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

**The Eurasia Proceedings of Educational & Social Sciences (EPESS), 2016**

**Volume 5, Pages 190-195**

**ICRES 2016: International Conference on Research in Education and Science**

# PREDICTION OF STUDENTS' SUCCESS IN MATHEMATICS BY A CLASSIFICATION TECHNIQUE VIA POLYHEDRAL CONIC FUNCTIONS

Nur Uylas Sati
Muğla Sıtkı Koçman Üniversitesi

**Abstract**: There has been a lot of work that has been already done using data mining in educational institutes and organizations and due to great success, the people are getting more and more interested in this field. In this paper a not long ago developped polyhedral conic functions classification algorithm is applied to a dataset of student performance in mathematics. Implemantations are made in MATLAB and WEKA. Results are shown in tables. This method can be applied to various datasets related with education. It will be helpfull for all educational fields.

***Keywords***: Educational data mining, classification, polyhedral conic functions, mathematics education

## Introduction

Data mining, also called Knowledge Discovery in Databases (KDD), is the field of discovering novel and potentially useful information from large amounts of data. Data mining has been applied in a great number of fields, including marketing, bioinformatics, medicine, business, education, management etc. Data mining uses many techniques such as supervised and unsupervised classification, decision trees, neural networks, naive bayes, clustering and many others. All these data mining techniques were expressed by Alpaydın (2012). Educational data mining is a field that exploits statistical, machine-learning, and data-mining (DM) algorithms over the different types of educational data (Romero & Ventura, 2010). Its main objective is to analyze these types of data in order to resolve educational research issues (Barnes, Desmarais, Romero & Ventura, 2009).
Educational data mining (also referred to as “EDM”) is defined as the area of scientific inquiry centered around the development of methods for making discoveries within the unique kinds of data that come from educational settings, and using those methods to better understand students and the settings which they learn in (McGaw, Peterson & Baker, 2010 ). EDM uses prediction, clustering, relationship mining, discovery with models, distillation of data for human judgment methods. All these methods are expressed in (Mc Gaw et al., 2010). Oxford, UK: Elsevier). A large scaled literature review of various significant researches in the area of EDM ranging from Year 2002 to 2014 was presented in Thakar, Mehta and Manisha (2015).Prediction method have been used in this paper. The goal of prediction method is developing a model which can infer a single aspect of the data (predicted variable) from some combination of other aspects of the data (predictor variables). The key applications of this method are detecting student behaviors (e.g. gaming the system, offtask behavior, slipping); developing domain models; predicting and understanding student educational outcomes (Mc Gaw et al., 2010). In this paper we benefit from classification via polyhedral conic functions to predict whether the student will pass or fail the mathematics course in terms of specific attributes. In methods and procedures section, a previously proposed separation algorithm via polyhedral conic functions have been given and expressed in detail. Later on with minor changes another algorithm have been constructed and proposed. Description and preprocessing of the dataset has been given. Implemantations have been made in Weka and Matlab. In results and findings section, obtained results and comparements have been proposed in tables. And finally conclusion has been made in the last section.



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

*Corresponding author: Nur Uylas Sati-Mail: nursati@mu.edu.tr

## Methods and Procedures

In this section firstly a previously proposed separation algorithm via polyhedral conic functions have been indicated. Later on, modifications in this algorithm have been expressed and the used algorithm have been proposed. Finally the dataset used in this study have been explained in detail.

### Separation via Polyhedral Conic Functions

Polyhedral conic functions have recently been presented to separate two finite point sets in $IR^n$ (Gasimov & Öztürk, 2006). Definition 1 and Lemma 1 given below were proposed in Gasimov and Öztürk (2006). And also a separation algorithm via PCFs was presented in the same paper.

**Definition 1:** A function $g: IR^n \times IR$ is called polyhedral conic if its graph is a cone and all its level sets, $S_\alpha = \left\{x \in IR^n : g(x) \le \alpha\right\}, \alpha \in IR$ are polyhedrons.

Given $w, a \in IR^n, \xi, \gamma \in IR, w'x = w_1x_1 + ... + w_nx_n$ is a scalar product of, *w* and *x* $\|x\|_1 = |x_1| + ... + |x_n|$ is a $l_1$ norm of the vector $x \in IR^n$, a polyhedral conic function $g_{(w,\xi,\gamma,a)} : IR^n \to IR$ defined as

$$g_{(w,\xi,\gamma,a)} : IR^n \to IR = w'(x-a) + \xi\|x-a\|_1 - \gamma$$

**Lemma 1:** A graph of the function $g_{(w,\xi,\gamma,a)}$ defined in (1) is a polyhedral cone with a vertex at $(a, -\gamma) \in IR^n \times IR$. This cone is called a polyhedral conic set and *a* its center.

**Algorithm 1**. PCF Algorithm.
Let *A* and *B* be given sets containing *m* and *p* *n*-dimensional vectors, respectively:
$A = \left\{a^i \in R^n, i \in I\right\}, \ B = \left\{b^j \in R^n, j \in J\right\}$ where $I = \left\{1, ..., m\right\}, J = \left\{1, ..., p\right\}$.

Step 0.( Initialization step) $l$=1, $I_l = I, A_l = A$ and go to Step 1.

Step 1. Let be $a^l$ an arbitrary point of $A_l$ . Solve subproblem $P_l$ .

$$(P_l) \quad \min\left(\frac{y' e_{|I_l|}}{|I_l|}\right)$$

$$w'(a^i - a^l) + \xi\|a^i - a^l\|_1 - \gamma + 1 \le y_i, \quad \forall i \in I_l,$$

$$-w'(b^j - a^l) - \xi\|b^j - a^l\|_1 + \gamma + 1 \le 0, \quad \forall j \in J,$$

$$y = (y_1, ..., y_m) \in R^m_+, w \in R^n, \xi \in R, \gamma \ge 1$$

Let $w^l, \xi^l, \gamma^l, y^l$ be a solution of $(P_l)$ and let

$$g_l(x) = g_{(w^l, \xi^l, \gamma^l, a^l)}(x)$$

and go to Step 2.

Step 2. Let $I_{l+1} = \left\{i \in I_l : g_l(a^i) + 1 > 0\right\}, A_{l+1} = \left\{a^i \in A_l : i \in I_{l+1}\right\}, l = l + 1$ and if $A_l \ne \emptyset$ go to Step 1.

Step 3. Define the function $g$(x) (separating the sets A and B) as

$$g(x) = \min_l g_l(x)$$

and stop.

### Modifications

Even though separation is hardly depends on the vertex of the cone, the initilization point (vertex) is chosen arbitrarily in step 1. In the same paper to solve this problem a new modified one was proposed. In this modified algorithm, for every point(data taken as vertex) the minimization problem in step 2 was solved and the one that separates the maximum number of points was chosen as the vertex point. But this modification is proper just when the set A (dataset) under consideration is not **too** large.We used clustering method to solve this problem because it is applicable to very large datasets. In clustering methods groups of objects that share common properties are formed (Kusiak, 2001). After applying clustering method, we assign the found center points of the clusters as the vertex points of PCFs.

### *Clustering Methods*

Various algorithms have been studied for clustering method (Anderberg, 1973). In this paper, both of respected, *k*-medoids and *k*- means algorithm is used.

***k*- medoids algorithm**
This method is proposed by Kauffmann and Rousseeuw (1990).

**Step1** Select $k$ initial points from the dataset called A.

**Step2** Every point in A is assigned to one of $k$ clusters due to center points selected.

**Step3** Due to clustering in Step2, figure out new center points of clusters and turn back to Step2 until no points change its cluster.

***k*-means algorithm**
This method is proposed by James MacQueen (1967).

**Step 1**. Choose a seed solution consisting of k centers (not necessarily belonging to A);

**Step 2**. Allocate data points to its closest center and obtain k-partition of A;

**Step 3**. Recompute centers for this new partition and go to Step 2 until no more data points change cluster.

After the modifications, constructed algorithm (called Algorithm 2) differs from Algorithm 1 just in Step 0(Initilization) and Step 1. This modified algorithm have been proposed as follows:

**Algorithm 2**

***Step 0.( Initialization step)*** Apply a clustering algorithm on set of *A*. Let *s* be the number of clusters and *l=1* . $I_l=I$.

***Step 1***. Let $a^l$ be the center of *l* th cluster . Solve subproblem $P_l$ .

$$(P_l) \quad \min\left(\frac{y' e_{|I_l|}}{|I_l|}\right)$$

$$w'(a^i - a^l) + \xi\|a^i - a^l\|_1 - \gamma + 1 \le y_i, \quad \forall i \in I_l,$$
$$-w'(b^j - a^l) - \xi\|b^j - a^l\|_1 + \gamma + 1 \le 0, \quad \forall j \in J,$$
$$y = (y_1, ..., y_m) \in R_+^m, w \in R^n, \xi \in R, \gamma \ge 1$$

Let $w^l, \xi^l, \gamma^l, y^l$ be a solution of $(P_l)$. Let

$$g_l(x) = g_{(w^l, \xi^l, \gamma^l, a^l)}(x)$$

***Step 2.*** If $l < s$, let *l=l+1* , $I_l = \left\{ i \in I_{l-1} : g_{l-1}\left(a^i\right) > 0 \right\}$
and go to *Step 1*.

***Step 3.*** Define the function *g(x)* (separating the sets *A* and *B*) as

$$g(x) = \min_l g_l(x)$$

and stop.

## Dataset Preparations

The dataset used in this study was received from UCI Machine Learning Repository (Cortez P.,2008) and it is called Student Performance Dataset. It approaches 395 students achievement in secondary education of two Portuguese high schools. The dataset is provided regarding the performance in mathematics subject.

### *Data Selection and Transformation*

The used data attributes are given in Table 1.

Table 1. Dataset attributes

| | | |
|---|---|---|
| 1 sex | student's sex | (binary: 'F' female or 'M' male) |
| 2 age | student's age | (numeric: from 15 to 22) |
| 3 address | student's home address type | (binary: 'U' urban or 'R' rural) |
| 4 famsize | family size | (binary: 'LE3' less or equal to 3 or 'GT3' greater than 3) |
| 5 Pstatus | parent's cohabitation status | (binary: 'T' living together or 'A' apart) |
| 6 Medu | mother's education | (numeric: 0 none, 1 primary education (4th grade), 2 5th to 9th grade, 3 secondary education or 4 higher education) |
| 7 Fedu | father's education | (numeric: 0 none, 1 primary education (4th grade), 2 5th to 9th grade, 3 secondary education or 4 higher education) |
| 8 Mjob | mother's job | (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other') |
| 9 Fjob | father's job | (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other') |
| 10 traveltime | home to school travel time | (numeric: 1 <15 min., 2 15 to 30 min., 3 30 min. to 1 hour, or 4 >1 hour) |
| 11 studytime | weekly study time | (numeric: 1 <2 hours, 2 2 to 5 hours, 3 5 to 10 hours, or 4 >10 hours) |
| 12 schoolsup | extra educational support | (binary: yes or no) |
| 13 higher | wants to take higher education | (binary: yes or no) |
| 14 internet | Internet access at home | (binary: yes or no) |
| 15 romantic | with a romantic relationship | (binary: yes or no) |
| 16 freetime | free time after school | (numeric: from 1 very low to 5 very high) |
| 17 Dalc | workday alcohol consumption | (numeric: from 1 very low to 5 very high) |
| 18 Walc | weekend alcohol consumption | (numeric: from 1 very low to 5 very high) |
| 19 health | current health status | (numeric: from 1 very bad to 5 very good) |
| 20 absences | number of school absences | (numeric: from 0 to 93) |
| 21 success | Success of student in mathematics | (binary: fail 0 pass 1) |

Some changes have been made for implemantation. These changes are given detailed below.

All binary attributes (1,3,4,5,12,13,14,15) are changed as {0 or 1}.

Mother's job and Father's job nominal attributes (8,9) are changed as {0 for at home, 1 for teacher, 2 for others}.
Feature selection is applied and the most ineffective attributes have been found by InfoGainAttributeEval in WEKA and they have been removed to increase the performance of the algorithm.
In Figure 1, attribute selection output after InfoGainAttributeEval function is shown, as can be seen, 0 ranked attributes are "age" and "absences" so these two attributes are removed from the dataset. Accordingly, the dataset used in implementation consists of 395 instances and 19 attributes including class atrribute.

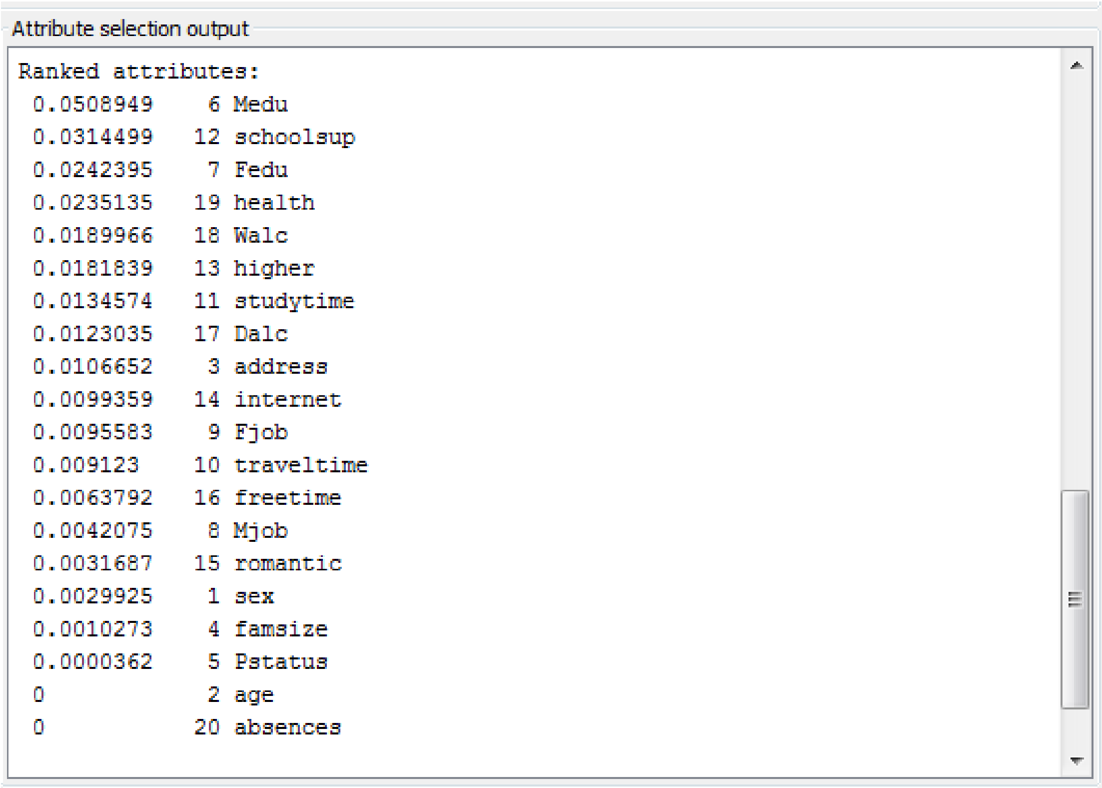

Figure 1. InfoGainAttributeEval function output

## Results and Findings

The accuracy and crossvalidation results of Algorithm 2 with *k*-means and Algorithm 2 with *k*-medoids with different *k* values, and also for comparison Classification Via Clustering in WEKA have been given respectively in Table 2. The implemantations have been made in MATLAB. Accuracy value is the ratio between the number of well classified data (students) to the number of training set elements (students). In testing phase for evaluation of learning algorithms' performance, 10-fold crossvalidation method is used.

Table 2. Algorithms' accuracy and crossvalidation results

| | Algorithm2 with *k*-means | | | Algorithm2 with *k*-medoids | | | Class. Via Cluster. |
|---|---|---|---|---|---|---|---|
| | *k*=2 | *k*=5 | *k*=10 | *k*=2 | *k*=5 | *k*=10 | |
| Accuracy % | 65.82 | 64.55 | 62.78 | 45.72 | 50.36 | 63.50 | 51.89 |
| Crossvalidation % | 59.23 | 58.92 | 56.90 | 40.34 | 46.61 | 59.88 | 55.44 |

In cross-validation, the dataset D is randomly split into 10 mutually exclusive subsets (the folds) D1,D2,..,D10 of approximately equal size. The inducer is trained and tested 10 times; each time t $\in\{1, 2,...,10\}$ , it is trained on D\Dt and tested on Dt. The cross validation estimate of accuracy is the overall number of correct classifications, divided by the number of instances in the data set (Kohavi, 1995). In this direction, crossvalidation results are more important then the accuracy results because desired goal is to predict the future data not the existing known data. When we discuss the results from Table 2. It can be told that Algorithm 2 with *k*=10-medoids is best one. Also as can be seen from the Table 2 results can change depending on the selection of the *k* value in the used clustering algorithm but we cannot make a certain approach. Approximations about *k* value can change depending on the used dataset properties.

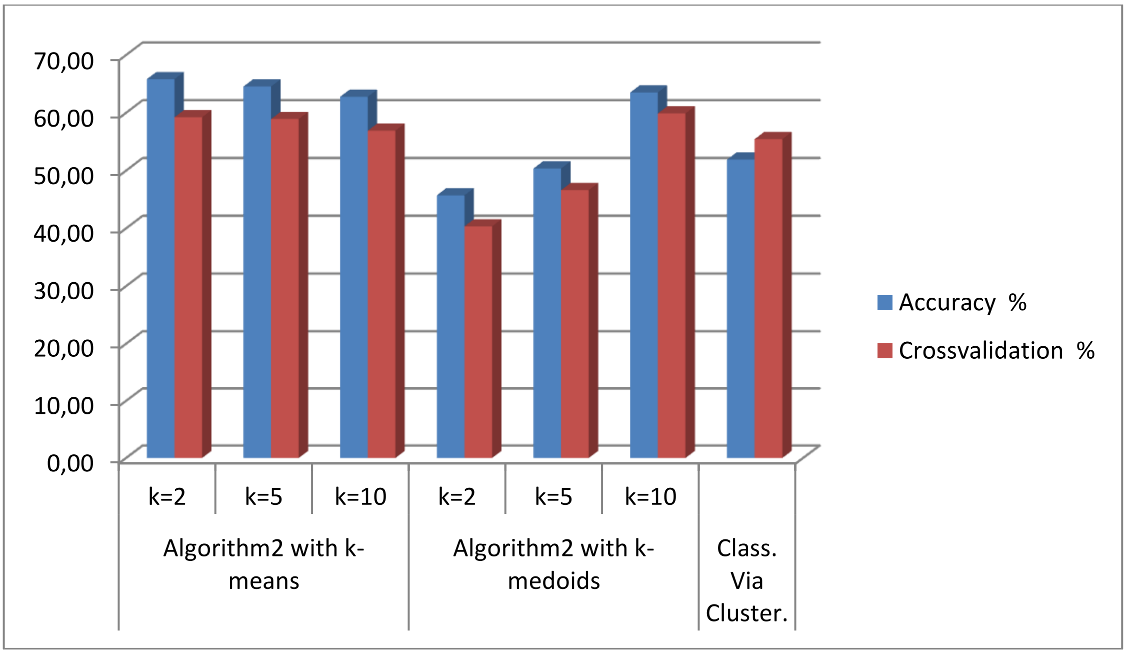


Figure 2. Graphical represantation of algorithms' accuracy and crossvalidation results

## Conclusion

In this paper, we propsed an existing classification model with minor changes, for prediction of students' success in mathematics. We can say with a specific accuracy in terms of the used algorithm, a student can fail or pass mathematics course. In paralel with these results, educationists or students can manage the period to increase the success. The suggested algorithm can be used in various fields of education with different datasets for binary classification. And also this algorithm can be developped for multi-class classification problems such as students assignment to A,B,C,.. classes in institutes. This subject can be studied in future works.

## References

Alpaydın, E. (2010). Introduction To Machine Learning. *The MIT Press Cambridge*, Massachusetts London, England.

Anderberg, M. R. (1973). Cluster Analysis for Applications. *Academic Press*.

Baker, R.S.J.d. (2010) Data Mining for Education. In McGaw, B., Peterson, P., Baker, E. (Eds.) *International Encyclopedia of Education (3rd edition), vol. 7, pp. 112-118. Oxford, UK: Elsevier.*

Barnes, T., Desmarais, M., Romero, C., & Ventura, S. (2009) *(Eds.)* Educational Data Mining 2009*: 2nd International Conference on Educational Data. Cordoba, Spain. July 1-3.*

Cortez P. (2008). Students' Performance Data Set. *UCI repository of machine learning databases*. Technical report, Department of Information and Computer Science, University of California, Irvine, available online at: https://archive.ics.uci.edu/ml/datasets/Student+Performance

Gasimov, R.N. & Öztürk, G. (2006). Separation via Polyhedral Conic Functions. *Optimization Methods and Software, 21/ 4 :527-540.*

Kaufman, L. & P, J, Rousseeuw. (1990). Finding Groups in Data. *John Wiley & Sons, New York.*

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection, *International Joint Conference on Artificial Intelligence*.

Kusiak, A. (2001). Data Analysis: Models and Algorithms. *Proc. SPIE Vol. 4191, pp. 1-9.*

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, Vol. 40, No. 6, pp. 601-618*

Thakar, P. & Mehta, A. & Manisha. (2015). Performance Analysis nad Prediction in Educational Data Mining: A Research Travelogue. *International Journal of Computer Applications 110(15): 60-68.*