Je-LKS
Journal of e-Learning and Knowledge Society
The Italian e-Learning Association Journal
PEER REVIEWED PAPERS
LEARNING ANALYTICS: FOR A DIALOGUE BETWEEN TEACHING PRACTICES AND EDUCATIONAL RESEARCH
Vol. 15, n.3, 2019
ISSN: 1826-6223 | eISSN: 1971-8829

UNIVERSITY DROPOUT PREDICTION THROUGH EDUCATIONAL DATA MINING TECHNIQUES: A SYSTEMATIC REVIEW

Francesco Agrusti, Gianmarco Bonavolontà, Mauro Mezzini

Roma Tre University
francesco.agrusti@uniroma3.it, gianmarco.bonavolonta@uniroma3.it, mauro.mezzini@uniroma3.it

Keywords: Dropout, Higher Education, Systematic Review, Educational Data Mining, Algorithms

The dropout rates in the European countries is one of the major issues to be faced in a near future as stated in the Europe 2020 strategy. In 2017, an average of 10.6% of young people (aged 18-24) in the EU-28 were early leavers from education and training according to Eurostat's statistics. The main aim of this review is to identify studies which uses educational data mining techniques to predict university dropout in traditional courses. In Scopus and Web of Science (WoS) catalogues, we identified 241 studies related to this topic from which we selected 73, focusing on what data mining techniques are used for predicting university dropout. We identified 6 data mining classification techniques, 53 data mining algorithms and 14 data mining tools.

for citations:
Agrusti F., Bonavolontà G., Mezzini M. (2019), *University Dropout Prediction through Educational Data Mining Techniques: A Systematic Review,* Journal of e-Learning and Knowledge Society, v.15, n.3, 161-182. ISSN: 1826-6223, e-ISSN:1971-8829
DOI: 10.20368/1971-8829/1135017

1 Introduction

One of the goals in the Europe 2020 strategy is to have at least 40% of adult (30-34 years-old) complete higher education (Vossensteyn *et al.*, 2015). Therefore, in several different field of study there is an increasing interest in reducing dropout and improving academic retention in higher education approaches for achieving this goal, which is regarded as crucial for building the high-level skills useful to foster productivity and social justice in Europe. In Europe, according to the 2016 report by the Organization for Economic Cooperation and Development (OECD) dropout rates ranged between 30% and 50%. In Italy, the enrolment rate of 20-24 years-old is one of the lowest among OECD countries (33.7 %, rank 31/40) (OECD, 2016).

Academic *retention* can be defined as the continuous participation of the student in the university's educational path until its end. Retention can be also conceptualized from the point of view of the student (in this case it is called *persistence*) representing the student motivation to achieve his or her academic goals, first of all obtaining of the degree (Hagedorn, 2005). Persistence is also the period of time in which a student remains enrolled at the university and it could be considered as a prerequisite, a necessary condition, even if not sufficient, of university success. When students leave university before achieving their intended goals, they could be labelled as *dropout* students. In this way, retention and dropout phenomena are then described as two side of the same coin; but when "something goes wrong" diverse and more complex failure scenarios may occur, which can be summarized as follows:

- Permanent leaving of studies (*drop-out*), it can be classified into *early* and *late drop-out* (respectively at the second year of enrolment or in subsequent course years).
- Transfer from one bachelor program to another in the same or in another university (*transfer*).
- Different forms of delay (in Italian language *fuoricorsismo*, *out-of-schooling*) that can be defined as a time extension of the forecasted time required to obtain the degree.

The general phenomenon that includes this type of criticalities is defined as *attrition*: "the diminution in numbers of students resulting from lower student retention" (Hagedorn, 2005, p. 6).

The term dropout, unfortunately, is recognized by Astin (1971), Tinto (1987), Bean (1990) and others as one of the more often misused labels for an educational descriptor in literature. Bean (*ibidem*) points out that a dropout student could return and transform his or her status in a "non-dropout" one.

Nevertheless, we will use in this review the university dropout definition by Søgaard Larsen & Dansk Clearinghouse (2013, p. 18): "withdrawal from a university degree program before it has been completed".

In this scenario, there is an increasing interest in the early prediction of student dropout, trying to predict its rates in the most precise manner possible. The main objective of this paper is to provide an overview of the educational data mining techniques that have been used to predict dropout rates in studies of the last decade.

Educational data mining is the use of data mining (also called knowledge discovery in database - KDD) applied in educational field in order to extract meaningful information, patterns and relationships among variables stored in a huge educational data set (Bala & Ojha, 2012; Koedinger, D'Mello, McLaughlin, Pardos, & Rosé, 2015; Mohamad & Tasir, 2013; Romero & Ventura, 2007; Shahiri, Husain, & Rashid, 2015). The useful information may be used to predict dropout causes and finally to improving student persistence preventing identified causes (i.e. providing teachers a dropout student dashboard to improve their teaching approach). Previous literature reviews on educational data mining have covered different topics such as intelligent tutoring systems, learning analytics, student modelling, prediction of student performance and several others. But none of these studied the university dropout except one, but with a limited time frame and only 67 identified papers (Alban & Mauricio, 2019).

2 Methodology

In order to perform this review, we considered the procedures described by Kitchenham in her technical report called "Procedures for Performing Systematic Reviews" (2004). First of all, we proposed three research questions in order to determine the aspects that have been developed to predict university student dropout, stated as follows:

- What data mining techniques have been used to predict university dropout in traditional (face-to-face) courses?
- Which data mining algorithms were used?
- Which data mining tools were used?

Book sections, conference papers and journal articles were reviewed in above mentioned catalogues. To identify relevant documents, we have used the advanced search engines provided by Scopus and WoS respectively. The Scopus query used in *advanced search* function is described in appendix 1.

Through this query we selected all the English documents that had the words: dropout, drop-out, dropping out, attrition, higher education, university, college, data mining, neural network, bayesian, artificial intelligence, AI in the title, abstract and keywords. In addition, with the boolean operators we excluded the documents that did not respond to the research questions. After this step, we applied the selection criteria (Table 1) to refine the final search.

Table 1
SELECTION CRITERIA

Inclusion	Exclusion
Documents including data mining-based university dropout prediction. Documents presenting metrics to assess the quality of predictive models of university dropout. Documents answering research questions.	Documents about dropout's prediction that are not related to the university level in attendance (exclusion of primary, secondary and postgraduate education and all distance learning courses). Documents that do not use data mining techniques. Documents that do not report research data and metrics and where the methodology and techniques used have not been explained

The same methodology has been used for the selection of documents on WoS with small differences in the query due to the different syntax of the search engine. The WoS query used is described in appendix 1.

3 Identified documents

The selection process was completed by deleting the duplicate documents (listed both on Scopus and WoS) with a result of 73 documents selected: 36 documents from Scopus and 37 documents from WoS (Table 2). Figure 1 presents the increasing number of selected studies during timeline (we did not specify any time range in the search query, nether on Scopus or WoS). The first selected document is from 1999, the last one is from 2019.

It is crucial to notice that the number of selected studies has a notable increment since 2014.

Table 2
SELECTED DOCUMENTS

Source	Identified documents	Selected documents
SCOPUS	144	36
WoS	97	37
TOTAL	241	73

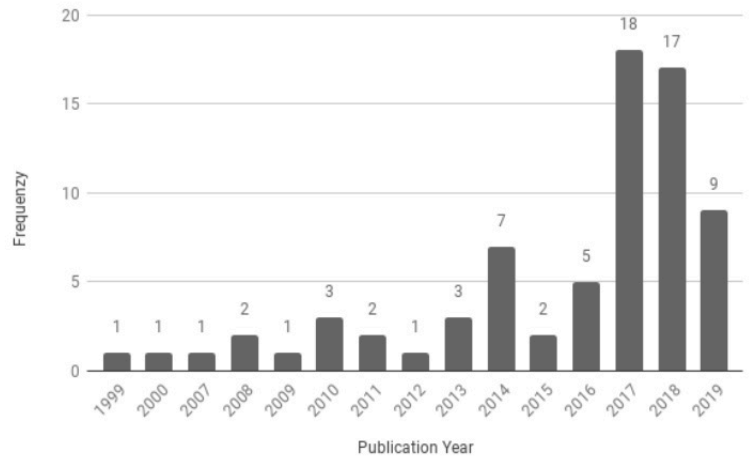

Fig. 1 - Temporal trend of selected publications.

From the selected documents, we identified three aspects regarding university dropout prediction: data mining techniques, algorithms and tools. As stated above, data mining technique is part of the process of converting raw data into useful information, from data pre-processing to postprocessing of data mining results (Tan *et al.*, 2005). We identified six classification techniques: *Decision Tree*, *K-Nearest Neighbor*, *Support Vector Machines*, *Bayesian Classification*, *Neural Networks*, *Logistic regression*, and on *miscellanea* class for other techniques.

3.1 k-Nearest Neighborhood

The *k-Nearest Neighborhood* is a simpler classifier based on the idea that an object O can be classified by taking the class of the object which is most similar to O. First of all we need to find an objective way to measure the similarity. This can be accomplished by decoding all the object in the training set as a numerical real valued vectors $\boldsymbol{x} \in \mathbb{R}^n$ where n is the number of features of each object. Then we can use any distance function defined in the n-dimensional space of reals like for example the euclidean distance function, in order to give an objective measure that states how similar two objects are. The object C in the training set having the smallest distance to O will be the nearest to O and we will give to O its class. Another strategy could be to take the set S of the first k objects nearest to O. Then take the class of which most the objects in S belong breaking ties arbitrarily.

3.2 Decision Tree

Let $U=\{A_1,\ldots,A_n\}$ be a set of attributes or features of a set Ω of objects. Decision Tree (DT) is a directed acyclic rooted tree. To each node i of the DT is associated a single attribute A_i of U and a subset of objects in Ω. The association to subset of Ω to each node is recursively done as follows. The root node contains all the objects in Ω. Let i be internal node and S_i be the subset of Ω associated to i. For every different value v_i of the attribute A_i there is a child C_j of i and the set of objects associated to C_j are the object of S_i for which the value of attribute A_i is V_j. A node is a leaf if the set of objects associated to it contains objects all of the same class. The classification of an object O is made on the following way. Starting from the root we inspect each node i until we reach a leaf. At that point, to O is given the class of the object associated to the leaf. At a generic internal node i we inspect the value v_j of the attribute A_i of O and then, we continue the traverse of the DT in the child C_j of i.

3.3 Bayesian Networks and Bayesian classifiers

Bayesian Networks (BN) are one of the most effective tool for the classification task (Pearl, 1988). Let $U=\{A_1, \ldots , A_n\}$ be a set of discrete random variables. We call the set of all the possible different values the variable A_i can take, the *domain* of A_i. A BN describes a joint probability distribution of the set of random variables over U both qualitatively and quantitatively by using a directed acyclic graph (DAG) and a set of parameters. Formally a BN $B=(G,\Theta)$ where G is a DAG whose vertex set is U and Θ contains the parameters of the network in the form $\Theta=\{\Theta|A \in U\}$ where $\Theta_A = P(A|\pi_A)$ where π_A is the set of parents of A in G and $P(A|\pi_A)$ represent the probability distribution of A given its parents π_A. Based on this, we can decompose the joint probability distribution as

$$P(\boldsymbol{U}) = \prod_{\boldsymbol{A} \in \boldsymbol{U}} P(A|\,\Pi_{\boldsymbol{A}}) \qquad (1)$$

Without loss of generality suppose that A_1 is the random variable specifying the class label of a group of objects. In a naive Bayesian Classifier, a strong assumption is made that every distinct attribute A_i and A_j, $i,j>1$ are conditionally independent given A_1. Therefore the joint probability distribution of U (1) can be expressed as

$$P(\boldsymbol{U}) = \prod_{\boldsymbol{1<i\leq n}} P(A_i|\,\mathrm{A}_1)$$

Which simplify greatly the network and the prediction queries.

3.4 Perceptrons, SVM and Neural Networks

In the brain or in the nervous system of a living organism each neuron is composed by a body, called the *soma*, a set of *dendrites* and an *axon*. Both dendrites and axon are filaments that extrudes from the soma. The dendrites resemble the roots of a tree and act collectively as input to the neuron cell while the axon bring a signal to other neuronal cells by using the axon terminals called *Synapsys*.

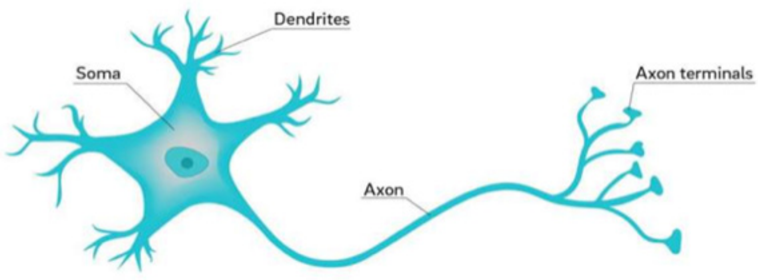

Fig. 2 - A neuron cell.

We assume that if a neuron has *n* dendrites then there are *n* possible different signal in input to the neuron and there is only one output signal transported in output by the axon to other neurons. If each of the input signals has strength x_i, $i=1,\dots,n$ where x_i is a real number, we may assume that the neuron transforms each signal by multiplying it with a weight w_i. Then the sum of the transformed signals can be used by the axon to transmit a signal to other neurons. If we denote by $\boldsymbol{x} \in \mathbb{R}^n$ the input signals and by $\boldsymbol{w} \in \mathbb{R}^n$ the weights associated to each dendrite the signal the axon will transmit can be computed by the following function

$$f(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \sum_{i=1}^{n} x_i w_i + b > 0 \\ -1 & \text{otherwise} \end{cases} \qquad (2)$$

where *b* is a real number called the *bias parameter*. What we obtained here is sometimes called a *perceptron*. Clearly a single perceptron can be used as a binary classifier. In other words, if we think to a single neuron as a binary classifier which can be *activated,* or it can be *deactivated* when it receives some input *x* then the perceptron mimics the behavior of a neuron.

Therefore, if we represent an object *O* by a real value vector *x* of features then we can use a single perceptron as a classifier in order to recognize if the object *O* belong or not to a particular class.

While perceptrons can be used as binary classifiers, there are cases in which we want to classify an object among different classes. For example, the relatively simple nervous system of a bird should be able to classify if an object is a car, is an insect or is a tree. In this case we can use a stack of perceptrons and obtain what we call a Support Vector Machine (SVM) (Hearst, 1998). If we use *D* perceptrons in an SVM we can imagine that all the perceptrons take the same input object but each perceptrons is specialized to be activated only for a certain class and not for the other. The function we obtain is a *D*-dimensional vector $\boldsymbol{y} \in \mathbb{R}^D$

$$\boldsymbol{y} = \boldsymbol{W}\boldsymbol{x} + \boldsymbol{b} \qquad (3)$$

where W is a $D * n$ matrix of real numbers and b is a *D*-dimensional vector of real numbers. The index of the component of y which take the maximum value will be taken as the number of the class predicted by the SVM. For example, if the possible class are {*Car*, *Insect*, *Tree*} we have that D= 3 and if $y=(0.2, 6.4, -3.7)$ then we may conclude that the class with maximum score is the class 2 that is x is an insect.

The problem with perceptron and in general with the SVM is that they work well if the class of objects are linearly separable, that is if there exists for each class a hyperplane that separates the class from all the other classes. To overcome the problem of classification when the space of the classes is not linearly separable at the end of each perceptron a nonlinear sigmoidal function is applied. Then the output so obtained is sent to another perceptron. The output produced by the last perceptron can be expressed as

$$y = \sum_{i=1}^{n} \alpha_i \sigma(\boldsymbol{w}_i \boldsymbol{x} + b_i) \qquad (4)$$

where w_i is the n-dimensional vector of weights of the i-th perceptron. Such type of classifier is called *Neural Networks*. Cybenko (1989) proved that the above formula can be used to compute any classification function.

A Neural Network is a mathematical object used to roughly mimics the functions of the neurons in a nervous system. Contrary to the classic paradigm of computer programming, in which the programmer needs to have a complete knowledge of the problem to be solved in order to design a correct algorithm, like for example in (Malvestuto, Mezzini, & Moscarini, 2011; Mezzini, 2010, 2011, 2012, 2016, 2018; Mezzini & Moscarini, 2015, 2016), in order to implement a Neural Network the programmer need not to understand the meaning and the mechanism behind the phenomenon to be classified and uses the Neural Network as a black box.

4 Results

Table 3 summarizes the total identified and selected documents by classification techniques. Approximately 67% (49 out of 73 documents) used Decision tree classifiers. Bayesian Classification hold the second highest frequency of use with approximately 49%, then Neural Networks with approximately 40% and Logistic regression with approximately 34%. Support Vector Machines, Miscellanea and K-Nearest Neighbour are used respectively with approximately 23%, 15% and 12%.

Table 3
CLASSIFICATION TECHNIQUES

Techniques	Frequency
Decision Tree [1,3,4,10,11,12,13,15,16,18,19,20,22,24,26,27,28,30,33,38,39,40,43,44,48,49,60,61,62,63,64,65,66,69,70,71,72,73,76,82,84,85,86,88,90,91,93,95,98]	49
Bayesian Classification [3,13,15,17,20,25,27,28,30,34,35,38,39,40,44,46,47,59,60,64,65,67,69,70,71,73,76,78,79,81,86,90,93,94,95,98]	36
Neural Networks [1,2,5,13,16,18,21,26,27,28,29,38,39,41,42,43,45,64,65,66,73,77,83,85,86,93,94,96,98]	29
Logistic regression [3,11,15,16,18,20,26,33,45,48,62,64,64,66,69,70,75,79,82,85,86,94,95,97,98]	25
Support Vector Machines [1,13,18,19,20,21,33,38,39,40,48,71,72,79,85,94,98]	17
Miscellanea [4,15,43,49,60,65,72,75,82,95,98]	11
K-Nearest Neighbour [3,19,20,27,28,64,72,95,98]	9

In addition, we have identified the specific algorithms used in the selected documents, grouped by classification techniques with the result of 53 algorithms: 19 for Decision Tree (Table 4), 11 for Bayesian Classification (Table 5), 6 for Neural Networks (Table 6), 3 for Logistic regression (Table 7), 4 for Support Vector Machines (Table 8), 8 for Miscellanea (Table 9). Instead, we have not identified specific algorithms for K-Nearest Neighbour. Unfortunately, not all the selected documents cited explicitly the algorithms used.

Table 4
DECISION TREE ALGORITHMS

Algorithm	Frequency
C4.5 (j48) [4,13,93,22,27,28,38,24,44,49,60,65,15,71,72,73,76,82,88,91]	20
Random Forest [3,15,26,33,40,43,62,64,70,71,85,95]	12
C5.0 [3,16,18,24,93,98]	6
CART [11,15,24,63,49,38]	6
CHAID [10,61,63,98]	4
ID3 [4,84,93]	3
Random Tree [60,49,90]	3
Gradient Boosting Tree [40,64]	2
ADTree [65,49]	2
AdaBoost [20,64]	2
Decision Forest [94,98]	2
Decision Jungle [94,98]	2
Gradient Boosted Trees [40,64]	2
Boosted Decision Tree [94,98]	2
Decision Table [39,44]	2
EM5.3 [66]	1
Rpart [3]	1
Ctree [3]	1
REPTree [49]	1

Table 5
BAYESIAN CLASSIFICATION ALGORITHMS

Algorithm	Frequency
Naïve Bayes [3,13,20,25,27,28,34,38,39,40,44,60,64,65,67,69,70,71,73,76,79,86,90, 93,95]	25
Bayesian Network [15,30,34,35,65,81,93]	7
TAN [17,35,34]	3
K2 [34,35,67]	3
PC [34,35,67]	3
Bayesian Profile Regression [78,79]	2
Markov chains [46,47]	2
Bayes Point Machine [94,98]	2
Bayesian binary quantile regression [59]	1
Gaussian Naive Bayes algorithm [70]	1
AutoClass [67]	1

Table 6
NEURAL NETWORK ALGORITHMS

Algorithm	Frequency
Multilayer perceptron [5,16,18,29,38,39,45,73,83,77,85]	11
Radial Basis Function [21,65]	2
Fuzz-ARTMAP neural network [41,42]	2
Self-organizing map (SOM) [96]	1
Adaptive Network based Fuzzy Inference System (ANFIS) [2]	1
Probabilistic neural network [45]	1

Table 7
SUPPORT VECTOR MACHINE ALGORITHMS

Algorithm	Frequency
Averaged perceptron [94]	2
Polinomial kernel [39]	1
RBF kernel 39]	1
Least-Square Support Vector Classification [21]	1

Table 8
LOGISTIC REGRESSION ALGORITHMS

Algorithm	Frequency
Iterative Logistic Regression [95]	1
Logit [15]	1
Generalized Linear Model [64]	1

Table 9
MISCELLANEA ALGORITHMS

Algorithm	Frequency
ONE R [15,49,60,65]	4
K-means [4,75,82]	3
JRip [15,49]	2
Random guess [95]	1
Gradient boosting machine [43]	1
Ridor [49]	1
Quest [98]	1
EUSBoost [72]	1

In order to answer to our third research question ("Which data mining tools were used?") we identified each tool used in selected documents. *Data mining* *tool* refers to software used to extract, process and analyze the data. Only 46 out of 73 selected searches present the tools used, therefore we have identified 14 tools summarized in Table 10.The results highlight that the most widely used tools were WEKA, SPSS and R.

Table 10
DATA MINING TOOLS

Tool	Frequency
WEKA [4,13,15,30,49,38,44,60,82,73,86,88,90,91]	14
SPSS [1,5,10,16,19,45,61,77,98]	9
R [3,24,25,43,59,78,79,85]	8
Rapid Miner [1,12,22,60,76]	5
Elvira [34,35,67]	3
H2O [43,64]	2
SAS [66,97]	2
Watson Analytics [69,70]	2
Azure Machine Learning [94,98]	2
Matlab [2]	1
Orange3 [60]	1
Statistica [83]	1
NeuralWorks Professional II/PLUS [45]	1

Conclusion and future developments

This paper presents a systematic literature review on educational data mining techniques used to predict university dropout in traditional courses. We identified 241 studies related to this topic from which we selected 73 papers accordingly to above mentioned inclusion and exclusion criteria. We identified six classification techniques: Decision Tree, K-Nearest Neighbour, Support Vector Machines, Bayesian Classification, Neural Networks, Logistic regression (plus one category for minor techniques called “Miscellanea”).

The educational data mining technique which presented a higher frequency of use is Decision tree (67%), followed by Bayesian Classification (49%), Neural Networks (40%) and Logistic regression (34%).

Moreover, we identified 14 data mining tools used in the studies, highlighting that the most used ones are WEKA, SPSS and R.

It is of high evidence that university dropout prediction is of elevated interest for academic researchers' community and that highly precision techniques are being developed to address this crucial issue. However, we did not find any study about dropout and Convolutional Neural Network (CNN), a very efficient algorithm more frequently used in image recognition researches.

As further developments we intend to analyse the selected documents more in detail, trying to answer to the following questions:

- Which predictive model evaluation metrics were presented in the research?
- What are the levels of reliability reached by the techniques presented in the research?

In conclusion, this systematic review on predicting dropout rates has motivated us to carry out further research to be applied in higher educational data mining field in order to monitoring students' performance in a systematic and even more automated way.

Acknowledgements

Research group is composed by the authors of the contribution that was edited in the following order: F. Agrusti (1, 5), G. Bonavolontà (2, 3, 4 except 3.1 - 3.4), M. Mezzini (3.1 – 3.4).

REFERENCES

Adejo, O. W., & Connolly, T. (2018). Predicting student academic performance using multi-model heterogeneous ensemble approach. *Journal of Applied Research in Higher Education*, *10*(1), 61–75.

Adil, M., Tahir, F., & Maqsood, S. (2018). Predictive Analysis for Student Retention by Using Neuro-Fuzzy Algorithm. In *2018 10th Computer Science and Electronic Engineering Conference (ceec)* (pp. 41–45).

Ahuja, R., & Kankane, Y. (2017). Predicting the probability of student's degree completion by using different data mining techniques. *2017 Fourth International Conference on Image Information Processing (ICIIP)*, 1–4. https://doi.org/10.1109/ICIIP.2017.8313763

Al-shargabi, A. A., & Nusari, A. N. (2010). Discovering Vital Patterns From UST Students Data by Applying Data Mining Techniques. In V. Mahadevan & Z. Jianhong (Eds.), *2010 2nd International Conference on Computer and Automation Engineering (iccae 2010), Vol 2* (pp. 547–551).

Alban, M., & Mauricio, D. (2019). Neural networks to predict dropout at the universities. *International Journal of Machine Learning and Computing*, *9*(2), 149–153. https://doi.org/10.18178/ijmlc.2019.9.2.779

Alban, M., & Mauricio, D. (2019). Predicting University Dropout through Data Mining: A systematic Literature. *Indian Journal of Science and Technology*, *12*(4), 1–12. https://doi.org/10.17485/ijst/2019/v12i4/139729

Astin, A. W. (1971). Predicting academic performance in college: Selectivity data for 2300 American colleges.

Bala, M., & Ojha, D. D. B. (2012). STUDY OF APPLICATIONS OF DATA MINING TECHNIQUES IN EDUCATION. *Vol. No.*, (1), 10.

Bean, J. P. (1990). Using retention research in enrollment management. *The Strategic Management of College Enrollments*, 170–185.

Bernardo, A., Cervero, A., Esteban, M., Tuero, E., Casanova, J. R., & Almeida, L. S. (2017). Freshmen program withdrawal: Types and recommendations. *Frontiers in Psychology*, *8*(SEP). https://doi.org/10.3389/fpsyg.2017.01544

Burova, S., Meyer, D., Doube, W., & Apputhurai, P. (2014). Predicting Undergraduate Onsite Student Withdrawals Based On Enrolment, Progress, And Online Student Data (T. Loster & T. Pavelka, Eds.).

Castro R., L. F., Espitia P., E., & Montilla, A. F. (2018). Applying CRISP-DM in a KDD process for the analysis of student attrition. *Communications in Computer and Information Science*, *885*, 386–401. https://doi.org/10.1007/978-3-319-98998-3_30

Costa, E. B., Fonseca, B., Santana, M. A., de Araujo, F. F., & Rego, J. (2017). Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. *Computers in Human Behavior*, *73*, 247–256. https://doi.org/10.1016/j.chb.2017.01.047

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, *2*(4), 303–314. https://doi.org/10.1007/BF02551274

Dekker, G. W., Pechenizkiy, M., & Vleeshouwers, J. M. (2009). *Predicting students drop out: A case study*. 41–50. Retrieved from Scopus.

Delen, D. (2011). Predicting student attrition with data mining methods. *Journal of College Student Retention: Research, Theory and Practice*, *13*(1), 17–35. https://doi.org/10.2190/CS.13.1.b

Delen, D., Topuz, K., & Eryarsoy, E. (2019). Development of a Bayesian Belief Network-based DSS for predicting and understanding freshmen student attrition. *European Journal of Operational Research*. https://doi.org/10.1016/j.ejor.2019.03.037

Delen, Dursun. (2010). A comparative analysis of machine learning techniques for student retention management. *Decision Support Systems*, *49*(4), 498–506. https://

doi.org/10.1016/j.dss.2010.06.003

Dharmawan, T., Ginardi, H., & Munif, A. (2018). *Dropout Detection Using Non-Academic Data*. Presented at the Proceedings - 2018 4th International Conference on Science and Technology, ICST 2018. https://doi.org/10.1109/ICSTC.2018.8528619

Gopalakrishnan, A., Kased, R., Yang, H., Love, M. B., Graterol, C., & Shada, A. (2017). A Multifaceted Data Mining Approach to Understanding what Factors Lead College Students to Persist and Graduate.

Guner, N., Yaldir, A., Gunduz, G., Comak, E., Tokat, S., & Iplikci, S. (2014). Predicting Academically At-Risk Engineering Students: A Soft Computing Application. *Acta Polytechnica Hungarica*, *11*(5), 199–216.

Gustian, D., & Hundayani, R. D. (2017). Combination of AHP Method With C4.5 in The Level Classification Level Out Students.

Hagedorn, L. S. (2005). How to define retention. *College Student Retention Formula for Student Success*, 90–105.

Hasbun, T., Araya, A., & Villalon, J. (2016). Extracurricular activities as dropout prediction factors in higher education using decision trees. In J. M. Spector, C. C. Tsai, D. G. Sampson, Kinshuk, R. Huang, N. S. Chen, & P. Resta (Eds.), *2016 Ieee 16th International Conference on Advanced Learning Technologies (icalt)* (pp. 242–244).

Hegde, V., & Prageeth, P. P. (2018). Higher Education Student Dropout Prediction and Analysis through Educational Data Mining.

Hoffait, A.-S., & Schyns, M. (2017). Early detection of university students with potential difficulties. *Decision Support Systems*, *101*, 1–11. https://doi.org/10.1016/j.dss.2017.05.003

Iam-On, N., & Boongoen, T. (2014). Using Cluster Ensemble to Improve Classification of Student Dropout in Thai University. In 2014 Joint 7th International Conference on Soft Computing and Intelligent Systems (scis) and 15th International Symposium on Advanced Intelligent Systems (isis) (pp. 452–457).

Iam-On, N., & Boongoen, T. (2017). Improved student dropout prediction in Thai University using ensemble of mixed-type data clusterings. *International Journal of Machine Learning and Cybernetics*, *8*(2), 497–510. https://doi.org/10.1007/s13042-015-0341-x

Karamouzis, S. T., & Vrettos, A. (2008). An Artificial Neural Network for Predicting Student Graduation Outcomes. In *Wcecs 2008: World Congress on Engineering and Computer Science* (pp. 991–994).

Khasanah, A. U., & Harwati. (2017). *A Comparative Study to Predict Student's Performance Using Educational Data Mining Techniques*. *215*. https://doi.org/10.1088/1757-899X/215/1/012036

Kitchenham, B. (2004). Procedures for Performing Systematic Reviews (Keele University. Technical Report TR/SE-0401).

Koedinger, K. R., D'Mello, S., McLaughlin, E. A., Pardos, Z. A., & Rosé, C. P. (2015). Data mining and education: Data mining and education. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(4), 333–353. https://doi.org/10.1002/wcs.1350

Kondo, N., Okubo, M., & Hatanaka, T. (2017). *Early Detection of At-Risk Students Using Machine Learning Based on LMS Log Data*. 198–201. https://doi.org/10.1109/IIAI-AAI.2017.51

Lacave, C., & Molina, A. I. (2018). Using Bayesian Networks for Learning Analytics in Engineering Education: A Case Study on Computer Science Dropout at UCLM. *International Journal of Engineering Education*, *34*(3), 879–894.

Lacave, C., Molina, A. I., & Cruz-Lemus, J. A. (2018). Learning Analytics to identify dropout factors of Computer Science studies through Bayesian networks. *Behaviour & Information Technology*, *37*(10–11), 993–1007. https://doi.org/10.1080/0144929X.2018.1485053

M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt and B. Scholkopf (1998). Support vector machines. *IEEE Intelligent Systems and their Applications*, 13, 18-28. doi: 10.1109/5254.708428

Malvestuto, F. M., Mezzini, M., & Moscarini, M. (2011). Computing simple-path convex hulls in hypergraphs. Information Processing Letters, 111(5), 231–234. https://doi.org/10.1016/j.ipl.2010.11.026

Manhães, L. M. B., Da Cruz, S. M. S., & Zimbrão, G. (2014a). The impact of high dropout rates in a large public brazilian university a quantitative approach using educational data mining. 3, 124–129. Retrieved from Scopus.

Manhães, L. M. B., Da Cruz, S. M. S., & Zimbrão, G. (2014b). *WAVE: An architecture for predicting dropout in undergraduate courses using EDM*. 243–245. https://doi.org/10.1145/2554850.2555135

Manrique, R., Nunes, B. P., Marino, O., Casanova, M. A., & Nurmikko-Fuller, T. (2019). *An analysis of student representation, representative features and classification algorithms to predict degree dropout*. 401–410. https://doi.org/10.1145/3303772.3303800

Martinho, V. R. C., Nunes, C., & Minussi, C. R. (2013a). An intelligent system for prediction of school dropout risk group in higher education classroom based on artificial neural networks. 159–166. https://doi.org/10.1109/ICTAI.2013.33

Martinho, V. R. C., Nunes, C., & Minussi, C. R. (2013b). *Prediction of school dropout risk group using neural network*. 111–114. Retrieved from Scopus.

Martins, L. C. B., Carvalho, R. N., Carvalho, R. S., Victorino, M. C., & Holanda, M. (2017). *Early prediction of college attrition using data mining* (X. Chen, B. Luo, F. Luo, V. Palade, & M. A. Wani, Eds.).

Mashiloane, L., & Mchunu, M. (2013). Mining for marks: A comparison of classification algorithms when predicting academic performance to identify 'students at risk'. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8284 LNAI*, 541–552. https://doi.org/10.1007/978-3-319-03844-5_54

Mason, C., Twomey, J., Wright, D., & Whitman, L. (2018). Predicting Engineering Student Attrition Risk Using a Probabilistic Neural Network and Comparing Results with a Backpropagation Neural Network and Logistic Regression. *Research in Higher Education*, *59*(3), 382–400. https://doi.org/10.1007/s11162-017-9473-z

Massa, S., & Puliafito, P. P. (1999). An application of data mining to the problem of the University students' dropout using Markov chains. In J. M. Zytkow & J. Rauch (Eds.), *Principles of Data Mining and Knowledge Discovery* (Vol. 1704, pp. 51–60).

Massa, S., & Puliafito, P. P. (2000). Data mining in temporal sequences: A technique based on MC. In N. Ebecken & C. A. Brebbia (Eds.), *Data Mining Ii* (Vol. 2, pp. 289–298).

Mayra, A., & Mauricio, D. (2018). Factors to Predict Dropout at the Universities: A case of study in Ecuador. In Proceedings of 2018 Ieee Global Engineering Education Conference (educon). Emerging Trends and Challenges of Engineering Education (pp. 1238–1242).

Meedech, P., Iam-On, N., & Boongoen, T. (2016). Prediction of Student Dropout Using Personal Profile and Data Mining Approach. In K. Lavangnananda, S. PhonAmnuaisuk, W. Engchuan, & J. H. Chan (Eds.), *Intelligent and Evolutionary Systems, Ies 2015* (Vol. 5, pp. 143–155).

Mezzini, M. (2010). On the complexity of finding chordless paths in bipartite graphs and some interval operators in graphs and hypergraphs. Theoretical Computer Science, 411(7), 1212–1220. https://doi.org/10.1016/j.tcs.2009.12.017

Mezzini, M. (2011). Fast minimal triangulation algorithm using minimum degree criterion. Theoretical Computer Science, 412(29), 3775–3787. https://doi.org/10.1016/j.tcs.2011.04.022

Mezzini, M. (2012). Fully dynamic algorithm for chordal graphs with O(1) query-time and O(n2) update-time. Theoretical Computer Science, 445, 82–92. https://doi.org/10.1016/j.tcs.2012.05.002

Mezzini, M. (2016). On the geodetic iteration number of the contour of a graph. Discrete Applied Mathematics, 206, 211–214. https://doi.org/10.1016/j.dam.2016.02.012

Mezzini, M. (2018). Polynomial time algorithm for computing a minimum geodetic set in outerplanar graphs. Theoretical Computer Science, 745, 63–74. https://doi.org/10.1016/j.tcs.2018.05.032

Mezzini, M., & Moscarini, M. (2015). On the geodeticity of the contour of a graph. Discrete Applied Mathematics, 181, 209–220. https://doi.org/10.1016/j.dam.2014.08.028

Mezzini, M., & Moscarini, M. (2016). The contour of a bridged graph is geodetic. Discrete Applied Mathematics, 204, 213–215. https://doi.org/10.1016/j.dam.2015.10.007

Mezzini, M., Bonavolontà, G., & Agrusti, F. (2019). Predicting university dropout by using convolutional neural networks. In INTED2019.

Mohamad, S. K., & Tasir, Z. (2013). Educational Data Mining: A Review. *Procedia - Social and Behavioral Sciences*, *97*, 320–324. https://doi.org/10.1016/j.sbspro.2013.10.240.

Mollica, C., & Petrella, L. (2017). Bayesian binary quantile regression for the analysis of Bachelor-to-Master transition. *Journal of Applied Statistics*, *44*(15), 2791–2812. https://doi.org/10.1080/02664763.2016.1263835

Moscoso-Zea, O., Vizcaino, M., & Luján-Mora, S. (2017). *Evaluation of methods* *and algorithms of educational data mining*. Presented at the 2017 Research in Engineering Education Symposium, REES 2017. Retrieved from Scopus.

Moseley, L. G., & Mead, D. M. (2008). Predicting who will drop out of nursing courses: A machine learning exercise. *Nurse Education Today*, *28*(4), 469–475. https://doi.org/10.1016/j.nedt.2007.07.012

Murakami, K., Takamatsu, K., Kozaki, Y., Kishida, A., Kenya, B., Noda, I., … Nakata, Y. (2019). *Predicting the Probability of Student Dropout through EMIR Using Data from Current and Graduate Students*. 478–481. https://doi.org/10.1109/IIAI-AAI.2018.00103

Mustafa, M. N., Chowdhury, L., & Kamal, M. S. (2012). Students Dropout Prediction for Intelligent System from Tertiary Level in Developing Country.

Nagy, M., & Molontay, R. (2018). *Predicting Dropout in Higher Education Based on Secondary School Performance*. 000389–000394. https://doi.org/10.1109/INES.2018.8523888

Nandeshwar, A., Menzies, T., & Nelson, A. (2011). Learning patterns of university student retention. *Expert Systems with Applications*, *38*(12), 14984–14996. https://doi.org/10.1016/j.eswa.2011.05.048

Olinsky, A., Schumacher, P., & Quinn, J. (2016). An Expanded Assessment of Data Mining Approaches for Analyzing Actuarial Student Success Rate. *International Journal of Business Analytics*, *3*(1), 22–44. https://doi.org/10.4018/IJBAN.2016010102

Oviedo, B., Moral, S., & Puris, A. (2016). A hierarchical clustering method: Applications to educational data. *Intelligent Data Analysis*, *20*(4), 933–951. https://doi.org/10.3233/IDA-160839

Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Perez, B., Castellanos, C., & Correal, D. (2018). *Applying Data Mining Techniques to Predict Student Dropout: A Case Study*. Presented at the 2018 IEEE 1st Colombian Conference on Applications in Computational Intelligence, ColCACI 2018 - Proceedings. https://doi.org/10.1109/ColCACI.2018.8484847

Pérez, B., Castellanos, C., & Correal, D. (2018). Predicting student drop-out rates using data mining techniques: A case study. *Communications in Computer and Information Science*, *833*, 111–125. https://doi.org/10.1007/978-3-030-03023-0_10

Ram, S., Wang, Y., Currim, F., & Currim, S. (2015). *Using big data for predicting freshmen retention*. Presented at the 2015 International Conference on Information Systems: Exploring the Information Frontier, ICIS 2015. Retrieved from Scopus.

Ramentol, E., Madera, J., & Rodríguez, A. (2019). Early detection of possible undergraduate drop out using a new method based on probabilistic rough set theory. *Studies in Fuzziness and Soft Computing*, *377*, 211–232. https://doi.org/10.1007/978-3-030-10463-4_12

Rocha, C. F., Zelaya, Y. F., Sánchez, D. M., & Pérez, A. F. (2017). *Prediction of university desertion through hybridization of classification algorithms*. *2029*, 215–222. Retrieved from Scopus.

Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, *33*(1), 135–146. https://doi.org/10.1016/j.eswa.2006.04.005

Sajjadi, S., Shapiro, B., McKinlay, C., Sarkisyan, A., Shubin, C., & Osoba, E. (2017). *Finding Bottlenecks: Predicting Student Attrition with Unsupervised Classifier*.

Santoso, L. W., & Yulia. (2019). The Analysis of Student Performance Using Data Mining. *Advances in Intelligent Systems and Computing*, *924*, 559–573. https://doi.org/10.1007/978-981-13-6861-5_48

Sarker, F., Tiropanis, T., & Davis, H. C. (2014). *Linked Data, Data Mining and External Open Data for Better Prediction of at-risk Students* (I. Kacem, P. Laroche, & Z. Roka, Eds.).

Sarra, A., Fontanella, L., & Di Zio, S. (2018). *Identifying Students at Risk of Academic Failure Within the Educational Data Mining Framework*. https://doi.org/10.1007/s11205-018-1901-8

Serra, A., Perchinunno, P., & Bilancia, M. (2018). Predicting student dropouts in higher education using supervised classification algorithms. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10962 LNCS*, 18–33. https://doi.org/10.1007/978-3-319-95168-3_2

Shahiri, A. M., Husain, W., & Rashid, N. A. (2015). A Review on Predicting Student's Performance Using Data Mining Techniques. *Procedia Computer Science*, *72*, 414–422. https://doi.org/10.1016/j.procs.2015.12.157

Shiratori, N. (2017). *Modeling dropout behavior patterns using Bayesian Networks in Small-Scale Private University* (T. Matsuo, N. Fukuta, M. Mori, K. Hashimoto, & S. Hirokawa, Eds.). New York: Ieee.

Shyamala, K., & Rajagopalan, S. P. (2007). Mining student data to characterize drop out feature using clustering and decision tree techniques. *International Journal of Soft Computing*, *2*(1), 150–156. Retrieved from Scopus.

Siri, A. (2014). Predicting students' academic dropout using artificial neural networks. Scopus.

Sivakumar, S., Venkataraman, S., & Selvaraj, R. (2016). Predictive modeling of student dropout indicators in educational data mining using improved decision tree. *Indian Journal of Science and Technology*, *9*(4), 1–5. https://doi.org/10.17485/ijst/2016/v9i4/87032

Solis, M., Moreira, T., Gonzalez, R., Fernandez, T., & Hernandez, M. (2018). *Perspectives to Predict Dropout in University Students with Machine Learning*. Presented at the 2018 IEEE International Work Conference on Bioinspired Intelligence, IWOBI 2018 - Proceedings. https://doi.org/10.1109/IWOBI.2018.8464191

Sultana, S., Khan, S., & Abbas, M. A. (2017). Predicting performance of electrical engineering students using cognitive and non-cognitive features for identification of potential dropouts. *International Journal of Electrical Engineering & Education*, *54*(2), 105–118. https://doi.org/10.1177/0020720916688484

Tan, P.-N., Steinbach, M., Kumar, V. (2005). Introduction to Data Mining. Addison

Wesley. ISBN: 0321321367.

Timaran Pereira, R., & Caicedo Zambrano, J. (2017). *Aplication of Decision Trees for Detection of Student Dropout Profiles* (X. Chen, B. Luo, F. Luo, V. Palade, & M. A. Wani, Eds.).

Tinto, V. (1987). Leaving college: Rethinking the causes and cures of student attrition. ERIC.

Vila, D., Cisneros, S., Granda, P., Ortega, C., Posso-Yépez, M., & García-Santillán, I. (2019). Detection of desertion patterns in university students using data mining techniques: A case study. *Communications in Computer and Information Science*, *895*, 420–429. https://doi.org/10.1007/978-3-030-05532-5_31

Villwock, R., Appio, A., & Andreta, A. A. (2015). Educational Data Mining with Focus on Dropout Rates. *International Journal of Computer Science and Network Security*, *15*(3), 17–23.

Vossensteyn, J. J., Kottmann, A., Jongbloed, B. W. A., Kaiser, F., Cremonini, L., Stensaker, B., … Wollscheid, S. (2015). *Dropout and completion in higher education in Europe: Main report*. https://doi.org/10.2766/826962

Wang, Z., Zhu, C., Ying, Z., Zhang, Y., Wang, B., Jin, X., & Yang, H. (2018). Design and Implementation of Early Warning System Based on Educational Big Data. In *2018 5th International Conference on Systems and Informatics (icsai)* (pp. 549–553).

Zea, L. D. F., Reina, Y. F. P., & Molano, J. I. R. (2019). Machine Learning for the Identification of Students at Risk of Academic Desertion. *Communications in Computer and Information Science*, *1011*, 462–473. https://doi.org/10.1007/978-3-030-20798-4_40

Zhang, L., & Rangwala, H. (2018). Early identification of at-risk students using iterative logistic regression. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10947 LNAI, 613–626. https://doi.org/10.1007/978-3-319-93843-1_45

Zhang, Y., Li, Y., You, F., & Xu, X. (2010). Withdrawal prediction using the blackboard learning management system through SOM. 340–344. Retrieved from Scopus.

Zhuhadar, L., Daday, J., Marklin, S., Kessler, B., & Helbig, T. (2019). Using survival analysis to discovering pathways to success in mathematics. *Computers in Human Behavior*, *92*, 487–495. https://doi.org/10.1016/j.chb.2017.12.016

Zuka, R., Krasts, J., & Rozevskis, U. (2017). Using Data Mining Technology For Student Data Analysis.

APPENDIX 1

Queries used in Scopus and WoS.

Scopus

(TITLE-ABS-KEY (dropout OR drop-out OR "drop out" OR "dropping out" OR "attrition") AND TITLE-ABS-KEY ("higher education" OR "university" OR "college") AND TITLE-ABS-KEY ("data mining" OR "neural network" OR "bayesian" OR "artificial intelligence" OR "AI")) AND (EXCLUDE (DOCTYPE, "er")) AND (LIMIT-TO (LANGUAGE, "English")) AND (EXCLUDE (EXACTKEYWORD, "E-learning") OR EXCLUDE (EXACTKEYWORD, "MOOCs") OR EXCLUDE (EXACTKEYWORD, "On-line Education") OR EXCLUDE (EXACTKEYWORD, "On-line Analytical Processing") OR EXCLUDE (EXACTKEYWORD, "Online") OR EXCLUDE (EXACTKEYWORD, "Virtual Learning Environment") OR EXCLUDE (EXACTKEYWORD, "Image Classification") OR EXCLUDE (EXACTKEYWORD, "Image Processing") OR EXCLUDE (EXACTKEYWORD, "Images Classification") OR EXCLUDE (EXACTKEYWORD, "Gene Cluster") OR EXCLUDE (EXACTKEYWORD, "Gene Deletion") OR EXCLUDE (EXACTKEYWORD, "Gene Ontology") OR EXCLUDE (EXACTKEYWORD, "Genetic Selection") OR EXCLUDE (EXACTKEYWORD, "Genetic Variation") OR EXCLUDE (EXACTKEYWORD, "Genetics") OR EXCLUDE (EXACTKEYWORD, "MOOC") OR EXCLUDE (EXACTKEYWORD, "Distance Education") OR EXCLUDE (EXACTKEYWORD, "Distance Higher Education") OR EXCLUDE (EXACTKEYWORD, "Distance Learning") OR EXCLUDE (EXACTKEYWORD, "Distance Learning Course") OR EXCLUDE (EXACTKEYWORD, "Open And Distance Learning") OR EXCLUDE (EXACTKEYWORD, "Massive Open Online Course") OR EXCLUDE (EXACTKEYWORD, "Massive Open Online Course (MOOC)") OR EXCLUDE (EXACTKEYWORD, "Multi-MOOC") OR EXCLUDE (EXACTKEYWORD, "Multivariate Time Series") OR EXCLUDE (EXACTKEYWORD, "Segmented Images") OR EXCLUDE (EXACTKEYWORD, "Entrepreneurial Success") OR EXCLUDE (EXACTKEYWORD, "Breast Cancer") OR EXCLUDE (EXACTKEYWORD, "Immersive Technology") OR EXCLUDE (EXACTKEYWORD, "Web Services") OR EXCLUDE (EXACTKEYWORD, "Web-based Learning") OR EXCLUDE (EXACTKEYWORD, "Traffic Signs") OR EXCLUDE (EXACTKEYWORD, "Brain Tumor Segmentation") OR EXCLUDE (EXACTKEYWORD, "Vis-NIRS") OR EXCLUDE (EXACTKEYWORD, "Tinnitus Dropout") OR EXCLUDE (EXACTKEYWORD, "Amelogenesis Imperfecta"))

WoS

(TS=(dropout OR drop-out OR "drop out" OR "dropping out" OR "attrition") AND TS=("hiher education" OR "university" OR "college") AND TS= ("data mining" OR "neural network" OR "bayesian" OR "artificial intelligence" OR "AI") OR TI=(dropout OR drop-out OR "drop out" OR "dropping out" OR "attrition") AND TI=("higher education" OR "university" OR "college") AND TI= ("data mining" OR "neural network" OR "bayesian" OR "artificial intelligence" OR "AI")) AND LANGUAGE: (English)