

**Department of Information, Mechanical
and Electrical Engineering**
Shanghai Normal University
swan@shnu.edu.cn

Oct 4th, 2023

Dear Editors

We thank the reviewers for their all generous comments on the manuscript and have have edited the manuscript to address their concerns.

We believe that the manuscript is now suitable for publication in PeerJ.

A handwritten signature in black ink, appearing to be 'J. Wang' or similar, written in a cursive style.

Dr. Gong Wang
Associate Professor

On behalf of all authors.

Reviewer 1

Basic reporting

The article seems to be written in professional English.

The article references various studies, such as works by "Smith et al.", "Asif et al.", and "Latif et al.". This indicates a foundation in existing literature and provides context.

The article has sections like "Related Work", "Dataset Collection and Preprocessing", and "Methodology". It also references tables and figures. The raw data collection process is described, especially in the context of online student behavior during the pandemic.

The article mentions various methodologies and results.

The article provides definitions for various terms, especially in the context of the dataset.

Experimental design

The study aims to evaluate the influence of students' online learning data on their learning performance, especially during the COVID-19 pandemic.

The study proposes three research questions:

RQ 1: Can the chosen model effectively predict using online learning data?

RQ 2: Which features have the most predictive power, and how do they differ from those in other studies?

RQ 3: Is the model behavior perceived as reasonable by course teachers in practice?

The study mentions obtaining written informed consent from all participants before enrollment. The research seems to be conducted rigorously, using various machine learning algorithms and feature selection techniques.

The study details the machine learning techniques used, including Naive Bayes, LibSVM, Multi-Layer Perception (MLP), and others. It also describes feature selection techniques, such as filter-based feature ranking and subset selection. The use of the WEKA tool for applying EDM techniques is also mentioned.

Validity of the findings

The study focuses on student learning behaviors during the COVID-19 pandemic, evaluating the performance of various techniques. The research questions posed aim to understand the effectiveness of the chosen model, the significance of features affecting academic performances, and the model's perceived reasonability by course teachers.

The study collects datasets from online courses during the COVID-19 pandemic and evaluates the performance of each technique. The dataset construction and preprocessing are detailed, and various machine learning algorithms and feature selection techniques are employed.

The paper concludes by summarizing its main contributions, such as conducting a study on student learning behaviors during the pandemic, collecting datasets, and evaluating the performance of each technique. The paper also highlights the change in significant features affecting academic performances during the pandemic.

Additional comments

Strengths:

- 1. The study is timely, focusing on student learning behaviors during the COVID-19 pandemic, a topic of significant interest given the shift to online learning.*
- 2. The article employs various machine learning algorithms and feature selection techniques, showcasing a thorough approach.*
- 3. The study mentions obtaining written informed consent, indicating ethical considerations in the research process.*
- 4. The dataset construction and preprocessing are detailed, providing transparency in the data gathering process.*
- 5. The study poses clear and relevant research questions, guiding the research's direction.*

Weaknesses:

- 1. While the study mentions various methodologies and results, a more in-depth analysis of each method's performance might enhance the article's value.***

The characteristics of different algorithms lead to different results. We have supplemented and analyzed the performance of these algorithms in *Discussion*. ([Page 10-11 Line 424-449 in revise_tracked.pdf](#))

- 2. A comparison with pre-pandemic student behaviors might provide more insights into the changes brought about by the pandemic.***

Before the outbreak of the COVID-19 pandemic, because the teaching mode was traditional classroom-teaching and most of the online courses had not been built and covered, students did not form the habit of online learning behavior on the online platform, which lead to the online data was so sparse that it was impossible to conduct students' performance prediction and other work. We have compared the important features in RQ2 to those of other datasets in the literature listed in Appendix.

Reviewer: Maheswari Raja

Basic reporting

The researchers/ author have made significant steps towards the future education system, which is much appreciated. But the following points were suggested and listed to improve the standard of the article.

1. *Require improvement in the standard of writing and presenting the article.*

We have checked up and corrected.

2. *A few more recent articles (2023,2022) related to the objective of the work shall be reviewed and included. To validate the requirement of this work in 2023 as the entire world has rejoined to normal life from new-normal.*

We have added some articles(2023,2022) related in Related Work to validate the requirement of this work in 2023. It also can provide suggestion for next pandemic. (Page 4 Line 151-155, Line 158-162 in [revise_tracked.pdf](#))

3. *The dataset taken 228 for Course A and 129 for Course B looks very small where the machine learning model itself might not required.*

We complemented the experiments and compared the results on datasets extended. Big datasets can better improve algorithm performance and validation effectiveness. In machine learning and artificial intelligence, small datasets are important as well as big datasets. For problems with special conditions, such as the early stage of the pandemic or an outbreak, studies on small datasets can sometimes give some information. (Page 10-11 Line 424-449 in [revise_tracked.pdf](#))

4. *As the data set is small, the model prediction rate looks high. Need to prove the maintenance of same score even with higher data set.*

We have constructed the datasets extended by including more student online learning data of Course A and Course B during the pandemic. The performance of the same algorithms on the datasets extended remains almost consistent. (Page 8-9 Line 328-341 with Table 8 and Table9 in [revise_tracked.pdf](#))

5. *Insufficient results with very less courses begin considered (just two courses - Course A and Course B).*

Courses in higher education is various, and there may be different degrees of variability between different types of courses (e.g., programming courses versus English courses), so we chose the same type of English courses for our study. In addition, the English courses used as object and datasets are relatively few in related studies.

Experimental design

The knowledge gap under investigation should be improved and highlighted.

We have added the knowledge gap in the first paragraph of this section. The new datasets mentioned above are different from other datasets in terms of the category of features included, i.e., it has a greater emphasis on being relevant to student learning and being adjustable. (Page 7 Line 270-271 in [revise_tracked.pdf](#))

The author needs to validate how the study helps to bridge the gap in the physical education system with respect to online education.

In terms of the task completion rate, teachers in online teaching check students' completion rate and quality, instead of collecting students' homework regularly to understand the degree of students' knowledge assessment. For stage ability assessment, the difference between online teaching and offline teaching actually lies in the different media used (paper/electronic), but sometimes online teaching is more inclusive of the form in which students submit their works. For the teaching atmosphere, online teaching can further provide space for students to freely combine and play their abilities, so as to better teach students in accordance with their aptitude. The important features in our study can provide reference for physical education, or become one of the teaching tools. We also have added the discussion to validate in *Discussion*. (Page 12 Line 526-533 in [revise_tracked.pdf](#))

The author needs to investigate the goals in an exhaustive way with a higher technical caliber. The research must have been carried out in accordance with the current requirements in a normal environment without Pandemic situation.

The datasets were newly constructed, so we first effectively validated these algorithms on the datasets. Moreover, these machine learning algorithms are excellent, and MOEFC has not appeared in previous studies. We also have added the analysis about the performance in the *Discussion*.

Unfortunately, the research couldn't be carried out in a normal environment now, because it takes a lot of time to re-collect student informed consent and student information on a large scale. We will take this in the future research plan.

Validity of the findings

Much more system validation and verification with respect to increased courses, and data set review comments to evidence the performance, efficiency, and accuracy has to be included.

- We have constructed the extended datasets. (Page 5 Line 192-200 in [revise_tracked.pdf](#))
- We have used the same algorithms to implement prediction on the extended datasets and almost all algorithms maintain or even improve predictive performance, evidencing the validity. (Page 8-9 Line 328-341 with Table 8 and Table 9 in [revise_tracked.pdf](#))
- We have further discussion about the experiment result and supplement the Finding in *Discussion*. (Page 10-11 Line 424-449 in [revise_tracked.pdf](#))

The findings are stated clearly and they are presented in a good manner representing the summary of the article.

Additional comments

It is really appreciated that the authors/researchers have taken important measures towards the betterment of online education in the future. However, the mentioned suggestions may raise the article's quality.