IJICIS, Vol.22, No.3, 124-137
DOI: 10.21608/ijicis.2022.146420.1195

International Journal of Intelligent Computing and Information Sciences

https://ijicis.journals.ekb.eg/

PREDICTING STUDENTS' PERFORMANCE USING AN ENHANCED AGGREGATION STRATEGY FOR SUPERVISED MULTICLASS CLASSIFICATION

Mohamed Farouk Yacoub*
Department of Information Systems, Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt.
mohamed_yacoub2020@yahoo.com

Huda Amin Maghawry
Department of Information Systems, Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt.
huda_amin@cis.asu.edu.eg

Nivin A. Helal
Department of Information Systems, Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt.
nivin.atef@cis.asu.edu.eg

Sebastián Ventura Soto
Department of Computer Sciences and Numerical Analysis, University of Cordoba, Spain.
sventura@uco.es

Tarek F. Gharib
Department of Information Systems, Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt.
tfgharib@cis.asu.edu.eg

Received 2022-06-22; Revised 2022-07-30; Accepted 2022-08-01

Abstract: *Predicting students performance efficiently became one of the most interesting research topics. Efficiently mining the educational data is the cornerstone and the first step to make the appropriate intervention to help at-risk students achieve better performance and enhance the educational outcomes. The objective of this paper is to efficiently predict students' performance by predicting their academic performance level. This is achieved by proposing an enhanced aggregation strategy on a supervised multiclass classification problem to improve the prediction accuracy of students' performance. Two binary classification techniques: Support Vector Machine (SVM) and Perceptron algorithms, have been experimented to use their output as an input to the proposed aggregation strategy to be compared with a previously used aggregation strategy. The proposed strategy improved the prediction performance and achieved an accuracy, recall, and precision of 75.0%, 76.0%, and 75.48% using Perceptron, respectively. Moreover, the proposed strategy outperformed and achieved an accuracy, recall, and precision of 73.96%, 73.93%, and 75.33% using SVM, respectively.*

Keywords: *Machine Learning, AI, Students' performance prediction, Educational data mining, Multiclass Classification, Supervised Learning.*

* Corresponding author: Mohamed Farouk Yacoub
Department of Information Systems, Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt.
E-mail address: mohamed_yacoub2020@yahoo.com

1. Introduction

Educational systems contain massive educational data such as students' behavior, students' enrollment, results, and attendance. These data could be investigated to enhance the learning and educational outcomes [1]. Educational Data Mining (EDM) has become very important to discover knowledge and hidden patterns that can help decision makers to take the suitable actions for enhancing the educational process [2].

Data mining deals with the educational research topic to discover and evaluate various crucial indicators from the data [3]. EDM can reduce the students' drop-out rate. Besides, it can provide the academic organizations with knowledge to develop knowledge-based strategies and make on time decisions regarding at-risk students.

Efficient students' performance prediction is a challenge and an interesting topic in EDM research area [4]. It helps educators to follow the students' performance to determine those at-risk and who need the help [5]. Students' performance prediction is useful to determine the features, behaviors, and hidden patterns and relations that may affect the students' performance [6][7]. There are two perspectives in research, the first perspective is to identify a predictive model to efficiently predict students' performance. The second one is to find the features that affect the performance of students [4].

The performance of the classifiers may be different based on two factors, the number of the dataset features and the number of classes to be predicted [8]. Various machine learning, especially supervised learning techniques can be experimented to predict at-risk students. Besides, they can be used to identify the factors and patterns that may be relevant to the existing drop-out rate [9]. Therefore, educational data mining (EDM) research field is very important for the educational knowledge discovery and the decision making [10].

Therefore, the objective of this paper is to propose an efficient students' performance classification model. This is achieved by proposing an enhanced aggregation strategy of the results generated from the binary classifiers after applying the One-vs-Rest (OvR) technique

[11] on supervised multiclass classification problem. Support Vector Machine (SVM) [12] and Perceptron [13] classifiers have been applied. Experimental results proved the effectiveness of the proposed aggregation strategy.

The following sections of the paper are organized as follows: section 2 presents the related work, section 3 describes the methodology, experiments and results are discussed in section 4, and finally conclusion is discussed in section 5.

2. Related Work

Students' performance prediction became a necessity educational task. However, it is considered as a challenge as there are a lot of influencing factors and indicators that may affect the student performance. The reputation of the educational organizations or institution can be negatively affected by the increased number of low performance students. Detecting students who are at-risk as early as possible is very important to provide the convenient guidance to those students to help them achieve better results and performance [14] [6] [15]. Many studies applied clustering and classification techniques in the EDM research area trying to get an enhanced predictive model for students' performance prediction. There are many research papers that focus on one of two major perspectives for enhancing the learning process outcomes. The first perspective is achieving a contribution in enhancing the prediction performance using a combination of an existing techniques and pay more attention to the pre-processing stage that may open an opportunity to enhance the proposed model [16].

The second perspective is mining the educational data and finding the most important features in a dataset to be used for building the predictive model and enhancing the students' behavior and influencing factors in the students' performance [17]. Xing et al. [18] implemented deep learning techniques to improve the online learning outcomes and performance in terms of temporal dropout prediction concept. They achieved an accuracy of 90.8% in the first week and an accuracy of 96.1% in the week number 7.

Bindhia et al. [19] proposed a hybrid data mining approach to predict the students who are vulnerable to fail or will not get good performance. The hybrid model contains clustering and classification techniques. Moreover, they achieved an accuracy of 75.47%. Uzel et al. [20] applied various machine learning and deep learning techniques such as Multilayer Perceptron (MLP), Random Forest (RF), Naïve Bayes (NB), Decision Tree (J48), and voting method. Moreover, to identify the relations between the dataset features, they applied the Apriori algorithm.

According to Galar et al. [21], the One-vs-One (OvO) strategy is one of the most used techniques to solve the multi-class classification problems. Regarding the OvO and One-vs-Rest (OvR) approaches, the multi-class problem is divided into binary classification problems to be easier to be solved.

Multiclass classification problems are challenging in machine learning. Therefore, the class binarization method became widely used to implement multiclass classification problems by converting it to multiple binary classifiers. Moreover, NeuroEvolution, like NeuroEvolution of Augmenting Topologies (NEAT), became widely used to generate the artificial neural networks using evolutionary algorithms [25]. Gao et al. [26] applied class binarization techniques to a neuroevolution algorithm to generate neural networks for the multiclass classification problem. They proposed a new methodology that is based on applying Error-Correcting Output Codes (ECOC) to redesign the class binarization strategies. Moreover, the proposed method achieved higher accuracy with lower variance. However, there is a lack in paying more attention to the aggregation strategies. Therefore, this paper proposes an enhanced aggregation strategy of the results generated from the binary classifiers after applying the OvR technique on a supervised multiclass classification problem.

3. Methodology

This research aims to efficiently predict students' performance by predicting their academic performance level. This is achieved by proposing an enhanced aggregation strategy of the results generated from the binary classifiers after applying the One-vs-Rest (OvR) technique on a supervised multiclass classification problem. The results of the proposed aggregation strategy were compared with the results from the traditional argmax aggregation strategy [22]. Argmax aggregation strategy is a function that returns the max value of a list of predicted probabilities argument. The first subsection presents the dataset description in detail. For good understanding of the dataset, the attributes were visualized, presented, and the statistical analysis was discussed. In the second subsection, we describe in detail the proposed aggregation strategy.

3.1. Dataset

The dataset used in this study is "Students' Academic Performance Dataset" [23]. It is gathered from a learning management system (LMS). The dataset contains 480 records or students from different educational levels with 16 features. The attributes are categorized into 10 academic attributes ("Viewing announcements", "Educational Stages", "Grade Levels", "Section ID", "Topic", "Semester", "Discussion groups", "Raised hand", "Visited resources, and "Student Absence Days"), 4 personal attributes ("Gender", "Parent Answering Survey", "Parent responsible for student", and "Parent School Satisfaction"), and 2 demographic attributes ("Place of birth" and "Nationality"). Based on these attributes, the academic level is considered as the performance indicator. It is categorized into three classes: "Low-Level" which includes 127 data instances, "Middle-Level" which includes 211 data instances, and "High-Level" which includes 142 data instances.

Pre-processing is an essential process for any data set. It includes data cleaning and transformation. The dataset has been preprocessed and prepared to be appropriate for applying various machine learning and classification techniques. At first, data was converted from nominal to numerical values. Secondly, some features were reshaped to be within a certain range using standardization method. Figure 1 shows a sample of dataset attributes and data samples.

Dataset features were visualized to understand their distribution, as shown in Figures 2 to 5. Figures 2, 3, 4, and 5 show the Box Plots of "Raise hands", "Visited resources", "Announcements view" and "Discussion" features, respectively. Regarding the "Raise hand" and "Visited resources" features as shown in the figures 2 and 3 respectively, there is a significant difference between the two groups "Low" and "High" classes through data samples. However, there is a small difference between the two groups "Low" and "Middle" classes and between the two groups "High" and "Middle" classes through data samples, which is an indication for the medium effect of these features on the target feature for the prediction model. Regarding the figure 4 that is related to "Announcements view" feature, there is a small effect between the two groups "Low" and "High" classes and between the two groups "Low" and "Middle" classes through data instances. As shown in figure 5 that is related to "Discussion" feature, there is a moderate effect between the two groups "Low" and "High" classes in the dataset samples.

As shown in Figures 2 and 5 related to "Raise hands" and "Discussion" features respectively, the data of "Middle" and "High" classes are symmetric, but in "Low" class, the data is skewed. As shown in Figures 4 and 5 related to "Visited resources" and "Announcements view" features respectively, the data of "Low", "Middle" and "High" classes are symmetric.

	gender	NationalITy	PlaceofBirth	StageID	GradeID	SectionID	Topic	Semester	Relation	raisedhands	VisITedResources	AnnouncementsView	Discussion	ParentAnsweringSurvey	ParentschoolSatisfaction	StudentAbsenceDays
470	0	11	5	1	5	0	8	0	0	0.81	0.868687	0.877551	0.408163	0	1	0
471	0	11	5	1	5	0	8	1	0	0.78	0.828283	0.795918	0.530612	0	1	0
472	0	11	11	1	5	0	11	0	0	0.80	0.878788	0.755102	0.683673	0	1	0
473	0	11	11	1	5	0	11	1	0	0.85	0.888889	0.806122	0.704082	0	1	0
474	1	5	5	1	5	0	10	0	0	0.02	0.070707	0.040816	0.071429	1	0	1
475	1	5	5	1	5	0	10	1	0	0.05	0.040404	0.051020	0.071429	1	0	1
476	1	5	5	1	5	0	11	0	0	0.50	0.777778	0.142857	0.275510	1	0	0
477	1	5	5	1	5	0	11	1	0	0.55	0.747475	0.255102	0.285714	1	0	0
478	1	5	5	1	5	0	8	0	0	0.30	0.171717	0.142857	0.571429	1	0	1
479	1	5	5	1	5	0	8	1	0	0.35	0.141414	0.234694	0.622449	1	0	1

Figure 1: Snapshot from the Students' Academic Performance Dataset.

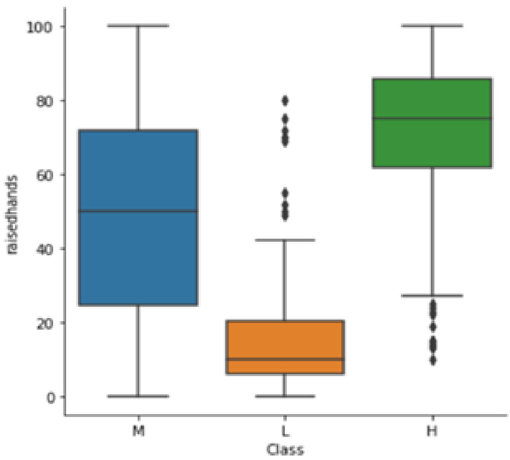

Figure 2: Box Plot for "raised hands" attribute

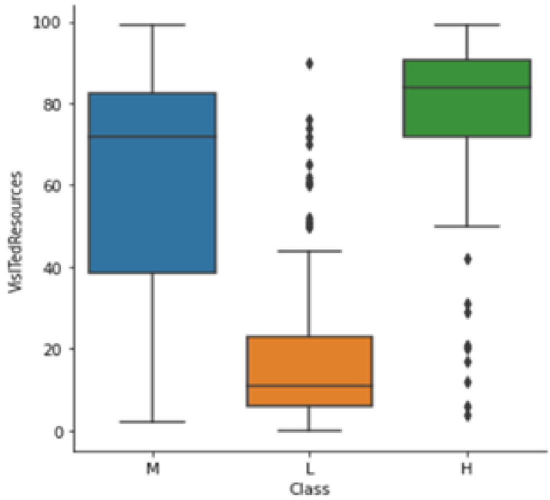

Figure 3: Box Plot for "Visited Resources" attribute

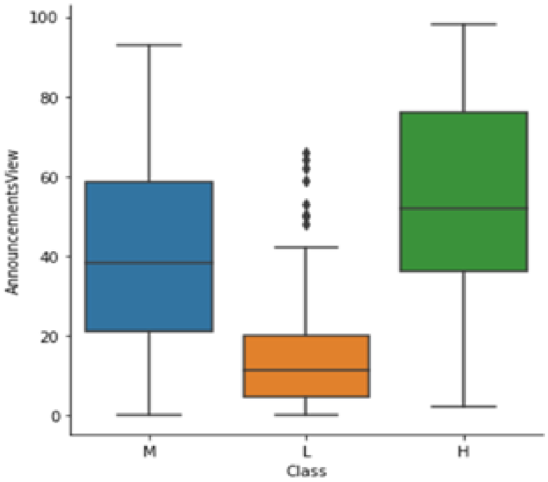

Figure 4: Box Plot for "Announcements View" attribute

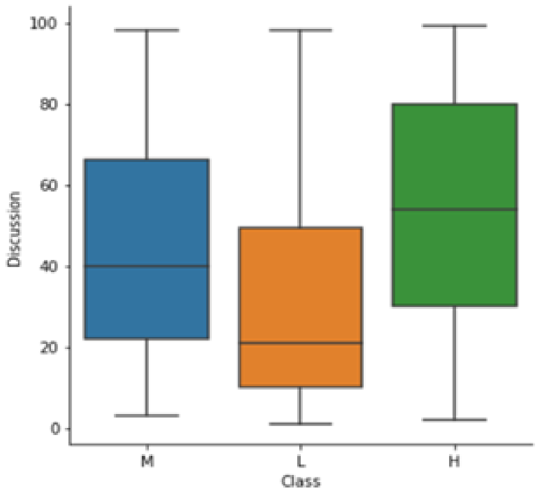

Figure 5: Box Plot for "Discussion" attribute

3.2. Proposed Aggregation Strategy

The main idea is to use binary classifiers for a multiclass classification problem. The reason is that predicting only two classes rather than three produces better results. Therefore, we applied the One-vs-Rest (OvR) heuristic method to use binary classification algorithms for a multi-class classification problem. Therefore, we proposed an enhanced aggregation strategy of the generated results from the binary classifiers after applying the One-vs-Rest (OvR) approach on a supervised multiclass classification problem.

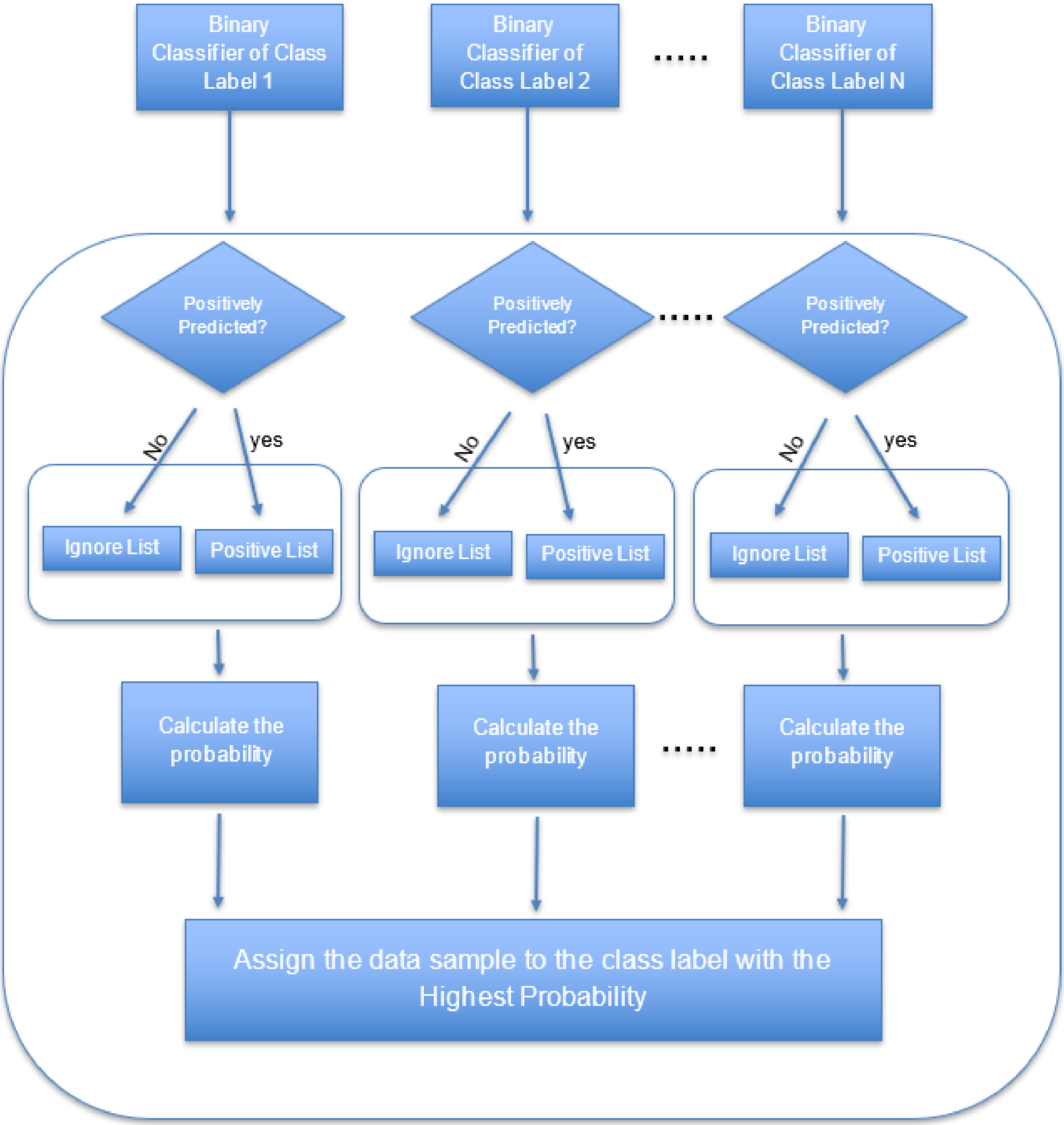

Figure 6: The Proposed Aggregation Strategy Architecture

As shown in figure 6, the proposed strategy for the aggregation consists of two steps. The first step contains binary classifiers. The number of binary classifiers is as shown in equation (1).

$$\textit{Number of binary classifiers} = N \qquad \text{equ. (1)}$$

Where N is the number of class labels in the multiclass classification problem.

Each of the binary classifiers is corresponding to a specific class label. The data sample is classified using each classifier once. If the predicted class label of the data sample is the same as the classifier class label, then the classifier output will be value of 1 and will be added to "Positive List". This means that the binary classifier predicted this data sample positively with a certain probability. However, if the classifier predicted output is -1, then this means that data sample does not belong to this class of the binary classifier with a certain probability and it and will be added to "Ignore List". The probability is calculated as shown in equation (2).

$$P = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots)}} \qquad \text{equ. (2)}$$

Where P is the Probability, B_o, B_1, and B_2 are the coefficients, and X_1 and X_2 are the independent variable values.

Therefore, the first step is about generating a list of data samples which positively predicted from each classifier along with their corresponding probability of prediction. This means that the data sample is positively classified and belongs to the corresponding class label of the binary classifier.

The second step is to assign each data sample to the class label with the highest probability across all classifiers. Regarding the negative prediction, if the classifier value is -1, then the data sample is ignored and would not be added to the list of positively predicted data samples and will be added to "Ignore List". It will be an input to this step, that selects the final class label based on the highest probability of prediction. Also, if the data sample is negatively predicted by all binary classifiers, it will be assigned to a class label with the highest prediction probability from the "Ignore List".

4. Experiments and Results

4.1 Experiment Environment

Anaconda Navigator [24] was used to simplify the packages management and deployment. The implementation language was Python using pandas, numpy, and sklearn libraries. The experiments were performed on a machine with 2.60 GHz processor, 16 GB memory, and Windows 10 64-bit operating system.

4.2 Experiments

To evaluate the proposed aggregation strategy, SVM and Perceptron classifiers were used to compare their aggregation results with the argmax. The proposed aggregation strategy approach was evaluated using the accuracy measure, as shown in equation (3).

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \qquad \text{equ. (3)}$$

Where: TP = True positive; FP = False positive; TN = True negative; FN = False negative.
According to dataset, the performance indicator was the success levels that were classified to three categories. The dataset has been split into 80% for training and 20% for testing. Therefore, the test dataset consists of 96 new data samples to be predicted. All experiments were carried out using the 10-fold cross validation:

- Low-Level contains 127 of the data instances.
- Middle-Level contains 211 of the data instances.
- High-Level contains 142 of the data instances.

First, each classifier correspondent to a certain class label was evaluated. Both SVM and Perceptron classifiers were used. The results are shown in table 1 in terms of accuracy, recall, and precision metrics.

Table 1: Comparison of SVM and Perceptron binary classifiers in terms of accuracy, recall, and precision metrics for each class label

Method	**Class Label**	**Accuracy**	**Recall**	**precision**
Perceptron	**Low**	**90.63%**	**84.62%**	**81.48%**
	Medium	55.21%	2.38%	33.33%
	High	77.08%	28.57%	**80.0%**
SVM	**Low**	89.58%	80.77%	80.77%
	Medium	**59.38%**	**40.48%**	**54.84%**
	High	**84.38%**	**64.29%**	78.26%

Then, the proposed aggregation strategy was compared to the commonly used argmax approach [22]. The approaches were applied to aggregate the results after applying the OvR approach. Both SVM and Perceptron were used as binary classifiers. The results are shown in table 2.

Table 2: Comparison of the proposed aggregation strategy versus the argmax strategy [22] in terms of accuracy, recall, and precision metrics using SVM and Perceptron as binary classifiers

Method	**Aggregation Strategy**	**Accuracy**	**Recall**	**precision**
Perceptron	Argmax [22]	73.96%	74.42%	75.03%
	Proposed Strategy	**75.0%**	**76.0%**	**75.48%**
SVM	Argmax [22]	72.92%	72.74%	74.43%
	Proposed Strategy	**73.96%**	**73.93%**	**75.33%**

As shown in table 2, the proposed aggregation strategy outperformed the commonly used aggregation strategy, argmax strategy [22] using both classifiers, Perceptron and SVM classifiers. The proposed strategy achieved an accuracy of 75.0%, recall of 76.0%, and precision of 75.48% using Perceptron classifier. However, the results achieved using argmax aggregation strategy was an accuracy of 73.96%, recall of 74.42%, and precision of 75.03%. Using SVM classifier, the proposed strategy achieved an accuracy of 73.96%, recall of 73.93%, and precision of 75.33% which is higher than the results achieved by the argmax aggregation strategy with an accuracy of 72.92%, recall of 72.74%, and precision of 74.43%.

5. Conclusion

One of the main objectives of educational data mining is to efficiently and accurately predict students' academic performance as early as possible to help them and make the appropriate customized intervention from the educational organization. In this paper, we proposed an enhanced aggregation strategy of the results generated from the binary classifiers after applying the One-vs-Rest (OvR) approach on a supervised multiclass classification problem. Two binary classifiers, Support Vector Machine (SVM) and Perceptron have been experimented to use their output as an input to the proposed aggregation strategy. The results were compared with a previous commonly used aggregation strategy, argmax aggregation strategy, in terms of the accuracy measure. The proposed aggregation strategy outperformed the argmax strategy using both classifiers, Perceptron and SVM classifiers. The proposed strategy achieved an accuracy of 75.0% using Perceptron classifier. This was higher than the result achieved using the argmax aggregation strategy with an accuracy of 73.96%. The proposed strategy achieved an accuracy of 73.96% using SVM classifier. The result is higher than the achieved result by the argmax aggregation strategy using the same classifier with an accuracy of 72.92%. For future work, new datasets can be used by the proposed aggregation strategy and this proposed strategy could be a base for further enhancements and improvements to propose better aggregation strategies.

References

1. A. Dutt, M. A. Ismail, and T. Herawan, "A systematic review on educational data mining", IEEE Access, IEEE Xplore, 2017.
2. A. Narvekar, V. Menezes, O. P. Angle, M. Lotlekar, S. Naik, and A. P urohit, "Students performance prediction using data mining techniques",International Journal for Research in Applied Science and Engineering Technology (IJRASET), 2020.
3. A. S. Olaniyi, S. Yakub, H. Moshood, S. Ibrahim, and A. Nathaniel, "Student's performance analysis using decision tree algorithms", Anale: Seria Informatic˘a. Computer Science Series, 2017.
4. A. Khan, S. K. Ghosh, "Student performance analysis and prediction in classroom learning: A review of educational data mining studies", Educ Inf Technol 26, 205–240, Springer, 2020.
5. M. Swathi, K. L. Soujanya, R. Suhasini, "Review on predicting student performance", International Conference on Communications and Cyber Physical Engineering (pp.1323-1330), Springer, 2020.
6. A. Namoun, A. Alshanqiti, "Predicting student performance using data mining and learning analytics techniques: a systematic literature review", Applied Sciences, 11(1), 237, 2021.
7. L. M. Zohair, "Prediction of student's performance by modelling small dataset size", International Journal of Educational Technology in Higher Education, 16-27, 2019.
8. R. Ghorbani, R. Ghousi: Comparing Different Resampling Methods in Predicting Students' Performance Using Machine Learning Techniques. In: IEEE Access, vol. 8, pp. 67899-67911, 2020.
9. W. F. W. Yaacob, N M. Sobri, S. A. M. Nasir, N. D. Norshahidi, W. Z. W. Husin, and K. Shaalan: Predicting Student Drop-Out in Higher Institution Using Data Mining Techniques. IOPScience, 2020.
10. A. Narvekar, V. Menezes, O. P. Angle, M. Lotlekar, S. Naik, A. Purohit: Students Performance Prediction using Data Mining Techniques. International Journal for Research in Applied Science and Engineering Technology (IJRASET), 2020.
11. C. M. Bishop: Pattern Recognition and Machine Learning, Springer, 2006.
12. C. CORTES and V. VAPNIK: Support-Vector Networks. Machine Learning. 20, 273-297, 1995.
13. Y. Freund and R. E. Schapire: Large Margin Classification Using the Perceptron Algorithm. Machine Learning, 37(3):277-296, 1999.
14. A. L. Sønderlund, E. Hughes, and J. Smith: The efficacy of learning analytics interventions in higher education: A systematic review. British J. Edu.Technol., pp. 1–25, 2018.
15. C. Romero and S. Ventura: Data mining in education. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 3, no. 1, pp. 12–27, 2013.
16. M. M. Rahman, Y. Watanobe, T. Matsumoto, R. U. Kiran, and K. Nakamura: Educational Data Mining to Support Programming Learning Using Problem-Solving Data. IEEE Access, vol. 10, pp. 26186-26202, 2022.
17. M. Kumar, C. Sharma, S. Sharma, N. Nidhi, and N. Islam: Analysis of Feature Selection and Data Mining Techniques to Predict Student Academic Performance. 2022 International Conference on Decision Aid Sciences and Applications (DASA), pp. 1013-1017, 2022.
18. W. Xing, D. Du: Dropout Prediction in MOOCs: Using Deep Learning for Personalized Intervention. Journal of Educational Computing Research, 2019.

19. B. K. Francis, S. S. Babu: Predicting Academic Performance of Students Using a Hybrid Data Mining Approach. Journal of Medical Systems, 2019.
20. V. N. Uzel, S. S. Turgut, and S. A. Özel: Prediction of Students' Academic Success Using Data Mining Methods. 2018 Innovations in Intelligent Systems and Applications Conference (ASYU), pp. 1-5, 2018.
21. M. Galar, A. S. Fernández, E. Barrenechea, H. Bustince, and F. Herrera: Dynamic classifier selection for One-vs-One strategy: Avoiding non-competent classifiers. ELSEVIER, Volume 46, Issue 12, Pages 3412-3424, 2013.
22. D. Ferger: A continuous mapping theorem for the argmax-functional in the non-unique case. Statistica Neerlandica, Vol. 58, nr. 1, pp. 83–96, 2004.
23. P. Cortez, [Students' Academic Performance Dataset]. https://www.kaggle.com/aljarah/xAPI-Edu-Data, 2008.
24. R. Phelps, M. Krasnicki, R. A. Rutenbar, L. R. Carley, and J. R. Hellums: Anaconda: robust synthesis of analog circuits via stochastic pattern search. IEEE, 1999.
25. Z. Gao and G. Lan: A NEAT-based multiclass classification method with class binarization. Genetic and Evolutionary Computation Conference, pp 277–278, 2021.
26. G. Lan, Z. Gao, L. Tong, and T. Liu: Class binarization to neuroevolution for multiclass classification. Neural Comput & Applic, Springer, 2022.