PERJ Vol. 5 No.2 (2023), 97-106

ISSN 2685 - 6190 (p)
ISSN 2714 - 7746 (e)

Physics Education Research Journal

Department of Physics Education
Faculty of Science and Technology
UIN Walisongo Semarang-Indonesia

Predicting Physics Students' Achievement Using In-Class Assessment Data: A Comparison of Two Machine Learning Models

Purwoko Haryadi Santoso[1,2*], Hayang Sugeng Santosa[3], Edi Istiyono[1,4], Haryanto[1], Heri Retnawati[1]

[1]Graduate School of Educational Research and Evaluation, Universitas Negeri Yogyakarta, Indonesia
[2]Department of Physics Education, Universitas Sulawesi Barat, Indonesia
[3]Graduate School of Educational Research and Evaluation, Universitas Muhammadiyah Prof. Dr. HAMKA, Indonesia
[4]Department of Physics Education, Universitas Negeri Yogyakarta, Indonesia

ARTICLE INFO

Article history:
Submitted : February 27th, 2023
Revised : May 19th, 2023
Accepted : July 11th, 2023

Keywords:
Achievement Prediction; Assessment; Machine Learning; Physics Education; Students

ABSTRACT

Data is the primary source to scaffold physics teaching and learning for teachers and students, mainly reported through in-class assessment. Machine learning (ML) is an axis of artificial intelligence (AI) study that immensely attracts the development of physics education research (PER). ML is built to predict students' learning that can support students' success in an effective physics achievement. In this paper, two ML algorithms, logistic regression and random forest, were trained and compared to predict students' achievement in high school physics (N = 197). Data on students' achievement was harvested from in-class assessments administered by a physics teacher regarding knowledge (cognitive) and psychomotor during the 2020/2021 academic year. Three assessment points of knowledge and psychomotor were employed to predict students' achievement on a dichotomous scale on the final term examination. Combining in-class assessment of knowledge and psychomotor, we could discover the plausible performance of students' achievement prediction using the two algorithms. Knowledge assessment was a determinant in predicting high school physics students' achievement. Findings reported by this paper recommended open room for the implementation of ML for educational practice and its potential contribution to supporting physics teaching and learning.

Introduction

Since the emergence of the integration of information and communication technology in physics learning, educational institutions and academic processes have been sources of complex and large digital data; not only data collected through the class assessment process but learning channeled through the learning management system is one channel for collecting learning data that can be produced at any time and is large (Santoso et al., 2022). The emergence of data of this type and large size can be used to support learning, as attempted by the Educational Data Mining (EDM) and Learning Analytics (LA) research groups (Santoso & Munawanto, 2020). These two groups were created so that we can be more aware of the existence of data in the educational process. EDM and LA were born when data was a potential resource to be addressed in the current era.

Regarding the functions of EDM and LA, artificial intelligence (AI) technology is the forerunner to the birth of these two research groups (Fynn et al., 2022). AI studies have created predictive technology processed through machine learning (ML) for several purposes, including education (Leitner et al., 2017). One of the ML algorithms' analytical tasks is analyzing state predictions beyond model training ((Shafiq et al., 2022; Albreiki et al., 2021). ML algorithms can learn patterns from extensive training data to provide predictions to support learning decisions (Romero & Ventura, 2020). Although the application of ML in

*Correspondence email: purwokoharyadi.2021@student.uny.ac.id
doi: 10.21580/perj.2023.5.2.14127

physics education is still relatively new, several studies predicting physics learning achievement have been carried out since 2019 (Aikenhead, 2023; Lin et al., 2023). The results of the ML study report some success in supporting physics learning, but further studies still need to be carried out.

Most ML models built by physics education researchers are random forests and logistic regression as tested. Previous researchers could predict student learning achievement with an accuracy of up to 80% using these two algorithms. Teachers can use this information as feedback that can be given to students to support student learning (Atmam & Mufit, 2023; Ndoa & Anastasia, 2022). According to the constructivist approach, effective physics learning provides students with independent learning, which is supported by the teacher's activeness in monitoring learning, which the results of ML predictions can generate.

The research presented in this article aims to explore assessment data collected by teachers to predict student physics learning achievement. The data is processed to train an ML model based on logistic regression and random forest algorithms, as has been done by many physics education researchers above. The prediction performance produced in the two ML models was then evaluated and compared to the most optimal one.
To guide the investigation, three research questions were posed as follows.
Problem 1. How is the prediction performance of physics learning outcomes produced by the logistic regression model?
Problem 2. How is the prediction performance of physics learning outcomes produced by the random forest model?
Problem 3. What does data contribute most to predicting physics learning outcomes?
One of the implications of this article aims to open a discussion space in physics education to support physics learning through ML predictive technology.

Machine Learning

Machine learning (ML) artificial intelligence (AI) is a branch that focuses on utilizing data and algorithms to imitate how a knowledge machine constantly improves its capabilities. ML differs from statistical approaches that rely heavily on probability theory for hypothesis testing. The terms often used in ML studies are training and testing that do not aim to generalize to a population. However, ML is trained to make predictions from input data not used in model training (Aikenhead, 2023).

Two divisions of ML learning types are most widely cited in the literature, namely unsupervised and supervised (Susilawati et al., 2021)—the difference between the two lies in the labels or targets available in the training dataset. Unsupervised algorithms are more exploratory because interpretation of the results still requires the role of the model user. An example of an unsupervised ML algorithm is clustering. In contrast, supervised ML methods have predetermined targets before model training. ML is trained to predict one of the labels contained in the target variable. Several tasks are included in the supervised learning type (Bloor & Santini, 2023). One of them that is often applied in education is classification. The topic of classification, namely the prediction of student learning, is the basis for the birth of the field of ML studies in the field of education, namely educational data mining (EDM) and learning analytics (LA). Even though they have slightly different terminology, there is a common goal of supporting student learning.

Several ML algorithms are often used to predict student learning achievement. A systematic review of the history of algorithms that have been used has been synthesized, such as decision trees, random forests, k-nearest neighbors, support vector machines, naive Bayes, logistic regression, and artificial neural networks (Sekeroglu et al., 2021). Due to the limitations of the study focus addressed in this article, only two ML algorithms as the basis for developing ML models were used in predicting learning achievement reported in this article.

Logistic Regression

Although its name is mentioned in the term 'regression,' which is another task in the supervised learning type; logistic regression is mainly used for tasks. This algorithm has the same properties as linear regression. The difference lies in using the dependent variable in categorical form, where the sigmoid function is used to classify the class of the dependent variable.

Random Forest

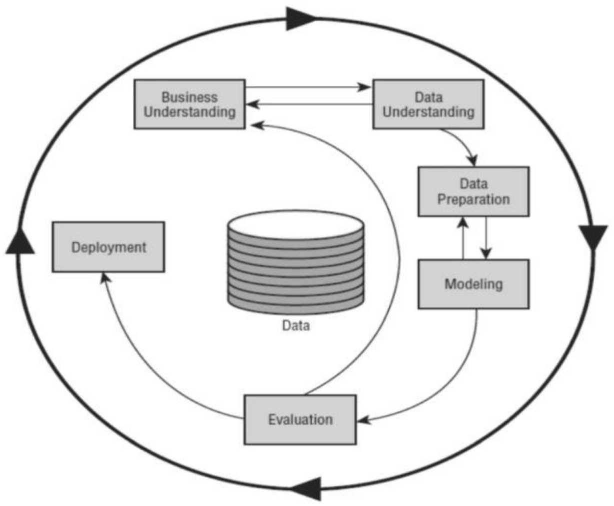
This model is one of a family of decision tree algorithms for carrying out classification tasks on categorical data. Random forest is a model from a group of tree-based algorithms that grows several trees (decision trees) to determine classification decisions for a class of targets as an ensemble. This method is also known as optimization of decision trees by considering the regression average of trees that grow as an ensemble.

Methods

Research Design

This research explores training two types of ML models using the framework suggested by The Cross Interprofessional Standard Data Mining Process (CRISP-DM) in data science research (Chapman, 2000). Six stages structure this framework, illustrated in Figure 1.

Figure 1
Work Cycle of Data Science Research According To CRISP-DM



1. Business understanding.
   The purpose of applying ML in this research is to perform classification tasks. Classification tasks are one of the most researched topics by EDM and LA groups (Sekeroglu et al., 2021). Prediction of students' physics learning achievement is an example of a classification task in ML research.
2. Data understanding.
   Two types of data obtained through classroom assessments are used to predict physics learning outcomes—namely, teacher assessment data on aspects of knowledge and skills by the Indonesian physics curriculum syllabus. The class assessment data in this research comes from recorded academic data obtained by a teacher from one of the state schools in Indonesia in the 2020/2021 academic year. Student assessments are measured using assessment instruments developed and validated by teachers based on the class X high school physics curriculum reference by the current COVID-19 pandemic conditions. There are three essential physics competencies that teachers focus on during this period, including the nature of physics, measurement, and vectors.
3. Data preparation.
   Based on data from physics teachers, 197 students are enrolled in physics learning. Students are divided into five classes X from the science department. However, there are cases where students do not take part in the assessments carried out by the teacher even though the teacher has carried out additional assessment sessions for the students. However, there are still students who cannot join for several reasons. The teacher agreed that a grade of zero should be given in that case. Data that is ready to be saved in CSV format.
4. Modeling.
   The next stage is training the ML model through logistic regression and random forest algorithms. The proportion of sharing 75% of training data and 25% of test data was chosen in this research. Modeling was carried out in R using several packages such as 'caret' (Hochberg et al., 2018), 'randomForest,' 'lme4' (Bates et al., 2015), 'caTools', 'pROC' (Starr et al., 2020), 'varSelRF', 'pscl', 'MASS', as well as 'ggplot2', 'ggthemes', 'RColorBrewer', 'Rmisc', and 'car' for data visualization purposes.
5. Evaluation.
   The ML model that has been trained is then used to make predictions on the test data at this stage. The prediction performance of the ML model on the test data is then tabulated using the confusion matrix shown in Table 1. Several evaluation metrics are used in evaluating ML classification models based on the number calculated in the confusion matrix. Several evaluation metrics that are often used in ML studies predicting learning outcomes include accuracy, precision, sensitivity, specificity, F1-measure, and area under the curve (AUC) (Darling-Hammond et al., 2020), which are calculated based on the formula described in Equation 1 to Equation 6.

Table 1
Confusion Matrix

	Actual+	Actual -
Prediction +	True positive (TP)	False positive (FP)
Prediction -	False negative (FN)	True negative (TN)

$$Accuracy = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \tag{1}$$

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \tag{2}$$

$$Sensitivity = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \tag{3}$$

$$\mathit{Specificity} = \frac{\mathrm{TN}}{\mathrm{TN+FP}} \tag{4}$$

$$\mathit{F1\text{-}measure} = 2 \cdot \frac{presisi \cdot recall}{presisi + recall} \tag{5}$$

Then, the 'stepAIC' function measures how important a variable (knowledge or skills) is in the logistic regression model and 'varImpPlot' for the random forest model.

6. Deployment.
   In this research, this stage was not carried out because the study development process was still being carried out.

Result and Discussions

This research aims to train a prediction model for physics learning achievement with two algorithms: logistic regression (Problem 1) and random forest (Problem 2). A comparison of prediction performance between the two ML methods based on six evaluation metrics (accuracy, precision, sensitivity, specificity, F1-measure, and AUC) is reported in Table 2.

Six models have been drilled using logistic regression and random forest algorithms. The first step is establishing a baseline where the target variable (student learning achievement) is predicted as "1" or more from the minimum completeness during the final test. This model can be used as a comparison of the fifth model that was explored later. This baseline model detects the extent of model accuracy in the case of pure guessing. Both logistic regression and random forest show the same results in the first model.

Then, another fifth model was created to follow the development of student learning in each basic physics competency. The last two models were created to measure the contribution of knowledge and skills aspects in predicting physics learning achievement with two ML algorithms.

First, model accuracy increases as basic competency (BC) variables increase, reaching the highest peak at basic competency 3. However, model performance based on just one measure of accuracy can cause several problems (Zabriskie et al., 2019). Table 1 above is the confusion matrix produced when testing the ML model using test data. This contingency table will then calculate the six model evaluation metrics reported in Table 2.

Second, the accuracy aspect must also be complemented by a measure of precision, namely the extent to which predictions have been successfully made from the total TP and FP predictions in the confusion matrix, as explained in equation (2) above. In contrast to the accuracy defined in measurement principles in physics, the precision described in equation (2) is explicitly used to see the proportion of accurate positive predictions (TP) to false optimistic predictions (FP). The larger the FP results, the smaller the precision. The baseline model did not find precision values because TP and FP were not found in this group. The model precision value in Table 2 increases as the variables involved in the model increase. This result indicates good results because it corresponds to the most significant increase in accuracy values at BC 3.

Table 2
Performance of Predicting Physics Learning Achievement Using Logistic Regression and Random Forest

Model	Logistic Regression						*Random Forest*					
	Acc	Prec	Sens	Spec	F1	AUC	Acc	Prec	Sens	Spec	F1	AUC
Baseline	0.714	-	0.000	1.000	-	0.500	0.714	-	0.000	1.000	-	0.500
BC 1	0.735	1.000	0.071	1.000	0.133	0.727	0.714	0.500	0.071	0.971	0.125	0.498
BC 2	0.796	0.833	0.357	0.971	0.500	0.862	0.816	0.778	0.500	0.943	0.609	0.872
BC 3	0.857	0.769	0.714	0.914	0.741	0.951	0.878	0.786	0.786	0.914	0.786	0.945
K only	0.857	0.769	0.714	0.914	0.741	0.951	0.878	0.786	0.786	0.914	0.786	0.947
S only	0.878	0.900	0.643	0.971	0.750	0.876	0.816	0.692	0.643	0.886	0.667	0.909

Note: Acc : *accuracy*, Prec : *precision*, Sens : *sensitivity*, Spec : *specificity*, F1 : *F1-measure*, AUC : *area under curve*, BC : basic competence, K : Knowledge, S : Skills

The TP and TN cases are what we expect in studies predicting student learning achievement. An accurate prediction means there is a match between what is predicted and what happens. The good or bad performance predicted by the teacher during the learning process will correlate with student learning achievement at the end of learning (Chen et al., 2018; Le et al., 2022; Lu et al., 2021). However, cases of FP and FN may occur, and these two results are not expected to predict learning outcomes. Both cases can reduce students' learning motivation or lead to excessive self-confidence (Rubie-Davies, 2006; Shengnan et al., 2018). Therefore, considering only the TP and TN numbers as evaluated through the above accuracy and precision metrics can be problematic in their interpretation. We need to review other numbers that can understand the extent to which the model produces FP and FN results in predicting student physics learning achievement, namely through sensitivity and specificity measures.

Motivation learning can be inaccurate, and prediction reports influence them. For example, the FN results in this case cause students who are high performers (+) to be predicted as students with poor learning achievement (-). Then, excessive self-confidence can be caused by FP prediction reports, which report that students who perform poorly (-) are instead predicted to have exemplary learning achievements (+) as has been found (Jeong et al., 2021; Vasalou et al., 2021), interpreting the metrics involving the number of FN and FP above must be adjusted to the context of predicting learning outcomes aimed in this research.

Nevertheless, researchers argue that if FP creates higher student self-confidence, it should have a better impact on the physics learning process than the negative effect caused by the FN case. This argument can be supported by the self-regulated learner (SRL) theory proposed by (Kind, 2013; Zimmerman et al., 2014). Positive feedback teachers give can increase students' self-confidence, so it impacts students' learning independence according to the SRL theoretical framework. Therefore, the case of predicting physics learning achievement should place more emphasis on prediction performance, which can minimize cases of FN, which can reduce students' confidence in learning physics. Sensitivity or recall measures are more suitable to pay attention to in cases where we want to minimize FN cases in our prediction results. The sensitivity value will be more significant when the FN value is smaller, or what we expect in Equation 3. Thus, reviewing a more considerable sensitivity value can be an option in evaluating the model with the most optimal prediction performance to predict student learning achievement.

Third, BC 3 in Table 2 shows satisfactory prediction performance based on sensitivity. Using the skill aspect can have a sensitivity of 90% in the logistic regression algorithm. These results exceed those found by Zabriskie et al. (2019), whose model only achieved an accuracy level of 80% using only five variable assessment points.

Fourth, specificity measures should be expected to be the opposite of those found by sensitivity. The focus on predicting learning outcomes is more aimed at minimizing expectation errors, which reduces student learning motivation in the case of FN. Based on Table 2 above, we can see that BC 3 is a model with good performance relevant to the three previous measures: accuracy, precision, and sensitivity.

Fifth, we will use the proposed F1 measure to improve the ML classification performance measures described by the four previously used measures. F1 measure can overcome uneven class distribution in the training dataset (Luo et al., 2021; Yang et al., 2017). For example, the predicted learning outcomes of students categorized as above the minimum completeness value or "1" are much higher than those classified as "0". Based on this measure, the BC3 model is still superior to the other two. The fifth measure offered by the F1 measure further strengthens the argument provided by the four previous evaluation metrics that the more formative assessment variables are implemented, the more the prediction performance provided by ML will improve.

Lastly is the area under curve (AUC), calculated via receiver operating characteristic (ROC). As a general rule, the higher the AUC, the more the ML prediction model can differentiate the class of each target. This means that the prediction model is increasingly accurate in predicting “1” as “1” or “0” as “0”. A good ML model has an AUC close to 1. Based on the results shown in Table 2, the BC 3 (vector material) model is the best model according to AUC.

Based on the six measures reported by the evaluation metrics above, it can be concluded that the ML prediction model, which involves all formative assessments up to BC 3, is the most optimal in predicting student physics learning outcomes for both the logistic regression and random forest algorithms. Researchers found no significant difference in the prediction performance results reported by logistic regression and random forest.

These results are from previous research they are ducted by (Semenikhina et al., 2020). The five assessment points found that they could produce prediction models with up to 80% accuracy with just three formative assessment data. These results have implications for learning purposes in that the predictive information provided by ML can become input or feedback that students can immediately use to improve learning. Simultaneously, teachers can also involve these results as a reflection of the learning they carry out in improving the effectiveness of physics learning. Through ML support, physics classes can be accepted by students more effectively. However, experimental studies must be conducted to test this conjecture.

Apart from looking at the performance of ML prediction results between logistic regression and random forest, this research also investigates what factors contribute most to the optimal prediction results. The last two models were deliberately created to look at each aspect of the assessment carried out by teachers, namely knowledge and skills. Based on the six evaluation metrics, the performance of the logistic regression and random forest models did not show significant differences when predicting student learning outcomes using aspects of knowledge or knowledge alone. Except that AUC reported a higher difference in logistic regression performance for the knowledge aspect than for the skill aspect. Even by using the features indicated by the knowledge aspect alone, the ML model can show performance in predicting learning outcomes with performance that is close to the AUC value shown by the BC 3 model. This initial information illustrates that the knowledge aspect still significantly contributes to the success of studying high school physics found in this research.

This finding may be challenged by several opinions of previous researchers who suggested the importance of aspects of science process skills in physics learning, including what was found (Lin et al., 2023; Yang et al., 2017), which still uses non-cognitive variables in its physics learning outcomes prediction model. However, this problem is outside the research question, which is the aim of this research. Critical studies with more comprehensive data are highly recommended to open up space for scientific discussion in the physics learning assessment field.

The 'stepAIC' function is a method that is often used to select features in classification models. By what has been described in the method above, this function has been used to investigate the contribution of each feature involved in the model. In this presentation, 'stepAIC' is applied to the BC 3 model because we have found that this model shows the most optimal performance in predicting physics learning achievement according to the previous discussion. However, it should be emphasized that 'stepAIC' is not intended to improve model fit to data or more optimal performance of ML prediction results. This function is used to simplify the model without significantly affecting the prediction performance of the ML model. Therefore, AIC only measures the amount of information loss when one of the features is omitted from the model. AIC is an abbreviation of Akaike Information Criteria.

The final stage of the 'stepAIC' calculation concluded that three variables could be simplified for the KD3 model with a lower AIC value of 76.56 from the previous one of 82.17 by involving all class assessment data including assessment of knowledge and skills aspects. These three variables are displayed in Table 3 below, which displays the coefficients from the logistic regression equation, which has been simplified using the 'stepAIC' method. The explanation of the variable code 'P.PH3.1' is that the leftmost letter 'P' represents the knowledge aspect, then 'PH' means the method used, namely daily assessment, and '3.1' is the basic physics competency code that is being assessed by the teacher, namely the vector by Indonesian physics learning class X curriculum. We can see that all aspects of knowledge are reported from the three variables summarized by 'stepAIC'. These results can confirm and strengthen what was reported by the previous AUC measure that the knowledge aspect is a determining factor in predicting physics learning achievement in this study. In addition, basic vector competence is the most critical aspect in this case.

Table 3

Estimation Results of KD3 Logistic Regression Model Coefficient Parameters in Final 'stepAIC'

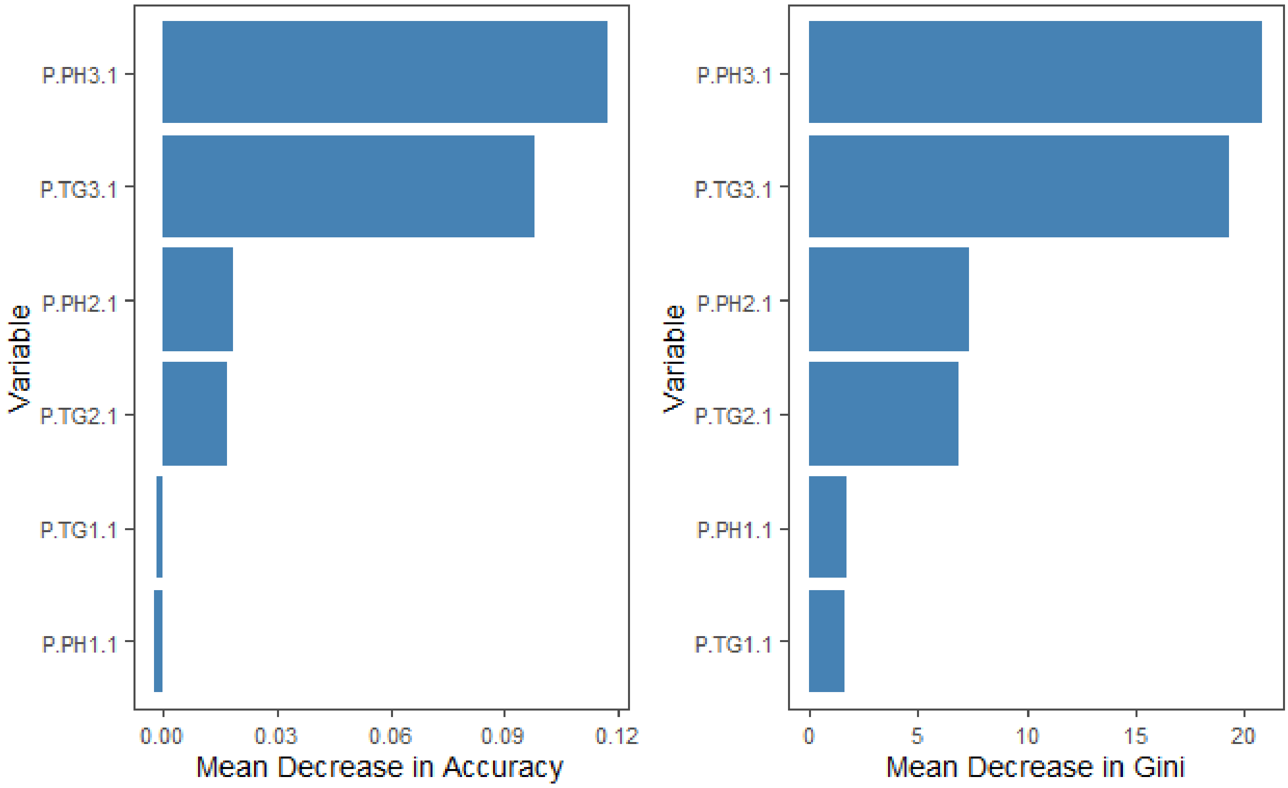
Intercept	P.PH3.1	P.TG3.1	P.TG2.1
-11.1255	0.07216	0.07039	0.02768

As a reinforcement of the AUC and stepAIC findings above, the 'varImpPlot' function is used in Figure 2 above to extract the most essential variables in the model by visualizing the decrease in the mean of the accuracy metric and Gini index. Both relative report results that are not that different from the visualization of a horizontal bar chart arranged from the most critical variables to the least important. Based on previous findings, three knowledge variables were accurately reported as having the most significant contribution to model 3. Both AUC, stepAIC, and feature importance from random forest concluded that these three aspects were determining factors in this study's predicted results of physics learning achievement.

The research results reported in this study have built two ML models to predict students' physics learning achievement with two algorithms that achieved entirely satisfactory performance. The next stage of this research is further testing the ML model through training data with different contexts or setting several hyperparameters of the logistic regression and random forest models, especially for data up to KD3, which is proven to perform best according to the above findings. Apart from that, deployment has not been carried out in this research, which should be an inseparable part of a data science project, according to CRISP-DM. The opportunity to continue the studies designed in this research is still open to carry out these stages or test other ML models.

Figure 2

Feature Importance of the KD3 Model with the Random Forest Algorithm



Conclusions

The development of technology and information-based learning is a channel that can produce large and complex data. One data that physics teachers often measure is the class assessment reported to students at the end of learning in the knowledge and skills aspects of the Indonesian physics learning curriculum. In this research, two supervised ML models, namely logistic regression and random forest, have been trained using academic data from physics lessons conducted by a teacher. Six ML models were created by considering the development of student learning in the three essential physics competencies and investigating the contribution of each assessment data used to predict the most optimal physics learning achievement. Logistic regression and random forest did not show significant differences in performance to predict learning achievement in physics classes. The three assessment points of the basic physics competency of class X high school can be predicted optimally if there is an assessment of the final physics competency, namely vector. The knowledge assessment aspect is proven to substantially influence the performance of predicting physics learning outcomes using logistic regression and random forest. The results of this research can recommend one method teachers can use to design physics learning with student feedback.

References

Aikenhead, G. S. (2023). Humanistic school science: Research, policy, politics and classrooms. *Science Education*, *107*(2), 237–260. https://doi.org/https://doi.org/10.1002/sce.21774

Albreiki, B., Zaki, N., & Alashwal, H. (2021). A Systematic Literature Review of Student' Performance Prediction Using Machine Learning Techniques. *Education Sciences*, *11*(9). https://doi.org/10.3390/educsci11090552

Atmam, P. L., & Mufit, F. (2023). *Using Adobe Animated CC in Designing Interactive Multimedia Based on Cognitive Conflict on Parabolic Motion Materials*. *8*(1), 64–74.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bloor, T., & Santini, J. (2023). Modeling the Epistemic Value of Classroom Practice in the Investigation of Effective Learning. *Science & Education*, *32*(1), 169–197. https://doi.org/10.1007/s11191-021-00298-9

Chapman, P. (2000). *CRISP-DM 1.0: Step-by-step data mining guide*. https://api.semanticscholar.org/CorpusID:59777418

Chen, J., Wang, M., Grotzer, T., & Dede, C. (2018). Using a three-dimensional thinking graph to support inquiry learning. *Journal of Research in Science Teaching*, *55*. https://doi.org/10.1002/tea.21450

Darling-Hammond, L., Flook, L., Cook-Harvey, C., Barron, B., & Osher, D. (2020). Implications for educational practice of the science of learning and development. *Applied Developmental Science*, *24*(2), 97–140. https://doi.org/10.1080/10888691.2018.1537791

Domínguez Romero, E., & Bobkina, J. (2021). Exploring critical and visual literacy needs in digital learning environments: The use of memes in the EFL/ESL university classroom. *Thinking Skills and Creativity*, *40*, 100783. https://doi.org/https://doi.org/10.1016/j.tsc.2020.100783

Fynn, A., Adamiak, J., & Young, K. (2022). A Global South Perspective on Learning Analytics in an Open Distance E-learning (ODeL) Institution. In P. Prinsloo, S. Slade, & M. Khalil (Eds.), *Learning Analytics in Open and Distributed Learning: Potential and Challenges* (pp. 31–45). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-0786-9_3

Hochberg, K., Kuhn, J., & Müller, A. (2018). Using Smartphones as Experimental Tools – Effects on Interest, Curiosity, and Learning in Physics Education. *Journal of Science Education and Technology*, *27*. https://doi.org/10.1007/s10956-018-9731-7

Jeong, J. S., González-Gómez, D., & Cañada-Cañada, F. (2021). How does a flipped classroom course affect the affective domain toward science course? *Interactive Learning Environments*, *29*(5), 707–719. https://doi.org/10.1080/10494820.2019.1636079

Kind, P. M. (2013). Establishing Assessment Scales Using a Novel Disciplinary Rationale for Scientific Reasoning. *Journal of Research in Science Teaching*, *50*(5), 530–560.

Le, B., Lawrie, G. A., & Wang, J. T. H. (2022). Student Self-perception on Digital Literacy in STEM Blended Learning Environments. *Journal of Science Education and Technology*, *31*(3), 303–321. https://doi.org/10.1007/s10956-022-09956-1

Leitner, P., Khalil, M., & Ebner, M. (2017). Learning Analytics in Higher Education---A Literature Review. In A. Peña-Ayala (Ed.), *Learning Analytics: Fundaments, Applications, and Trends: A View of the Current State of the Art to Enhance e-Learning* (pp. 1–23). Springer International Publishing. https://doi.org/10.1007/978-3-319-52977-6_1

Lin, R., Yang, J., Jiang, F., & Li, J. (2023). Does teacher's data literacy and digital teaching competence influence empowering students in the classroom? Evidence from China. *Education and Information Technologies*, *28*(3), 2845–2867. https://doi.org/10.1007/s10639-022-11274-3

Lu, K., Yang, H. H., Shi, Y., & Wang, X. (2021). Examining the key influencing factors on college students' higher-order thinking skills in the smart classroom environment. *International Journal of Educational Technology in Higher Education*, *18*, 1. https://doi.org/10.1186/s41239-020-00238-7

Luo, M., Sun, D., Zhu, L., & Yang, Y. (2021). Evaluating scientific reasoning ability: Student performance and the interaction effects between grade level, gender, and academic achievement level. *Thinking Skills and Creativity*, *41*, 100899. https://doi.org/https://doi.org/10.1016/j.tsc.2021.100899

Narvaez Rojas, C., Alomia Peñafiel, G. A., Loaiza Buitrago, D. F., & Tavera Romero, C. A. (2021). Society 5.0: A Japanese Concept for a Superintelligent Society. *Sustainability*, *13*(12). https://doi.org/10.3390/su13126567

Ndoa, Y. A. A., & Anastasia, D. P. (2022). *Development of An Android-Based Physics E-Book with A Scientific Approach to Improve The Learning Outcomes of Class X High School Students on Impulse and Momentum Materials*. *18*(December), 107–121. https://doi.org/10.15294/jpfi.v18i2.30824

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery*, *10*(3), e1355. https://doi.org/https://doi.org/10.1002/widm.1355

Rubie-Davies, C. M. (2006). Teacher expectations and student self-perceptions: Exploring relationships. *Psychology in the Schools*, *43*(5), 537–552. https://doi.org/https://doi.org/10.1002/pits.20169

Santoso, P. H., Istiyono, E., & Haryanto. (2022).

Physics Teachers’ Perceptions about Their Judgments within Differentiated Learning Environments: A Case for the Implementation of Technology. *Education Sciences*, *12*(9). https://doi.org/10.3390/educsci12090582

Santoso, P. H., & Munawanto, N. (2020). Approaching Electrical Circuit Understanding with Circuit Builder Virtual Laboratory. *Jurnal Ilmiah Pendidikan Fisika Al-Biruni*, *9*(2), 259–269. https://doi.org/10.24042/jipfalbiruni.v9i2.5976

Sari, D., Herlina, K., Viyanti, V., Andra, D., & Safitri, I. (2023). E-module Newton's Law of Gravity based Guided Inquiry to Train Critical Thinking Skills. *Physics Education Research Journal, 5*(1), 13-20. doi:https://doi.org/10.21580/perj.2023.5.1.11657

Sekeroglu, B., Abiyev, R., Ilhan, A., Arslan, M., & Idoko, J. B. (2021). Systematic Literature Review on Machine Learning and Student Performance Prediction: Critical Gaps and Possible Remedies. *Applied Sciences*, *11*(22). https://doi.org/10.3390/app112210907

Semenikhina, O., Yurchenko, A., & Udovychenko, O. (2020). Formation of skills to visualize of future physics teacher: results of the pedagogical experiment. *Physical and Mathematical Education*, *23*(1), 122–128. https://doi.org/10.31110/2413-1571-2020-023-1-020

Shafiq, D. A., Marjani, M., Habeeb, R. A. A., & Asirvatham, D. (2022). Student Retention Using Educational Data Mining and Predictive Analytics: A Systematic Literature Review. *IEEE Access*, *10*, 72480–72503. https://doi.org/10.1109/ACCESS.2022.3188767

Shengnan Wang, C. M. R.-D., & Meissel, K. (2018). A systematic review of the teacher expectation literature over the past 30 years. *Educational Research and Evaluation*, *24*(3–5), 124–179. https://doi.org/10.1080/13803611.2018.1548798

Starr, C., Hunter, L., Dunkin, R., Honig, S., Palomino, R., & Leaper, C. (2020). Engaging in science practices in classrooms predicts increases in undergraduates' STEM motivation, identity, and achievement: A short-term longitudinal study. *Journal of Research in Science Teaching*, *57*. https://doi.org/10.1002/tea.21623

Susilawati, S., Azizah, N. A. N., & Kusuma, H. H. (2021). Investigating differences in project activities and student digital literacy between learning through electronic workbench and PhET Simulation. *Jurnal Ilmiah Pendidikan Fisika Al-Biruni*, *10*(2), 299–311. https://doi.org/10.24042/jipfalbiruni.v10i2.10008

Vasalou, A., Benton, L., Ibrahim, S., Sumner, E., Joye, N., & Herbert, E. (2021). Do children with reading difficulties benefit from instructional game supports? Exploring children's attention and understanding of feedback. *British Journal of Educational Technology*, *52*(6), 2359–2373. https://doi.org/https://doi.org/10.1111/bjet.13145

Yang, C., Lan, S., Shen, W., Huang, G. Q., Wang, X., & Lin, T. (2017). Towards product customization and personalization in IoT-enabled cloud manufacturing. *Cluster Computing*, *20*. https://doi.org/10.1007/s10586-017-0767-x

Zimmerman, L., Spillane, S., Reiff, P., & Sumners, C. (2014). Comparison of Student Learning about Space in Immersive and Computer Environments. *Journal Review Astronomical Education and Outreach*, *1*(1), A5–A20.