

# Statistical and Clustering Based Rules Extraction Approaches for Fuzzy Model to Estimate Academic Performance in Distance Education

Osman Yildiz & Abdullah Bal  
Yildiz Technical University, TURKEY

Sevinc Gulsecen  
Istanbul University, TURKEY

Received 25 May 2014; accepted 17 March 2015; published 25 March 2015

The demand for distance education has been increasing at a rapid pace all around the world. This, in turn, places a special importance on the need for the development of more distance education systems. However, there is an alarming rise in the number of distance education students that drop out of the system without asking for any help. The present study focuses on forming three fuzzy-based models through K-Means, C-Means and subtractive clustering. The models are designed to predict students' year-end academic performance based on the 8-week data kept in the learning management system (LMS). Next, the models are evaluated in terms of their accuracy in order to determine the most suitable one. Then, the data was analyzed through various statistical methods and the results were compared. The model provides invaluable information regarding the students' year-end success or failure by analyzing the data on Basic Computer Skills, a course included in the curriculum for sophomores at a local university. Thanks to such information, those who are likely to drop out can be determined and accordingly, the institution can start to take measures to encourage students not to drop out early in the semester, which, in turn, can increase the extent to which distance education can be successful. The present study will hopefully decrease the number of students that drop out of distance education systems.

*Keywords:* Distance education, subtractive, k-means, fuzzy c-means, clustering, academic performance.

## INTRODUCTION

In recent years, the educational process has been characterized by a notable shift from conventional teaching to online education. The underlying reasons for the transformation to distance education include easy access, flexibility, individual learning and strong feedback (Chou & Liu, 2005). Distance education

systems present online educational contents visually and orally. Such systems are continuously updated and can be accessed everywhere regardless of the students' location. Properties such as online forums and chat rooms make the process student-centric. Moreover, with different educational interfaces and modules, distance education programs offer an education system that is geared toward the specific needs of each learner.

However, despite the above-mentioned advancements and advantages of distance education there is still a high rate of dropouts among students enrolled in distance education programs. A study by Education Dynamics focused on online learning attempted to find out the reasons for high dropout rate in distance education. The study identified five main

Correspondence to: Osman Yildiz,  
Department of Informatics, Yildiz Technical  
University, A2009, Davutpasa,  
Esenler, Istanbul, Turkey  
E-mail: oyildiz@yildiz.edu.tr  
doi: 10.12973/eurasia.2015.1356a

### ***State of the literature***

- Educational Data Mining (EDM) theory aims to reveal unknown information and make the best use of it. The ability to predict the academic performance of students will help to contribute to the success of students.
- Variables such as demographic data, assignments and test results and student participation in forums are used to predict students' academic performance.
- A variety of data mining methods such as artificial neural networks, general algorithm, decision trees, Support Vector Machines, and naive Bayes are used.

### ***Contribution of this paper to the literature***

- This study presents a new mathematical model that predicts students' year-end academic performance in distance education systems.
- The results of this study have shown that by using only five variables such as recency, frequency, monetary, midterm exam and quiz results this method is able to predict students' year-end academic performance with very high accuracy.
- Data set, which was obtained through an 8-week study, was proven to be invaluable to both instructors and the administrators of the institution. In the simplest sense this information could help school administrators to decrease the dropout rate significantly in distance education programs.

reasons, namely financial challenges (41%), life events (32%), health issues (23%), lack of personal motivation (21%) and lack of faculty interaction (21%) (Education Dynamics, 2013). Among these reasons, lack of personal motivation and lack of faculty interaction are the main issues that can be resolved by organizations that provide distance education. The sooner such organizations can identify these issues, the sooner they can take relevant measures and therefore encourage students not to drop out from their education and increase their achievements.

Another reason for dropping out, which was not mentioned in the study by Education Dynamics, is the lack of observation. It is rarely observed in conventional education systems, because in traditional teaching environments the teacher is able to observe students' behaviors and take remedial measures accordingly. However, it is impossible in distance education.

Distance education is commonly delivered on a platform called the learning management system (LMS). This platform hosts a huge amount of data. All of the students' actions are closely monitored and recorded as

logs in the database. An analysis of such data can yield invaluable information (Zafra & Ventura, 2009). In this study, an LMS platform called the Moodle was chosen because it has previously been used as a LMS platform for sharing useful information, documentation, and knowledge management in research projects and provided important benefits for researchers (Psycharis, Chalatzoglidis, & Kalogiannakis, 2013).

The present study was designed to form a fuzzy-based model to process the LMS data of 337 students. Recency, Frequency, and Monetary (RFM) is a method widely used in the advertising industry for analyzing customer profiles. In short, RFM is used to assign various scores to three basic questions: the period of time that passed after the last transaction made by the customer, the frequency at which he/she makes a transaction, and how much money he/she spends on his/her transactions. The scores are then used to create a profile for that particular customer (Wei, Lin & Wu, 2010). The present study is based on a similar principle and uses a particular dataset to find answers to questions such as how much time passed after the students last log on to the LMS, the frequency at which he/she logs on to the system, and how much time he/she spends online on the system. Since the behaviors of the students in the LMS are subject to ambiguities, a mathematical model was established by using a fuzzy logic, which is a method that is considered to produce the best results in the face of ambiguities. The statistical methods for the model include outlier data analysis and data normalization. The model has five inputs and one output. Three of the inputs are recency, which represents the number of days that passed before a student logs on to the course after it has been uploaded to the system, frequency, which stands for the frequency at which a student logs on to the system, and monetary, which shows the amount of time spent in the system. The two remaining inputs are an online quiz administered to the students in the 4th week and a paper-based midterm exam in the 8th week. The output is the student's year-end academic performance.

In his paper entitled "Fuzzy Model Identification Based on Cluster Estimation", Chiu identified the algorithms for subtractive clustering and showed how cluster centers could be determined through these algorithms. With these cluster centers, he focused on establishing the rules for Takagi-Sugeno fuzzy modeling and finding the parameters for these rules (Chiu, 1994).

One disadvantage of fuzzy sets is that rules are established by experts and researchers are developing new strategies to overcome this issue. In their article entitled "Generation of Fuzzy Rules with Subtractive Clustering" published in the *Jurnal Teknologi* in 2005, Priyono et al. established a model through Chiu's subtractive clustering, calculated the limit fit value using

a genetic algorithm, and interpreted their results (Priyono, Muhammad Ridwan, & Atiq, 2005).

Similarly, in a report on fuzzy rules designed for the behaviors of small mobile robots, presented in 1997, Kim and Kong established fuzzy rules with Chiu's method and demonstrated how they could be used for mobile robots (Kim & Kong, 1997).

Moreover, in her "Introduction to Five Data Clustering Algorithms" Moertini provided information about K-means clustering, fuzzy C-means clustering, mountain clustering, subtractive clustering and partition simplification fuzzy C-Means clustering.

Based on the hybridization of fuzzy C-means clustering and subtractive clustering, two methods commonly used in fuzzy clustering algorithms, "A Modified Hybrid Fuzzy Clustering Algorithm for Data Partitions" provides experimental results (Hossen, Rahman, Sayeed, Samsuddin, & Rokhani, 2011). It presents the differences between the clustering without the hybrid method and the one with the hybrid method.

In their study, Yildiz, Bal and Gulsecen (2013) established a model designed to measure distance education students' academic performance through Mamdani fuzzy model. The results were compared via classical fuzzy, expert fuzzy, and gene-fuzzy models. The authors based their study on six-week data on a total of 218 participants and three variables. The accuracy rate was around 82% (Yildiz, Bal, & Gulsecen, 2013).

Lykourantzou et al. (2009) assessed their results obtained through three different methods and predicted via multiple genetic algorithms whether a student would drop out of a course or school. The study involved test results, project assessments and demographic data (Lykourantzou, Giannoukos, Nikolopoulos, Mpardis, & Loumos, 2009).

In 2007, Vandamme et al. classified students as "low-risk", "medium-risk" and "high-risk" groups in reference to their demographics, socio-economic background and academic background. In this way, they used neural networks method to predict who would fail in a course or drop out of school (Vandamme, Meskens, & Superby, 2007).

In addition, in their study in 2006, Kalles and Pierrakes used a genetic algorithm and decision trees to predict distance education students' academic performance. In another study, Zafra and Ventura (2009) used multiple instance genetic algorithms to predict whether students would pass or fail in a course. The study was based on the students' scores in quizzes, assignments and their activities on forums.

In a 2013 study conducted by Borkar ve Rajeswari, rules to predict the correlation between unit test, university result and graduation were obtained by using two variables such as the assignments and attendance rate of 60 students taking Master of Computer

Application class. This study on Educational Data Mining was found to be beneficial to academic performance of students (Borkar & Rajewari, 2013).

In another study carried out in 2013, the authors tried to predict the students' grades. To achieve this goal, they determined whether the demographic or educational data sets had more predictive power. The variables were modeled using different data mining methods (Ramesh, Porkavi & Ramar, 2013).

Educational Data Mining (EDM) theory aims to reveal unknown information and make it useful in the process of education. Being able to predict the academic performance of students help contribute to the success of students. Tekin in a study done in 2014 predicted the overall GPAs of students using their grades during the first three years by employing a variety of data mining methods. Based on this method, the students that need more support can be identified. This piece of information is very precious since it helps to contribute to the meeting of the educational needs of students (Tekin, 2014).

The main disadvantage of the above-mentioned studies is that the data used to evaluate the academic performance of students was collected in an extensive period of time. In our study, we aimed to address this issue by considerably shortening the time for data collection.

## **METHODOLOGY**

### **Sample**

The study was conducted on a total of 337 students registered to Basic Computer Sciences, an online course, offered at Yildiz Technical University during the 2011-2012 and 2012-2013 academic years. Since 24 students did not participate in any of the activities in the distance education system, they were excluded from the sample. While 218 students were registered to the course during the 2011-2012 academic year, the remaining 95 took the course during the 2012-2013 academic year. The former group of students was divided into two datasets, namely 70% as a training dataset and 30% as a test dataset. The remaining 95 students were assigned as verification dataset. The demographics of the participants were not included in the data analysis. The study had five inputs and one output. The data on recency, frequency and monetary were obtained from the Moodle, the distance education platform on which the course was delivered. The six-week data were obtained from the Moodle in the form of a log file. The log file had almost 75 thousand lines. The values associated with recency, frequency and monetary were calculated for each student through software designed on Matlab. The fourth input was the scores of the participants in the quiz administered online on the Moodle in week 4.

Within the scope of the course, the students were required to take three online quizzes, two midterm exams and one final exam throughout the term. The distribution of these examinations by their contribution to the year-end academic performance was as follows: three online quizzes made up 20%, two midterm exams made up 40%, and one final exam made up 40%.

A fuzzy-based model was used to predict the distance education students' year-end academic performance. The data were subject to clustering algorithms with the results being used to establish the Sugeno type fuzzy model. The clustering methods were K-Means, fuzzy C-means and subtractive.

### Fuzzy Logic

Human beings experience a number of problems in their daily lives and attempt to overcome them on the basis of the information and experiences they have already acquired. Some of these problems are clear-cut and easy to identify; therefore, it is also easier to handle them. On the other hand, it is relatively harder to deal with problems that involve ambiguities or are not fully identified.

A fuzzy set is identified by assigning a value to each relevant element and the value represents the degree of its membership to the set in mathematical terms. The value refers to the extent to which the element belongs to the concept represented by the fuzzy set. Therefore, each element has varying degrees of membership, which are expressed in real numbers ranging from 0 and 1. Full membership and lack of membership are represented in the fuzzy set by 1 and 0 respectively (Sari, Murat, & Kirbali, 2013).

Two types of models are commonly used in fuzzy logic, namely Mamdani and Takagi-Sugeno fuzzy models. Mamdani fuzzy model is widely used since it is suitable for human behaviors and can easily be established. It consists of three main steps. The first step is fuzzification, which is the process where inputs in the system are blurred and each input is assigned a value of membership ranging from 0 to 1. The second stage is where rules are processed. Here, rules are derived in the form of "if then". Inputs are handled in accordance with the rule table. The third stage, defuzzification involves transforming fuzzy values into actual values.

On the other hand, Takagi-Sugeno fuzzy logic model, which was first introduced in 1985, is derived as follows:

If  $x_1$  is  $A_1$  and ... and  $x_k$  is  $A_k$

Then  $y = p_0 + p_1 \cdot x_1 + \dots + p_k \cdot x_k$

characterized by "and" connective and linear equation. Where  $x_1, \dots, x_k$  are variables composing the premises of implications.  $A_1, \dots, A_k$  are membership functions of the fuzzy sets in the premises, abbreviated as premise

parameters.  $p_0, \dots, p_k$  are parameters in the consequences. (Takagi & Sugeno, 1985; Mathworks, 2013).

The fact that Sugeno output membership functions are either linear or constant is what significantly distinguishes Sugeno from Mamdani. Another difference between the two is the consequents of their fuzzy rules; therefore, there is a corresponding difference between their aggregation and defuzzification procedures (Sivanandam, Sumathi, & Deepa, 2007).

The underlying reason behind the use of Takagi-Sugeno fuzzy logic model in the present study is that the model can arrange the intervals and that rule formation is not based on interpretation but governed by the model itself. Through clustering analysis, the dataset is divided into sets of elements with similar characteristics. With the formation of membership functions and rules by the model, inputs are entered into the system and outputs are produced.

The basic principle of fuzzy clustering is to partition the data into fuzzy clusters and to make sure that one particular part of the system behavior is symbolized by each cluster. One can find the antecedent sections of the fuzzy rules after transmitting clusters onto the input space; in this case, the consequent parts of the rules can be simple functions. One rule of Sugeno fuzzy model is represented by one cluster accordingly (Priyono et al., 2005).

### Identifying the Parameters Using Least-Square Estimation

When certain input values  $x_1^0, x_2^0, \dots, x_n^0$  are given to the input variables  $x_1, x_2, \dots, x_n$ , the conclusion from the kth rule(1) in a Takagi-Sugeno model is a crisp value  $w^k$ :

$$w^k = p_0^k + p_1^k x_1^0 + \dots + p_n^k x_n^0 \quad (1)$$

$p_0^k, p_1^k, \dots, p_n^k$  are the optimal consequent parameters.

With a certain rule firing strength (weight) defined as

$$\alpha^k = \mu_1^k(x_1^0) \cap \mu_2^k(x_2^0) \cap \dots \cap \mu_n^k(x_n^0)$$

Where  $\mu_1^k(x_1^0), \mu_2^k(x_2^0), \dots, \mu_n^k(x_n^0)$  are membership grades in the kth rule. The symbol  $\cap$  is a conjunction operator. The output model is computed (using weighted average aggregation) as

$$w = \frac{\sum_{m=1}^k \alpha^k w^k}{\alpha^k} \quad (2)$$

Suppose

$$\beta^k = \frac{\alpha^k}{\sum_{k=1}^m \alpha^k} \quad (3)$$

Where  $\beta^k$  is the matching weight of the k-th fuzzy rule. Then formula 1 can be converted into a linear least-square estimation problem, as

$$w = \sum_{k=1}^m \beta^k w^k \quad (4)$$

For a group of w data vectors, the equations can be obtained as:

$$\begin{aligned} w^1 &= \beta_1^1 (p_0^1 + p_1^1 x_1 + \dots + p_n^1 x_n) + \\ &\beta_1^2 (p_0^2 + p_1^2 x_1 + \dots + p_n^2 x_n) + \dots + \\ &\beta_1^m (p_0^m + p_1^m x_1 + \dots + p_n^m x_n) \\ w^2 &= \beta_2^1 (p_0^1 + p_1^1 x_1 + \dots + p_n^1 x_n) + \\ &\beta_2^2 (p_0^2 + p_1^2 x_1 + \dots + p_n^2 x_n) + \dots + \\ &\beta_2^m (p_0^m + p_1^m x_1 + \dots + p_n^m x_n) \\ &\dots \\ w^w &= \beta_w^1 (p_0^1 + p_1^1 x_1 + \dots + p_n^1 x_n) + \\ &\beta_w^2 (p_0^2 + p_1^2 x_1 + \dots + p_n^2 x_n) + \dots + \\ &\beta_w^m (p_0^m + p_1^m x_1 + \dots + p_n^m x_n) \end{aligned}$$

This equation can be represented as:

$$\begin{bmatrix} w^1 \\ \vdots \\ w^m \end{bmatrix} = \begin{bmatrix} \beta_1^1 x_1 & \beta_1^1 x_2 & \dots & \beta_1^1 x_n & \beta_1^1 & \dots & \beta_1^m x_1 & \beta_1^m x_2 & \dots & \beta_1^m x_n & \beta_1^m \\ \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots \\ \beta_w^1 x_1 & \beta_w^1 x_2 & \dots & \beta_w^1 x_n & \beta_w^1 & \dots & \beta_w^m x_1 & \beta_w^m x_2 & \dots & \beta_w^m x_n & \beta_w^m \end{bmatrix} \begin{bmatrix} p_1^1 \\ \vdots \\ p_n^1 \\ p_0^1 \\ \vdots \\ p_1^m \\ \vdots \\ p_n^m \\ p_0^m \end{bmatrix} \quad (5)$$

Using the standard notation  $AP=W$ , this becomes a least square estimation problem where A is a constant matrix (known), w is a matrix of output values (known) and P is a matrix of parameters to be estimated (Ren, 2006).

$$P = (A^T A)^{-1} A^T W \quad (6)$$

### Clustering Algorithms

Clustering algorithms are unsupervised learning methods that generate subgroups on the basis of similarities in a cluster (Oliveira & Pedrycz, 2007). The objective is to divide the data set to as many clusters as possible using certain criteria and on the basis of the differences or similarities between the variables. The present study used K-means, fuzzy C-means clustering and subtractive clustering methods.

### K-Means Clustering

K-means is one of the most commonly used unsupervised learning methods. It was first described by Macqueen in 1967. In this clustering algorithm, each datum can belong to one single cluster. Central points are also the center of the clusters. The number of clusters is specified in advance. Formula 1 presents the total sum of squares for the distance from the elements of the cluster to the cluster center (Gan, Ma, & Wu, 2007). The error function is represented as quadratic error function. The cluster with the lowest value yields the best result.

$$E = \sum_{i=1}^k \sum_{x \in C_i} d(x, \mu(C_i)) \quad (7)$$

where  $C_1, C_2, \dots, C_k$ , are discrete clusters,  $\mu(C_i)$  stands for the cluster center, and  $d(x, \mu(C_i))$  represents the distance between x and  $\mu(C_i)$ , the cluster center. Even though there are many different methods for calculating the distance, it is Euclidian method that is generally used.

K-means algorithm goes on until one has a specified k number of clusters that are as dense and separate as possible. The algorithm does not stop as long as it finds differences between elements of the cluster.

### Fuzzy C-Means Clustering

This method is the most commonly known and used of all fuzzy division techniques. Unlike K-means, it allows elements to belong to more than one cluster. As is the case in fuzzy logic, each element has a value of membership that varies from one to zero. Each element's membership to all clusters has a total value of 1. The closer one element is to one cluster center, the higher value of membership it has to that cluster center. Fuzzy c-means algorithm is a method based on the objective function, too. The algorithm is operated to set back and minimize the objective function, which is the generalization of the least squares method (Moertini, 2002).

$$J_m = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \|x_i - c_j\|^2 \quad 1 \leq m < \infty \quad (8)$$

where m is any real number greater than 1,  $u_{ij}$  is degree of membership of  $x_i$  in the cluster j,  $x_i$  is the ith of d-dimensional measured data,  $c_j$  is the d-dimension center of the cluster, and  $\|*\|$  is any norm expressing the similarity between any measured data and the center. Where the membership matrix U is randomly selected and algorithm is started. The next step is to calculate central vectors in accordance with the formula (Oliveira & Pedrycz, 2007).

$$c_j = \frac{\sum_{i=1}^N u_{ij}^m x_i}{\sum_{i=1}^N u_{ij}^m} \quad (9)$$

A comparison is made between the new matrix U, which is restructured with the following formula in accordance with the newly-calculated cluster centers, and the old matrix U.

$$u_{ij} = \frac{1}{\sum_{i=1}^c \left( \frac{\|x_i - c_i\|}{\|x_i - c_k\|} \right)^{2/(m-1)}} \quad (10)$$

where  $\|x_i - c_i\|$  is the Distance from point i to current cluster centre j,  $\|x_i - c_k\|$  is the distance from point i to other cluster centers k.

This iteration will stop when  $\max \left( |u_{ij}^{(k+1)} - u_{ij}^k| \right) < \varepsilon$ , where  $\varepsilon$  is a termination criterion between 0 and 1, whereas k are the iteration steps. This procedure converges to a local minimum or a saddle point of  $J_m$ .

### Subtractive Clustering

Subtractive clustering is based on the measurement of the density of data points. The more neighbors a point has, the more density it has. That point is chosen as center point. In this method, each point is a candidate for being a cluster center; therefore, the density is calculated as follows:

$$P_i^* = \sum_{j=1}^n e^{-\alpha \|x_i - x_j\|^2} \quad (11)$$

$$\alpha = \frac{\gamma}{r_a} \quad (12)$$

$P_i^*$  is the potential- value i-data as a cluster centre

$\alpha$  is the weight between i-data to j-data

$x$  is the data point

$\gamma$  is variables (commonly set 4)

$r_a$  is a positive constant called cluster radius

If a data point has many neighbors, then it has a high density accordingly. The point with  $P_i^*$ , the highest density, will be  $c_1 = (d_1, e_1)$ , the first cluster center.

$$P_i^* = P_i^* - P_k^* e^{-\beta \|x_i - c_k\|^2} \quad (13)$$

$$\beta = \frac{4}{r_b^2} \quad (14)$$

$$r_b = r_a * \eta \quad (15)$$

$P_i^*$  is the new potential- value i-data

$P_k^*$  is the potential- value i-data as a cluster centre

$\alpha$  is the weight between i-data to j-data

$c$  is the cluster center of data

$\beta$  is the weight of i, data to cluster centre

$r_b$  is the distance between cluster centre

$\eta$  is the quash factor

The point with the highest density,  $P_k^*$ , will be a cluster center as long as it satisfies the following condition:

$$\frac{d_{min}}{r_a} + \frac{P_k^*}{P_1^*} \geq 1$$

$d_{min}$  is the minimal distance between  $c_1$  and all previously found cluster centers. The data point is still taken as the next cluster center  $c_2$ . Unless a potential cluster center meets the above described criteria, it is rejected as a cluster center and its potential is decreased to 0. In that case, the new possible cluster center is the data point with the next highest potential  $P_k^*$ , which is subject to a retest. Until the following criterion is satisfied, the clustering goes on.

$$P_k^* < \varepsilon P_1^*$$

where  $\varepsilon$  is a small fraction (Chiu S. L., 1994).

### Outlier Data Analysis

The outlier data analysis was performed via three methods, namely Mahalanobis Distance, Cook Distance, and Leverage Point (Franklin & Thomas, 2000; Field, 2009).. The methods were employed on SPSS 17. The analysis was conducted on the training dataset and test dataset for the 218 students registered to the course during the 2011-2012 academic year. Next, the analysis was carried out on the data on 95 students taking the same course during the 2012-2013 academic year.

A total of 15 pieces of data, which were identified through three different methods, were excluded from the dataset. The exclusion of the outlier data meant that the training dataset and test dataset, which contained 218 subjects, were now based on 203 participants. Similarly, the number of pieces of data included in the verification dataset was reduced to 85 after the outlier data were excluded.

### Data Normalization

In clustering analysis, it was assumed that data has normal distribution. Thus, the Shapiro Wilk test was performed for the dataset to be included in clustering analysis. Table 1 presents the results of the Shapiro Wilk test.

The results of the Shapiro Wilk test indicated that the  $p$  value was higher than 0.05, confirming  $H_0$  hypotheses for recency, frequency, and quiz. In other words, the data for these independent variables had normal distribution at a confidence level of 95%. On the other hand, the results suggested that  $H_0$  was disproved for monetary and midterm exam. In other words, the data for them did not have normal distribution at a confidence level of 95%. There are several techniques for transforming data that do not have normal distribution. Calculating the square root, one of such techniques, was performed and tested on

**Table 1.** Test of Normality

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Recency	.059	143	.200*	.989	143	.309
Frequency	.080	143	.026	.977	143	.015
Monetary	.076	143	.043	.957	143	.000
Quiz	.081	143	.022	.974	143	.009
Midterm Exam	.083	143	.017	.957	143	.000
SqrtMidterm1	.054	143	.200*	.975	143	.009
AP	.062	143	.200*	.982	143	.055

\*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

**Table 2.** The Accuracy Ratios by the Number of Clusters in the Training Data

Number of (Clusters)	2	3	4	5	6	7	8	9
Accuracy Ratio	87.44	87.91	88.83	88.87	89.63	89.39	90.93	90.11

**Table 3.** The Cluster Centers Identified with K-Means Clustering

Cluster Center	Recency	Frequency	Monetary	Midterm Exam	Quiz
1	20.51	3.82	3.76	51.64	5.91
2	14.92	5.16	4.40	80.31	6.98
3	14.41	4.22	3.98	42.94	4.25
4	10.67	5.31	5.01	67.65	6.91
5	21.15	4.48	4.15	65.25	5.67
6	13.00	2.13	2.61	27.89	4.27
7	8.97	3.77	3.59	51.91	5.64
8	12.80	5.34	4.93	95.71	8.87

**Table 4.** The Gaussian Curve Equations for the Variables Specified through K-Means Clustering.

No	Recency	Frequency	Monetary	Midterm Exam	Quiz
$\mu_1$	$e^{-\frac{1}{2}\left(\frac{x-20.51}{5.13}\right)^2}$	$e^{-\frac{1}{2}\left(\frac{x-3.82}{2.12}\right)^2}$	$e^{-\frac{1}{2}\left(\frac{x-3.76}{1.29}\right)^2}$	$e^{-\frac{1}{2}\left(\frac{x-51.64}{15.20}\right)^2}$	$e^{-\frac{1}{2}\left(\frac{x-5.91}{1.77}\right)^2}$
$\mu_2$	$e^{-\frac{1}{2}\left(\frac{x-14.92}{5.13}\right)^2}$	$e^{-\frac{1}{2}\left(\frac{x-5.16}{2.12}\right)^2}$	$e^{-\frac{1}{2}\left(\frac{x-4.40}{1.29}\right)^2}$	$e^{-\frac{1}{2}\left(\frac{x-80.31}{15.20}\right)^2}$	$e^{-\frac{1}{2}\left(\frac{x-6.98}{1.77}\right)^2}$
$\mu_3$	$e^{-\frac{1}{2}\left(\frac{x-14.41}{5.13}\right)^2}$	$e^{-\frac{1}{2}\left(\frac{x-4.22}{2.12}\right)^2}$	$e^{-\frac{1}{2}\left(\frac{x-3.98}{1.29}\right)^2}$	$e^{-\frac{1}{2}\left(\frac{x-42.94}{15.20}\right)^2}$	$e^{-\frac{1}{2}\left(\frac{x-4.25}{1.77}\right)^2}$
$\mu_4$	$e^{-\frac{1}{2}\left(\frac{x-10.67}{5.13}\right)^2}$	$e^{-\frac{1}{2}\left(\frac{x-5.31}{2.12}\right)^2}$	$e^{-\frac{1}{2}\left(\frac{x-5.01}{1.29}\right)^2}$	$e^{-\frac{1}{2}\left(\frac{x-67.65}{15.20}\right)^2}$	$e^{-\frac{1}{2}\left(\frac{x-6.91}{1.77}\right)^2}$
$\mu_5$	$e^{-\frac{1}{2}\left(\frac{x-21.15}{5.13}\right)^2}$	$e^{-\frac{1}{2}\left(\frac{x-4.48}{2.12}\right)^2}$	$e^{-\frac{1}{2}\left(\frac{x-4.15}{1.29}\right)^2}$	$e^{-\frac{1}{2}\left(\frac{x-65.25}{15.20}\right)^2}$	$e^{-\frac{1}{2}\left(\frac{x-5.67}{1.77}\right)^2}$
$\mu_6$	$e^{-\frac{1}{2}\left(\frac{x-13.00}{5.13}\right)^2}$	$e^{-\frac{1}{2}\left(\frac{x-2.13}{2.12}\right)^2}$	$e^{-\frac{1}{2}\left(\frac{x-2.61}{1.29}\right)^2}$	$e^{-\frac{1}{2}\left(\frac{x-27.89}{15.20}\right)^2}$	$e^{-\frac{1}{2}\left(\frac{x-4.27}{1.77}\right)^2}$
$\mu_7$	$e^{-\frac{1}{2}\left(\frac{x-8.97}{5.13}\right)^2}$	$e^{-\frac{1}{2}\left(\frac{x-3.77}{2.12}\right)^2}$	$e^{-\frac{1}{2}\left(\frac{x-3.59}{1.29}\right)^2}$	$e^{-\frac{1}{2}\left(\frac{x-51.91}{15.20}\right)^2}$	$e^{-\frac{1}{2}\left(\frac{x-5.64}{1.77}\right)^2}$
$\mu_8$	$e^{-\frac{1}{2}\left(\frac{x-12.80}{5.13}\right)^2}$	$e^{-\frac{1}{2}\left(\frac{x-5.34}{2.12}\right)^2}$	$e^{-\frac{1}{2}\left(\frac{x-4.93}{1.29}\right)^2}$	$e^{-\frac{1}{2}\left(\frac{x-95.71}{15.20}\right)^2}$	$e^{-\frac{1}{2}\left(\frac{x-8.87}{1.77}\right)^2}$

recency and midterm exam, which indicated that the p values were higher than 0.05. Thus, they were incorporated into the dataset with normal distribution.

**RESULTS**

The present study attempted to predict the distance education students' academic performance through three different clustering methods. The findings revealed by K-means, fuzzy C-means and subtractive clustering methods are presented below.

**K-Means Clustering Method**

Table 2 presents the accuracy ratios of the model that was formed on the basis of the cluster centers identified according to K-means clustering. The findings suggest that the best accuracy ratio was provided by the model with eight cluster centers. Table 3 presents the cluster centers of this model, which had five inputs, formed according to K-means clustering. Gaussian curve is used in Sugeno fuzzy logic. The curve is comprised of two parameters (Fuzzy Logic Membership Function, 2012).

$$gauss(x; c; \sigma) = e^{-\frac{1}{2}(\frac{x-c}{\sigma})^2} \quad (16)$$

where c stands for the center of the curve whereas  $\sigma$  represents standard deviation.

Gaussian curves were formed in accordance with the cluster centers presented in Table 4. The following

is the Gaussian curve for the cluster center 1 for Recency.

$$gauss(x; 20.51; 5.13) = e^{-\frac{1}{2}(\frac{x-20.51}{5.13})^2}$$

A Gaussian curve equation was identified for each of the variables on the basis of the cluster centers specified through K-means clustering. The Gaussian curve equations for the input variables are presented in Table 4.

The membership functions are presented in Figure 1. If recency membership function graph depicted in figure 1 is closely examined, it could be seen that there are 8 curves, each of which is composed of cluster centers found by using K-means clustering algorithm. In Sugeno fuzzy logic, the identification of membership functions needs to be followed by the identification of functions for rules. The rules in Sugeno fuzzy logic are in the form of if-then. In the if section, input variables are entered into the system, whereas the then section includes a system where the rules exist. The results for each rule are finalized through formula 17.

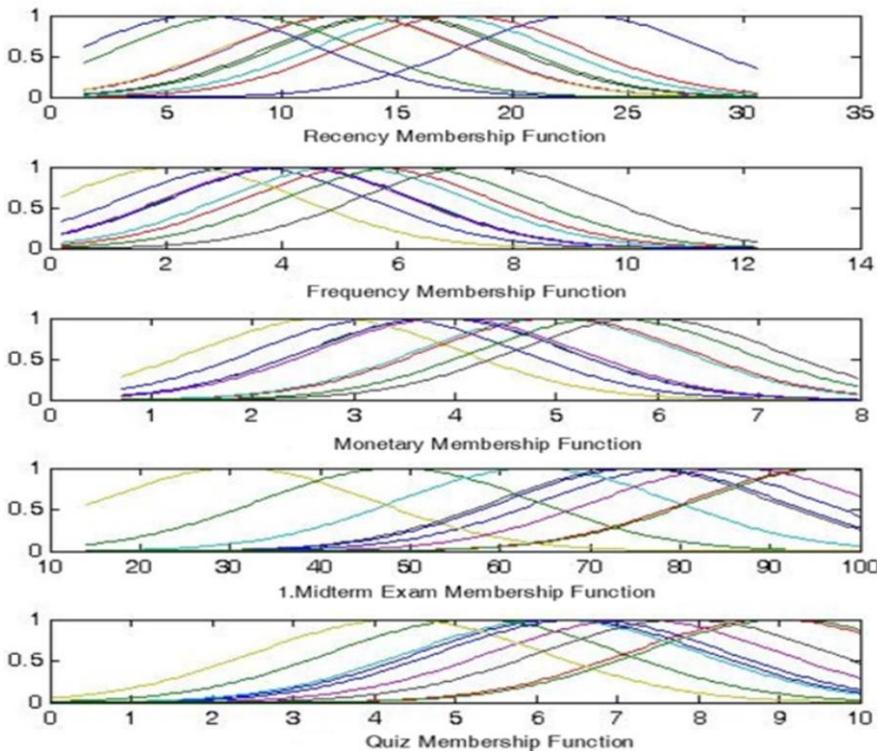
*If  $x_1 \in Recency$  and  $x_2 \in Frequency$  ve  $x_3$*

*$\in Monetary$  and  $x_4$*

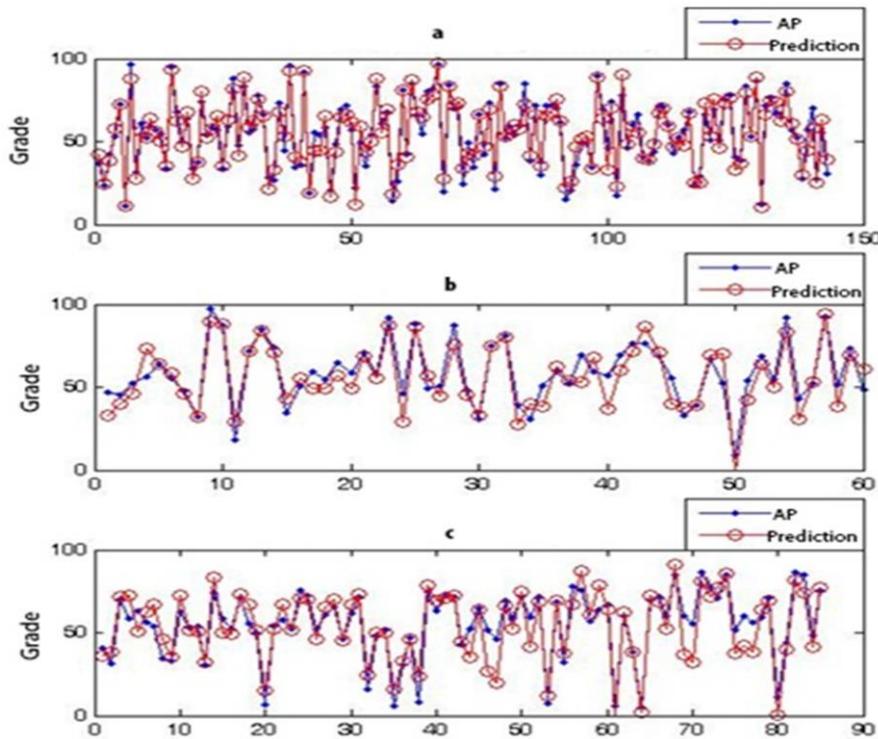
*$\in Midterm$  and  $x_5 \in Quiz$*

*then  $Z \rightarrow w^k = \beta_0^k + \beta_1^k x_1 + \beta_2^k x_2 +$*

$$\beta_3^k x_3 + \beta_4^k x_4 + \beta_5^k x_5 \quad (17)$$



**Figure 1.** The graphs of the membership functions found through K-Means clustering



**Figure 2.** Comparison between AP and predictions for the training dataset (a), test dataset (b) and verification dataset (c) in the fuzzy model established via K-means clustering

The parameters for the rules were identified through the least squares method. An analysis of the data in the same way yielded equations for eight rules.

$$w^1 = 145.78 - 0.70x_1 - 18.89x_2 + 7.80x_3 - 3.10x_4 + 6.17x_5$$

$$w^2 = 23.29 - 1.47x_1 - 1.02x_2 - 3.02x_3 + 0.26x_4 + 3.02x_5$$

$$w^3 = -52.41 + 3.87x_1 + 4.51x_2 + 0.82x_3 - 0.33x_4 + 1.45x_5$$

$$w^4 = -1.26 + 0.50x_1 - 1.10x_2 - 3.66x_3 + 0.98x_4 + 4.27x_5$$

$$w^5 = 292.38 - 2.17x_1 + 9.26x_2 - 31.74x_3 - 0.55x_4 + 8.02x_5$$

$$w^6 = -5.03 - 0.86x_1 + 2.51x_2 - 0.54x_3 + 0.44x_4 + 4.98x_5$$

$$w^7 = -13.01 + 4.60x_1 - 1.47x_2 - 7.82x_3 + 0.54x_4 + 4.16x_5$$

$$w^8 = 58.99 + 1.83x_1 - 0.42x_2 - 0.61x_3 - 0.36x_4 + 5.62x_5$$

Figures 2-a, 2-b and 2-c present the year-end academic performance (AP) of the distance education students and the results obtained from the fuzzy model for the training dataset, test dataset, and verification dataset, respectively.

The model, which was formed on the basis of the cluster centers identified via K-means clustering, had an accuracy ratio of 90.93%.

When the test dataset, whose purpose was to test the training dataset, was incorporated into the model, the accuracy ratio was 84%. When the verification dataset was incorporated into the model, the accuracy ratio was 80.21%.

On the other hand, the analysis that did not exclude the outlier data from the system suggested that the accuracy ratios for the training dataset and verification dataset were 90.03% and 70.91%, respectively. In other words, the exclusion of the outlier data led to a significant difference in the model.

### Fuzzy C-Means Clustering

The accuracy ratios of the model that was formed on the basis of the cluster centers were identified according to fuzzy C-means clustering.

The findings suggest that the best accuracy ratio was provided by the model with nine cluster centers. The model, which was formed on the basis of the cluster centers identified via fuzzy C-means clustering, had an

accuracy ratio of 91.09%. The cluster centers of this model, which had five inputs, were formed according to fuzzy C-means clustering.

Figure 3 presents the membership functions for the inputs of the model with nine cluster centers

Table 5 presents the parameters for the rules calculated through the least squares method as a result of the membership functions identified via fuzzy C-means clustering.

Figures 4-a, 4-b and 4-c present the year-end

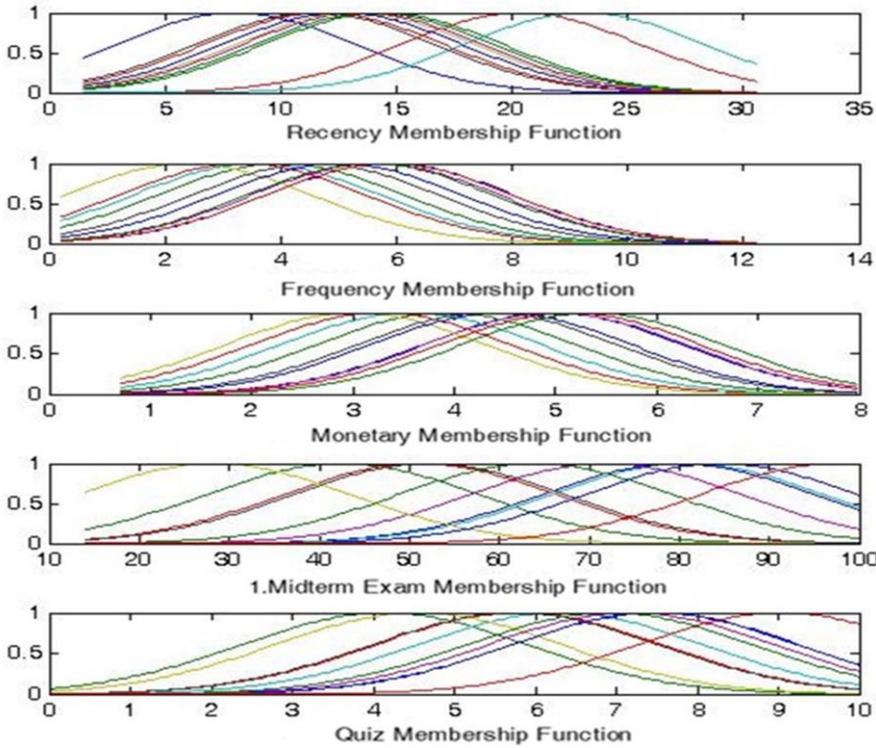


Figure 3. The graphs of the membership functions for Fuzzy C-Means clustering

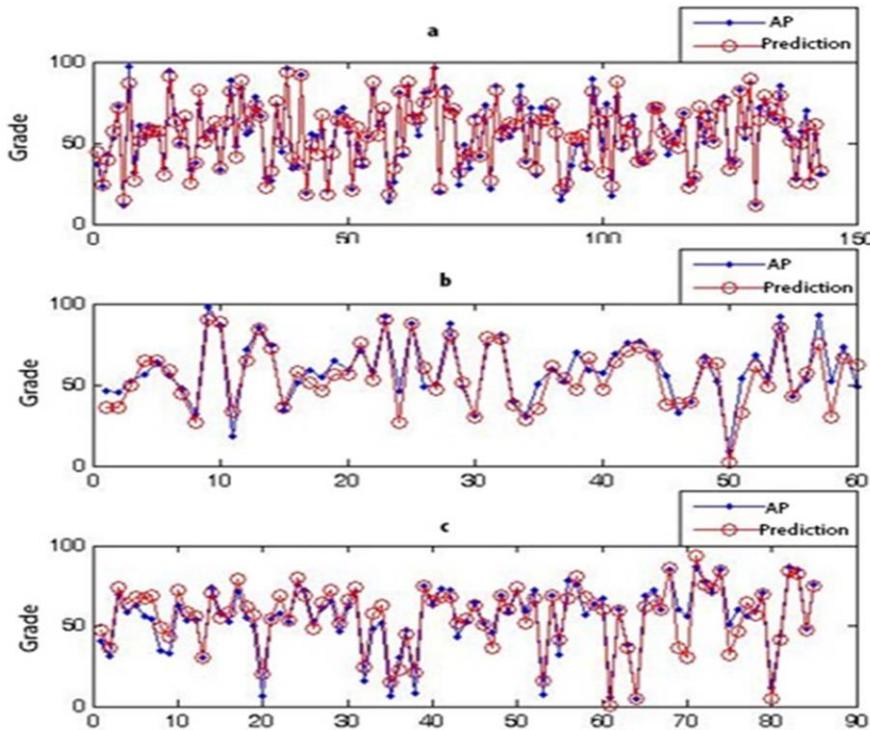


Figure 4. Comparison between the academic performance (AP) and predictions for the training dataset (a), test dataset (b) and verification dataset (c) in the fuzzy model established via fuzzy C-means clustering

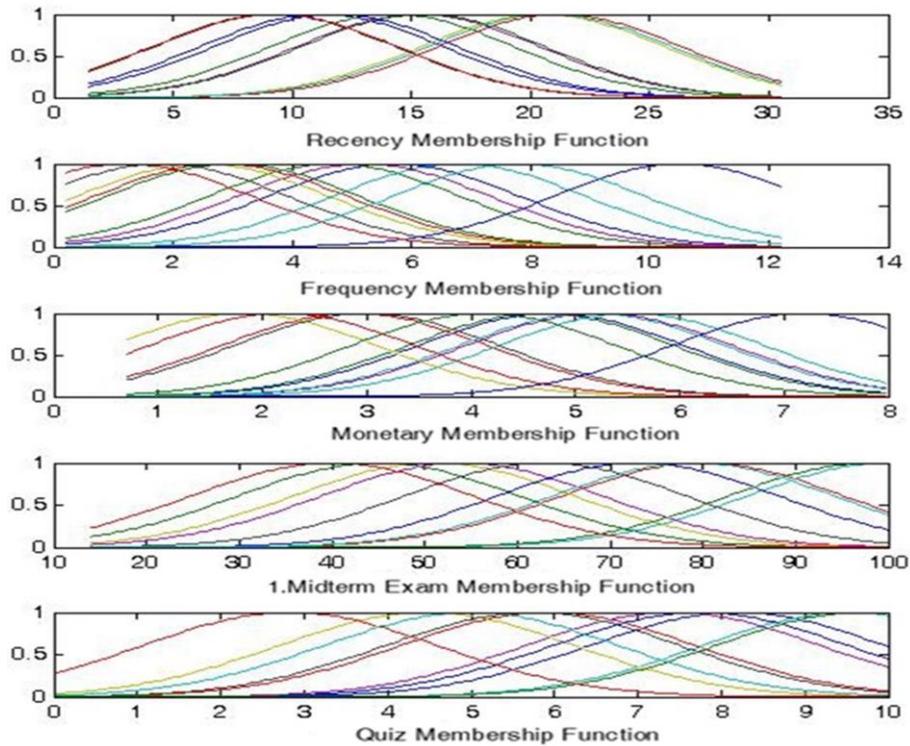


Figure 5. The membership functions in reference to Subtractive Clustering

Table 5. The Parameters for the Rules.

Rule No	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$
1	-156.10	-0.05	9.94	-4.71	2.62	1.68
2	215.18	1.27	-7.50	-0.54	-1.70	0.21
3	-244.46	-0.79	1.92	1.78	0.64	27.48
4	16.91	2.00	-3.77	9.94	-0.89	9.56
5	-41.12	0.73	1.65	9.30	0.05	1.60
6	-22.20	0.11	-2.92	8.54	0.40	8.14
7	2.43	-0.01	-0.69	-0.08	0.43	2.28
8	3.15	0.09	8.19	-5.65	0.33	-1.55
9	91.53	4.74	-8.62	-25.80	0.22	4.76

academic performance of the distance education students and the results obtained from the fuzzy model for the training dataset, test dataset, and verification dataset, respectively.

When the test dataset, whose purpose was to test the training data, was incorporated into the model, the accuracy ratio was 88.07%. When the verification dataset was incorporated into the model, the accuracy ratio was 83.94%.

On the other hand, the analysis that did not exclude the outlier data from the system suggested that the accuracy ratios for the training dataset and verification dataset were 91.15% and 65.14%, respectively. In other

words, the exclusion of the outlier data led to a significant difference in the model.

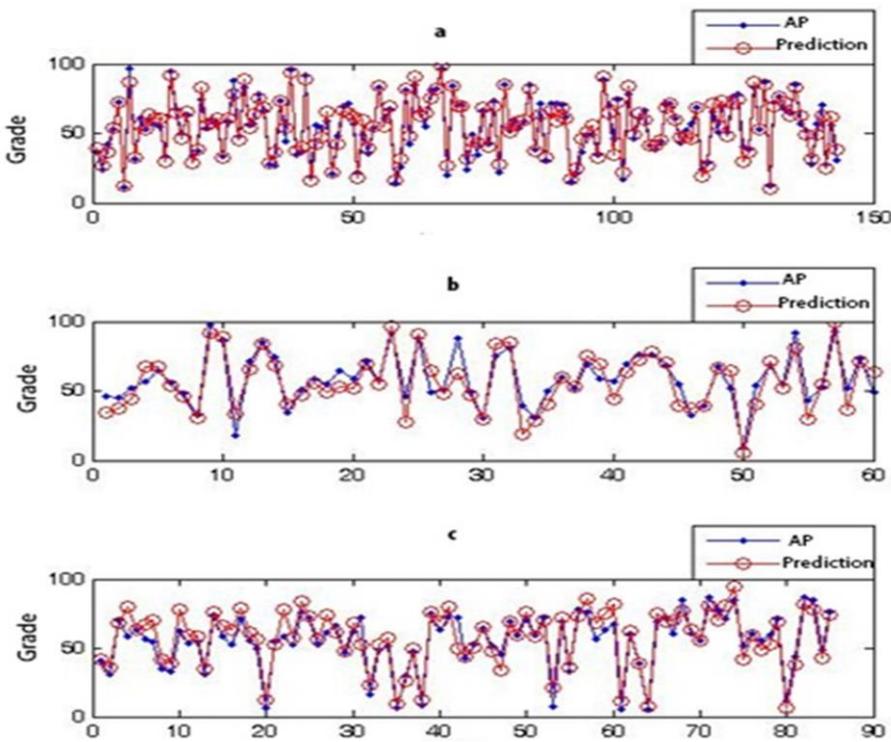
### Subtractive Clustering

Unlike K-means and fuzzy C-means, one does not have to enter the number of clusters in subtractive clustering. With its algorithm, the method determines the number of clusters on its own.

The algorithm identified 11 cluster centers in subtractive clustering. Figure 5 presents the membership functions formed in reference to the cluster centers.

**Table 5.** The Parameters for the Rules

Rule No	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$
1	198.42	-1.50	1.16	-9.80	-0.89	-1.81
2	1.45	0.85	1.69	-2.12	0.13	3.18
3	-9.98	1.39	5.90	-5.53	0.28	4.01
4	362.30	2.98	-11.14	-4.14	-1.54	-5.66
5	-86.32	-0.24	-0.40	7.84	0.49	10.75
6	50.80	-1.96	-1.81	7.56	-0.11	3.72
7	-17.35	3.12	-3.60	-4.03	0.65	4.09
8	-95.29	-0.56	1.27	2.00	1.01	8.34
9	-93.02	1.45	0.69	-1.02	-0.55	21.75
10	28.40	-1.49	-2.12	0.00	0.12	-0.79
11	19.71	0.13	5.15	-1.07	-0.18	2.65



**Figure 6.** Comparison between AP and predictions for the training dataset (a), test dataset (b) and verification dataset (c) in the fuzzy model established via subtractive clustering

Table 6 presents the parameters for the rules calculated on the basis of the membership functions revealed through subtractive clustering.

Figures 6-a, 6-b and 6-c present the year-end academic performance (AP) of the distance education students and the results obtained from the fuzzy model for the training dataset, test dataset, and verification dataset, respectively.

The model, which was formed on the basis of the cluster centers identified via subtractive clustering, had an accuracy ratio of 91.33%.

When the test dataset, whose purpose was to test the training data, was incorporated into the model, the accuracy ratio was 86.28%. When the verification dataset was incorporated into the model, the accuracy ratio was 83.24%.

**Table 6.** The Accuracy Ratios for Numerical Predictions of AP

Accuracy Ratio (%)						
Methods	Number of Clusters	Training Dataset	Test Dataset	Verification Dataset	Mean	
K-Means	8	90.93	84.00	80.21	85.05	
Fuzzy C-Means	9	91.09	88.07	83.94	87.70	
Subtractive	11	91.33	86.28	83.24	86.95	

**Table 8.** The Accuracy Ratios for the Classification of AP as Passed or Failed

Methods	Number of Clusters	Training Dataset	Test Dataset	Verification Dataset	Mean
K-Means	8	92.31	83.33	90.59	88.74
Fuzzy C-Means	9	93.01	88.33	95.29	92.21
Subtractive	11	93.71	86.67	96.47	92.28

On the other hand, the analysis that did not exclude the outlier data from the system suggested that the accuracy ratios for the training dataset and verification dataset were 89.75% and 76.40%, respectively. In other words, the exclusion of the outlier data led to a significant difference in the model.

**Comparison of the Results**

Table 7 presents the accuracy ratios for the training dataset and test dataset for all the three methods. In terms of the mean scores, the best result was produced by fuzzy C-means clustering.

Table 8 presents the accuracy ratios for the categorical classification of the students as passed or failed. In terms of the mean scores, the best result was produced by subtractive clustering. The best result was quite close to the one produced by Fuzzy C-Means clustering.

**DISCUSSION AND CONCLUSIONS**

The present study was carried out with the data on the students who were registered to an online course called Basic Computer Sciences. Further studies are needed to collect more data and on different courses and make comparisons with this study. The present study did not take the demographics of the participants into account. Further studies could establish a model considering their demographics, with results being discussed in comparison with those of the present study. More studies are planned to use alternative methods to the one used in the present study, such as the least squares method that will improve the results by predicting the parameters for the rules of the model.

The studies in the literature have commonly focused on predicting students’ AP in reference to the results of the tests they take during the semester, project assessments, demographics, academic background, socio-economic background, quiz results, assignment results and activities on platforms like forums (Lykourantzou, Giannoukos, Nikolopoulos, Mpardis, & Loumos, 2009; Vandamme, Meskens, & Superby, 2007; Zafra & Ventura, 2009; Kalles & Pierrakeas, 2006). The crucial difference between the present study and others is that the data were collected through the logs kept in the LMS on which the classes were taught. The data involved recency, frequency, monetary and quiz results. Only the paper-based midterm exam was entered into the system by hand. The model presented by the present study can easily predict distance education students’ year-end AP with minimum performance and a high accuracy ratio.

The extraction in fuzzy logic is generally based on two types, namely Mamdani and Takagi-Sugeno. In their study in 2013, Yildiz, Bal and Gulsecen used Mamdani extraction. Fuzzy logic-based models established via Mamdani extraction involve classical fuzzy logic, expert fuzzy, and gene-fuzzy logic, which is based on the optimization of the intervals for membership functions (Yildiz, Bal, & Gulsecen, 2013). In that study, their models could predict AP by 65% to 81%, and the best result was produced by gene-fuzzy logic hybrid model. In the present study, on the other hand, the results were obtained from Takagi-Sugeno extraction. The rules for the model were identified through subtractive clustering, K-means clustering, and fuzzy C-means clustering. The model was based on three datasets, namely the training dataset, test dataset, and verification dataset. In terms of the mean scores, the best result was produced by fuzzy C-means clustering. The mean accuracy ratio was

87.70%, suggesting that attempts to predict year-end AP of the distance education students were subject to an error margin of almost 12%. The prediction was based on 8-week data. The numerical prediction was followed by the categorization of the students as passed or failed. In this respect, the best result was produced by subtractive clustering, and the accuracy ratio was 92.28%. In conclusion, by using the aforementioned methods we were able to predict distance education students' year-end academic performance in terms of their grades and pass-fail categorization with high accuracy. However, future studies analyzing data from other online courses with more participants are needed to corroborate our results.

## REFERENCES

- Borkar, S. & Rajeswari, K. (2013). Predicting students' academic performance using education data mining, *International Journal of Computer Science and Mobile Computing*, 2(7), 273-279.
- Chiu, S. L. (1994). Fuzzy model identification based on cluster estimation. *Journal of Intelligent and Fuzzy Systems*, 2, 267-278.
- Chou, S.W., & Liu, C.H. (2005). Learning effectiveness in a web-based virtual learning environment: a learner control perspective. *Journal of Computer Assisted Learning*, 21, 65-76. doi: 10.1111/j.1365-2729.2005.00114.x
- EducationDynamics. *EducationDynamics Survey: Many Online Learners Never Seek Help Before Dropping Out*. Retrieved June 01, 2013, from <http://www.educationdynamics.com/News/Press-Releases/2009-Press-Releases/EducationDynamics-Survey--Many-Online-Learners-Neve.aspx>
- Field, A. (2009). *Discovering statistics using SPSS*. Dubai: Sage .
- Franklin, S., & Thomas, S. (2000). Robust multivariate outlier detection using Mahalanobis' distance and modified Stahel-Donoho estimators. *Proceedings of the Second International Conference on Establishment Surveys*, Buffalo, 697-706.
- Fuzzy Logic Membership Function. *A Research Note*, Retrieved May 01, 2013, from <http://www.bindichen.co.uk/post/AI/fuzzy-inference-membership-function.html>.
- Gan, G., Ma, C., & Wu, J. (2007). *Data clustering theory, algorithms, and application*. Virginia: American Statistical Association and the Society for Industrial and Applied Mathematics.
- Hossen, J., Rahman, A., Sayeed, S., Samsuddin, K., & Rokhani, F. (2011). A modified hybrid fuzzy clustering algorithm for data partitions. *Australian Journal of Basic & Applied Sciences*, 5(8).
- Kalles, D., & Pierrakeas, C. (2006). Analyzing student performance in distance learning with genetic algorithms and decision trees. *Applied Artificial Intelligence*, 20(8), 655-674.
- Kim, M.-S., & Kong, S.-G. (1997). Fuzzy rule generation using the clustering algorithm for cooperative behavior of small robots. *Proceedings of the 2nd Asian Control Conference*, Seoul, 623-626.
- Lykourentzou, I., Giannoukos, I., Nikolopoulos, V., Mpardis, G., & Loumos, V. (2009). Dropout prediction in e-learning courses through the combination of machine learning techniques. *Computers & Education*, 950-965.
- Mathworks. *What Is Sugeno-Type Fuzzy Inference?*. Retrieved May 10, 2013, from <http://www.mathworks.ch/ch/help/fuzzy/what-is-sugeno-type-fuzzy-inference.html>
- Moertini, V. S. (2002). Introduction to five clustering algorithms. *Integral*, 87-96.
- Oliveira, J. V., & Pedrycz, W. (2007). *Advances in fuzzy clustering and its application*. Chichester: John Wiley & Sons, Ltd.
- Priyono, A., Ridwan, M., Ahmad Jais Allias, R. A., Rahmat, O., Hassan, A., & Ali, M. A. (2005). Generation of fuzzy rules with subtractive clustering. *Jurnal Teknologi*, 143-153.
- Psycharis, S., Chalatzoglidis, G., & Kalogiannakis, M. (2013). Moodle as a learning environment in promoting conceptual understanding for secondary school students. *Eurasia Journal of Mathematics, Science & Technology Education*, 9(1), 11-21.
- Ramesh, V., Parkavi, P. & Ramar, K. (2013). Predicting student performance: a statistical and data mining approach, *International Journal of Computer Applications*, 63(8), 35-39.
- Ren, Q., 2006, *Type-2 Takagi-Sugeno-Kang fuzzy logic modeling using subtractive clustering*, Thesis (Ph.d), University of Montreal.
- Sari, M., Murat, Y. Ş., & Kırbalı, M. (2013). Fuzzy modelling approach and application. Retrieved May 05, 2013, from <http://www3.dpu.edu.tr/akademik/fbe/wp-content/uploads/2005/12/09.pdf>
- Sivanandam, S., Sumathi, S., & Deepa, S. (2007). *Introduction to fuzzy logic using Matlab*. New York: Springer.
- Takagi, T., & Sugeno, M. (1985). Fuzzy identification of systems and its applications to modelling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, 15, 116-132.
- Tekin, A., (2014). Early prediction of students' grade point averages at graduation: a data mining approach, *Eurasian Journal of Educational Research*, 54, 207-226.
- Vandamme, J. -P., Meskens, N., & Superby, J. -F. (2007). Predicting academic performance by data mining methods. *Education Economics*, 405-419.
- Wei, J., Lin, S., & Wu, H., (2010). A review of the application of RFM model. *African Journal of Business Management*, 4(19), 4199-4206.
- Yildiz, O., Bal, A., & Gulsecen, S. (2013). Improved fuzzy modelling to predict the academic performance of distance education students. *The International Review of Research in Open and Distance Learning*, 14(5), 144-165.
- Zafra, A., & Ventura, S. (2009). Predicting student grades in learning management systems with multiple instance genetic programming. *Educational Data Mining*, 307-314.

