0.83	0.33	0.4	0.33	0.66	1.0

Determining the number of clusters

A DEC model comes out with the combination of AE and K-means machine learning methods (Xie et al., 2016). The first and the most important parameter is the identification of number k that refers to the number of clusters. In our study, identification of k number was predicted by calculating (Wang et al., 2017) silhouette value, one of the most popular and efficient methods. Silhouette value is figuring a measurement's concordance to its own cluster with a value between -1 and +1 when compared to other clusters. A high value shows that the measurement compromises with its own clusters but displays a bad match with adjacent clusters. Silhouette values that come up when values between 2 and 9 are taken for k for K-means algorithm are given on Table 2. When Table 2 was analysed, best silhouette values were found as 0.1151 and 0.122, and value 3 was selected for k of the best two values in our study.

Table 2. Silhouette values

k	K-means silhouette value
2	0.12282
3	0.11516
4	0.09622
5	0.10219
6	0.08557
7	0.09291
8	0.09416
9	0.08694

Clustering approach in this study was implemented through open source access Keras library by considering the steps that Xie et al. (2016) stated in their studies. Steps that form the algorithm are showed in Figure 2 (Chollet, 2017).

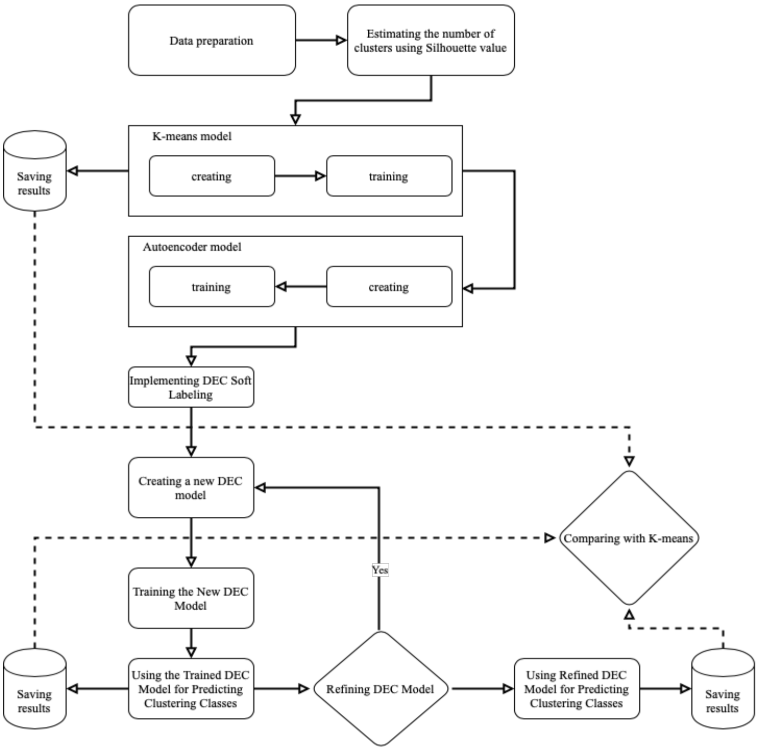

Data preparation
Estimating the number of clusters using Silhouette value
K-means model
creating
training
Saving results
Autoencoder model
training
creating
Implementing DEC Soft Labeling
Creating a new DEC model
Comparing with K-means
Yes
Training the New DEC Model
Saving results
Using the Trained DEC Model for Predicting Clustering Classes
Refining DEC Model
Using Refined DEC Model for Predicting Clustering Classes
Saving results

Figure 2. Model

Data Collection Tool

This section contains information about the participants, the questionnaire used for data collection and the explanations of the datasets.

Participants

The study group of the research consists of 127 students who were studying at the Department of Computer Education and Instructional Technology at Gaziosmanpaşa University Faculty of Education and who took Programming Languages I and Programming Languages 2 in the 2017-2018 academic year.

Collection of Data

The data were collected through a questionnaire consisting of 20 questions about the demographic characteristics of the students, Programming Languages I and Programming Languages II course success grades, created by the researchers. Toplanan veri sonucu bir veri seti oluşturulmuştur. A dataset was formed as a result of the collected data.

Dataset

The attributes and value ranges of the dataset were presented in Table 3. The attributes were divided into two groups to represent the student's demographic characteristics and course success. "Programming Languages I Course Success Grades " and "Programming Languages 2 Course Success Grades" fields belong to the course success group. Other attributes are those included in the demographic group.

Table 3. Attributes and Value Ranges of the Dataset

Attributes	Value Ranges
Gender	Female, Male
Grade	1, 2, 3, 4
Age	17-29, 20-22, 23-25, 26+
High School Graduation	General High School, Anatolian High School, Anatolian Teacher High School, Vocational High School, Anatolian Vocational High School, Religious Vocational High School
Vertical Transfer Examination Situation	Yes, No
Existence of the Personal Computer	Yes, No
Programming Experience (Year)	0-2, 3-5, 6-8, 9+
Living place of Family	Village, District, Province
Family's Monthly Income	0-750 TL, 751-1500 TL, 1501-2250 TL, 2251-3000 TL, 3001 TL +
Having Internet at the Place of Residence	Yes, No
Homework Research Way	Internet, Book, Internet and Book, Internet and Family, Internet-Book-Family
Place of Residence	Credit and Dormitories Institution, Special Domitories, Student House (One Person), Student House (With Friends), With Family
Foreign Language Level (English)	None, Low, Medium, Good, Very Good
Order of Preferring the Department	0-5, 6-10, 11+
Mother Education Status	Illiterate, Literate, Elementary School, Secondary School, High School, University, Postgraduate
Father Education Status	Illiterate, Literate, Elementary School, Secondary School, High School, University, Postgraduate
Mother Occupation Status	Officer, Employee, Housewife, Teacher, Freelancer, Retired, Farmer
Father Occupation Status	Officer, Employee, Teacher, Freelancer, Retired, Farmer, Other
Programming Languages I Course Success Grades	0-30, 31-59, 60-75, 76+
Programming Languages II Course Success Grades	0-30, 31-59, 60-75, 76+

In the dataset table, the success grades of the students in Programming Languages I and II, which are vocational programming courses in the Computer Education and Instructional Technology department, have been taken into consideration. Success of a student was defined as having 60-75 and 76+ range values. The high school programs that students graduate are a determining indicator of their previous education in programming languages. Vertical Transfer Examination transition status and university preference rankings give an idea about the level of consciousness about the department.

The statistical descriptive information of the dataset that emerged as a result of this research conducted with a questionnaire containing 20 different attributes for 127 different students were given in Table 4.

Table 4. Descriptive Statistical Datas

	Count	Mean	Std	Min	Max
Gender	127	1.49606	0.50196	1.0	2.0
Grade	127	2.85826	0.82350	1.0	4.0
Age	127	2.29133	0.63132	1.0	4.0
High School Graduation	127	3.40944	1.77439	1.0	6.0
Vertical Transfer Examination Situation	127	1.96062	0.19524	1.0	2.0
Existence of the Personal Computer	127	1.00787	0.08873	1.0	2.0
Programming Experience (Year)	127	1.71653	0.85354	1.0	4.0
Living place of Family	127	1.64566	0.74030	1.0	3.0
Family's Monthly Income	127	3.19685	1.20869	1.0	5.0
Having Internet at the Place of Residence	127	1.07874	0.27039	1.0	2.0
Homework Research Way	127	2.46456	1.32014	1.0	5.0
Place of Residence	127	2.44094	1.13146	1.0	5.0
Foreign Language Level (English)	127	3.40157	1.35845	1.0	5.0
Order of Preferring the Department	127	1.85826	0.85192	1.0	3.0
Mother Education Status	127	3.37795	2.22514	1.0	7.0
Father Education Status	127	3.81102	2.19932	1.0	7.0
Mother Occupation Status	127	3.05511	0.64619	1.0	7.0
Father Occupation Status	127	4.03937	1.90821	1.0	7.0
Programming Languages I Course Success Grades	127	3.02362	0.69538	1.0	4.0
Programming Languages II Course Success Grades	127	2.67716	0.95851	1.0	4.0

In order to gather information about students' internet access and utilization trends, the questions of access to information methods, foreign languages and the existence of an online connection in their place of residence were included in the survey. In order to explain the socio-cultural and socio-economic status of the parents, family income, residence centers, literate status and occupational information were added to the study. Monthly income of the household ranges, Turkey Statistics Institute for the period of the study was determined using the data.

Information on Demographic Characteristics

This section contains information about the answers given by the students participating in the research to the questionnaire questions. In analyzing the demographic characteristics of the students in the study, the findings were presented by calculating the percentage (%) and frequency (f) values. The frequency value is the number of students corresponding to each answer; the percentage value refers to the ratio of the frequency value to the total number of students.

Table 5. Gender Frequency

Gender	Frequency (f)	Percent (%)
Female	64	50.4
Male	63	49.6
Total	127	100.0

When the distribution of the participants by gender in Table 5 is examined, it was seen that 50.4% (f = 64) are female participants and 49.6 (f = 63) are male participants. Accordingly, it was seen that the participants are almost equally distributed in terms of gender.

Table 6. Grade Frequency

Grade	Frequency (f)	Percent (%)
1	1	0.8
2	50	39.4
3	42	33.1
4	34	26.8
Total	127	100.0

When the distribution of the participants according to their grade attributes is examined in Table 6, it was seen that 0.8% (f = 1) are first grade, 39.4% (f = 50) are second grade, 33.1% (f = 40) are third grade and 26.8%. (f = 34) consisted of fourth grade participants.

Table 7. Age Frequency

Age	Frequency (f)	Percent (%)
17-29	4	3.1
20-22	90	70.9
23-25	25	19.7
26+	8	6.3
Total	127	100.0

When the distribution of the participants by age is examined in Table 7, 3.1% (f = 4) were in the 17-29 age range, 70.9% (f = 90) in the 20-22 age range, 19.7% (f = 25) in the 23-25 age range and% 6.3 of them (f = 8) were observed to be participants who were 25 and over.

Table 8. High School Graduation Frequency

High School Graduation	Frequency (f)	Percent (%)
General High School	14	11.1
Anatolian High School	52	40.9
Anatolian Teacher High School	2	1.6
Vocational High School	2	1.6
Anatolian Vocational High School	41	32.3
Religious Vocational High School	16	12.6
Total	127	100.0

When the distribution of the participants by high school graduation type is examined in Table 8, it was seen that the majority are Anatolian High School (%40,9, f=52) and Anatolian Vocational High School (%32,3, f=41) graduates.

Table 9. Vertical Transfer Examination Situation Frequency

Vertical Transfer Examination Situation	Frequency (f)	Percent (%)
Yes	5	3.9
No	122	96.1
Total	127	100.0

In Table 9, it was seen that 3.9% (f=5) of the participants started the department after taking the Vertical Transfer Exam.

Table 10. Existence of the Personal Computer Frequency

Existence of the Personal Computer	Frequency (f)	Percent (%)
Yes	126	99.2
No	1	0.8
Total	127	100.0

When Table 10 is examined, it was seen that only 1 person does not have a computer. This indicates that most of the participants use computers.

Table 11. Programming Experience Frequency

Programming Experience (Year)	Frequency (f)	Percent (%)
0-2	65	51.2
3-5	37	29.1
6-8	21	16.5
9+	4	3.1
Total	127	100.0

When Table 11 is examined, it was seen that 65 people (51.2%) have 0-2 years of programming experience. It was determined that only 4 people (3.1%) have programming experience of 9 or more years.

Table 12. Living place of Family Frequency

Living place of Family	Frequency (f)	Percent (%)
Village	65	51.2
District	42	33.1
Province	20	15.7
Total	127	100.0

When Table 12 is examined, it was seen that the families of 65 (51.2%) of the participants live in the village. It was determined that 42 participants (33.1%) lived in the district and 20 participants (15.7%) lived in the province.

Table 13. Family's Monthly Income Frequency

Family's Monthly Income	Frequency (f)	Percent (%)
0-750 TL	6	4.7
751-1500 TL	40	31.5
1501-2250 TL	28	22.0
2251-3000 TL	29	22.8
3001 TL +	24	18.9
Total	127	100.0

When Table 13 is examined, it was seen that the monthly family income of the participants differs in terms of the amount of Turkish Lira. It can be said that the monthly family income of only 6 (%4.7) participants is between 0 and 750 TL, and this amount is not enough to support a family. Also, the same can be said for the 751 and 1500 TL range and the 1501-2250 TL range.

Table 14. Having Internet at the Place of Residence Frequency

Having Internet at the Place of Residence	Frequency (f)	Percent (%)
Yes	117	92.1
No	10	7.9
Total	127	100.0

When Table 14 is examined, it was determined that 92.1% (f=117) of the participants have Internet at the Place of Residence.

Table 15. Homework Research Way Frequency

Homework Research Way	Frequency (f)	Percent (%)
Internet	51	40.2
Book	4	3.1
Internet & Book	39	30.7
Internet & Family	28	22.0
Internet - Book - Family	5	3.9
Total	127	100.0

When Table 15 is examined, it was seen that most of the participants researched their homework using the internet and other tools.

Table 16. Place of Residence Frequency

Place of Residence	Frequency (f)	Percent (%)
Credit and Dormitories Institution	35	27.6
Special Domitories	28	22.0
Student House (One Person)	40	31.5
Student House (With Friends)	21	16.5
With Family	3	2.4
Total	127	100.0

When Table 16 is examined, it was determined that only 3 of the participants live with their families in terms of place of residence.

Table 17. Foreign Language Level Frequency

Foreign Language Level	Frequency (f)	Percent (%)
None	1	0.8
Low	52	40.9
Medium	15	11.8
Good	13	10.2
Very Good	46	36.2
Total	127	100.0

When Table 17 is examined, it was determined that 40.9% (f=52) of the participants have a low level of foreign language, and 36.2% (f=46) have a very good foreign language knowledge.

Table 18. Order of Preferring the Department Frequency

Order of Preferring the Department	Frequency (f)	Percent (%)
0-5	56	44.1
6-10	33	26.0
11+	38	29.9
Total	127	100.0

In Table 18, the order of the participants to choose the department they studied according to the score they obtained after taking the university entrance exam is shown. According to Table 2, the number of people who chose the department they study in the 0-5 ranking range is 56 (44.1%).

Table 19. Mother Education Status Frequency

Mother Education Status	Frequency (f)	Percent (%)
Illiterate	50	39.4
Literate	1	0.8
Elementary School	21	16.5
Secondary School	7	5.5
High School	8	6.3
University	34	26.8
Postgraduate	6	4.7
Total	127	100.0

When the mother education status of the participants was examined (Table 19.), it was determined that 50 (39.4%) people were illiterate. It was seen that only the mothers of 6 (4.7%) people have postgraduate education.

Table 20. Father Education Status Frequency

Father Education Status	Frequency (f)	Percent (%)
Illiterate	30	23.6
Literate	5	3.9
Elementary School	40	31.5
Secondary School	2	1.6
High School	2	1.6
University	30	23.6
Postgraduate	18	14.2
Total	127	100.0

When the father education status of the participants was examined (Table 20.), it was determined that 30 (23.6%) people were illiterate. It was seen that only the fathers of 18 (4.7%) people have postgraduate education.

Table 21. Mother Occupation Status Frequency

Mother Occupation Status	Frequency (f)	Percent (%)
Officer	2	1.6
Employee	25	17.3
Housewife	85	66.9
Teacher	5	3.9
Freelancer	3	2.4
Retired	2	1.6
Farmer	5	3.9
Total	127	100.0

When the occupational status of the mothers of the participants is examined in Table 21, it was seen that 66.9% (f=85) of them work as housewifes.

Table 22. Father Occupation Status Frequency

Father Occupation Status	Frequency (f)	Percent (%)
Officer	14	11.6
Employee	10	7.9
Teacher	37	29.1
Freelancer	14	11.0
Retired	24	18.9
Farmer	4	3.1
Other	24	18.9
Total	127	100.0

When the occupational status of the fathers of the participants is examined in Table 22, it is seen that only 4 (3.1%) people worked as farmers. In addition, it was determined that the father of 37 (29.1%) participants worked as a teacher.

Table 23. Programming Languages I Course Success Grades

Programming Languages I Course Success Grades	Frequency (f)	Percent (%)
0-30	6	4.7
31-59	11	8.7
60-75	84	66.1
76+	26	20.5
Total	127	100.0

When Table 23 is examined, it was seen that 66.1% (f=84) of the participants completed the Programming Languages I course in the 60-75 success grade range and 20.5% (f=26) in the 76 and above success grade range.

Table 24. Programming Languages II Course Success Grades

Programming Languages II Course Success Grades	Frequency (f)	Percent (%)
0-30	16	12.6
31-59	37	29.1
60-75	46	36.2
76+	28	22.0
Total	127	100.0

When Table 24 is examined, it was seen that 36.2% (f=46) of the participants completed the Programming Languages II course in the 60-75 success grade range and 22.0% (f=28) in the 76 and above success grade range.

FINDINGS

K-Means Algorithm

The number of samples in the 3 clusters formed after the data set with the K-means algorithm is 38 (cluster-1), 64 (cluster-2) and 25 (cluster-3) (Figure 3). In terms of gender, 100% of the samples were male in cluster-1 and cluster-3, and 100% in cluster-2 were determined as male. In contrast to cluster-1, where 2nd and 3rd grade students are closely dominant, 2nd grade students came together dominantly in cluster-2 with 79.2%, and 4th grade students in cluster-3 with 48%.)

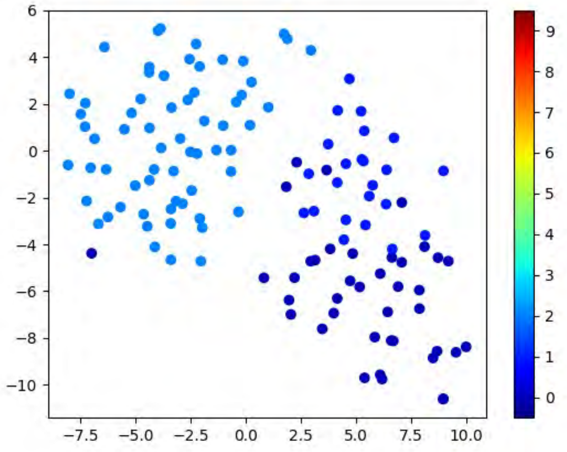


Figure 3. Clusters with K-Means Algorithm

A small number of students aged 26 and over represent 48% of cluster-3 and does not contain an age-determining value for the other two clusters. When examined in terms of graduated high schools, 34.2% of the students included in cluster-1 were distributed to vocational high schools, 48.3% of those included in cluster-2 were distributed to vocational high schools, and 64% of those included in cluster-3 were distributed to vocational high schools. In terms of the status of the Vertical Transfer Examination, 100% of the samples in cluster-1 have marked no.

While 47.4% of the samples of cluster-1 live in the province, this rate increases to 54.7% in cluster-2. In cluster-3, the number of families living in the province and district center is equal and has a share of 48%. Cluster-1 and cluster-2 show similarities in terms of monthly socio-economic income of families. In homework research way, cluster-1 and cluster-2 show similarities, and as the dominant value, the rate of searching homework from the internet is 50% and 43.8% in the cluster, respectively. 48% of the students who make up cluster-3 use books in homework research and this ratio represents the dominant rate. Examining the place of residence of the samples in cluster-3, it was determined that 64% of them continued their education together with their families.

60.5% of the samples in cluster-1 chose the Computer Education and Instructional Technology department among their first 5 choices. For cluster-3, this rate is 56%, and it is decisive with 11 and later placement in the preference ranking. The education status of the mother includes options with dominant rates in cluster-2 and cluster-3 with 53.1% and 44%. These options are primary school for cluster-1 and secondary school for cluster-3, respectively. High school option is the determining ratio in cluster-1 with 34.2% in the educational status of the father.

In terms of mother occupational status, being a housewife constitutes a large proportion in all instances and seems to be predominantly distributed in three clusters. In the father's occupation status, it appears that the pension choice is dominant for cluster-1 and cluster-3.

Considering their success in the programming languages course, cluster-2 and cluster-3 differ from cluster-1 in terms of the number of students 76 and above. When Programming Languages II is accepted as an indicator in terms of success and failure, it is seen that approximately 88% of the students in cluster-3 received successful scores for 3 clusters formed by k-means. It was determined that this ratio is approximately 62% in cluster-2 and approximately 31% for cluster-1. Focusing on cluster-2 and cluster-3 for successful students and cluster-1 data for unsuccessful students will allow us to produce more determinant values.

Attributes Affecting Differentiation Between Sets Using K-means Algorithm

Since the data weren't show normal distribution, the Kruskal Wallis test was applied to determine the variables that affect the formation of differentiation between clusters. Kruskal Wallis test results on demographic information that affect sorting of clusters in line with the data obtained from the students according to K-means are given on Table 25.

Table 25. Kruskal Wallis Test for K-Mean Clusters

	Clusters	n	Mean Ranks	Sd	X^2	p
Gender	0	38	96.00	2	126.00	.000
	1	64	32.50			
	2	25	96.00			
Grade	0	38	58.46	2	5.23	.073
	1	64	61.88			
	2	25	77.86			
Age	0	38	65.13	2	5.61	.061
	1	64	58.95			
	2	25	75.22			
High School Graduation	0	38	81.54	2	15.53	.000
	1	64	59.58			
	2	25	48.66			
Vertical Transfer Examination Situation	0	38	66.50	2	2.72	.256
	1	64	63.52			
	2	25	61.42			
Existence of the Personal Computer	0	38	63.50	2	4.08	.130
	1	64	63.50			
	2	25	66.04			
Programming Experience (Year)	0	38	47.37	2	15.10	.001
	1	64	68.05			
	2	25	78.90			
Living place of Family	0	38	67.68	2	0.66	.718
	1	64	62.57			
	2	25	62.06			
Family’s Monthly Income	0	38	59.16	2	2.52	.284
	1	64	63.16			
	2	25	73.52			
Having Internet at the Place of Residence	0	38	70.70	2	8.27	.016
	1	64	60.98			
	2	25	61.54			
Homework Research Way	0	38	56.84	2	9.02	.011
	1	64	61.07			
	2	25	82.38			
Place of Residence	0	38	60.93	2	12.64	.002
	1	64	73.55			
	2	25	44.22			
Foreign Language Level (English)	0	38	49.66	2	10.12	.006
	1	64	68.10			
	2	25	75.30			
Order of	0	38	52.61	2	16.26	.000

Preferring the Department	1	64	61.57			
	2	25	87.54			
Mother Education Status	0	38	67.42	2	7.86	.019
	1	64	56.21			
	2	25	78.74			
Father Education Status	0	38	60.79	2	0.66	.717
	1	64	64.25			
	2	25	68.24			
Mother Occupation Status	0	38	60.76	2	3.72	.156
	1	64	67.59			
	2	25	59.74			
Father Occupation Status	0	38	59.18	2	2.55	.280
	1	64	63.05			
	2	25	73.74			
Programming Languages I Course Success Grades	0	38	44.83	2	21.97	.000
	1	64	70.15			
	2	25	77.40			
Programming Languages II Course Success Grades	0	38	42.66	2	25.97	.000
	1	64	67.38			
	2	25	87.80			

According to the results of the analysis, it was identified that the effect of the demographic information obtained from the students differed in a meaningful way that came up as a result of K-means algorithm implemented for clustering. It was determined that the demographic information that affected this significant differentiation in the cluster was Gender ($X^2_{(sd=2,\ n=127)}$=126.00, p<.05), High School Graduation ($X^2_{(sd=2,\ n=127)}$=15.53, p<.05), Programming Experience ($X^2_{(sd=2,\ n=127)}$=15.10, p<.05), Having Internet at the Place of Residence ($X^2_{(sd=2,\ n=127)}$=8.27, p<.05), Homework Research Way ($X^2_{(sd=2,\ n=127)}$=9.02, p<.05), Place of Residence ($X^2_{(sd=2,\ n=127)}$=12.64, p<.05), Foreign Language ($X^2_{(sd=2,\ n=127)}$=10.12, p<.05), Order of Preferring the Department ($X^2_{(sd=2,\ n=127)}$=16.26, p<.05), Mother Education Status ($X^2_{(sd=2,\ n=127)}$=7.86, p<.05), Programming Languages I Course Success Grades ($X^2_{(sd=2,\ n=127)}$=21.97, p<.05), Programming Languages II Course Success Grades ($X^2_{(sd=2,\ n=127)}$=25.97, p<.05). It was found that this demographic information affected the sorting of three different clusters that came up as a result of K-means algorithm in a meaningful way.

DEC Algorithm

The number of samples in the 3 clusters formed after the data set with the DEC algorithm is 40 (cluster-1), 23 (cluster-2) and 64 (cluster-3) (Figure 4). In terms of gender, 100% of the samples were male in cluster-1 and cluster-2, and 100% in cluster-3 were determined as female. In contrast to cluster-1, where 2nd and 3rd grade students are closely dominant, 4nd grade students came together dominantly in cluster-2 with 47.8%, and 2th grade students in cluster-3 with 48.3%.

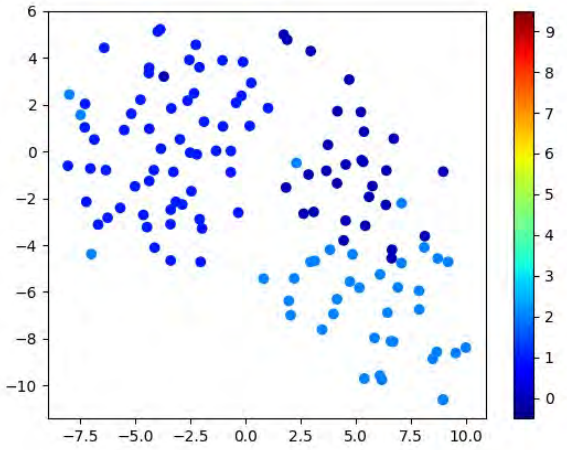


Figure 4. Clusters with DEC Algorithm

A small number of students aged 26 and over represent 10% of cluster-1 and does not contain an age-determining value for the other two clusters. When examined in terms of high schools graduated, 32.5% of the students included in cluster-1 are vocational high school graduates and 30% are general high school graduates. When these ratios were examined, high school for cluster-1 appears to be the differential value against other clusters. 43.2% of those included in cluster-2 were distributed to vocational high schools. 43.8% of those included in cluster-3 were distributed to vocational high schools and 34.4% to anatolian vocational high schools.

The living place rates of the families of the samples in cluster-1 in the city and district center are close to each other. This rate increases in favor of living in the province center with 65.2% in cluster-2. In cluster-3, the number of families living in the province center has a share of 54.7%.

The clusters are similar in terms of monthly socio-economic income of the families. In homework research way, the rate of searching homework from the internet for cluster-1 is 50%. 22.5% of the samples receive research support from family and educators. Internet and family support rate in homework research way for cluster-2 together is approximately 83%. The dominant value in homework research way for cluster-3 was found to be the internet with 43.8%. In Cluster-3, the place of residence of the samples is decisive with the Credit and Dormitories Institution option with a rate of 32.8%. This rate is 17.5% and 13% for cluster-1 and cluster-2, respectively.

47.5% of the samples in cluster-1 chose the Computer Education and Instructional Technology department among their first 5 choices. For cluster-3, this rate is 46.9% and it is the highest rate. For cluster-2, these ratios are evenly distributed and appear to be ineffective. The education status of the mother is in cluster-1, primary school, high school and secondary school options are equally dominant with 20%. For Cluster-2, the rate of those whose mother education status is secondary school is 39.1% and it represents the most dominant value. For Cluster-3, primary school is the option with the highest sample and this is 53.1%. In the educational status of the father, high school in cluster-1, secondary school-high school in cluster-2 and primary school-high school in cluster-3 appear as the dominant values.

In terms of mother occupational status, being a housewife constitutes a large proportion in all instances and seems to be predominantly distributed in three clusters. In cluster-2, all samples are homogeneously separated from the other two clusters in terms of mother occupational status. In the father occupational status, the situation appears to be dominant in cluster-1 and cluster-2 retirement, officer-teacher, and farmer for cluster-3.

Considering their success in the programming languages course, cluster-1 and cluster-3 differ from cluster-2 in terms of the number of students who have 60 or more successful scores. In terms of success and failure, when the Programming Languages II course was accepted as an indicator, it was determined that approximately 47.8% of the students in cluster-2 for the 3 clusters formed with DEC have received high successful scores. It was determined that this ratio was 61% in cluster-3 and 31% for cluster-1. Although there were slight differences with k-means for successful students, DEC includes successful students in cluster-2 and cluster-3. cluster-1 contains less successful student data. Focusing on these data will allow us to produce more determinant values.

Attributes Affecting Differentiation Between Sets Using DEC Algorithm

Since the data weren't show normal distribution, the Kruskal Wallis test was applied to determine the attributes that affect the formation of differentiation between clusters. Kruskal Wallis test results on demographic information that affect sorting of clusters in line with the data obtained from the students according to DEC are given on Table 26.

Table 26. Kruskal Wallis Test for DEC Clusters

	Clusters	n	Mean Ranks	Sd	X^2	p
Gender	0	40	96.00	2	126.00	.000
	1	23	96.00			
	2	64	32.50			
Grade	0	40	57.81	2	6.86	.032
	1	23	80.67			
	2	64	61.88			
Age	0	40	63.18	2	8.34	.015
	1	23	79.50			
	2	64	58.95			
High School Graduation	0	40	77.84	2	9.98	.007
	1	23	52.24			
	2	64	59.58			
Vertical Transfer Examination Situation	0	40	66.50	2	3.09	.214
	1	23	60.98			
	2	64	63.52			
Existence of the Personal Computer	0	40	63.50	2	4.52	.104
	1	23	66.26			
	2	64	63.50			
Programming Experience (Year)	0	40	42.65	2	30.57	.000
	1	23	89.85			
	2	64	68.05			
Living place of Family	0	40	73.41	2	6.44	.040
	1	23	51.61			
	2	64	62.57			
Family’s Monthly Income	0	40	59.94	2	2.16	.340
	1	23	73.41			
	2	64	63.16			
Having Internet at the Place of Residence	0	40	70.11	2	7.43	.024
	1	23	61.76			
	2	64	60.98			
Homework	0	40	58.20	2	7.89	.020

Research Way	1	23	82.24			
	2	64	61.07			
Place of Residence	0	40	62.80	2	15.57	.000
	1	23	39.52			
	2	64	73.55			
Foreign Language Level (English)	0	40	46.45	2	18.26	.000
	1	23	83.11			
	2	64	68.10			
Order of Preferring the Department	0	40	63.40	2	1.52	.467
	1	23	71.80			
	2	64	61.57			
Mother Education Status	0	40	71.08	2	6.39	.041
	1	23	73.37			
	2	64	56.21			
Father Education Status	0	40	59.93	2	1.26	.532
	1	23	70.39			
	2	64	64.25			
Mother Occupation Status	0	40	59.41	2	3.90	.142
	1	23	62.00			
	2	64	67.59			
Father Occupation Status	0	40	55.96	2	6.92	.031
	1	23	80.61			
	2	64	63.05			
Programming Languages I Course Success Grades	0	40	48.13	2	15.83	.000
	1	23	74.50			
	2	64	70.15			
Programming Languages II Course Success Grades	0	40	46.05	2	19.82	.000
	1	23	85.83			
	2	64	67.38			

According to the results of the analysis, it was identified that the effect of the demographic information obtained from the students differed in a meaningful way that came up as a result of DEC algorithm implemented for clustering. It was determined that the demographic information that affected this significant differentiation in the cluster was Gender ($X^2_{(sd=2,\ n=127)}$=126.00, p<.05), Grade ($X^2_{(sd=2,\ n=127)}$=6.86, p<.05), Age ($X^2_{(sd=2,\ n=127)}$=8.34, p<.05), High School Graduation ($X^2_{(sd=2,\ n=127)}$=9.98, p<.05), Programming Experience ($X^2_{(sd=2,\ n=127)}$=30.57, p<.05), Living place of Family ($X^2_{(sd=2,\ n=127)}$=6.44, p<.05), Having Internet at the Place of Residence ($X^2_{(sd=2,\ n=127)}$=7.43, p<.05), Homework Research Way ($X^2_{(sd=2,\ n=127)}$=7.89, p<.05), Place of Residence ($X^2_{(sd=2,\ n=127)}$=15.57, p<.05), Foreign Language Level ($X^2_{(sd=2,\ n=127)}$=18.26, p<.05), Mother Education Status ($X^2_{(sd=2,\ n=127)}$=6.39, p<.05), Father Education Status ($X^2_{(sd=2,\ n=127)}$=6.92, p<.05), Programming Languages I Course Success Grades ($X^2_{(sd=2,\ n=127)}$=15.83, p<.05) and Programming Languages II Course Success Grades ($X^2_{(sd=2,\ n=127)}$=19.82, p<.05). It was found that this demographic information affected the sorting of three different clusters that came up as a result of DEC algorithm in a meaningful way.

CONCLUSION

Course success is of great importance for students during their education life. Students take the grades they take from exams, the knowledge experienced in their personal life processes, the vital processes of themselves and their families are an important factor for the success of the course. With this type of data, students' success for the courses can be predicted. Information resulting from the processing of data can provide a variety of information to both students and educators. Processing these data with educational data mining and artificial intelligence applications is an important development that can improve today's educational environments.

With the machine learning method, it is possible to classify the course success according to the various information obtained from the students with the prediction processes for the students' course success and the machine learning algorithms used. Student clusters that emerge with the algorithms used can offer various advantages to educators. Various recommendations can be made for all educational processes, from the teaching method to the course material used.

In this study, it was aimed to discover students' course success by using unsupervised machine learning algorithms. With the study, a model was proposed to both educators and instructional designers. For this purpose, factors affecting students' course success were determined and students were divided into various clusters regarding these factors. The demographic characteristics and exam grades obtained from the students in clustering and the success processes of the students for the programming lessons were tried to be determined by various algorithms.

It was determined that the students were divided into 3 clusters with the K-means algorithm, which is one of the unsupervised machine learning algorithms used. When the factors affecting the clustering of students with the K-means algorithm was examined, it was determined that the Gender, High School Graduation, Programming Experience, Having Internet at the Place of Residence, Homework Research Way, Place of Residence, Foreign Language Level, Order of Preferring the Department, Mother Education Status, Programming Languages I Course Success Grades and Programming Languages II Course Success Grades factors affect the course success.

In addition, it was determined that the students were divided into 3 clusters with the DEC algorithm, which is one of the unsupervised machine learning algorithms used. When the factors affecting the clustering of students with the DEC algorithm was examined, it was determined that the Gender, High School Graduation, Programming Experience, Living place of Family, Having Internet at the Place of Residence, Homework Research Way, Place of Residence, Foreign Language Level, Mother Education Status, Father Education Status, Programming Languages I Course Success Grades and Programming Languages II Course Success Grades factors affect the course success.

In both algorithms, Gender, High School Graduation, Programming Experience, Having Internet at the Place of Residence, Homework Research Way, Place of Residence, Foreign Language Level, Mother Education Status, Programming Languages I Course Success Grades and Programming Languages II Course Success Grades attributes are determined as common factors. With these attributes, various information can be given to both teachers and students about course success. In order to make the programming education-training processes more effective and efficient, with these factors can be predicted by predicting the success of the course.

Various factors affected both algorithms. It was found that 11 attributes in the K-means algorithm and 14 attributes in the DEC algorithm affect the clustering of students. According to these results, it can be said that the clusters obtained from the DEC algorithm are better than the K-means algorithm. The greater the differentiation of attributes, the more information to the educators and students can give about course success. According to these attributes, it can be ensured that the methods used and applied in the lessons can be determined and the success of the students can be determined. Such attributes can be used to predict course success with machine learning methods, and more effective and efficient learning environments can be designed in advance in line with the data obtained from students. In line with the data obtained from the students, attributes and clusters, the course success of the students can be determined in advance with the machine learning method and various precautions and decisions regarding the course processes can be taken in advance.

REFERENCES

Alpaydın, E. (2004). *Introduction to Machine Learning*. London: The MIT Press.

Baker, T., & Smith, L. (2019). Educ-AI-tion rebooted. Exploring the future of artificial intelligence in schools and colleges. Nesta Foundation https://media.nesta.org.uk/documents/Future_of_AI_and_education_v5_WEB.pdf

Chollet, F. (2017). *Deep Learning with Python*. Shelter Island: Manning Publications, 384.

Dalyan, T. (2006). *Makine öğrenmesinde 1R algoritması ve ikinci kuralın (2R) oluşturulması*. [Master's thesis]. http://dspace.kocaeli.edu.tr:8080/xmlui/bitstream/handle/11493/979/197908.pdf?sequence=1&isAllowed=y

Dong, G., Liao, G., Liu, H., & Kuang, G. (2018). A review of the autoencoder and its variants: A comparative perspective from target recognition in synthetic-aperture radar images. IEEE Geoscience and Remote *Sensing Magazine, 6*(3), 44-68.

Hussain, M., Zhu, W., Zhang, W., & Abidi, S. M. R. (2018). Student engagement predictions in an e-learning system and their impact on student course assessment scores. *Computational intelligence and neuroscience, 2018*. https://doi.org/10.1155/2018/6347186

International Educational Data Mining Society, (2020, October). *Educational Data Mining.* https://educationaldatamining.org

Kazu, İ.Y., & Özdemir, O. (2009). Öğrencilerin Bireysel Özelliklerinin Yapay Zekâ İle Belirlenmesi (Bulanık Mantık Örneği). *Akademik Bilişim*, Harran Üniversitesi, Şanlıurfa.

Koitka, S., & Friedrich, C. M. (2016, September). Traditional Feature Engineering and Deep Learning Approaches at Medical Classification Task of ImageCLEF 2016. *In CLEF (Working Notes)* (pp. 304-317).

Lykourentzou, I., Giannoukos, I., Mpardis, G., Nikolopoulos, V., & Loumos, V. (2009). Early and dynamic student achievement prediction in e-learning courses using neural networks. Journal of the American *Society for Information Science and Technology, 60*(2), 372-380.

Na, S., Xumin, L., & Yong, G. (2010, April). Research on k-means clustering algorithm: An improved k-means clustering algorithm. In *2010 Third International Symposium on intelligent information technology and security informatics* (pp. 63-67). IEEE.

Ren, Y., Hu, X., Shi, K., Yu, G., Yao, D., & Xu, Z. (2018, August). Semi-supervised denpeak clustering with pairwise constraints. In *Pacific Rim International Conference on Artificial Intelligence* (pp. 837-850). Springer, Cham.

Robinson, C., Yeomans, M., Reich, J., Hulleman, C., & Gehlbach, H. (2016, April). Forecasting student achievement in MOOCs with natural language processing. In *Proceedings of the sixth international conference on learning analytics & knowledge* (pp. 383-387).

Romero, C., & Ventura, S. (2010). Educational data mining: a review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40*(6), 601-618.

Siemens, G. (2010). What are learning analytics?. http://www.elearnspace.org/blog/2010/08/25/what-are-learning-analytics/

Vandamme, J. P., Meskens, N., & Superby, J. F. (2007). Predicting academic performance by data mining methods. *Education Economics, 15*(4), 405.

Wang, F., Franco-Penya, H. H., Kelleher, J. D., Pugh, J., & Ross, R. (2017, July). An analysis of the application of simplified silhouette to the evaluation of k-means clustering validity. In *International Conference on Machine Learning and Data Mining in Pattern Recognition* (pp. 291-305). Springer, Cham.

Xie, J., Girshick, R., & Farhadi, A. (2016, June). Unsupervised deep embedding for clustering analysis. In *International conference on machine learning* (pp. 478-487).

Yukselturk, E., Ozekes, S., & Türel, Y. K. (2014). Predicting dropout student: an application of data mining methods in an online education program. *European Journal of Open, Distance and e-learning, 17*(1), 118-133.

Yusof, N., Zin, N. A. M., Yassin, N. M., & Samsuri, P. (2009, December). Evaluation of Student's Performance and Learning Efficiency based on ANFIS. In *2009 International Conference of Soft Computing and Pattern Recognition* (pp. 460-465). IEEE.