Predictive Machine Learning Approach for Complex Problem Solving Process Data Mining

Aleksandar Pejić* and Piroska Stanić Molcer**

*University of Szeged, Doctoral School of Computer Science, Árpád tér 2, 6720 Szeged, Hungary; pejic@inf.u-szeged.hu

**Subotica Tech College of Applied Sciences, Marka Oreškovića 16, 24000 Subotica, Serbia; piroska@vts.su.ac.rs

Abstract: Problem-solving is considered to be an essential everyday skill, in professional as well as in personal situations. In this paper, we investigate whether a predictive model for a problem-solving process based on data mining techniques can be derived from raw log-files recorded by a computer-based assessment system. Modern informatics-based education relies on electronic assessment systems for evaluating knowledge and skills. OECD's PISA 2012 computer-based assessment database was used, which contains a rich problem-solving dataset. The dataset consists of detailed action logs and results for several problem-solving tasks. Two feature sets were extracted from the selected PISA 2012 Climate Control problem solving task: a set of time-based features and a set of features indicating the employment of the VOTAT problem-solving strategy. We evaluated both feature sets with six machine learning algorithms in order to predict the outcome of the problem-solving process, compared their performance and analyzed which algorithms yield better results with respect to the observed feature set. The approach presented in this paper can be used as a potential tool for better understanding of problem-solving patterns, and also for implementing interactive e-learning systems for training problem solving skills.

Keywords: machine learning; data mining; predictive model; problem solving

1 Introduction

The first Programme for International Student Assessment (PISA) assessment was conducted in the year 2000 and since then it is repeated every three years measuring the scholastic performance of 15 years old students in reading, mathematics, and science literacy. The number of participating countries has reached 80 in the year 2018, with more than 540,000 students taking part in the assessment worldwide. A full set of responses from individual students, school principals and parents for each PISA assessment since the year 2000 is available for researchers to engage in their own analysis of the data. Nowadays PISA is one

of the most recognized international large-scale educational assessments and continues to inspire many research projects in various scientific fields.

The computer-based PISA assessment datasets are particularly significant source of information for scientists researching cognitive skills and cognitive processes of students [1-4] because besides the results of tests and questionnaire answers, these datasets also contain steps taken while working on solving a given task. Cognitive skills and processes can be studied in the field of cognitive infocommunications, where it can be analyzed how cognitive processes can co-evolve with infocommunications methods [5]. The computer-generated PISA datasets were also found very interesting by informatics researchers, especially in the fields of data mining and machine learning, where exploring of new methods for extracting information and predicting the successfulness of task solving are the most popular research topics [6-8]. Because the PISA 2012 CBA (computer-based assessment) problem-solving dataset captures in detail the sequences of actions taken by the students while performing complex problem solving, it is especially suitable for extensive analyses of behavioral processes that underlie successful and unsuccessful performance [4].

The computer-based instrument for problem-solving assessment was designed to contain tasks based on real-world situations and to engage students' higher-level cognitive processes [9]. This method of measurement allows analyzing each step of the applied problem-solving strategy. The exploration of cognitive processes allows identifying the possible mistakes of human cognitive systems, especially the preprocessing type of mistakes, such as pattern recognition. It also allows observing the mistakes of higher-level cognitive processes, like problem-solving and reasoning. With the employment of Assistive Technology and intelligent games [10] in cognitive skills training it is possible to bridge the gap between capabilities and expectations, if not completely then at least to some extent [11]. Examination of cognitive capabilities also gives the opportunity to put into use adequate learning games from the socio-cognitive ICT storehouse, for which the difficulty levels are estimated in advance [12].

A significant research topic in the problem-solving field of educational computer-based assessment is the analysis of behavioral data and strategic behavior while solving a complex problem. Most of the research in this area builds on the work of Tschirgi [13], who investigated the differences in reasoning between adults and students while trying to solve a task based on the manipulation of variables. The given task was set in an everyday situation, presented as a short story containing a specific problem to solve. For each story three different answers were offered, where each answer represented a distinct approach, or rather a strategy for solving the task. The three strategies embedded into the answers were vary-one-thing-at-a-time (VOTAT), hold-one-thing-at-a-time (HOTAT), and change-all (CA). Tschirgi noted that subjects employing a specific strategy are often not aware of its logical structure. Recent research has extensively investigated the use of the VOTAT strategy in complex problem-solving tasks set in a computer-based

environment [4, 14-15]. In [14] it was examined whether the prior knowledge of VOTAT strategy from a pen-and-paper environment is relevant for solving an unfamiliar problem in a computer-based environment. While they found that the prior knowledge of a strategy is important, it is not sufficient for application on an unfamiliar problem in a new environment. The effective use of strategic behaviors while solving complex problems was researched in [15]. Their study examined the use of VOTAT and NOTAT (vary no-thing-at-a-time) strategies, finding that students employed an adaptive behavior: when the chosen strategy was effective, students used it with increasing rates, and when it was ineffective, the strategy was used with decreasing rates. VOTAT as an optimal exploration strategy for the Climate Control task on the PISA 2012 computer-based assessment was investigated in [4]. Their results provide important insight for our study, as well as for the research field of complex problem-solving. It was stated in [4] that the application of VOTAT strategy is an indicator of a broader set of strategic competencies and that it increases the overall proficiency in problem-solving.

In this study we are examining the Climate Control problem-solving task from the PISA 2012 computer-based assessment using data mining techniques. Two sets of features were engineered by extracting time spent on different activities while working on the task, and by extracting actions taken while working on the task which employed the VOTAT problem-solving strategy. We compared six different techniques for predicting the successfulness of solving the given problem-solving task and analyzed the importance of the extracted features. Furthermore, we investigated extensively the deep learning algorithm, as it proved to be the best fit for both feature sets. In accordance with this, the following research questions were formulated:

1) Which machine learning algorithm is best suited for predicting the outcome of the Climate Control problem solving task from the PISA 2012 computer-based assessment, considering our datasets are assembled from raw log-file databases?

2) Can the feature set constructed from raw log-files by extracting the actions employing the VOTAT strategy while working on the problem-solving task serves as a predictor for the outcome of the task?

3) Can the feature set constructed from raw log-files by extracting time spent on activities while working on the problem-solving task serves as a predictor for the outcome of the task?

4) Is it possible to further enhance the prediction performance by optimizing the machine learning algorithm most fitting to work with both feature sets?

Section 2 of this paper describes the chosen problem-solving task from PISA and the initial dataset. Section 3 introduces the machine learning algorithms, feature extraction, and assembly of the final dataset. Research results are discussed in Section 4, while Conclusions are drawn in the last section of the paper.

2 PISA Problem Solving Task

The objective of the Climate Control task of the PISA 2012 problem solving computer-based assessment is to discover how to operate a new air conditioner appliance, given that it was delivered without any instructions. The air conditioner unit has three input control sliders (top, central, and bottom), which affect two output parameters - temperature and humidity. The student has two parts of the screen available to work with: the top screen part is graphically representing the air conditioner unit with control sliders (Figure 1), where the behavior of the air conditioner can be explored, and the bottom screen part where the student has to draw the relation between the input and the output variables. The task is considered to be solved correctly when the student draws the exact relation diagram (Figure 2).

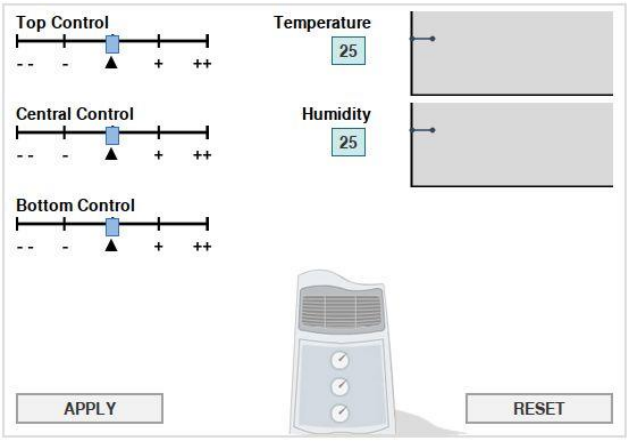

Figure 1
Climate Control problem solving task in PISA 2012 computer-based assessment

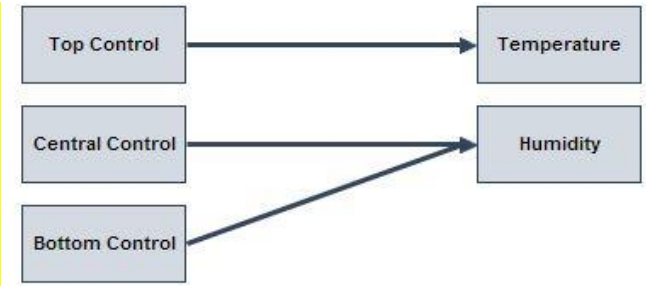

Figure 2
Relation diagram in Climate Control problem solving task

The Climate Control task itself consists of two questions, but in this work, we have considered only the first question (PISA Unit Item Code CP025Q01), which is described above. The problem-solving computer-based assessment of PISA 2012 was undertaken by 44 countries. In the initial preprocessing phase, we have extracted from the PISA dataset only the entries relevant for the Climate Control task. The initial Climate Control CP025Q01 study dataset was congregated from two files: the scored cognitive item response data file which contained assessment results for each individual student, and the problem-solving data files. The latter files contained a record of all steps taken while working on the task: the position of each control slider, which button was clicked, and what actions were taken to draw the relation diagram. For every entry in the data files a sequence order number and a timestamp were assigned.

3 Methods, Features, and Final Dataset

In this research, six machine learning models are built and compared for predicting the students' climate control problem-solving success. These models were built using the following classification algorithms: Naïve Bayes, logistic regression, deep learning, decision tree, random forest, and gradient boosted trees. Classification by machine learning is a widely adopted approach occurring in scientific research focusing on many areas of modern life [16-18]. It is also commonly used in educational research. Educational data mining as an emerging discipline often utilizes machine learning methods in a broad range of research subjects from an educational setting, involving examination of students' performance. Commonly referenced analysis methods are decision trees [19-20]. An extensive review of related research was given in [21]. A comprehensive study of recent research interests, problems, and techniques associated with data mining in education was conducted in [22]. Classification based on data mining techniques, namely decision trees were used in [23-24] to find the factors which influence students' success at the most. Classification and regression trees were used in [25] to predict student performance from activity data on Moodle learning platform, and also for analyzing large scale data from OECD educational indicators and PIRLS (Progress in International Reading Literacy Study) Curriculum Questionnaire in [26]. Classification by support vector machine model on PIRLS data is reported in [27]. Data from the results of TIMSS (Trends International Mathematics and Science Study) was used for classification and prediction of students' successes using machine learning techniques in [28-31].

Classification techniques have various advantages and disadvantages, they behave differently, and generally, their performance highly depends on the dataset. The six chosen classification techniques were evaluated in terms of accuracy, classification error, recall, F-measure, the area under the ROC curve, and runtime.

All of the models were tested in prediction performance. In [32] the PISA assessment dataset was used to compare regression models and neural networks as predictor models. The artificial neural network with two layers had better performance in prediction than regression models. In [33] classification of results of mathematics in PISA 2012 assessment was described. The observed dataset contained mathematics assessment results of Turkey, and decision tree models were built for classification. A logistic regression model was used in [34] to reveal which features have an impact on the success of the reading assessment in PISA 2009. Various models of decision trees are among the most used techniques for examining PISA data sets [35-37]. Often used are also logistic regression models [38-39]. Many papers compare different techniques to find the one yielding the best prediction performance with the given dataset [40-41]. Results of the PISA 2012 mathematics assessment for Finland and for all other countries participating in PISA are analyzed by machine learning methods in [42]. In [43] the mathematics assessment results were analyzed by machine learning methods in the sample of Australia. The database from Turkey's PISA 2015 scientific literacy assessment was used to compare the performance of data mining methods in [44].

This research uses and processes the raw log-file databases for PISA 2012 computer-based items, and analyzes the direct actions of students to uncover the relations between the strategies and the cognitive, problem-solving capabilities of students. All countries participating in the PISA 2012 problem-solving assessment were taken into account. In [4] it was shown, that the VOTAT combination of the switches is in correlation with the success of the problem-solving task. Here, the features used as predictors based on the VOTAT strategy are: V_1, V_2, V_3, and V_4 (1) (2) (3) (4). The number of control slider apply events where two control sliders were set to initial position, and only one control slider was set to a different position, related to all apply events were computed for all three control sliders, the top, the central and the bottom control slider. *Nb* is the number of apply actions when only the bottom switch is changed, others are 0. *Nt* is the number of apply actions when only the top switch is changed, others are 0. *Nc* is the number of apply actions when only central switch is changed, others are 0. As the fourth feature, a total number of apply action is taken (*AN*).

$$V_1 = Nb / AN \tag{1}$$

$$V_2 = Nt / AN \tag{2}$$

$$V_3 = Nc / AN \tag{3}$$

$$V_4 = AN \tag{4}$$

Figures 3, 4 and 5 represent histograms of relative usage of the respective control on the air conditioner device in the Climate Control problem solving task of PISA 2012 computer-based assessment. Figures compare the number of students who finished the task successfully (TRUE), and those who did not (FALSE).

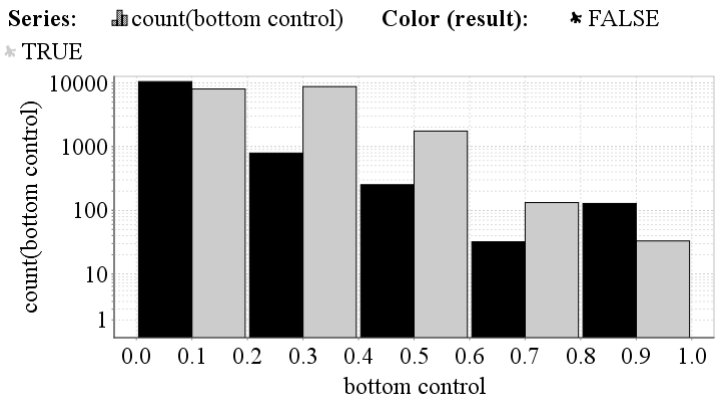

Figure 3
Histogram of relative usage of bottom control based on feature V_1

The horizontal axis shows the percentage ratios of the respective control apply steps on a normalized scale, relative to the total number of apply actions in the problem-solving process. The vertical axis represents the number of students on a logarithmic scale.

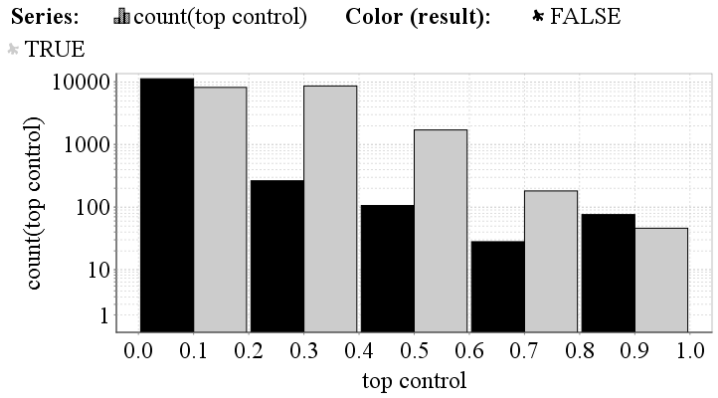

Figure 4
Histogram of relative usage of top control based on feature V_2

Evaluation of histograms reveals that students who tended to experiment with all three controls evenly had the most success in solving the task. This is noticeable when observing histogram bars at a value of 0.3 on the vertical axis, which represents 30% of the relative usage of the respective control. The TRUE/FALSE ratio of successfully solving the task is here the highest. Students who concentrated on experimenting with two controls have also achieved high rate of success.

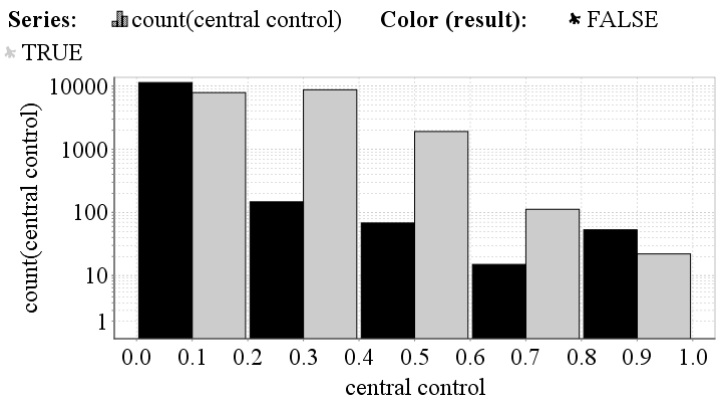

Figure 5
Histogram of relative usage of central control based on feature V_3

We have also considered whether a particular VOTAT-based feature is a good predictor for the given model. There was no single feature equally important across all models. For the Naïve Bayes and the logistic regression models, the relative usage of the central control had the biggest impact on the successful prediction outcome of the model. In more complex models, for instance, the deep learning, the bottom control had the highest importance, while the central control was the least important.

The climate control problem-solving workflow can be divided into three phases: first, reading the instructions of the task and thinking about the problem before taking any action, second, experimenting with the control switches, and third, drawing the relation diagram to explain the relation between input and output variables. These phases were analyzed in the dimension of time. Paradata, a by-product of computer-based data collection was extracted from the problem-solving log-files to get information about features related to the dimension of time [45-46]. The time spent on reading the task can be a good predictor, feature F_1 (5), since students who carefully read the instructions have more chance to finish successfully than students who do superficial reading and do not gain enough insight to the problem. In [47] regression model was built to predict the task completion success, where the regression coefficient showed that the longer the reading time, the bigger the likelihood of success. The second time-related feature was the time spent on experimenting with the switches. This measure was divided by the entire time spent on experimenting and drawing the results by a particular student - feature F_2 (6). The third feature has been computed as the time spent on actions while drawing the results on a diagram, divided by time spent on experimenting and drawing – feature F_3 (7). The entire time normalized through the whole number of students participating in the task was the fourth feature F_4 (8). In [47] the speed of actions was investigated in the PISA climate problem-solving task, and a high correlation was found with the success of the completion.

The number of apply actions divided by the time spent on experimenting was represented with the fifth feature F_5 (9).

$$F_1 = (Faats - Fts)/(Lts - Faats) \tag{5}$$

$$F_2 = (Laats - Faats)/(Lts - Faats) \tag{6}$$

$$F_3 = (Lts - Fdts)/(Lts - Faats) \tag{7}$$

$$F_4 = (T - \min(T_1 : T_k))/(\max(T_1 : T_k) - \min(T_1 : T_k)) \tag{8}$$

$$F_5 = AN/(Laats - Faats) \tag{9}$$

Timestamps used were: First timestamp (*Fts*), First apply action timestamp (*Faats*), Last apply action timestamp (*Laats*), First diagram timestamp (*Fdts*), Last timestamp (*Lts*). The number of apply actions is *AN*, while the number of samples is *k*.

Further analysis of the time-based features was conducted by constructing several histograms. Figure 6 depicts the distribution of feature F_1, which is relative time spent on reading the instructions for the problem-solving task with regards to the time spent on experimenting and drawing the relation diagram.

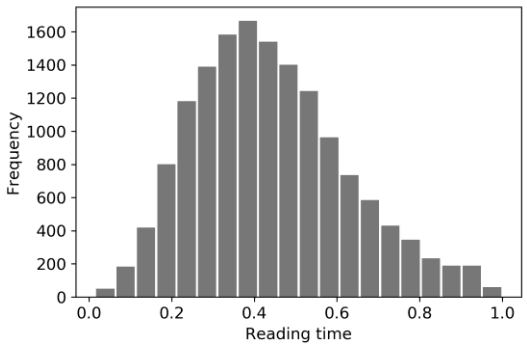

Figure 6

Histogram of relative time spent on reading based on feature F_1

The reading time was 60% shorter than the time spent on experimenting and drawing, thus the reading time was about 30% of the entire spent time on task in the case of most of the students.

Figure 7 is showing the distribution of time spent on experimenting with the control switches relative to the entire time spent on experimenting and drawing the relation diagram.

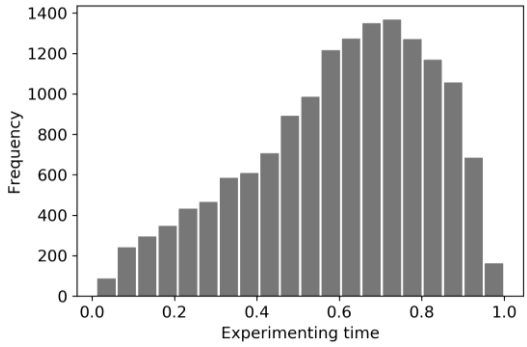

Figure 7

Histogram of relative time spent on experimenting based on feature F_2

Figure 8 depicts the distribution of time spent on drawing the relation diagram for the problem-solving task with regards to the total time spent on experimenting and drawing. These histograms are based on features F_2 and F_3.

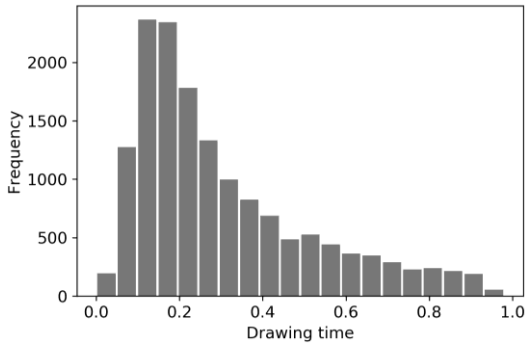

Figure 8

Histogram of relative time spent on drawing based on feature F_3

Figure 9 shows the distribution for feature F_4, namely the time spent on experimenting with control switches and drawing the relation diagram, normalized across the entire set of samples.

Figure 10 depicts the distribution of the number of clicks onto the Apply button in the unit of time. The horizontal axis of the histogram represents the number of clicks per second. According to the histogram, the rate of clicking the apply button is very low. The vast majority of the students carefully considered their next action before clicking the apply button, hence at least a few seconds passed between two clicks to apply button.

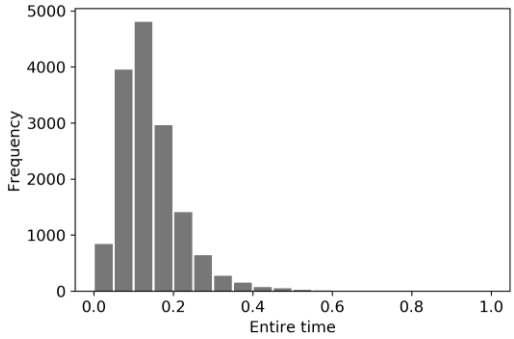

Figure 9
Histogram of entire time spent on task based on feature F_4

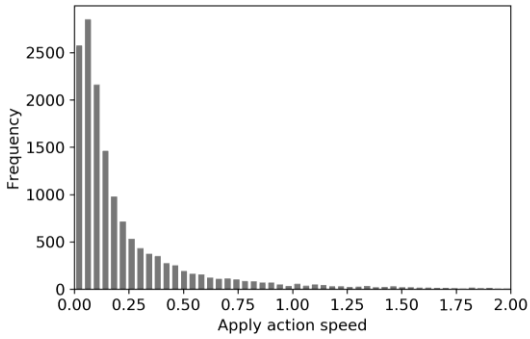

Figure 10
Histogram of apply action speed based on feature F_5

The resulting dataset was combined with the PISA file containing information about the success of the task. In the prediction preprocess phase, the success of students was split into two classes. Upon cleaning the data, downsampling was done to reach balance a between the two classes. After blending the two datasets, preprocessing and eliminating corrupted entries the final dataset had 15194 entries left.

4 Results

The performances of six predictor models are evaluated by accuracy, AUCROC, class recall, class precision, F-measure, classification error, and runtime.

The performance of models using the VOTAT-based set of features is shown in Table 1. In the Naïve Bayes model, feature V_3 (central control usage ratio) had the biggest impact on the classification. As the relative ratio of using the central

control increased, the probability of successfully completing the problem-solving task became higher. The logistic regression model also showed that the most important feature was the percentage of central control steps.

The default deep learning model has 2 hidden layers with 50-50 neurons and rectifier activation in each of two hidden layers, with Bernoulli distribution function and cross-entropy loss function. The decision tree model was optimized. After optimization, it was found that the highest performance of the model is reached at tree depth = 7. Random forest and gradient boosted trees models outperformed all other models in terms of accuracy, both achieving 94.1%. However, their computing runtime was by far the longest among all models.

Table 1

Performance comparison of models using VOTAT-based features

Model	Measures					
	Accuracy	*Runtime (s)*	*AUC*	*Class. Error*	*Recall*	*F-measure*
Naïve Bayes	87,8%	45,0	0,920	13,2%	92,5%	88,5%
Logistic Regression	89,7%	47,0	0,933	10,3%	93,3%	93,3%
Deep Learning	93,9%	81,0	0,947	6,1%	99,7%	94,3%
Random Forest	94,1%	153,3	0,950	5,9%	99,9%	94,5%
Decision Tree	93,4%	48,0	0,950	6,6%	99,9%	93,8%
Gradient Boosted Trees	94,1%	291,1	0,950	5,9%	99,8%	94,5%

Decision tree and deep learning achieved also very good accuracies, 93.4% and 93.9% respectively, with much shorter computing runtimes. Further performance comparison of the two algorithms has shown that deep learning achieves better classification results with *F-measure* value of 94.3%.

Table 2 contains performance measures for the used prediction models based on time-based features.

Table 2

Performance comparison of models using time-based features

Model	Measures					
	Accuracy	*Runtime (s)*	*AUC*	*Class. Error*	*Recall*	*F-measure*
Naïve Bayes	69.8%	49.0	0.743	30.2%	69.7%	70.3%
Logistic Regression	74.3%	41.2	0.743	25.7%	74.3%	74.1%
Deep Learning	77.1%	98.9	0.847	22.9%	77.1%	77.1%
Random Forest	69.8%	171.5	0.740	30.2%	69.5%	69.7%
Decision Tree	73.1%	42.8	0.780	26.9%	73.1%	73.1%
Gradient Boosted Trees	75.7%	311.3	0.830	24.3%	75.6%	75.6%

The deep learning model has achieved the best results with time-based features in all classification performance measurement categories. Computing runtime expectedly took about twice the time compared with the fastest models – logistic regression, decision tree and Naïve Bayes. However, we have focused rather on the classification performance than the model speed, and experimented further with optimization of the neural network in order to achieve even better results. An active field of research in the field of machine learning is automatization of neural network structure optimization [48-50]. In our case, the suboptimal neural network structure was searched by a genetic algorithm using a fitness function defined as the sum of *AUCROC* and *F-measure* values.

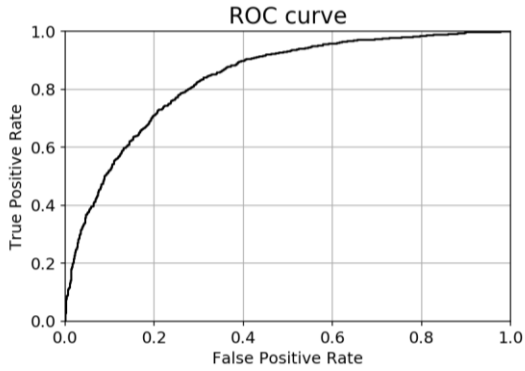

Figure 11
ROC of the neural network optimized by genetic algorithm

In the case of time-based features, the number of generations was set to 10, the number of solutions per population was set to 20 and the number of parents mating to 4, the best structure with the lowest number of neurons had 18 neurons in the first hidden layer and 14 neurons in the second hidden layer. The number of epochs to train the neural network was set to 190. In case when the number of generations was set to 20, solution per population set to 20 and the number of parents mating to 8, the best structure with the lowest number of neurons was with 28 neurons in the first hidden layer and 26 neurons in the second hidden layer. The number of epochs to train the neural network was set to 190. In both cases, the AUCROC was 0.857 and the accuracy 77.9%. The ROC curve of the deep learning model is shown in Figure 11.

When the number of epochs was set to a low value (50 epochs), the best individual was that with the largest number of neurons: in the first hidden layer 38 and in the second hidden layer 22 neurons.

In the case of VOTAT features, the best structure was a network with 20 neurons in the first hidden layer and 23 neurons in the second hidden layer. The accuracy was 94.9%, the AUCROC 0.958. Table 3 summarizes the performance of the optimized deep learning models.

Table 3
Performance comparison of optimized deep learning models

Feature set	Measures				
	Accuracy	*AUC*	*Class. Error*	*Recall*	*F-measure*
VOTAT-based	94.9%	0.958	5.1%	94.8%	94.9%
Time-based	77.9%	0.857	22.1%	77.9%	77.9%

Conclusions

In this study we have built six different machine learning models for predicting the success of students in the Climate Control problem-solving task from PISA 2012. The prediction models were tested with two distinct feature sets: first based on features indicating the use of the VOTAT problem-solving strategy, second based on time-based features. A dataset constructed from a raw log-file database was used for measuring performance of a Naïve Bayes, a logistic regression, deep learning, a random forest, a decision tree, and a gradient boosted trees prediction model. *F-measure* was used to evaluate the performance of these models. Four related research questions were deliberated.

Our first research question was "Which machine learning algorithm is best suited for predicting the outcome of the Climate Control problem-solving task from the PISA 2012 computer-based assessment, considering our datasets are assembled from raw log-file databases?". To answer this question, we have evaluated each model with both datasets and computed the respective *F-measure* value for comparison. Overall, the best suited machine learning algorithm for both of our feature sets is deep learning, whose *F-measure* score is 94.3% with VOTAT-based feature set, and 77.1% with the time-based feature set. The random forest and the gradient boosted trees model both scored slightly higher with the VOTAT-based feature set, 94.5%, but scored significantly lower with time-based features than deep learning, 69.7% and 75.6%. Generally, all six models performed well, especially with the VOTAT-based feature set.

The second research question was "Can the feature set constructed from raw log-files by extracting the actions employing the VOTAT strategy while working on the problem-solving task serve as a predictor for the outcome of the task?". The performance results in Table 1 clearly indicate that the employment of the VOTAT strategy is a strong predictor of the successful outcome of the problem-solving task. All six of the models showed high prediction accuracy with *F-measure* scores above 88.5% when evaluated with the dataset based on the VOTAT-based features.

Answer to the third research question, "Can the feature set constructed from raw log-files by extracting time spent on activities while working on the problem-solving task serve as a predictor for the outcome of the task?", is based on the results displayed in Table 2. Evaluation of the models with the time-based features

has also yielded good results in the context of predicting the outcome of the problem-solving task. While the overall *F-measure* scores are lower compared to the VOTAT-based feature set results, all models have scored between 69.7% and 77.1%. These results confirm the time-based feature set as a good predictor.

To answer the fourth research question, "Is it possible to further enhance the prediction performance by optimizing the machine learning algorithm most fitting to work with both feature sets?", we have taken into account the answer for the first research question, and optimized the neural network structure of our deep learning model for both feature sets using a genetic algorithm using a fitness function defined as the sum of *AUCROC* and *F-measure* values. After optimizing the deep learning model, the *F-measure* score has reached 77.9% with time-based features and 94.9% with VOTAT-based features, which is an increase of 0.8% and 0.6% respectively. Hence, it is possible to enhance the prediction performance by optimizing the machine learning algorithm.

Models built on time-based features might have more potential for further research and applications, as they could be used to enhance interactive e-learning environments. A predictive model using both types of features, VOTAT-based and time-based, could serve for building an effective learning environment with online assistance while improving attention [51] and training learning [52] and problem-solving skills using CogInfoCom supported education methods [53]. Based on the interaction times with the learning environment, or the absence of VOTAT strategy employment, the computer-based learning system could advise to take a specific action or change the strategy in order to increase the likelihood of successfully solving the problem at hand.

As future work, we plan to further study the possibilities of improving the outcome prediction of the PISA problem solving tasks, by evaluating different problem-solving tasks, enhancing the data pre-processing of time-based features and fine-tuning the deep learning model.

References

[1] M. Csernoch and P. Bíró: First year students' attitude to computer problem solving, Proceedings of IEEE Int. Conf. on Cognitive Infocommunications, 2017, pp. 225-230

[2] P. Intaros, M. Inprasitha and N. Srisawadi: Students' problem solving strategies in problem solving - mathematics classroom, Procedia - Social and Behavioral Sciences, No. 116, 2014, pp. 4119-4123

[3] P. Tóth: Learning Strategies and Styles in Vocational Education, Acta Polytechnica Hungarica, Vol. 9, No. 3, 2012, pp. 195-216

[4] S. Greiff, S. Wüstenberg, and F. Avvisati: Computer-generated log-file analyses as a window into students' minds? A showcase study based on the PISA 2012 assessment of problem solving, Computers & Education, Vol. 91, 2015, pp. 92-105

[5] Baranyi P., Csapo A., Sallai Gy.: Cognitive Infocommunications (CogInfoCom) Springer, 2015

[6] A. Pejic and P. S. Molcer: Exploring data mining possibilities on computer based problem solving data, Proceedings of IEEE Int. Symp. Intelligent Systems and Informatics, 2016, pp. 171-176

[7] X. Qiao and H. Jiao: Data mining techniques in analyzing process data: A didactic, Front. Psychol., Vol. 9, 2018, pp. 1-11

[8] C. H. Yu, C. Kaprolet, A. Jannasch-Pennell and S. DiGangi: A Data Mining Approach to Comparing American and Canadian Grade 10 Students' PISA Science Test Performance, Journal of Data Science, Vol. 10, 2012, pp. 441-454

[9] OECD, Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006, 2006, Paris: OECD Publishing

[10] C. Sik-Lanyi et al.: How to develop serious games for social and cognitive competence of children with learning difficulties, in Proceedings of IEEE Int. Conf. on Cognitive Infocommunications, Debrecen, 2017, pp. 321-326

[11] L. Izsó: The significance of cognitive infocommunications in developing assistive technologies for people with non-standard cognitive characteristics: CogInfoCom for people with nonstandard cognitive characteristics, in Proceedings of IEEE Int. Conf. on Cognitive Infocommunications, Győr, 2015, pp. 77-82

[12] M. Szabó, K. D. Pomázi, B. Radostyán, L. Szegletes and B. Forstner: Estimating task difficulty in educational games, in Proceedings of IEEE Int. Conf. on Cognitive Infocommunications, Wroclaw, 2016, pp. 397-402

[13] J. Tschirgi: Sensible Reasoning: A Hypothesis about Hypotheses, Child Development, Vol. 51, No. 1, Mar., 1980, pp. 1-10

[14] S. Wüstenberg, M. Stadler, J. Hautamäki and S. Greiff: The Role of Strategy Knowledge for the Application of Strategies in Complex Problem Solving Tasks, Technology, Knowledge and Learning, Vol. 19, No. 1-2, 2014, pp. 127-147

[15] C. Lotz, R. Scherer, S. Greiff and J. R. Sparfeldt: Intelligence in action – Effective strategic behaviors while solving complex problems, Intelligence, Vol. 64, No. January, 2017, pp. 98-112

[16] Z. Fazekas, G. Balázs, P. Gáspár: ANN-based Classification of Urban Road Environments from Traffic Sign and Crossroad Data, Acta Polytechnica Hungarica, Vol. 15, No. 8, 2018, pp. 83-100

[17] S. Vadloori, Y. P. Huang and W. C. Wu: Comparison of Various Data Mining Classification Techniques in the Diagnosis of Diabetic Retinopathy, Acta Polytechnica Hungarica, Vol. 16, No. 9, 2019, pp. 27-46

[18] O. Solmaz, M. Ozgoren and M. H. Aksoy: Hourly cooling load prediction of a vehicle in the southern region of Turkey by Artificial Neural Network, Energy Conversion and Management, Vol. 82, 2014, pp. 177-187

[19] B. K. Baradwaj and S. Pal: Mining Educational Data to Analyze Students' Performance, International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6, 2011, pp. 63-69

[20] R. Asif, A. Merceron, S. A. Ali and N. G. Haider: Analyzing students' performance using educational data mining, Computers & Education, 2017

[21] A. M. Shahiri, W. Husain and N. A. Rashid: A Review on Predicting Student's Performance using Data Mining Techniques, Procedia Computer Science, Vol. 72, 2015, pp. 414-422

[22] B. Quadir, N. Chen and P. Isaias: Analyzing the educational goals, problems and techniques used in educational big data research from 2010 to 2018, Interactive Learning Environments, 2020

[23] F. Martínez Abad and A. A. Chaparro Caso López: Data-mining techniques in detecting factors linked to academic achievement, School Effectiveness and School Improvement, 2016, DOI: 10.1080/09243453.2016.1235591

[24] M. B. Sanzana, S. S. Garrido and C. M. Poblete: Profiles of Chilean students according to academic performance in mathematics: An exploratory study using classification trees and random forests, Studies in Educational Evaluation, Vol. 44, 2015, pp. 50-59

[25] K. Casey and D. Azcona: Utilizing student activity patterns to predict performance, International Journal of Educational Technology in Higher Education, Vol. 14, No. 4, 2017

[26] F. Alivernini: An exploration of the gap between highest and lowest ability readers across 20 countries, Educational Studies, Vol. 39, No. 4, 2013, pp. 399-417

[27] Y. Xiao and J. Hu: Assessment of Optimal Pedagogical Factors for Canadian ESL Learners' Reading Literacy Through Artificial Intelligence Algorithms, International Journal of English Linguistics, Vol. 9, No. 4, 2019, pp. 1-14

[28] Y. Jin Eun and R. Minjeong: TIMSS 2015 Korean Student, Teacher, and School Predictor Exploration and Identification via Random Forests, The SNU Journal of Education Research, Vol. 26, No. 4, 2017, pp. 43-61

[29] E. Filiz and E. Öz: Finding the Best Algorithms and Effective Factors in Classification of Turkish Science Student Success, Journal of Baltic Science Education, Vol. 18, No. 2, 2019, pp. 239-253

[30] S. K. Depren, Ö. E. Aşkin and E. Öz: Identifying the Classification Performances of Educational Data Mining Methods: A Case Study for TIMSS, Educational Sciences: Theory and Practice, Vol. 17, No. 5, 2017, pp. 1605-1623

[31] E. Filiz and E. Öz: Educational Data Mining Methods for TIMSS 2015 Mathematics Success: Turkey Case, Sigma Journal of Engineering and Natural Sciences, Vol. 38, No. 2, 2020, pp. 963-977

[32] S Benzer and R. Benzer: Examination of International Pisa Test Results with Artificial Neural Networks and Regression Methods, The Journal of Defense Sciences, Vol. 16, No. 2, 2017, pp. 1-13

[33] G. Aksu, Gökhan and C. O. Güzeller: Classification of PISA 2012 Mathematical Literacy Scores Using Decision-Tree Method: Turkey Sampling, Egitim ve Bilim, Vol. 41, No. 185, 2016, pp. 101-122

[34] K. Kasapoglu: A Logistic Regression Analysis of Turkey's 15-year-olds' Scoring above the OECD Average on the PISA'09 Reading Assessment, Educational Sciences, Theory and Practice, Vol. 14, No. 2, 2014, pp. 649-667

[35] F. Martínez-Abad, A. Gamazo and M. Rodríguez-Conde: Educational Data Mining: Identification of factors associated with school effectiveness in PISA assessment, Studies in Educational Evaluation, Vol. 66, 2020, 100875

[36] F. Alivernini, S. Manganelli and F. Lucidi: The last shall be the first: Competencies, equity and the power of resilience in the Italian school system, Learning and Individual Differences, Vol. 51, 2016, pp. 19-28

[37] C. H. Yu and S. Digangi: A data mining approach to compare American and Canadian Grade 10 students in PISA 2006 Science test performance, Journal of Data Science, Vol. 10, 2012, pp. 441-464

[38] S. K. Depren: Prediction of students' science achievement: An application of multivariate adaptive regression splines and regression trees, Journal of Baltic Science Education, Vol. 17, No. 5, 2018, pp. 887-903

[39] A. Gorostiaga and J. L. Rojo-Álvarez: On the use of conventional and statistical-learning techniques for the analysis of PISA results in Spain, Neurocomputing, 2015

[40] M. Ozel, S. Caglak and M. Erdogan: Are affective factors a good predictor of science achievement? Examining the role of affective factors based on PISA 2006, Learning and Individual Differences, Vol. 24, 2013, pp. 73-82

[41] J. M. Rajchert, T. Żułtak and M. Smulczyk: Predicting reading literacy and its improvement in the Polish national extension of the PISA study: The role of intelligence, trait- and state-anxiety, socio-economic status and school-type, Learning and Individual Differences, Vol. 33, 2014, pp. 1-11

[42] M. Saarela, J. ulent Yener, M. J. Zaki, and T. Kärkkäinen: Predicting Math Performance from Raw Large-Scale Educational Assessments Data: A Machine Learning Approach, in Proceedings of Int. Conf. on Machine Learning, Vol. 48, New York, 2016, pp. 1-8

[43] F. Gabriel, J. Signolet, and M. Westwell: A machine learning approach to investigating the effects of mathematics dispositions on mathematical literacy, Int. Journal of Research & Method in Education, Vol. 41, No. 3, 2018, pp. 306-327

[44] S. Büyükkıdık, B. Bakırarar, and O. Bulut: Comparing the Performance of Data Mining Methods in Classifying Successful Students with Scientific Literacy in PISA 2015, Computer Science, 2018, pp. 68-75

[45] F. Kreuter: Improving surveys with paradata: analytic uses of process information, Vol. 581, 2013, Wiley, Hoboken

[46] U. Kroehne, F. Golhammer: How to conceptualize, represent, and analyze log data from technology based assessments? A generic framework and an application to questionnaire items, Behaiormetrika, Vol. 45, 2018, pp. 527-56

[47] Y. Chen, X. Li, J. Liu and Z Ying: Statistical Analysis of Complex Problem-Solving Process Data: An Event History Analysis Approach, Frontiers in Psychology, Vol. 10, 2019, pp. 1-10

[48] B. van Stein, H. Wang and T. Bäck: Automatic Configuration of Deep Neural Networks with Efficient Global Optimization, Int. Joint Conf. on Neural Networks, 2018, pp. 1-7

[49] K. Park, D. Shin and S. Chi: Variable Chromosome Genetic Algorithm for Structure Learning in Neural Networks to Imitate Human Brain, Applied Sciences, Vol. 9, 2019, pp. 1-10

[50] B. Islam, M. Q. Raza: Optimization of neural network architecture using genetic algorithm for load forecasting, Proceedings of Int. Conf. on Intelligent and Advanced Systems, Kuala Lumpur, 2014, pp. 1-6

[51] J. Katona, A. Kovari: The Evaluation of BCI and PEBL-based Attention Tests. Acta Polytechnica Hungarica, Vol. 15, No. 3, 2018, pp. 225-249

[52] J. Katona, A. Kovari: Examining the Learning Efficiency by a Brain-Computer Interface System. Acta Polytechnica Hungarica, Vol. 15, No. 3, 2018, pp. 251-280

[53] A. Kovari: CogInfoCom Supported Education: A review of CogInfoCom based conference papers. 9th IEEE International Conference on Cognitive Infocommunications, 2018, pp. 233-236