An Assessment of Fitness of Undergraduates in BITZH by Using SMOTE and Machine Learning Algorithms

Shiyi Wang[1], Zejian Lin[1], Yanhui Huang[1], Chuangfeng Ma[1], Xindong Zhao[1] and Xiaoyu Wei[1, *]

[1]*Beijing Institute of Technology, Zhuhai, Zhuhai, China*
**Corresponding Author: xiaoyu.wei@zhuhai.Bryant.edu*

ABSTRACT

The physical fitness test is used as a tool to evaluate students' physical health. This paper proposed a assessment of physical fitness test based on machine learning (ML) to improve college students' awareness of physical fitness. In this paper, we collected the number of records of fitness test at about 120 thousands from undergraduates who come from Beijing institute of technology in Zhuhai. Firstly, we first classified students' physical fitness into five categories by using K-Means. Then, we resampled the dataset by using the synthetic minority over-sampling technique (SMOTE) to address the imbalance of dataset. This framework that constructed with ML methods, included DT, RF, GBN, LR, SVM, XGB. In addition, Voting combined with single model which can improve the accuracy of model. The model was evaluated by using these performance metrics, such as Macro-Precision, Kappa, and so on. The result of experiment shows that the precision of SVM is 99.54%, and the recall of this is 99.53%. At the same time, the ensemble model combined SVM with Voting have better performance than others. In conclusion, the model which build based on Voting and SVM can detect and predict the level of health effectively.

Keywords: *Physical Fitness, Machine Learning, Multiple Classifiers, Voting*

1. INTRODUCTION

Students' physical fitness test as a tool for the assessment of health. It is important for mentors and colleagues to know the level of physical fitness development of students [16]. However, the system for evaluating student physical fitness has some shortcomings. For example, the level of physical fitness has some missing health information and it does not reflect the unevenness of individual fitness. At the same time, the current physical fitness assessment method are more likely to focuses on objectivity. It is quite difficult to build a bridge from assessment to monitoring to other practical problems, which can no longer meet the new needs of assessment in the new era [6]. Therefore, we take into account the development trend of physical fitness evaluation values at home and abroad [9], it is vital to build a scientific and complete evaluation system of physical fitness test for colleague students based on theoretical foundations, such as developmental evaluation and general system theory. It is essential to ensure that every youngers what they should do and how to keep fit.

With time goes by, more and more people may pay more attention on the level of health and the development of them. Results show that 39.9% of colleague students have a negative attitude towards the National Student Physical Fitness Standards (2016 Revision) (the Standards) test. They are afraid about the physical fitness test themselves and ignoring their weakness, and do not purposely strengthening their exercise [14]. Therefore, the assessment of physical fitness test for students was established that enable universities to understand the level of physical development of individual students and propose improvement measurements to help them to be strong.

In this paper, the unsupervised learning was based on the K-Means algorithm [20] by inputting the dataset and classified into 5 categories. This study used the synthetic minority over-sampling technique (SMOTE) [7] to solve the imbalance of class found in the physical fitness tests dataset. Then, we carried out six machine learning (ML) algorithms, such as Decision Trees (DT) [11], Random

A. El-Hashash et al. (Eds.): IEIT 2022, ACSR 100, pp. 587–595, 2023.
https://doi.org/10.2991/978-94-6463-058-9_94

Forest (RF) [1], Gaussian Naive Bayes (GNB) [13], Logistics Regression (LR) [5], Support Vector Machine (SVM) [18], and Extreme Gradient Boosting (XGB) [4], to build the assessment of physical fitness test for students. Voting is used to make progress in multiple classifiers to increase the precision and robustness of the model. The major contribution of this research work can be summarized as follows:

- We have built a comprehensive assessment model based on K-means and various ML methods to help college students understand their level of health, physical fitness development, and physics.
- We implemented SMOTE to address the imbalance of data categories found in the physical fitness test dataset.
- We collected the number of records of fitness test for five years for about 120 thousand from undergraduates who come from the Beijing institute of technology in Zhuhai.
- We proposed multiple classifiers using Voting to rise the reliability of the model. In addition, we used multi-categorical performance metrics for comparative analysis of the models, using the following metrics: Macro-Precision, Macro-Recall, Macro-F1, Kappa, Hamming Loss.

The rest of the paper is organized as follows. Section 2 provides a literature review of previous work. Section 3 provides a background of the ML methods and datasets that were used in this paper. In Section 4, we conduct the experiments. Section 5 concludes the research.

2. RELATED WORK

In term of the imbalance of class and how to classify the type of samples, Chawla [3] proposed the SMOTE algorithm, which is a composite sampling algorithm for synthesizing data to address imbalanced class problems in 2002. Du Yunmei et.al [8] used NB to classify the physical fitness test data into four categories in 2018. The accuracy of the classifier reached 77.98%, which could achieve a certain probability sense of correct judgment on the physical fitness of college students. In 2021, Kou Lei et.al [12] used three ML algorithms, RF, GBDT, and the neural network algorithm to classify and predict the physical fitness of college students, which showed GBDT is the best model, and the accuracy of it can reach 96%. Later, Hao Linlin [10] proposed a model based on K-Means and BPNN to evaluate the physical fitness of university students, which classified the physical fitness of university students into eight categories. With different types having different physical characteristics, and the accuracy of its model reached at 94%.

Through the study of previous work, classification is the foundations of ML [2]. There are two main types of classification, one is binary classification, such as DT, GNB, LR, and SVM. Another is multi-classification, such as DT, RF, GNB, and SVM. In the real world, it often faces the problem of category imbalance and Hierarchical Multi-label Classification(HMC). Valiant, L.G. [17] proposed the Probably Approximately Correct (PAC) model in 1984, which is the first time of the definition of weak learning and strong learning. In other words, the accuracy of recognition is higher than random guess called the former one, and vice versa. The most representative methods in ensemble classifiers, such as Boosting, Bagging, and Stacking. XGB is a representative algorithm in Boosting, proposed by Chen Tianqi in 2016. He stated that the XGB is an algorithm based on Gradient Boosting Decision Tree (GBDT). XGB supports column sampling, which reduces overfitting and also reduces the regularization term to prevent overfitting. In 2017, Wang Jing et.al [19] published a paper in which SVM was used to build a prediction model of college students' physical fitness and a particle swarm algorithm was used to select the model parameters. This model overcomes in practice the imbalance in student class categories of the traditional model and achieves a prediction accuracy of 94.73%, improving the prediction of college student fitness.

In conclusion, previous studies have used a single ML or multiple classifiers to classify and predict the physical fitness of university students and compared finally. In this paper, K-Means was first used to classify the sample. Then, we selected the top 2 models among all models by the performance. Finally, we introduced several performance metrics which can evaluate these models and make up for the shortcomings of previous works which only use one model.

3. RESEARCH METHODOLOGY

3.1. Framework of the assessment of the fitness test

This paper proposed a assessment of physical fitness test for students based on SMOTE and ML. The process of this article is divided into five steps. Firstly, we need to process the dataset, such as missing values and outliers, and normalize by using MIN-MAX. Secondly, we plan to carry out K-Means algorithm and take the value of K according to the criterion of Sum of the Squared Errors(SEE). Then, we implemented SMOTE to address the imbalance of data categories. In addition, we intend to use DT, RF, GBN, LR, SVM, XGB to build the assessment. In the fourth step, we use five performance metrics, such as Marco-P, Marco-R, Marco-F1, Kappa, and Hamming Loss, to evaluate performance of model. Finally, we select the best ML model among the six models and combine them with Voting to improve the robustness and reliability of the models. The overall framework of the proposed the assessment of fitness for students is shown in Figure 1.

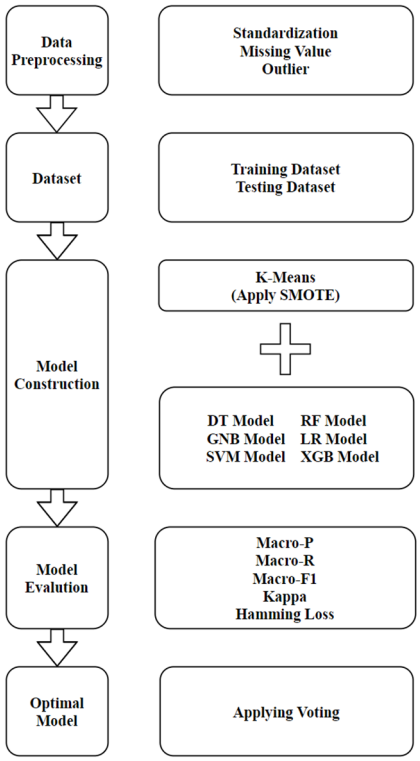

Figure 1: The overall framework of the proposed intelligent approach for assessment of the fitness test.

3.2. K-Means

The K-Means algorithm is a division-based algorithm in cluster analysis. It has the simplicity of thought, effectiveness, and ease of implementation. It is widely used in multiple areas of ML [15]. The K-Means divides a dataset containing multidimensional data points into multiple data subsets. Each division represents one class, and each class has a category center. The K-Means clustering algorithm selects the Euclidean distance as the similarity and distance judgment criterion and calculates the sum of squared distances from each point in the class to the center of the cluster. The goal of clustering is to minimize the sum of squares of the total distances of each class. The measures as shown in (1).

$$J(C_k) = \sum_{x_i \in C_k}^{n} \left\| x_i - \mu_k \right\|^2 \qquad (1)$$

In summary, the K-Means classify samples from an initial class division. Then, it assigns each data point to each class in order to reduce the total sum of squared distances.

3.3. OTHER MACHINE LEARNING ALGORITHMS

The aim of this paper is to predict the type of students by using machine learning models. Thus, we will introduce the details of several models that used in this, such as DT, RF, GNB, LR, SVM, and XGB.

DT is a non-parametric supervised learning method for classification and regression. It is a collection of nodes that can make decisions for some feature connected to certain classes. The purpose of the DT is to create a model that predicts the value of a target variable by learning simple decision rules from indicators. DT is also the basis of RF and XGB which is a part of the tree.

RF model is an ensemble learning model that can analyse data which have numerous features. In general, the algorithm selects the random subset of features for training once. At the same time, it can save time and implement easily compared with others. In addition, the algorithm also introduces randomness, which can avoid the overfitting phenomenon effectively.

GNB is a type of supervised ML. This algorithm assumes that all features are Gaussian distributed and independent of each other. It can apply both binary classification and multi-Classification. The algorithm uses density functions for continuous variable data. After that, its accuracy is higher than before [21].

LR can predict whether company have financial fraud or not by inputting many variables. In addition, it is not only can not be affected by slight multicollinearity but also it can analyse large data while using fewer resources.

SVM is a supervised ML that used for binary classification regression and classification subjects. The algorithm is very effective for data with many feature. Currently, there are two ways to solve classification problems by carrying out SVM. The first one is to construct several binary classifiers and combining them together. Another one is to directly consider the parameter optimization of all subclassifies simultaneously. This can avoid the neural network structure and local minima problems effectively and make progress in the performance.

XGB is combined with a categorical regression tree. Its idea is to keep adding trees and feature splitting to grow a tree, each time when adding a tree is equivalent to learning a new function to fit the residuals of the previous prediction. The final training is completed to get k trees, each tree will fall to a corresponding leaf node and a corresponding score, and finally, the corresponding score of each tree will be added up to the predicted value of the sample.

Instead of a separate machine learning algorithm, the ensemble classifier approach combines a given learning algorithm to obtain a more comprehensive integrated model. The underlying idea of ensemble classifier is that even if a particular weak classifier gets an incorrect prediction, other weak classifiers can correct the error back, allowing the algorithm to have a good strategy on data sets of all sizes. XGB is an important branch within the ensemble classifier boosting approach, an algorithm that provides a parallel tree boosting for fast and accurate problem-solving. It is an optimized distribution gradient

enhancement library that is efficient, flexible, and portable.

3.4. VOTING

Voting is a combination strategy for classification problems within an ensemble classifier. Its basic idea is to select the class with the highest output among all ML algorithms. It involves a combined model with at least two algorithms. Each algorithm makes its own predictions for each test sample, which requires a majority of valid votes to be approved. The algorithm as shown in (2).

$$H(x)=\begin{cases} c_j, & \sum_{i=1}^{T} h_i^j(x) > 0.5\sum_{k=1}^{N}\sum_{i=1}^{T} h_i^k(x) \\ refuse, & other \end{cases} \quad (2)$$

According to (2), T denotes that the number of classifiers and N denotes that the number of N categories, i.e. If the prediction result of category j by T classifiers is greater than half of the total voting result, it is predicted to be category j. Otherwise, it is rejected.

3.5. DATASET

This paper collected data from a university for five years from 2016 to 2020. The data was collected in strict accordance with the Standards for Student's physical fitness and technical specifications and graded by the rules of Standards, and the field quality control met the requirements. Therefore, this data has the completeness, accuracy and validity. In the following, we clean the data to facilitate the modeling and analysis later. This is done as follows:

- Removing some unnecessary fields from the data, such as class and name, and remaing only the gender and the score information of each test item.
- Delete all null values and zero values.
- The noise of the dataset is removed by using a quartile to avoid the impact of outlier data on the analysis.
- Min-Max was used to normalize the data and eliminate the effect of the magnitude.

3.6. SAMPLING METHOD

SMOTE is one of the most prominent techniques used to resolve the category imbalance found in the dataset. SMOTE connects data points with their nearest data points to generate class-specific data points. It is very important to avoid overfitting the sampled dataset during the training process.

3.7. PERFORMANCE METRICS

We researched the multi-classification problem. The performance criteria of multi-classification model prediction performance can be divided into two categories according to the aspects examined, one class is instance-based classification accuracy, and in this paper, we choose macro-precision, macro-recall, and macro-F1. Another type of ranking is based on tags, in this paper, we choose hamming loss.

When n confusion matrices are generated, Precision, recall, F1-measure are calculated separately for each confusion matrix. Then the respective averages are calculated to produce: Macro-P, Macro-R, and Macro-F1 as shown in (3),(4) and (5).

$$Macro-P=\frac{\overline{TP}}{\overline{TP}\times\overline{FP}} \quad (3)$$

$$Macro-R=\frac{\overline{TP}}{\overline{TP}\times\overline{FN}} \quad (4)$$

$$Macro-F1=\frac{2\times Macro-P\times Macro-R}{Macro-P+Macro-R} \quad (5)$$

True Positive(TP): Predict positive class to positive class number, true is 0, prediction is also 0.

False Negative(FN): Predict positive class to negative class number, true to 0, predicted to 1.

False Positive(FP): Predict the number of negative classes to positive classes, true to 1, predicted to 0.

True Negative(TN): Predict the negative class as the number of negative classes, the true is 1 and the prediction is also 1.

The Kappa coefficient is used as a common indicator of the credibility of a model. When Kappa is higher, it means that the credibility of model is higher than before. It is calculated by the formula as shown in (6).

$$Kappa=\frac{p_0-p_c}{1-p_c} \quad (6)$$

p_0 is the sum of the number of correctly classified samples in each category divided by the total number of samples. We assumed that the number of real samples in each category is a_1, a_2, a_3, a_4, a_5. The predicted number of samples for each category is b_1, b_2, b_3, b_4, b_5. The total number of samples is n. The function for this p_c as shown in (7).

$$p_c=\frac{a_1\times b_1+a_2\times b_2+\cdots+a_n\times b_n}{n^2} \quad (7)$$

Hamming Loss (HL) is concerned with the number of misclassified tags. This metric measures the degree of inconsistency between the predicted markers and the actual markers of the sample.The smaller the value of Hamming loss, the better the model effect. This function is shown in (8).

$$\text{HammingLoss}(x_i, x_j) = \frac{1}{|D|}\sum_{i=1}^{|D|}\frac{xor(x_i, x_j)}{|L|} \quad (8)$$

|D| means the total number of samples. |L| is equal to the total number of labels. x_i means the prediction of label. y_i represents the real label. In other words, it is the difference between the predicted label and the real label.

4. EXPERIMENTS AND RESULTS

4.1. EXPERIMENTAL SETUP

The experiment was conducted using Windows 10, x64 processor, software version Python 3.7.4, and python library including scikit-learn 1.0.2.

4.2. RESULTS AND DISCUSSIONS

We practiced K-Means in datasets and calculated distortions each class from 1 to 10. Based on that, we plotted the figure was shown in Figure 2.

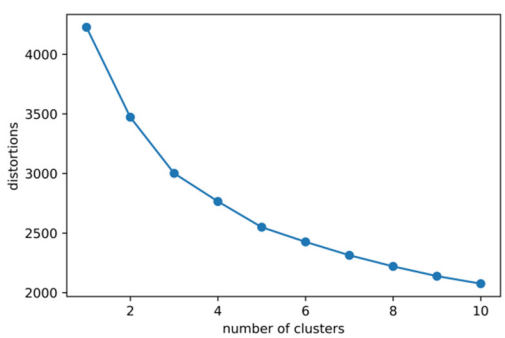

Figure 2: The results of the number of clusters by using K-Means.

As can be seen from Figure 2, the curve starts to flatten out when K is around 5. Thus, it is more appropriate to classify colleague students into 5 major categories of physical fitness. Then, we calculated the mean values of each indictor of each type as shown in Table 1. It is a good way to analyse the pros and cons of the level of health of students.

According to Table 1, category 0 boys have lower quality of body than other types, poor flexibility, low lung capacity, moderate levels of explosive power, low overall fitness, and poor levels of physical fitness. Category 1 boys is fatter. However, they have high lung capacity, low endurance, and a lack of explosive power. It is quite hard for them to do some physics sports because they are obseity.

Boys in category 2 are well-rounded, have high levels of endurance, flexibility, and explosiveness, and have a good level of physical fitness. Boys in this category are outstanding in all areas. The category 3 boys are mediocre students, which have the normal figure and below-average speed, flexibility, explosiveness, and endurance. Category 4 boys are explosive students who have a normal physique and good flexibility. In general, they are fast, flexible, above average, and have explosive lower limbs strength.

Table 1: MEAN VALUES OF EACH INDICTOR OF EACH TYPE.

Type	0	1	2	3	4
BMI	19.92	26.82	20.48	21.08	21.49
Body	5.66	9.85	14.17	10.46	17.88
Lungs	3699.44	4273.58	4045.91	3787.17	4130.63
50m	7.19	7.61	6.93	9.02	7.06
1000m	256.79	275.74	237.46	269.82	245.38
Pull	6.10	3.55	12.43	5.93	4.94
Jump	217.24	205.30	233.11	213.63	225.80

We used SMOTE to address the imbalance of the dataset. We divided the balanced dataset into a training set and a testing set according (80% and 20% respectively). DT, RF, GNB, LR, and SVM XGB models were used for training. Performance criteria were calculated for each model and the results were shown in Table 2.

Table 2: RESULTS OF DIFFERENT MODELS.

ML	Marco-P(%)	Marco-R(%)	Marco-F1(%)	Kappa(%)	HL (%)
DT	95.08	95.08	95.08	93.94	4.91
RF	97.62	97.62	97.62	97.03	2.38
GNB	94.98	94.91	94.90	92.50	5.08
LR	98.52	98.49	98.49	98.12	1.50
XGB	98.70	98.70	98.70	98.38	1.30
SVM	99.10	99.08	99.08	98.86	0.92

Table 2 shows that the model we have built is reasonable and valid. The Marco-P, Marco-R, Marco-F1, and Kappa values of each ML are high. All ML at over 94.98%, 94.91%, and 94.90% for Marco-P, Marco-R, and Marco-F1 respectively. A kappa value of these models at over 92.50% or more and a Hamming Loss of 5.0% or less. In addition, the accuracy of SVM and XGB is 99.12% and 98.71%. In conclusion, SVM and XGB have better performance than other four models. In addition, we plotted bar chart of the kappa values for the six models as shown in Figure 3.

The Figure 3 depicted that all models are outstanding. At the same time, 10-fold CV was performed on each of the six models to evaluate the performance of the models completely. The Marco-F1 values were used as the judging criteria. The results are shown in Figure 4.

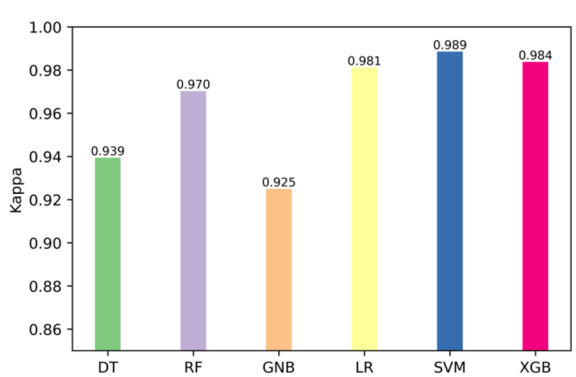

Figure 3: Kappa Value of Various Models.

The results of Figure 4 show that the F1 score of each model are above 93%, which indicate that the models are well fitted. The F1 score of the DT and GNB in the range between 93% and 95%. That of RF in the range from 97% to 98%. The F1 values of LR, SVM and XGB were able to maintain around 98%, while SVM having the highest F1 value and the best fit.

We carried out a Voting that gives the probability of classification in each category by each model and then takes the category with the highest probability as the predicted outcome. This will improve the shortcomings in the DT, RF, GNB, LR, SVM, and XGB and improve the accuracy of the models.

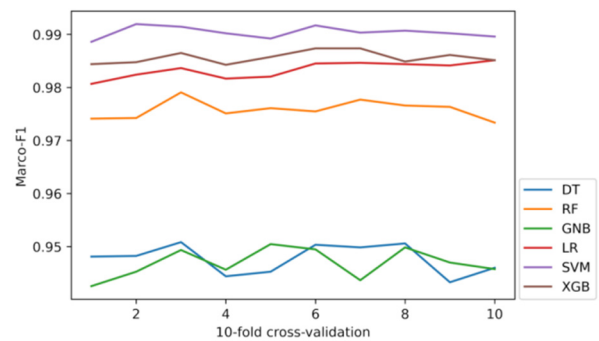

Figure 4: Macro-F1 of the proposed approach based on the 10-fold CV.

The experiment selected the SVM and XGB with the best evaluation parameters among the six ML models as the basic models and the other models as sub-model and the results are shown in Table 3.

Table 3 showed that the Marco-P, Marco-R, Marco-F1, and Kappa values of all the multiple classifiers were higher than single model. All multiple classifiers are exceed 95.08% for Marco-P, Marco-R, and Marco-F1 respectively. A kappa value of each multiple classifiers are more than 93.86% and a Hamming Loss of these are lower than 4.91% or less. In addition, RF+SVM, GNB+SVM, and LR+SVM can predict the type of sample more accurately than other models. To visualiztion, we plotted the histogram of Kappa values for the six combined models as shown in Figure 5.

Table 3: RESULTS OF VARIOUS MODELS BY USING VOTING.

ML	Marco-P(%)	Marco-R(%)	Marco-F1(%)	Kappa(%)	HL (%)
DT+SVM	95.08	95.08	95.08	93.86	4.91
DT+XGB	95.08	95.08	95.08	93.86	4.91
RF+SVM	99.58	99.58	99.58	99.48	0.42
RF+XGB	98.67	98.67	98.67	98.34	1.33
GNB+SVM	99.48	98.70	98.70	98.38	0.48
GNB+XGB	98.69	98.68	98.68	98.35	1.32
LR+SVM	99.54	99.53	99.53	99.42	0.47
LR+XGB	98.95	98.95	98.95	98.69	1.05

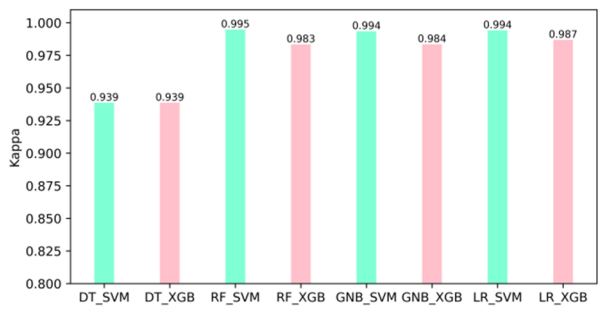

Figure 5: The histogram of Kappa values for the six combined models.

As shown in Figure 5, we know that the multiple classifier base on SVM is better than the multiple classifie base on XGB. In terms of validation, 10-fold CV was used for each multiple classifier and the Marco-F1 values were used as the main metric. The results are shown in Figure 6.

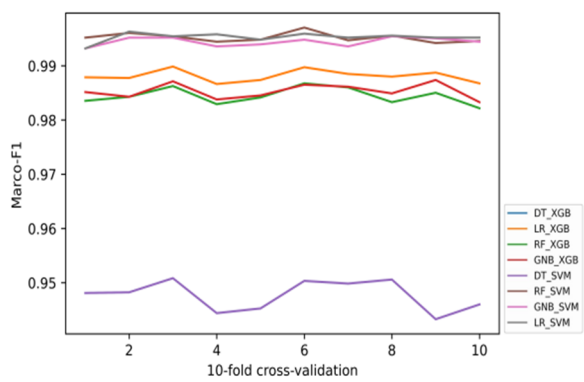

Figure 6: The Marco-F1 value of multiple models by using Voting.

Figure 6 shows a performance among each multiple classifier based on Marco-F1. The proposed approach reached the highest F1 score. Overall, RF+SVM, GNB+SVM, and LR+SVM have good performance among all.

5. CONCLUSION

This paper constructed the assessment of physical fitness tests for college students based on K-Means and various ML algorithms. Many standard models are used in the empirical evaluation, including DT, RF, GNB, LR, SVM, XGB. In addition, we also built ensemble algoritm by using Voting. During building, we chose SVM and XGB as basic model and combined others for training. Finally, we referred several performance metrics, such as, Macro-P, Macro-R, Macro-F1, Kappa, and Hamming Loss.

The ensemble classifier by using Voting has significantly improved compared with the initial model classification and prediction performance. The best model is the LR and SVM model. The precision of it is 99.54%. The recall of it is 99.53% recall. Macro-F1 is 99.53%. Kappa coefficient is 0.9942, and Hamming loss is 0.0047 which can predict the sample accurately and reliable. The results of the 10-fold CV show that the SVM model-based ensemble classifier effect is better overall than the XGB model-based ensemble classifier effect and the fitting effect of individual models.

For future work, the methodology studied in this paper will be extended to a comprehensive evaluation of college students' physical fitness as well as a recommendation system. The use of the recommendation system will be able to quickly provide substantial advice on individual student fitness, helping college students to enhance their physical fitness and improve their development in all aspects of physical shape, fitness, and function.

ACKNOWLEDGEMENTS

This project is supported by the 2021 Open Project of the Training Center of the University Ideological and Political Work Team of the Ministry of Education (South China Normal University) (No. SCNUKFYB067)

REFERENCES

[1] Breiman, L. (2001). Random Forests. Machine Learning, 45, 5-32.

[2] Brownlee, J. (2016) Machine Learning Mastery with Python: Understand Your Data, Create Accurate Models, and Work Projects End-to-End. Machine Learning Mastery, San Francisco.

[3] Chawla, N.V., et.al, (2002) Smote: Synthetic Minority Over-Sampling Technique. Journal of Artificial Intelligence Research, 16, 321-3 57.

[4] Chen, T.Q, et.al, (2016) XGBoost: A Scalable Tree Boosting System. The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, August 2016, 785-794.

[5] Cramer, J.S. (2002) The Origins of Logistic Regression. Tinbergen Institute Working Paper.

[6] Dai Xia, et.al, 2012. Rethinking and Optimizing the Evaluation Effectiveness of the National Standard for Student Physical Fitness - Construction of an Early Warning Mechanism for Student Physical Fitness. China Sports Technology. 3, 75-82.

[7] Dina Elreedy,et.al, 2019.A comprehensive analysis of synthetic minority oversampling technique (SMOTE) for handling class imbalance. Elsevier :Information Sciences, 32-64.

[8] Du Yunmei, et.al, 2018. Application of the plain Bayesian classification algorithm to the analysis of

university students' physical fitness. Journal of Sport. 25, 117-121.

[9] Guo Ruifan, 2019. A comparative study on the physical fitness assessment systems of Chinese and foreign adolescents. China Sports Technology. 55, 3-13.

[10] Hao Linlin, 2021. A K-Means clustering and BP neural network based model for evaluating the physical fitness type of university students. Contemporary Sports Technology. 11, 1-8.

[11] Kohavi, R. et.al, (2002) Decision Tree Discovery. In: Klosgen, W. and Zytkow, J.M., Eds., Handbook of Data Mining and Knowledge Discovery, Oxford University Press, New York, 267-276

[12] Kou Lei., et.al, 2021. A multidimensional data fusion-based early warning model for university students' physical fitness. Occupation and Health. 37, 92.

[13] Leonard, T. (1999). Bayesian methods: An analysis for statisticians and interdisciplinary researchers. Cambridge, U.K: Cambridge University Press.

[14] Liu Jianqiang, 2021. A survey and study on university students' awareness of the policy and attitudes towards testing the Student Physical Fitness Standards. Science and Wealth. 13, 156.

[15] MacQueen, J.B. (1967). Some Methods for Classification and Analysis of Multivariate Observations. In: Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics, University of California Press, Berkeley, 281-297.

[16] Meng Xianlu, et.al, 2021. Monitoring and intervention: Construction of a physical health development evaluation system for students. Chinese Journal of Education.

[17] Valiant, L. G. (1984). A theory of the learnable. Communications of the ACM, 1984,1134-1142.

[18] Vapnik, V.N. (1964). A note one class of perceptrons. Automation and Remote Control.

[19] Wang Jing. 2017. Machine learning based prediction and analysis of college sports performance. Modern Electronic Technology. 40, 116 -1 19.

[20] Wang Qian, et.al, 2012. A review of K-Means clustering algorithm research. Electronic Design Engineering. 20, 21-24.

[21] Wang shuangcheng, 2015. Gaussian kernel function-based dependency extensions for plain Bayesian classifiers. Control and Decision-Making. 30, 2280-2284.