

## Correction Note

**Manuscript ID:** IJiet-6223

**Title:** Improving SVM Classification Performance on Unbalanced Student Graduation Time Data Using SMOTE

Dear Reviewers

We have made improvements to our manuscript following the instructions and suggestions of the reviewers. About which parts we have revised and added writing, it is as explained below:

	Comments to Authors	Improvements or additions to the description (the correction is marked with a yellow highlight)	In-Page
	from Reviewer A		
1	Introduction - The introduction states the main topic and previews the structure of the paper. But the methodology used is not par with use case where the study has been applied. If we see the data for 6 semester is constantly lower/higher, human intelligence can also tell if graduation will be on time or not	<p>Predicting is not an easy task [29] [13]; difficulties arise due to considering several criteria as the basis for prediction or decision-making [30] [13]. Therefore, previous researchers emphasized that what often happens is inaccuracy in making decisions [29]. That is why there is a need for a system that can assist in predicting with reasonable accuracy the results. Machine learning can predict accurately [25].</p> <p>... who have completed their studies. Machine learning is useful for systematically predicting which students will graduate on time and who will be late for graduation based on variations in the 6-semester achievement index value which has a decimal value variation of 0.0 to 4.0. Students who have a good to very good achievement index are students who have a minimum achievement index of 2.0. Research data shows that not always students who excel and are very good will definitely graduate on time (see the data set in Table 2). Machine learning that implements data mining methods has intelligence that is able to reveal hidden patterns in big data [4] and can predict with high accuracy [5]</p>	<p>In the Introduction subsection, page 2<sup>nd</sup> column.</p> <p>In the Related Work subsection, page 3, 2<sup>nd</sup> column, in Data Collection subtitle</p>
2	Results – Accepted. But it would be better if we apply the same principal in some other data set where we are solving some complex problems.	Thank you very much for the valuable advice	
3	Conclusion – The conclusion of the paper is proper as per objective. But as mentioned it would have been much better and interesting to see the results where the data set is much complex to suit the need of SMOTE	Thank you very much for the valuable advice. The author suggests it in the Conclusion for further research as follows: ..... and research with more complex data sets to meet SMOTE needs	In the Conclusion subsection, page 5, 2 <sup>nd</sup> column

	Comments to Authors	Improvements or additions to the description (the correction is marked with a yellow highlight)	In-Page
	from Reviewer B		
1	Show the vectorization of each attribute especially those categorical data in the paper.	Attributes with categorical data types are converted to numeric data types before the oversampling process using SMOTE. The gender attribute has a categorical data type with categories 'L' and 'P', so the category 'L' becomes 0 and 'P' becomes 1.	In the Research Methodology subsection, page 4, 1 <sup>st</sup> column, in the Data Pre-processing subtitle
2	Present the ratio of the training and test sets in the paper	Before classification, the data-set is first divided into training and testing data using 10-fold cross-validation, divided into 10 data groups using python tools.	In the Research In Methodology subsection, page 4, 2 <sup>nd</sup> column, in the Classification Method subtitle
3	Present the hyperparameters of the SVM model in the paper	The parameters used in the SVM method are kernel RBF, C = 5, gamma = 2, and toll = 0.0001	In Methodology subsection, page 4, 2 <sup>nd</sup> column, in the Classification Method subtitle
4	Maybe add more tests like what if you are to add more skewness to your data set. Will the accuracy increase?	Thank you very much for the valuable advice. The use of Smote sampling reduces the skewness of the data distribution so that it can improve the performance of the classification method used [48] [49]	In the Result and Discussion subsection, page 5, 2 <sup>nd</sup> column
5	The increase in accuracy is not significant to conclude SMOTE in the classification. Add more methods to improve the accuracy like improving vectorization? add new attributes? or combine it with other algorithm.	Thank you for this valuable input. The authors place these inputs as suggestions for further research in the Conclusion subsection as follows: Finally, further research can also combine several ensemble learning-based methods with smote to get better accuracy with other datasets.	In the Conclusion subsection, page 5, 2 <sup>nd</sup> column

Hopefully, what we have done fulfills the wishes of the reviewers. Thank you very much.

Sincerely yours  
Authors