



ESJ Natural/Life/Medical Sciences

Prediction of Student Performance Using Rough Set Theory And Backpropagation Neural Networks

Bruno Cristos Madeira

Sakarya University, Institute of Natural Sciences, Information Systems
Engineering Department, Sakarya, Turkey

Tugrul Tasci

Numan Celebi

Sakarya University, Faculty of Computer & Information Sciences,
Information Systems Engineering Department, Sakarya, Turkey

[Doi:10.19044/esj.2021.v17n7p1](https://doi.org/10.19044/esj.2021.v17n7p1)

Submitted: 23 November 2020

Accepted: 12 February 2021

Published: 28 February 2021

Copyright 2021 Author(s)

Under Creative Commons BY-NC-ND

4.0 OPEN ACCESS

Cite As:

Madeira B.C., Tasci T. & Celebi N. (2021). *Prediction of Student Performance Using Rough Set Theory And Backpropagation Neural Networks* .European Scientific Journal, ESJ, 17(7), 1. <https://doi.org/10.19044/esj.2021.v17n7p1>

Abstract

With the rise of web-based education systems and the increased use of information systems in education institutions, the amount of data recorded on student performance and behavior has increased exponentially. Thus, bringing about a large number of contributions to the field of educational research, which in itself contributed to the further evolution off the field in the last two decades alone, with terms such as Educational Data Mining (EDM), Learning Analytics, Data-driven Education, Teaching Analytics and others being added to the literature. In this paper, we evaluate the usefulness of a model using Rough Set Theory (RST) and Backpropagation Neural Network (BPNN) in effectively predicting the students' overall performance. The dataset used consists of 10 different attributes and one decision factor belonging to 53 students collected from a language course which administers in-person education with the aid of an online platform for assignments. RST was implemented in order to reduce the number of attributes used as input in the neural network and the BPNN made an accurate prediction using only 5 of the initial attributes. Thus outperforming a model based solely on BPNN used on

the original dataset and reducing computational costs.

Keywords: Rough Set Theory, Feature selection, Backpropagation Neural Network, Learning Analytics, Educational Data Mining

Introduction

The use of e-learning materials, the wide adherence to massive open online courses (MOOCs), the establishment of cross institution databases and overall integration of Information technologies to education both private and public has led to the creation of large repositories of data. This exponential growth of recorded and available data, alongside the growing use of the internet and digital devices has enabled us to manipulate and analyze data at an unprecedented rate, (S. John walker, 2014). New types of educational environments play a large role in the gathering of huge amount of data about students (Romeo & Ventura, 2020). In the midst of this educational revolution many terms have come about, such as Academic Analytics, Institutional Analytics, Teaching Analytics, Data-Driven Decision-Making in Education, Big Data in Education, and Educational Data Science. The study at hand is associated better with Educational Data Mining, which concerns itself with the exploration of datasets of educational origin (Bakhshinategh, Zaiane, ElAtia, & Ipperciel, 2018), and Learning Analytics, which can be defined as the process of understanding and improving learning and the education environment through the mining of student related data(Lang, Siemens, Wise, & Gasevic, 2017). The use of diverse mining approaches in the educational environment has had many advantages that would have been difficult to attain through the manual analysis of the data. Examples of these applications are numerous and range from emotional learning analytics, where the importance of emotions during the learning process is studied (D'Mello, 2017), to writing analytics, where text mining and analytics tools are applied on text data from, chats, social networks, essays, and others (Ferreira-Mello, 2019). With the proper application of data mining to educational datasets it is possible to make a better allocation of resources in education environments, make the necessary improvements to the teaching system or better assess the learning process. It is in the scope of this contribution that this study is aimed at evaluating the efficiency of a Rough Set Theory and Backpropagation Neural Network based model in predicting the outcome of students in a language institute. The dataset used in this study was acquired from a language institute where the students undergo a 3 month a face to face English language education with the aid of an online platform for assignments which records their progress and activity throughout the course and breaks down their progress in terms of the language skills (further discussed in section 4). By nature of the process of language learning, determining what features define a successful student brings about

uncertainty and is regarded as subjective. The Rough Set Theory aspect of the model in this paper avoids subjective influence of the data in the mining process, unlike fuzzy set or Bayesian theory, which are based on prior knowledge, it is able to disclose hidden knowledge and reduce attributes only based on the dataset (Li, X. et al. 2019); It was used in this study due to its ability to reduce data dimension, eliminate redundant data and noise regardless of how incomplete or inconsistent the data provided is. However, due to the poor generalization ability of RST, the use of a non-linear algorithm with anti-noise ability was established and for this the BP neural network was elected due to its automated weight adjustment feature and the fact that its lack of ability to identify redundant data is compensated by the RST. Therefore, given the individual strengths and weaknesses of the algorithms mentioned, this paper puts forward the use of a model based on Rough Set Theory and Backpropagation Neural Network to predict the performance of students in a language learning environment. Through analyzing the results of this model, education institutes will be able to get a better understanding of any existent undesirable biases in their education system, determine crucial factors to student success and through the early prediction of student performance, make necessary interventions in the students' education process in order to maximize their gains. We hypothesized that the policy of the institute from which the data was collected will be made apparent in the results of the research, therefore having the features relating to language practice being established as the main factors to affect student performance.

The following sections are organized as follows. Section 2 describes related work in this field and where the algorithms used in this paper are employed. In section 3 the methods used in the study are discussed briefly. Section 4 entails a description of the dataset along with the model used. The implementation of the model is done in section 5, followed by discussion of the findings in section 6 and the presentation of the conclusion in section 7.

Related Work

The rapid development of technology and its wide use in daily life, the environments in which it has made itself useful is constantly on the rise. One of the areas that has seen the integration of technology to its environment is education. In Montreal, Canada in 2008, the first Educational Data Mining Conference was held, organized by the IEDM society and in 2011, the first Learning Analytics and Knowledge Conference was held in Banff, Canada, organized by SOLAR (Romeo & Ventura, 2020). These two areas, which emerged from independent conferences and communities have been the aim of many researches and contributions in the last two decades with over 9000 papers and main events about Educational Data Mining and Learning Analytics (EDM/LA) terms being registered in Google scholar between the

years 2000–2018. (Baker, 2015) claimed that the majority of traditional data mining techniques have already been applied in the educational domain. Nevertheless, given the continuous change in both educational and data mining environments, it is important to always seek viable and more cost effective alternatives to EDM/LA solutions.

Given the popularity of RST in data mining, its usefulness for the analysis of educational datasets has been made evident in researches such as (Saedudin et al., 2018), where given a dataset consisting of 1000 entries with 8 attributes each, referring to students from Telkom University, Indonesia. They found in an orderly manner the five most determinant attributes of the student performance. (Ramentol et al., 2019) reported that RST outperformed many state of the art algorithms in the early detection of undergraduate dropouts from the Informatics Engineering Department at the University of Camagüey, Cuba where the dropout rate of freshmen is about 25%. It was also reported that the model introduced in the study made data preprocessing redundant as it outperformed 8 well-known preprocessing methods. In studies where performance prediction was necessary, the application of neural networks was observed; as was the case in (Asif et al., 2017), where it was aimed to provide to teachers and study program directors information to improve their educational program by proposing a model which used a neural network to predict student performance at the end of a four-year university program using pre-university marks and marks belonging to the first and second year courses only, with a reasonable accuracy. (Berens et al., 2018) developed an Early Detection System (EDS) to identify potential dropouts in Germany using regression analysis, neural networks, decision trees and the AdaBoost algorithm. The results showed a prediction accuracy at the end of the first semester of 79% for the state university and 85% for the private university of applied sciences. After the fourth semester, the accuracy improved to 90% for the state university and 95% for the private university of applied sciences.

Methodology

Rough Set Theory (RST)

Widely implemented in the field of data mining, rough set theory (RST), proposed by (Pawlak, 1982), is based on the assumption that real-world knowledge is dependent on the accuracy of its representations. RST methods rely on the data at hand to establish the relations between the features of objects in an information system that displays uncertainty and the decisions they lead to. An information system is a table where rows are objects and columns are different attributes of said objects, Bin Yu, Mingjie Cai, Qingguo Li (2018). Information tables can be represented as follows:

$$S = (U, Q, V, f);$$

Where $U = \{x_1, x_2, x_3 \dots x_n\}$ is the set of objects to be analyzed, $Q = \{q_1, q_2, q_3 \dots q_n\}$ is the set of attributes or features of the objects from which the redundant ones are to be removed. V represents the domain of the attributes and $f(x, q)$ is the representation of the information function.

Indiscernibility relation

For $x, y \in U$ and $f(x, a) = f(y, a)$ it is said that x and y are indiscernible. For a specific set of attributes from set Q , called set A , a binary relation with set U classified as an indiscernibility relation and is represented as $IND(A)$. Elementary sets are equivalence classes in $Q/IND(A)$ or Q/A for short, $[x]_A$ is the equivalence class of $x \in U$.

Upper and Lower Approximation

Let $S = (U, Q, V, f)$ be an information system and $A \subseteq Q$. The \underline{A} lower approximation and \bar{A} upper approximations of $X \subseteq U$, denoted as $\underline{A}(X)$ and $\bar{A}(X)$, respectively, are defined as:

$$\underline{A}(X) = \cup \{x \in X \mid [x]_A \subseteq X\} \quad \text{and} \quad \bar{A}(X) = \cup \{x \in Q \mid [x]_A \cap X \neq \emptyset\}.$$

For any $X \subseteq U$, we can associate accuracy $C_A(X)$ to the approximation of X as:

$$C_A(X) = |\underline{A}(X)| \div |\bar{A}(X)|,$$

Where $|\underline{A}(X)|$ and $|\bar{A}(X)|$ denote the cardinality of $\underline{A}(X)$ and $\bar{A}(X)$, respectively. Evidently, $0 \leq C_A(X) \leq 1$. If $C_A(X) = 1$, X is a crisp set; otherwise, X is a rough set, with respect to A .

One of the main features of data reduction by RST is to eliminate attributes established as redundant and build reducts in order to create simpler models. Reducts are a minimal set of properties that allow for inseparability of the dataset. In other words, elementary sets in the information table are preserved by feature reduction by finding sub feature sets. A particular decision table may have multiple reducts, representing minimal quantities of sufficient information. The feature reduction operation is expressed as a calculation of reducts. If a particular feature is present in all the reducts, that feature is considered "core" and such features cannot be deleted from the information table.

Backpropagation Neural Network (BPNN)

A backpropagation neural network is a multi-layered neural network characterized by signal forward transmission and error feedback propagation. In these neural networks the input on one side of the network is processed through layers of neurons and an output is calculated through to the other side. Therefore, three layers can be established in these structures, the input layer, where the values to be fed into the network are entered; the hidden layer, responsible for the calculation of the output and where weights and biases

connecting the input and output layers are found; And an output layer where the final value calculated by the network is displayed, later on compared to a target value and from where the back propagation begins. A representation of a BPNN can be seen in Figure1.

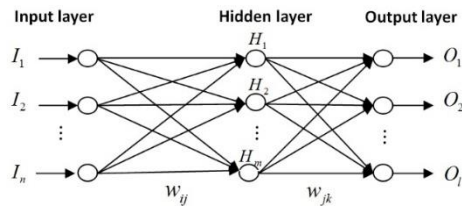


Fig1. Representation of a three layer BPNN

A basic three layer BPNN works as follows. Each in the input layer holds a value which is then passed on to the following layer, each and every one of the neurons in the hidden layer is attributed a value through the weighted sum of the neurons in the previous layer, a bias value is then added to this sum value in order to adjust the result and finally the result is put through an activation function, which determines the value that will be sent to the following layer, where the same process will take place again. The mathematical representation of this relation is given below.

$$H_j = f_1 \left(\sum_{i=1}^n w_{ij} I_i - v_j \right), j = 1, 2, \dots, m$$

$$O_k = f_2 \left(\sum_{j=1}^m w_{jk} H_j - v_k \right), k = 1, 2, \dots, l$$

Where I, H and O represent the input, hidden and output layers respectively, f_1 and f_2 the activation function(s), v the biases of each neuron, the weights of each connection between neurons given as w_{ij} and w_{jk} and the number of neurons in the hidden and output layers given by m and l respectively.

Once the value of the output established, a cost value is calculated as the difference between the expected value and the output. Represented as:

$$C_k = E_k - O_k, k = 1, 2, \dots, l$$

The next stage of the algorithm is backpropagation and is responsible for the learning process. In order to minimize the cost value (i.e. maximizing the accuracy of the neural network), the error is propagated backwards using a standard learning algorithm to adjust the weights and biases so that desirable accuracy levels are reached.

$$w_{ij} = w_{ij} + \eta H_j (1 - H_j) I_i \sum_{k=1}^l w_{jk} C_k$$

$$w_{ij} = w_{ij} + \eta H_j C_k$$

$$v_j = v_j + C_k \eta H_j (1 - H_j) I_i \sum_{k=1}^l w_{jk} C_k$$

$$v_k = v_k + \eta C_k$$

Where η , a value between 0 and 1 represents the learning rate.

Defining dataset and conceptual framework

The dataset for this study was provided by the Wimbledon Language Academy Sakarya-Turkey branch, a language institute that provides face to face English language education aided by an online platform for assignments. The system in place at the time that the dataset was collected was a practice oriented 3 month course with 4 quizzes equally distanced throughout the duration of the course, where students were expected to complete the automatically graded weekly assignments and participate in at least four specific English language skill oriented weekly activities. In order for a student to pass to the following level of education they were expected to have an overall score of 65 or higher. The overall score was calculated as 20% attributed to the average of the quizzes, 30% to the average score of the completed online assignments and the number of completed assignments, 10% to the teacher’s assessment of the student performance and classroom participation, 20% to a speaking evaluation done by the teacher at the last day of the course and 20% to the student’s participation in the auxiliary skill oriented activities. Given the subjective nature of the teacher’s assessment of the student, this attribute was not included in the study. As one of the aims of this study is to determine the underlying factors among student behavior and language skills that impact their overall performance, the amount of assignments completed in relation to the number of assignments given, average score of the assignments, total time spent on the online platform and marks for grammar, listening, reading and vocabulary skills from the automatically generated report from the online platform along with the overall score were acquired for the study. Hence, a dataset composed of 10 attributes, 1 decision and 53 observations was used. A sample of the dataset can be seen in Table 1.

#	Assign. completed	Assign. score	Time online	Gram.	Listen.	Read.	Vocab.	Quiz.	Speak.	Activit.	Final score
1	57/86	63%	03:06:47	88%	100%	100%	95%	76%	85%	100%	85
2	85/86	93%	08:43:49	90%	99%	97%	96%	76%	80%	100%	88
3	86/86	91%	07:25:00	87%	88%	96%	95%	95%	85%	100%	96
4	82/86	79%	05:38:24	81%	85%	61%	89%	41%	80%	100%	77
5	43/86	38%	01:44:31	54%	57%	100%	82%	56%	85%	100%	73
6	84/86	81%	06:07:56	77%	85%	69%	91%	50%	60%	100%	82

Table 1. Sample of the dataset

A representation of the proposed model is shown in Figure 2. The data was first collected from two separate repositories. The Assignments completed, Assignment scores, Total time spent online, Grammar, Listening, Reading and Vocabulary values were obtained from the online platform. The Quizzes, Speaking, Activity and Final score values were obtained from private data repository of the language institute. The two scores were manually appended through the use of the full name of the student.

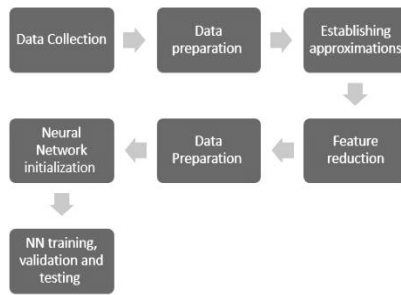


Fig 2. Framework of the proposed model

Implementation

Data preparation

In order for the information table to be loaded and processed in ROSE2, the attributes needed to be renamed for easier comprehension during the process and the data needed to be adapted so that the algorithm would be able to make clearer classifications. Therefore the attribute values were prepared for the Rough Set Theory algorithm as shown in Table 2.

Attribute	Definition	Attribute	Definition
Assignments Completed (A1)	0= 0-30	Activities (A10)	0= 0-30
	1= 31-55		1= 31-55
	2= 56-80		2= 56-80
	3= 81-100		3= 81-100
Assignment score, Grammar, Listening, Reading, Vocab., Quizzes, Speaking(A2, A4, A5, A6, A7, A8, A9)	0= 0-40	Final score (D)	0= 0-64
	2= 41-60		1= 65-80
	3= 61-80		3= 81-100
	4= 81-100		
Time Online (A3)	0= 0 min-1h40min		
	1= 1h41min-3h20min		
	2= 3h21min-5h		
	3= 5h-more		

Table 2. Attributes and decisions used in the study

Table 3 shows the first five records of the information table represented as they were fed into the RST algorithm.

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	D
X1	2	2	1	3	3	3	3	2	3	3	2
X2	3	3	3	3	3	3	3	2	3	3	2
X3	3	3	3	3	3	3	3	3	3	3	2
X4	3	3	3	3	3	2	3	1	3	3	1
X5	1	0	1	1	1	3	3	1	3	3	1

Table 3. First five records of the information table

Approximation classes and feature reduction

The analysis of the information table was carried out in the software ROSE2, created at the Laboratory of Intelligent Decision Support Systems of the Institute of Computing Science in Poznan, Poland, based on a fourteen-year experience in rough set based knowledge discovery and decision analysis. The ratio of the classification quality calculated by the ratios of the lower and upper cluster approaches of the data is 100% as shown in Table 4.

Quality of classification: 1.0000				
Approximations:				
Class	#of Objects	Lower Approximation	Upper Approximation	Accuracy
0	21	21	21	1.0000
1	18	18	18	1.0000
2	14	14	14	1.0000

Table 4. Approximation classes

Each object was shown to be a member of its decision class with an accuracy of 100%; therefore, we have perfect approximations given the fact that the cardinality of both lower and upper approximation values are equal.

After the establishment of the approximation classes the reduct sets were determined and are given in Table 5. Seven reducts were obtained, each with a length of five attributes, all extracted from the initial 10 attributes in the information table. Table 6 shows the core attributes along with the frequencies in which they appear in the reducts. As shown in the tables below, 7 different combinations of 9 attributes with the length of five were established, from these, in order to proceed to the initialization of the BPNN, the reduct composed of the core features with the highest frequencies was selected. Therefore, the reduced features for the BPNN were determined to be the assignment, reading skill, quizzes, speaking and activity scores (attributes A2, A4, A8, A9 and A10 respectively).

Table 5. Reduced feature sets (reducts)

No#	Reduct	Length
1	A4,A6,A8,A9,A10	5
2	A3,A6,A8,A9,A10	5
3	A2,A6,A8,A9,A10	5
4	A1,A6,A8,A9,A10	5
5	A1,A2,A8,A9,A10	5
6	A2,A4,A8,A9,A10	5
7	A2,A7,A8,A9,A10	5

Table 6. Core feature

No#	Attribute	Frequency	%Frequency
1	A1	2	28.57
2	A2	4	57.14
3	A8	7	100.00
4	A9	7	100.00
5	A10	7	100.00
6	A4	2	28.57
7	A6	4	57.14
8	A7	1	14.29
9	A3	1	14.29

Backpropagation neural network (BPNN)

The BPNN initialization, training and testing was done on MATLAB. For the purpose of simplicity and cost effectiveness, the traditional three-layer BPNN was used in this study. In order to better assess the reduction done by the RST algorithm, the prediction was initially done using all ten attributes and the results were compared with those of a neural network designed and trained based on the selected five attributes. In both cases, the number of neurons in the input layer were defined by the number of attributes the prediction is based on. In order to establish the number of neurons in the hidden layer, the cut-and-try method was employed, using the formula $m = \sqrt{n + l} + a$. Where m is the number of hidden neurons, the number of input and output neurons are denoted by n and l respectively and a , a constant, is established as $a \in [1,10]$ and $m < n-1$; thus, given a neural network with ten input and one output neurons, $m \in [5,9]$ and in the reduced neural network with five input neurons $m \in [3,4]$. As an activation function(used to calculate the layer output given its net input), the hyperbolic tangent sigmoid transfer function was employed due to its steeper derivatives in calculating the gradient compared to other activation functions considered for this study. During the fitting process, 70% of the observation was used for data training, 15% for validation and the remaining 15% was used to test the accuracy of the network. Although it requires more memory, the use of the Levenberg–Marquardt algorithm (LMA or LM) to train the neural network was seen as viable due to the simplicity of the network and the fact that it operates faster than the other training algorithms considered. The training process was interrupted automatically when the generalization stopped improving, indicated by an increase in the mean square error of the validation samples.

Results and Discussion

The progress of the training process was measured by a separate set of data in order to avoid data dependency, the same was done to the testing data.

In order to assess the improvement of the neural network’s accuracy, the mean square error (MSE), which is established by the squared difference between the outputs and expected values, was used in order to have a more error sensitive approach to the best network. The regression values, which measure the correlation between the outputs and expected values, were also recorded. These values range from 0 to 1, where 0 represents a random relationship and 1, a close relationship between outputs and target data. Results for the training of networks with ten input neurons and 5-9 neurons in the hidden layers are given in Table 7, while the results for the network with a reduced number of features has its results given in Table 8. For the sake of scalability the root of the average MSE was calculated.

# of neurons in hidden layer	RMSE average			Regression value		
	Training	Validation	Testing	Training	Validation	Testing
5	<u>1.67</u>	<u>5.1</u>	3.58	<u>0.982</u>	<u>0.983</u>	0.995
6	2.47	5.55	9.66	0.971	0.989	0.881
7	1.75	4.12	7.69	0.999	0.991	0.94
8	1.23	6.08	9.66	0.999	0.943	0.966
9	8.89	16.82	11.4	0.952	0.868	0.859

Table 7. Root mean square error and regression values for BPNN with 10 input neurons

# of neurons in hidden layer	RMSE average			Regression value		
	Training	Validation	Testing	Training	Validation	Testing
<u>3</u>	<u>2.94</u>	<u>5.49</u>	2.91	<u>0.995</u>	<u>0.985</u>	0.996
4	2.23	5.19	4.00	0.997	0.979	0.994

Table 8. Root mean square error and regression values for BPNN with 5 input neurons

With the regression values of the networks shown in both Tables 7 and 8, exhibiting such a strong relation between the output values of the neural networks and the expected values, it could be concluded that the BPNN is successful in predicting the outputs with the data provided to it. Comparing the performances of the neural networks, firstly amongst the models fed observations with ten attributes, the RMSE average value was at its lowest and the regression value was at its closest to 1 when the neural network was designed with five neurons in the hidden layer. Better results were also achieved when the network was designed with five input neurons and three neurons in the hidden compared to when four neurons in the hidden layer were used. Comparing these two well performing networks, the result for the network using the attributes selected by the RST for both regression and RMSE show to be better; therefore, the proposed model has proven to be successful in making an accurate prediction while reducing the amount of data initially made available in half.

Looking back at the results of the RST algorithm, the assignment, reading skill, quizzes, speaking and activity scores being determined as factors which have a direct impact on the students' performance at the end of the course and later on establishing that they can serve as a better input for the BPNN in making an accurate prediction of the student performance, validates the claims of the language institute from which the data was acquired to have a system based in language use and practice, given that only one of the five the attributes is exam or grammar based, this knowledge can be used by the administration of the institute to better allocate resources in order to maximize student success or by teachers, in order to aid students in risk of having a poor end of term performance by making an early assessment. If such were not the case and the results showed themselves to be different from the intended system that prioritizes daily use of the language, the administration would be urged to fix the undesired bias in the system, by making the necessary changes to reach the desirable outcome and make an assessment after a large enough number of students had passed through the adjusted new system to make another assessment.

Conclusion

In this study the end of course performance of students in a language institute was predicted using Rough Set Theory (RST) and Backpropagation Neural Network (BPNN). The design of the model was done taking into account the joint strengths and likelihood of eliminating each other's weaknesses seen in the two individual algorithms. The data was provided from a language institute where a 3 month long in-person English language course aided by an online platform used for assignments is administered. The dataset consists of 10 attributes and one decision of which only four attributes are used in the calculation of the overall average. Using RST in the ROSE2 program, the features were reduced, establishing the 5 attributes that have a higher impact on the overall student performance. Therefore, proving the initial hypothesis stated at the beginning of the study to be true. The following stage was to initialize the BPNN for both (original and reduced) datasets and compare the results to evaluate the accuracy of the proposed model in comparison to the analysis of the original dataset. The results showed that by using the reduced features it is possible to get a better result in the prediction of student performance. This model not only enables the administration of educational institutions to better assess the education system, but it also has the potential to aid teachers in having a more individual approach in helping students at risk. In future studies other factors related to student behavior and educational background could also be integrated to the study.

References:

1. Z. Pawlak, Rough set, *Int. J. Comput. Inf. Sci.* 11 (5) (1982) 341–356
2. S. John Walker, Big data: a revolution that will transform how we live, work, and think, *Int. J. Advertising* 33 (1) (2014) 181–183.
3. Y. Yao, Three-way decisions with probabilistic roughsets, *Inform.Sci.* 180(3) (2010)341–353.
4. Li, X., Jiang, Q., Hsu, M. K., & Chen, Q. (2019). Support or risk? Software project risk assessment model based on rough set theory and backpropagation neural network. *Sustainability (Switzerland)*, 11(17). <https://doi.org/10.3390/su11174513>
5. C.S. Son, Y.N. Kim, H.S. Kim, H.S. Park, M.S. Kim, Decision-making model for early diagnosis of congestive heart failure using rough set and decision tree approaches, *J.Biomed.Inform.* 45(5)(2012)999–1008.
6. P.J. Lingras, Y. Yao, Data mining using extensions of the rough set model, *J. Am.Soc.Inf.Sci.* 49(5)(1998)415–422
7. Priyanka Das, Asit Kumar Das, Janmenjoy Nayak, Danilo Pelusi, Weiping Ding, Group incremental adaptive clustering based on neural network and rough set theory for crime report categorization, *Neurocomputing*, 2019
8. Baker, R. S. (2015). *Big data and education* (2nd ed.). New York, NY: Teachers College, Columbia University.
9. Ferreira-Mello, R., André, M., Pinheiro, A., Costa, E., & Romero, C. (2019). Text mining in education. *WIREs: Data Mining and Knowledge Discovery*, 9(6), e1332
10. Andrzej Janusz [aut], Lala Septem Riza [aut], Dominik 'Slezak [ctb], Chris Cornelis [ctb], Francisco Herrera [ctb], Jose Manuel Benitez [ctb], Christoph Bergmeir [ctb, cre], Sebastian Stawicki [ctb], (2019) Package “RoughSets” (R documentation)
11. Lang, C., Siemens, G., Wise, A., & Gasevic, D. (2017). *Handbook of learning analytics*. SOLAR, Society for Learning Analytics and Research. New York, NY: SOLAR.
12. D'Mello, S. (2017). Emotional learning analytics. In *Handbook of learning analytics* (p. 115). New York, NY: SOLAR.
13. Deep Learning Toolbox Documentation - MathWorks Benelux. (n.d.). Retrieved May 3, 2020, from https://nl.mathworks.com/help/deeplearning/index.html?s_tid=CRUX_lftnav
14. Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers and Education*, 113, 177–194. <https://doi.org/10.1016/j.compedu.2017.05.007>

15. Berens, J., Schneider, K., Görtz, S., Oster, S., & Burghoff, J. (2019). Early Detection of Students at Risk - Predicting Student Dropouts Using Administrative Student Data from German Universities and Machine Learning Methods†. † *Journal of Educational Data Mining*, 11(3), 1–41. Retrieved from <https://jedm.educationaldatamining.org/index.php/JEDM/article/view/389>
16. Yang, T. Y., Brinton, C. G., Joe-Wong, C., & Chiang, M. (2017). Behavior-Based Grade Prediction for MOOCs Via Time Series Neural Networks. *IEEE Journal on Selected Topics in Signal Processing*, 11(5), 716–728. <https://doi.org/10.1109/JSTSP.2017.2700227>
17. Bakhshinategh, B., Zaiane, O. R., ElAtia, S., & Ipperciel, D. (2018). Educational data mining applications and tasks: A survey of the last 10 years. *Education and Information Technologies*, 23(1), 537–553. <https://doi.org/10.1007/s10639-017-9616-z>
18. Aldowah, H., Al-Samarraie, H., & Fauzy, W. M. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics and Informatics*, 37(April 2018), 13–49. <https://doi.org/10.1016/j.tele.2019.01.007>
19. Saedudin, R. R., Sutoyo, E., Kasim, S., Mahdin, H., & Yanto, I. T. R. (2018). Attribute selection on student performance dataset using maximum dependency attribute. *Proceeding - 2017 5th International Conference on Electrical, Electronics and Information Engineering: Smart Innovations for Bridging Future Technologies, ICEEIE 2017, 2018-Janua*, 176–179. <https://doi.org/10.1109/ICEEIE.2017.8328784>
20. Sarra, A., Fontanella, L., & Di Zio, S. (2019). Identifying Students at Risk of Academic Failure Within the Educational Data Mining Framework. *Social Indicators Research*, 146(1–2), 41–60. <https://doi.org/10.1007/s11205-018-1901-8>
21. Cazarez, R. L. U., & Martin, C. L. (2018). Neural networks for predicting student performance in online education. *IEEE Latin America Transactions*, 16(7), 2053–2060. <https://doi.org/10.1109/TLA.2018.8447376>
22. Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, (December 2019), 1–21. <https://doi.org/10.1002/widm.1355>
23. Ramasubramanian, P., Iyakutti, K., & Thangavelu, P. (2009). Enhanced data mining analysis in higher educational system using rough set theory. *African Journal of Mathematics and Computer Science Research*, 2(9), 184–188.

24. Cables, E. H., Lamata, M. T., & Verdegay, J. L. (2019). Uncertainty Management with Fuzzy and Rough Sets (Vol. 377). <https://doi.org/10.1007/978-3-030-10463-4>