JOURNAL OF E-LEARNING AND KNOWLEDGE SOCIETY
Vol. 17, No. 1 (2021), pp. 62-71

# Prediction of engineering students' virtual lab understanding and implementation rates using SVM classification

Teodora Hristova[a,1], Katerina Gabrovska-Evstatieva[b], Boris Evstatiev[c]

[a]*University of Mining and Geology "St. Ivan Rilski", Dept. of Electrical Engineering – Sofia (Bulgaria)*

[b]*University of Ruse Angel Kanchev, Dept. of Computer Science – Ruse (Bulgaria)*

[c]*University of Ruse Angel Kanchev, Dept. of Electronics – Ruse (Bulgaria)*

*(submitted: 28/12/2020; accepted: 8/7/2021; published: 13/7/2021)*

**Abstract**

In 2020 many universities were forced to switch to a distant form of education because of the COVID-19 lockdown. This was especially challenging for the engineering specialties, where laboratory and practical exercises are a fundamental part of the educational process. This study presents results from the electrical engineering education in two Bulgarian universities, where the Engine for Virtual Electrical Engineering Equipment was used as a tool for providing virtual labs. At the end of the semester the students were asked to fill in a survey, accounting for their learning program, years of studying, experience with virtual and real labs and the instructions delivery methods used. Data mining algorithms were utilized with the aim to predict students' rate of understanding and rate of implementation when dealing with virtual labs. Initially, a regression analysis model was created which achieved R-squared above 95%. However, the verification of the model showed an unsatisfactory prediction success rate of 37%. Next, SVM classification was utilized. The verification showed its success rates for predicting the rate of understanding and rate of implementation were 83% and 86%, respectively. This approach could be used to optimize the educational experience of students, using virtual labs, as well as for identification of students that might need additional support and instructions.

**KEYWORDS:** Classification, Prediction, Virtual Labs, Rate of Understanding, Rate of Implementation

**DOI**
https://doi.org/10.20368/1971-8829/1135420

**CITE AS**
Hristova, T., Gabrovska-Evstatieva, K., & Evstatiev, B. (2021). Prediction of engineering students' virtual lab understanding and implementation rates using SVM classification**.** *Journal of e-Learning and Knowledge Society*, *17*(1), 62-71. https://doi.org/10.20368/1971-8829/1135420

## 1. Introduction

Announced measures to protect the population during the COVID-19 pandemic in many countries have led to the closure of schools, universities, large shops, hotels and even factories. In the era of Internet technologies with developed communication channels, the education continued in various remote forms - asynchronous and synchronous, using video conferencing, sent documents, chat rooms, with various feedback between teachers and students (Dimitrov, 2020). Each teacher or school chose the most appropriate way according to the nature of the discipline, course, base, and personal training. Furthermore, teachers and professors had to decide when to organize online consultation, make sure students keep to the deadlines, prepare different assignments and verify the submitted ones in a timely manner (Mladenova et al., 2020). In this situation, difficulties arose with some engineering courses that required laboratory training of students. During these classes, students are commonly trained in a real environment to work with machines and equipment, which includes measuring, connecting circuits, tuning and starting motors, adjusting devices of apparatus.

With the announcement of isolation, teachers from technical universities had to quickly develop programs

[1] corresponding author - email: teodora@mgu.bg – address: Prof. Boyan Kamenov Str., Sofia (Bulgaria)

 

for electrical engineering education of specific laboratory exercises. The existing Engine for Virtual Electrical Engineering Equipment - EVEEE (University of Ruse Angel Kanchev, n.d.) was used to prepare and conduct laboratory exercises in several electrical engineering courses of the University of Ruse Angel Kanchev (URAK) and the University of Mining and Geology (UMG). EVEEE is a 2D environment, representing a 3D virtual reality, which includes all the phases a real laboratory exercise has: connecting the equipment together using virtual cables, plugging/unplugging elements, tuning the equipment and conducting measurements. Considering most of the students didn't have enough experience with virtual labs, they received instructions using different methods, such as video conferencing, text materials and recorded video instructions. The training was relatively successful according to the results shown by the students at the end of the semester. Nevertheless, considering the uncertainty of the COVID-19 situation, it was important to investigate the students' success rate when working with the virtual equipment. Therefore, it was necessary to assess the ease of understanding and ease of implementation by students when dealing with the virtual laboratory exercises using data mining algorithms (DMA). DMA are widely used in education for predicting dropouts (Kumar et al., 2017b), improving institutional effectiveness (Alturki et al., 2020), predicting student amelioration (Anoopkumar and Rahman, 2016), faculty performance analysis (Anoopkumar and Rahman, 2016) and others.

The educational data mining is an iterative process that includes raw data, application of data mining techniques, interpretation of the results and recommendations for the educational system (Romero and Ventura, 2007; Kumar, 2015). In Alyahyan and Düştegör (2020) was proposed a six-stage framework for implementation of data mining techniques to predict student success: data collection; data preparation; statistical analysis; data preprocessing; data mining implementation; and result evaluation.

Previous studies have shown that predicting students' performance can help universities provide timely measures in order to improve the success rate (Alyahyan and Düştegör, 2020). Anoopkumar and Rahman (2016) performed a review on the data techniques used to predict student amelioration from 2005 to 2015. Classification methods are leading significantly, followed by statistical methods and visualization analysis. In another study (Alturki et al., 2020) was performed a review on the prediction of higher education achievements. It showed that classification and regression algorithms are commonly used for forecasting students' achievements at course and degree level. In Kumar et al. (2017a) were investigated different data mining algorithms used for prediction of students' performance in education. According to their findings, the most commonly used algorithms are Decision trees, Naïve Bayes, Artificial neural networks, Rule based and K-Nearest neighbor (kNN). The reported maximum and minimum prediction accuracy vary depending on the algorithm but are generally within the range 55%-100%.

Numerous studies have used various data analysis methods to predict student's performance. Al luhaybi et al. (2018) used classification algorithms to predict students at high risk of failing a module. The results showed that the Naïve Bayes (NB) gives higher accuracy, compared to the C4.5 decision tree algorithm. In another study (Osmanbegovic and Suljic, 2012), the authors used three data mining algorithms to predict the success in a course and the performance of the learning methods. The study showed that the Naïve Bayes method achieved better results, compared to Multilayer Perceptron (MLP) and decision trees. The comparison was based on their predictive accuracy.

In Saa (2016) models for prediction of students' grades were created, based on training data collected through a survey. Multiple decision trees and Naïve Bayes algorithms were used. The average classification accuracy varied from 33.3% to 40%. In Bhutto et al. (2020) were used logistic regression and the Support Vector Machine (SVM) classification to predict student's academic performance. The results showed that the SVM algorithm has higher accuracy in the investigated cases. Kabakchieva (2013) compared six classification algorithms for predicting student's performance – C4.5, Naïve Bayes, Bayes Net, kNN 100, kNN 250 and the rule-based JRip. The most influencing factors were the admission score and the number of failures at the first-year exams. Nevertheless, all algorithms didn't perform very good with predictions rates varying in the range 52%-67%.

Alqurashi (2019) used regression analysis to investigate the relationship among four independent variables (online learning self-efficacy, learner-content interaction, learner-instructor interaction and learner-learner interaction). The results showed that the first three have a critical role on students' satisfaction and perceived learning. In another study. Tsiakmaki et al. (2020) tried to predict students at risk of failure using neural networks with different predictors, such as gender, course, pass/fail, page/folder views, assignment views, etc.

Other studies were aimed at engineering education. In Adekitan and Salau (2019) a data mining approach was used to validate the assumption that the performance of engineering students in the first three years is the most important for their final cumulative grade point average. The study used the program and the year of entry as predictors with different data mining algorithms. The highest accuracy of 89% was achieved using linear and quadratic regression models with coefficient of determination 0.955 and 0.957, respectively. In another study, Taodzera et al. (2017) used five classification algorithms to predict engineering students' success – decision trees, NB, SVM, artificial neural networks and linear regression. The predictors used were Math score, Physics score, ethnicity, school province and age. The prediction accuracy varied from 60% to 67%. In a similar study, Buenaño-Fernández et al. (2019) used data from the academic management system and decision trees classification to predict the pass/fail rates of different engineering courses. To the best of our knowledge, there are no studies investigating the rates of implementation of virtual labs in engineering education.

The aim of this study is to forecast the student's understanding and implementation rates when working with virtual laboratories in their engineering classes. It will be demonstrated that analysis of data obtained from a preliminary survey could be used to optimize the virtual labs learning experience, which is its main novelty. This will help students to successfully graduate with the acquisition of quality knowledge. The remaining of the paper is structured as follows: in section two is presented the structure of the questionnaire and the methodology for its analysis; in section three are presented a summary of the survey results, statistical analysis, the obtained forecasting models and a case study presenting its application; in the fourth section the obtained results are discussed and summarized.

## 2. Materials and Methods

### 2.1 The teaching methodology

The standard training for engineers in Bulgaria includes three types of frontal lessons: lectures, laboratory exercises and tutorials. During the labs, students work either in groups or individually and perform different tasks, such as connection of circuits, measurements of electric quantities (voltage, current power, resistance), verification of different laws and theorems (Kirchoff's laws, Faraday's law, Thevenin's theorem, etc.), comparison between experimental and theoretically expected values, etc. Thereafter, each student summarizes the obtained experimental results and makes appropriate conclusions in an individual report, which is given to the lecturer for verification and assessment.

This study presents results from the courses "General Electrical Engineering", "Theoretical Electrical Engineering" and "Electrical Measurements" which were taught remotely to students studying Mining engineering, Electrical engineering and Computer engineering. All instructions were delivered via several channels: video conferencing, e-learning websites, pdf files, pre-recorded video instructions, etc.

The implementation of the educational process during the spring semester of 2020 was done in several steps: needs analysis, preparation of teaching materials, selection of teaching methods, increase of competencies and selection of assessment methods (Evstatiev and Hristova, 2020).

In order to follow the standard training procedure, the labs training was implemented in the EVEEE environment, where the students can do the necessary tasks in a realistic environment. Furthermore, the implemented virtual equipment is an exact copy of the real one, which allows trainees to learn to work with the equipment. In the EVEEE environment student can implement a given circuit by plugging elements in a breadboard, connecting the equipment with virtual cables, tuning it up and observing its readings. Furthermore, the students had to write down, analyze and summarize the lab results in specially prepared online reports, which were implemented in Google Sheets. During the training the students gradually got used to the virtual environment, performed the set tasks and submitted electronic reports. The acquisition of useful skills for working with a specific software product has been previously reported as an advantage (Anastasova, 2016).

### 2.2 The questionnaire

In order to perform this study a questionnaire has been developed. It can be divided into two parts. The first part includes six questions aimed at identifying the profile of the participants (Table 1). The fifth and the sixth question were aimed at obtaining the previous experience of students in terms of using virtual and real laboratory equipment.

| № | Questions | Answers |
|---|---|---|
| 1 | In which university do you study? | Open question |
| 2 | What is the specialty of your study? | Open question |
| 3 | What is the form of your studying? | Single choice question:<br>Full-time student<br>Part-time student |
| 4 | How long have you been studying? | Single choice question:<br>First year student (1)<br>Second year student (2)<br>Third year student (3)<br>Fourth year student (4) |
| 5 | Have you used virtual labs before? | Yes/No question |
| 6 | Have you done similar laboratory exercises with real equipment? | Yes/No question |

**Table 1** - First group of questions regarding the profile of the participants.

The next group of questions was aimed at evaluating the experience of the students with the application of the EVEEE environment (Table 2). The first question is open and allows students to list the courses in which they have used the environment. The next three questions were aimed at identifying the types of synchronous and asynchronous instruction delivery methods, their tutors used - "recorded video instructions", "instructions in text form" and "audio/video conference instructions". The fifth and the sixth questions of the group were aimed at obtaining the student's understanding and implementation of the given instructions, with the answers varying from "Strongly disagree" to "Strongly agree".

| № | Questions | Type |
|---|---|---|
| 1 | In which courses have you used the EVEEE virtual environment? | Open question |
| 2 | Did your tutor use recorded video instructions? | Yes/No |
| 3 | Did your tutor use instructions in text format (PDF files, e-mails, etc.)? | Yes/No |
| 4 | Did your tutor deliver synchronous instructions using audio/video conferencing (Skype, ZOOM, BBB, etc.)? | Yes/No |
| 5 | Do you agree with the following statement: It was easy to understand how to work with the virtual equipment? | Agree/disagree question |
| 6 | Do you agree with the following statement: It was easy to implement the assigned tasks during the virtual laboratory exercises? | Agree/disagree question |

**Table 2** - Second group of questions regarding the experience with the EVEEE environment.

In order to analyze the survey results, the answers of the "Agree/Disagree" questions and of the "Yes/No" questions were given numerical values as shown in Table 3. Furthermore, numerical meanings were given to the student's specialty answers.

| Answers | Numerical value |
|---|---|
| *Agree…disagree questions* | |
| I strongly disagree | 1 |
| I disagree | 2 |
| Cannot decide | 3 |
| I agree | 4 |
| I strongly agree | 5 |
| *Yes/No questions* | |
| No | 1 |
| Yes | 2 |
| *What is your specialty question* | |
| General engineering | 1 |
| Computer systems and technologies (CST) | 2 |
| Electrical power engineering (EPE) | 3 |

**Table 3** - Corresponding numerical values of the single choice answers.

## 2.3 Data analysis

The goal of the study is to obtain a model that allows prediction of the student's rates of understanding and implementing virtual laboratories. This would allow us to understand which factors could be used to influence the success rate as well as to identify students which might need additional support with the labs. Therefore, nine factors were selected for further analysis (Table 4). Out of them, 7 are predictors, one is either predictor or target (the Ease of understanding - $u$) and one is a target (the Ease of implementation - $i$).

| Factor | Abbreviation | Role |
|---|---|---|
| Specialty | sp | Predictor |
| Years of studying | y | Predictor |
| Experience with virtual labs | ve | Predictor |
| Experience with real labs | re | Predictor |
| Application of recorded video instructions | r | Predictor |
| Application of text instructions | t | Predictor |
| Application of synchronous video conferencing | v | Predictor |
| Ease of understanding | u | Predictor/Target |
| Ease of implementation | i | Target |

**Table 4** - Predictors and targets of the analysis.

Our initial goal is to assess the correlations between the selected predictors and targets. This is done in two ways:

- By obtaining the average target values for the different values of the predictors – this would allow us to obtain preliminary information about the distributions of $u$ and $i$ depending on the values of the predictors;
- Using Pearson's correlation to obtain the "u" and "i" dependency on the other predictors – this would allow us to identify if strong correlations exist.

The next step is to obtain a model, which allows accurate prediction of the factors $u$ and $i$ using the available predictors. Two approaches are selected for investigation – multiple regression and classification. The aim is to compare the results of the two approaches in order to select the most effective teaching methods for laboratory virtual exercises. In both studies, the target function is the ability of students to cope with the assigned tasks.

### *Multiple regression model*

Regression analysis methods are designed for analysis of continuous variables in numerical form. In the investigated situation, the variables are heterogenous, and many of them are of type Boolean (the Yes/No questions), which is not very appropriate for multiple regression analysis. Therefore, a decision was taken that all variables should be normalized to take values in the range (0…1]. For example, the scale for the Agree/Disagree questions become: I strongly disagree – 0.2; I disagree – 0.4; Cannot decide – 0.6; I agree – 0.8; I strongly agree – 1. Similarly, the Yes/No questions become: No – 0.5; Yes – 1. Using the normalized values, a Fischer matrix is synthesized. After an appropriate multiple regression model is selected it is verified using the training data. Considering the model returns continuous data, it is categorized as shown in Table 5.

| Rule | Category |
|---|---|
| If $i_{mod} \geq 0.9$ | $i_{mod} = 1$ |
| If $i_{mod}. \geq 0.7$ and $i_{mod} < 0.9$ | $i_{mod} = 0.8$ |
| If $i_{mod}. \geq 0.5$ and $i_{mod} < 0.7$ | $i_{mod} = 0.6$ |
| If $i_{mod}. \geq 0.3$ and $i_{mod} < 0.5$ | $i_{mod} = 0.4$ |
| If $i_{mod} < 0.3$ | $i_{mod} = 0.2$ |

**Table 5 -** Categorization of the modelled multiple regression data.

### *Classification model*

Another data analysis is performed using the SVM classification algorithm. Unlike the regression analysis approach where the data should be numerical, the classification methods can accept both numerical and categorial predictors/targets. Considering the nature of the survey data, all variables are considered to be categorical. Once a classification model is trained, it is verified using the training data.

## 3. Results

### 3.1 Preliminary analysis

The study was performed at the end of the summer semester of 2020 in two Bulgarian universities where the EVEEE environment is used. All students were engaged in a formal education during the COVID-19 pandemic. Furthermore, ERASMUS+ students from France visiting RUAK also took part in the education process and survey.

Considering the different participants, the described questionnaire was developed in two languages – Bulgarian and English, to be used by the Bulgarian and French students, respectively. After the semester's end, they were asked to fill in the survey. The profile of the participants is summarized in Table 6. Students from several engineering programs participated, most of them studying Computer systems and technologies and Electrical power engineering. Furthermore, most of them were first year full-time students and 71% of them have not used virtual laboratories before.

Next, according to the developed questionnaire, the respondents answered six questions on their experience with the EVEEE environment. When asked for the methods, which their tutors used to deliver their instructions for implementing the virtual labs, 54% of them stated that recorded video instructions were used, 46% stated they received instructions in text form (PDF, etc.) and 23% of them were instructed using audio or video conferencing. It should be noted that this was a multiple-choice question and therefore many of the participants selected more than one answer. The survey results are summarized in Table 7 where the average understanding rate and the average implementation rate are shown for the different values of the predictors. The distributions of the $u$ and $i$ answers are presented in Figure 1.

| Category | Profile |
|---|---|
| University | UMG: 30<br>URAK: 36<br>ECE Paris (at RUAK): 4 |
| Specialty | General engineering: 14<br>CST: 32<br>EPE: 24 |
| Type of formal education | Full-time students: 46<br>Part-time student: 24 |
| Years of study | First year: 36<br>Second year: 26<br>Third year: 6<br>Fourth year: 2 |
| Experience with virtual labs | No: 50<br>Yes: 20 |
| Experience with real labs | No: 30<br>Yes: 40 |

**Table 6 -** Profile of the participants.

| Predictor | Answers | Average u | Average i |
|---|---|---|---|
| y | 1 | 4.17 | 3.78 |
| | 2 | 4.15 | 3.54 |
| | 3 | 4.33 | 4.33 |
| | 4 | 4.00 | 5.00 |
| sp | Other engineering | 4.08 | 3.67 |
| | CST | 4.19 | 3.88 |
| | EPE | 3.71 | 3.71 |
| r | No | 3.81 | 3.44 |
| | Yes | 4.47 | 4.05 |
| t | No | 4.32 | 3.89 |
| | Yes | 4.00 | 3.63 |
| v | No | 4.19 | 3.74 |
| | Yes | 4.13 | 3.88 |
| ve | No | 4.05 | 3.85 |
| | Yes | 4.44 | 4.11 |
| re | No | 4.47 | 4.00 |
| | Yes | 3.90 | 3.75 |

**Table 7 -** Average values of $u$ and $i$ for the different predictors.

Next, the Pearson's correlations between the predictors and the targets were investigated (Table 8). It can be seen that the understanding $u$ has a low correlation with the specialty $sp$, the experience with virtual labs $ve$ and the use of recorded video materials $r$. This shows that students are computer literate and able to cope successfully in a virtual environment, regardless of their specialty. The results also showed that the ease of implementation has a low correlation with the virtual labs experience and the use of recorded video materials. Furthermore, there was a medium correlation with the rate of understanding. In other words, in order to achieve good rate of implementation, the students should understand how to work with the virtual equipment. The last statement is expected but it also confirms that the ease of understanding could be used as a predictor for forecasting the ease of implementation. In general, there were no strong correlations between the predictors and targets. This indicates that it is necessary to create a more complex model in order to assess the specifics of the situation.

## 3.2 Multiple regression model

Previous authors reported that there is no universal tool when it comes to educational data mining (Slater et al., 2017). The different software tools are suited for different tasks, which is also the reason we selected different tools to create the two models. A multiple regression analysis has been implemented using the STATGRAPHICS software. According to the developed methodology, all predictors and targets were normalized to take values in the range (0…1]. In the initial modeling, the following variables were included independently of each other: year of study, specialty, previous experience and teaching methods to determine their correlation with the ability to perform the tasks. The models were adequate according to the theory with a coefficient of determination as high as 91%. In subsequent simulations, the variable teaching methods according to the specialty and our understanding according to the long-term experience were investigated.

The following solution candidate was selected for forecasting the rate of implementation $i$:

$$i = 0.462938 \cdot y^2 + 0.257946 \cdot ve \cdot \sqrt{u} + 0.178005 \frac{v}{ve} + 0.415603 \cdot u \cdot \sqrt{r}$$

The obtained continuous model has the following accuracy:

- R-squared - 95.75%;
- Standard Error – 0.171.

Considering the standard error is 0.171, the accuracy of the model might not be appropriate for the situation, which should be further investigated.

The statistical analysis for the coefficients of the multiple regression is presented in Table 9. Considering all P-values are lower than 0.05, the terms are statistically significant at the 95.0% confidence level.

The model shows that the ability of students to perform the assigned tasks corresponds in direct proportion to the understanding of the taught material. For parameters such as experience with virtual laboratories and years of experience at the university, no exact conclusion can be made, as these parameters are interrelated. It is also not possible to determine exactly the most effective method of teaching. Considering the model generates continuous data, it was rounded to the nearest category, as explained in the methodology. After the categorization, 63% of the samples were identified incorrectly, thus its success rate is only 37%. Therefore, the obtained multiple regression

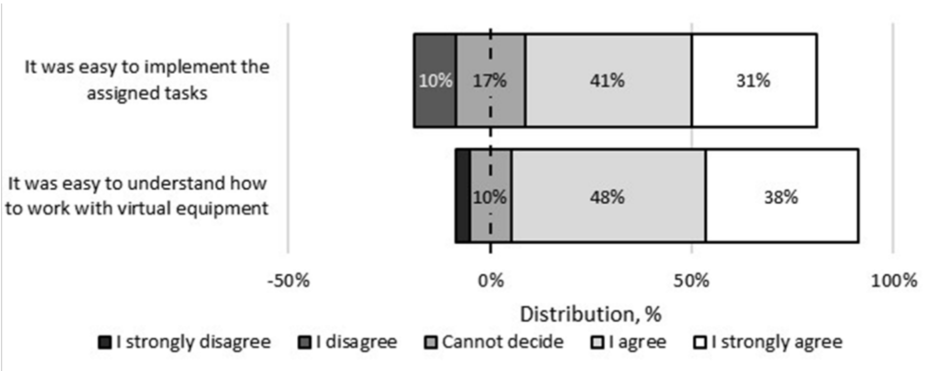


**Figure 1** - Overall distribution of the $u$ and $i$ indicators.

| | y | sp | ve | re | r | t | v | u |
|---|---|---|---|---|---|---|---|---|
| $u$ | 0.0092 | -0.27 | 0.22 | -0.088 | 0.35 | -0.17 | -0.027 | 1.0 |
| $i$ | 0.15 | 0.033 | 0.20 | 0.032 | 0.30 | -0.13 | 0.056 | 0.52 |

**Table 8** – Correlations between the predictors and the indicators.

| Parameter | Estimate | Standard Error | T Statistic | P-Value |
|---|---|---|---|---|
| y^2 | 0.462938 | 0.138778 | 3.33582 | 0.0022 |
| ve*sqrt(u) | 0.257946 | 0.117291 | 2.19919 | 0.0355 |
| v/ve | 0.178005 | 0.0594278 | 2.99531 | 0.0054 |
| u*sqrt(r) | 0.415603 | 0.175673 | 2.36577 | 0.0244 |

**Table 9** – Statistical results for the multiple regression coefficients

model is not appropriate for prediction of the student's implementation rate.

3.3 Classification model

Two classification models were generated using the developed methodology. Even though some of the predictors showed insignificant correlation with both $u$ and $i$, it was decided to use all of them:

- The dependency of the student's understanding rate on all available predictors: $u = f(sp, y, ve, re, r, t, v)$;
- The dependency of the student's implementation rate on all available predictors, including the understanding $u$: $i = f(sp, y, ve, re, r, t, v, u)$.

The SVM classifications was implemented with the Orange Data Mining software. The used parameters for the models are $C = 5.00$, $\varepsilon = 0.10$ and an RBF kernel.

The setup of the classification model and its verification in the Orange software are presented in Figure 2. For the $u$ model, the obtained precision and recall are 0.825 and 0.829, respectively, which means that 58 samples out of 70 were identified correctly (83% success rate). In a similar manner, a second SVM model was trained for $i$. The obtained precision and recall are 0.864 and 0.857, respectively, which means that 60 out of 70 samples were predicted correctly (86% success rate).

3.4 Case study

Out of the two models, the classification approach showed significantly better success rate at predicting student's ease of understanding and ease of implementation rates. Therefore, using the obtained SVM models a case study was implemented. It is assumed that 20 first year students studying "Computer systems and technologies" are taking virtual laboratory classes. Furthermore, it is assumed that a preliminary questionnaire provides information about their experience with virtual and real equipment. The test data is summarized in Table 10.

The goal of the case study is to obtain the optimal learning scenarios for the investigated students. Therefore, four scenarios are investigated:

- **Scenario 1:** Only text documents are used;
- **Scenario 2:** Text documents and video conferencing is used;
- **Scenario 3:** Recorded video instructions and text documents are used;
- **Scenario 4:** All three types of materials are used.

Using the trained SVM models were forecasted the expected results for the above scenarios. The results from the simulations are presented in Table 11.

The obtained results indicate that if such students are to be trained, the optimal scenarios are 3 and 4, i.e. recorded video instructions + text documents or all materials at once are used.

## 4. Discussion and Conclusions

In this study was made an attempt to predict the students' ease of understanding and ease of implementation of virtual labs in electrical engineering classes. It reflects results obtained in two Bulgarian universities during the 2020 spring lockdown due to COVID-19, when the traditional education process was forced into distant form. A questionnaire for the students was prepared and conducted at the end of the semester. A total of 70 students took part in the survey, including Erasmus students from France, which were in the University of Ruse when the lockdown occurred.

During the education process, different forms of delivering the lab instructions were used, including written text instructions (PDFs, etc.), recorded video instructions and video conferencing. The analysis of the results showed that on average, it was easy for the students to understand and implement the instructions – 86% were positive in terms of understanding and 72% in terms of the implementation. The survey results showed that recorded video instructions were the preferred form of content delivery, followed by video conferencing and text materials.

The survey questions were used to define two types of variables – predictors and targets. The selected predictors were the specialty, the years of studying, the previous experience with virtual labs, the previous experience with real labs, and the application of text instructions, recorded video instructions and video conferencing. The ease of understanding was selected as both a predictor and target and the ease of implementation as a target only.

The Pearson's correlation was used to investigate the dependencies between predictors and targets. The results showed that there is a medium correlation between the ease of understanding $u$ and the ease of implementation $i$, which means it is very important that students correctly understand the instructions for performing the virtual lab before they can implement it. This also shows that the ease of understanding is a very important predictor when forecasting the ease of implementation.

Furthermore, low correlations were obtained between $u$ and the students' specialty $sp$, previous experience with virtual labs $ve$, the application of recorded video materials. The ease of implementation $i$ also showed a low correlation with $sp$ and $ve$. This is probably caused by the better virtual experience of Computer systems and technologies students, compared to those from other specialties. Furthermore, the study results suggest that the recorded video instructions could be an important content delivery method when virtual labs are used.

Using the questionnaire data, two types of models were trained. The first one was based on multiple regression and achieved coefficient of determination R-squared above 95% when modelling the ease of implementation $i$. The model proves the need for students to understand the material in order to complete the tasks. The

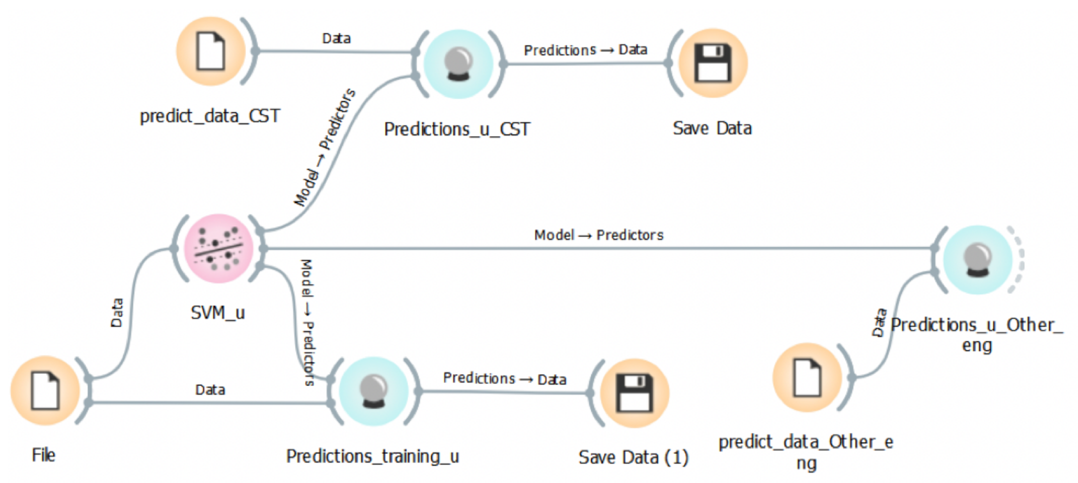


**Figure 2** - Setup of the SVM modelling, verification and prediction in the Orange 3 software.

| y | sp | ve | re |
|---|---|---|---|
| 1 | 2 | 1 | 2 |
| 1 | 2 | 2 | 2 |
| 1 | 2 | 1 | 2 |
| 1 | 2 | 1 | 2 |
| 1 | 2 | 1 | 2 |
| 1 | 2 | 2 | 1 |
| 1 | 2 | 1 | 1 |
| 1 | 2 | 1 | 1 |
| 1 | 2 | 2 | 2 |
| 1 | 2 | 2 | 2 |
| 1 | 2 | 1 | 1 |
| 1 | 2 | 1 | 2 |
| 1 | 2 | 2 | 2 |
| 1 | 2 | 1 | 1 |
| 1 | 2 | 1 | 1 |
| 1 | 2 | 2 | 1 |
| 1 | 2 | 2 | 2 |
| 1 | 2 | 1 | 1 |
| 1 | 2 | 2 | 2 |
| 1 | 2 | 1 | 2 |

**Table 10** - Input data for the case study.

| | 1 | 2 | 3 | 4 | 5 | Average |
|---|---|---|---|---|---|---|
| *Scenario 1* | | | | | | |
| Understanding | 12 | | 8 | | | 1,8 |
| Implementation | | 12 | 8 | | | 2.4 |
| *Scenario 2* | | | | | | |
| Understanding | 6 | | | 12 | 2 | 3.2 |
| Implementation | | 6 | | 14 | | 3.4 |
| *Scenario 3* | | | | | | |
| Understanding | | | | 12 | 8 | 4.4 |
| Implementation | | | | 18 | 2 | 4.1 |
| *Scenario 4* | | | | | | |
| Understanding | | | | 12 | 8 | 4.4 |
| Implementation | | | | 18 | 2 | 4.1 |

**Table 11** - Prediction results for the case study using the SVM model.

specialty's influence is inversely proportional, which can be explained by the fact that most students in electrical specialties have already conducted real laboratory exercises. Nevertheless, after the verification of the model with the training data, only 37% of the training data was categorized correctly. This means that the multiple regression approach is not appropriate in this situation.

The second model was implemented using SVM classification. Similarly, it was verified with the training data and scored 83% an 86% successful categorization rate, respectively for the ease of understanding $u$ and ease of implementation $i$ . Therefore, the SVM classification model has been selected as appropriate for forecasting student's performance with virtual labs.

Finally, a case study was conducted in order to demonstrate the application of the model. It was assumed that 20 first year Computer systems and technologies students should be trained using virtual labs. Furthermore, it is assumed that the students previous experience with virtual and remote labs is obtained using a preliminary questionnaire. For the needs of this case study, they were randomly selected.

The goal of the case study is to obtain the optimal content delivery methods to be used with these particular students in order to achieve optimal learning outcome. Therefore, four teaching scenarios were investigated: only text documents are used; text documents and video conferencing are used; recorded video instructions and text documents are used; all three types are used. The results from the case study showed that in this particular case the optimal approach would be to use scenario 3 and 4. Considering scenario 4 would require more efforts, the optimal scenario would be to use instructions in text form (PDF documents, etc.) and recorded video instructions.

The developed approach and model could be used to optimize the educational experience of students with virtual laboratories, which is the key novelty of this study. As demonstrated in the case study, it could be adopted by starting the course with a short survey in order to obtain the profile and previous experience of students with laboratory exercises. Furthermore, such approach could be used to identify students that might need additional instructions when dealing with virtual labs.

Considering the COVID-19 crisis is far from over, the authors of this article will continue to improve the accuracy of the selected SVM model by adding additional survey results to it. This would allow to increase the representativeness of the sample and hopefully will improve the prediction accuracy of the model. Furthermore, it would allow us to monitor the students' success rate when working with virtual labs.

## References


Adekitan A. I., Salau O. (2019), The impact of engineering students' performance in the first three years on their graduation result using educational data mining, Heliyon, 5(2), e01250.

Al luhaybi M., Tucker A., Yousefi L. (2018), The prediction of student failure using classification methods: A case study, Computer science and information technology (pp. 79–90). Academy & Industry Research Collaboration Center (AIRCC).

Alturki S. Hulpuş I., Stuckenschmidt H. (2020), Predicting Academic Outcomes: A Survey from 2007 Till 2018, Technology, Knowledge and Learning, https://doi.org/10.1007/s10758-020-09476-0.

Alqurashi E. (2019), Predicting student satisfaction and perceived learning within online learning environments, Distance Education, 40(1), 133–148.

Alyahyan E., Düştegör D. (2020), Predicting academic success in higher education: literature review and best practices, International Journal of Educational Technology in Higher Education, 17(3).

Anastasova Y., Anastasov D. (2016), The use of modern information technology in the training of students from University of Mining and Geology "St. Ivan Rilski", Proceedings of the 5th International Scientific and Technical Conference "Technologies and Practices in Underground Mining and Mine Construction", ISSN 1314-7056, 201-204.

Anoopkumar M., Rahman A. M. J. M. Z. (2016), A Review on Data Mining techniques and factors used in Educational Data Mining to predict student amelioration, 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE).

Bhutto S., Siddiqui I. F., Arain Q. A., Anwar M. (2020), Predicting students' academic performance through supervised machine learning, 2020 International Conference on Information Science and Communication Technology (ICISCT).

Buenaño-Fernández D., Gil D., Luján-Mora S. (2019), Application of Machine Learning in Predicting Performance for Computer Engineering Students: A Case Study, Sustainability, 11(10), 2833.

Dimitrov W. (2020), Studying SIEM in higher education is an important task in training cyber security professionals, 14th International Technology, Education and Development Conference, Valencia, Spain. 2-4 March, 2020, INTED2020 Proceedings, 348-357.

Evstatiev B. I., Hristova T. V. (2020), Adaptation of Electrical Engineering Education to the COVID-19

Situation: Method and Results, IEEE 26th International Symposium for Design and Technology in Electronic Packaging (SIITME), Pitesti, Romania, 304 - 308.

Kabakchieva D. (2013), Predicting student performance by using data mining methods for classification, Cybernetics and Information Technologies, 13(1), 61–72.

Kumar J. (2015), A Comprehensive Study of Educational Data Mining, International Journal of Electrical Electronics & Computer Science Engineering Special Issue - TeLMISR, ISSN: 2348-2273.

Kumar M., Singh A.J., Handa D. (2017a), Literature Survey on Student's Performance Prediction in Education using Data Mining Techniques, I.J. Education and Management Engineering, 6, 40-49.

Kumar M., Singh A.J., Handa D. (2017b), Literature Survey on Educational Dropout Prediction, I.J. Education and Management Engineering, 2, 8-19.

Mladenova Ts., Kalmukov Y., Valova I. (2020), Covid 19 – A Major Cause of Digital Transformation in Education or Just an Evaluation Test, TEM Journal, 9(3), 1163-1170.

Osmanbegovic E., Suljic M. (2012), Data Mining Approach for Predicting Student Performance, Economic Review – Journal of Economics and Business, 10(1).

Romero C., Ventura S. (2007), Educational Data Mining: a Survey from 1995 to 2005, Expert Systems with Applications, Elsevier, pp. 135-146.

Saa A. A. (2016), Educational Data Mining & Students' Performance Prediction, International Journal of Advanced Computer Science and Applications, 7(5), 212-220.

Slater S., Joksimović S., Kovanovic V., Baker R. S., Gasevic D. (2017), Tools for educational data mining: A review, Journal of Educational and Behavioral Statistics, 42(1), 85–106.

Taodzera T., Twala B., Carroll J. (2017), Predicting engineering student success using machine learning, Proceedings of the 4th Biennial Conference of the South African Society for Engineering Education, Cape Town.

Tsiakmaki M., Kostopoulos G., Kotsiantis S., Ragos O. (2020), Transfer Learning from Deep Neural Networks for Predicting Student Performance, Applied Science, 10, 2145.

University of Ruse Angel Kanchev. (n.d.), Engine for Virtual Electrical Engineering Equipment (EVEEE), Retrieved from https://eveee.uni-ruse.bg