*Advances in Social Science, Education and Humanities Research, volume 450*
**Proceedings of the 2020 International Conference on Advanced Education, Management and Information Technology (AEMIT 2020)**

# Research on Prediction of College Students' Performance Based on Support Vector Machine

Peng Wang*, Yinshan Jia

*School of Computer and Communication Engineering, Liaoning Shihua University, Fushun 113001, China*
**Corresponding author: 643397927@qq.com*

**ABSTRACT**
Using college entrance examination results to make accurate predictions of university course results can help students determine their efforts and help colleges and universities take effective measures to improve teaching quality. Based on the 2016 undergraduate college entrance examination results and university course scores of a general undergraduate college in China, a support vector machine was used to establish a college course performance prediction model, and cross-validation methods were used to obtain the best parameters and a reliable and stable model. Finally, in 2017 the model was applied to the college of computer science and technology major and communication engineering major performance forecast, the prediction accuracy rate reached 73.6%. The prediction results show that the support vector machine can accurately predict college course performance based on the college entrance examination results.
***Keywords:*** *Support vector machine, college entrance examination results, professional course results, prediction, cross validation*

## 1. INTRODUCTION

Learning is the job of each student, and academic performance is a staged evaluation of the student's learning situation, and it is also a major basis for evaluating teaching quality. With the rapid development of data science, more and more educational researchers have begun to focus on the analysis and mining of educational data. The purpose is to use the results of mining to help teachers improve teaching methods, help students improve the learning process, and help education managers optimize management decisions. Because there are certain differences between students, there are also certain differences in the situation of learning. If a student's college entrance examination results can be used to make predictions about the student's future college course results, the student can make a learning plan suitable for himself before or immediately after entering the school; teachers can also use the individual and overall forecast results, Appropriately change teaching strategies, improve teaching methods and teaching methods; the head teacher can strengthen individual or overall academic guidance based on the predicted results; decision-makers in the school's teaching management can adopt graded teaching and other methods based on the predicted results.

The earliest performance forecasting uses a manual-based forecasting method. Teachers or research experts manually collect data and estimate performance based on experience. Such methods are not only complicated to collect, but also time consuming. Later, prediction methods based on mathematical statistics appeared, such as least squares regression, gray models, etc.[1-2]. Such methods use mathematical modeling to predict student performance. However, the predictive ability of such methods for non-linear data is unsatisfactory, and it cannot well characterize the changes in student performance. In recent years, prediction methods based on machine learning have developed rapidly, such as Bayesian networks, support vector machines, and neural networks. This type of method has strong non-linear modeling capabilities [3-4].

Support vector machine is a new type of machine learning algorithm. It adopts the principle of structural risk minimization. Unlike the large sample requirements of neural networks, good generalization ability can still be obtained under small sample conditions. This paper uses support vector machines to build a college grade prediction model to the final grades of eleven courses such as advanced mathematics, university physics, compilation principles, and fiber optic communication for undergraduates majoring in computer science and technology at a university level of 16, and use cross-validation to find the best model parameters. The results show that the support vector machine improves the deficiencies of the general traditional model, and can achieve a more accurate prediction of college performance based on the college entrance examination results.

## 2. PREDICTION MODEL BASED ON MACHINE LEARNING ALGORITHM

### 2.1. Score Data and Preprocessing

The 2016 college computer science and technology and communication engineering students of a university were selected as the experimental data. Predict the results of the seven professional courses of 112 computer science and technology students and the four professional courses of 81 communication engineering students. The sample distribution of the original grades is shown in Figure 1. By normalizing the original data, the grades are divided into excellent (85-100), good (70-85), passing (60-70), and failing according to the score line (0-60). Four grades are used as the prediction category, and then a ten-fold cross-validation is applied to the sample to establish a college student performance prediction model.

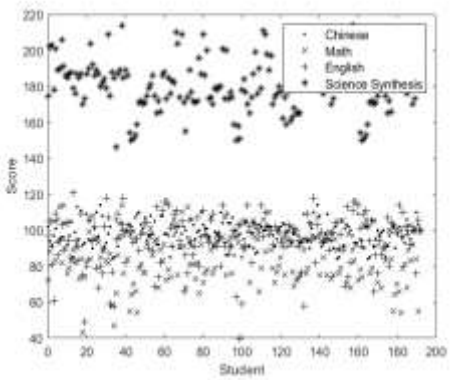


**Figure 1** Grade samples

### 2.2. Machine Learning Algorithms-Support Vector Machines

Support Vector Machines (SVM) is a machine learning algorithm for classification and prediction. It is a type of generalized linear classifier that performs binary classification of data by supervised learning. The maximum margin of the hyperplane[5-7]. SVM can perform non-linear classification through kernel methods, which is one of the common kernel learning methods[8].

SVM was proposed in 1964, and it developed rapidly after the 1990s and derived a series of improved and extended algorithms, which have been applied to pattern recognition problems such as portrait recognition and text classification[9-10].

Compared with neural networks and other traditional machine learning algorithms, support vector machines have fewer restrictions and no "overfitting" defects, which is very suitable for modeling and predicting the performance of small samples and nonlinear college students.

### 2.3. Model Training

Let university student grade sample set X be:

$$X=\{(x_1,y_1),(x_2,y_2),...,(x_n,y_n)\},\ i=1,2,...,n \qquad (1)$$

where i represents the ith sample, and the regression method of the support vector machine is specifically:

$$f(x)=w\phi(x)+b \qquad (2)$$

In the formula, both are parameters of support vector machine.

In order to find the most reasonable values of w and b, according to the principle of structural risk minimization, it is transformed into the following form:

$$\min \frac{1}{2}\|w\|^2+C\frac{1}{k}\sum_{1}^{n}\varepsilon(f(x_i)-y_i)$$
$$s.t.\quad \varepsilon(f(x_i)-y_i)=\begin{cases}|f(x_i)-y_i|-\varepsilon, & |\omega\cdot\phi(x)+b-y_i|\geq\varepsilon\\ 0, & |\omega\cdot\phi(x)+b-y_i|<\varepsilon\end{cases} \qquad (3)$$

where: C is the penalty parameter of the error, and the best value is found by the cross-validatio method, $\varepsilon(f(x_i)-y_i)$ is the regression error.

In order to simplify the solution process and reduce the computational complexity of modeling, the relaxation factors $\varepsilon$ and $\varepsilon^*$ are introduced:

$$\min_{w,b,\xi_i,\xi_i^*}\frac{1}{2}\|\omega\|^2+C\sum_{i=1}^{l}(\xi_i+\xi_i^*)$$
$$s.t.\begin{cases}y_i-w\cdot\phi(x)-b\leq\varepsilon+\xi_i, & \varepsilon_i\geq 0;\quad i=1,2,...,n\\ w\cdot\phi(x)+b-y_i\leq\varepsilon+\xi_i^*, & \varepsilon_i^*\geq 0;\quad i=1,2,...,n\end{cases} \qquad (4)$$

Using Lagrange multipliers $a_i$ and $a_i^*$ to further transform equation (4), we get:

$$\min_{a^{(*)}\in R^{2l}}\frac{1}{2}\sum_{i,j=1}^{n}(a_i^*-a_i)(a_j^*-a_j)k(x_i,x_j)+\varepsilon\sum_{i=1}^{n}(a_i^*-a_i)-\sum_{i=1}^{n}y_i(a_i^*-a_i) \qquad (5)$$

In the formula, $k(x_i,y_j)$ represents a kernel function.

The regression function of the support vector machine can be described as:

$$f(x)=\sum_{i=1}^{n}(a_i-a_i^*)(\phi(X_i),\phi(X))+b \qquad (6)$$

Select the RBF function as the kernel function, which is defined as:

$$k(x_i,x_j)=\exp\left(-\frac{\|x_i-x_j\|^2}{2\sigma^2}\right) \qquad (7)$$

where σ represents the parameter of RBF.

## 3. EXPERIMENT AND ANALYSIS

### *3.1. Experimental Results*

By using the support vector machine prediction model, a sample of 112 students from computer science and technology level 16 is used as the training set, a sample of 45 students from computer grade 17 is used as the test set; and a sample of 81 students from computer engineering major 16 is used as the training set. , Forty-five student samples at level 17 are used as test sets for classification and prediction. Each student 's college entrance examination scores correspond to each subject 's professional course scores. During model training, a professional course corresponds to a training model. Each model Corresponding to the prediction of the performance of each major course, according to the prediction of the performance of each major course by each model, the prediction categories of the scores of different major courses are obtained, and the accuracy of grade prediction of advanced mathematics courses is shown in the figure in Figure 2.The ordinates 1, 2, 3, and 4 represent excellent, good, passing, and failing grades respectively. If the sample data in the test set falls within a certain level interval, the grade sample is predicted to be that level, and the predicted level is not the same as the true level If they match, the prediction fails.

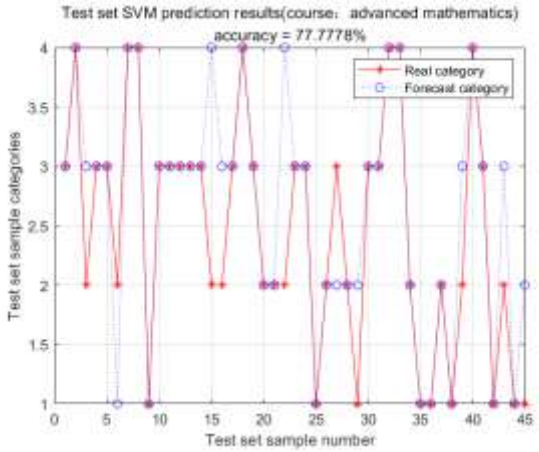


**Figure 2** Forecast results of advanced mathematics courses

### *3.2. Experimental Results*

The prediction accuracy of each course is shown in Table 1. According to the prediction accuracy, it can be found that the average prediction accuracy is 73.6%, which indicates that the model is more versatile and can be used in actual college performance prediction.

**Table 1** Statistics of college students' performance predictions

| Course | Prediction accuracy |
|---|---|
| advanced mathematics | 77.8% |
| college physics | 73.3% |
| discrete mathematics | 73.3% |
| compilation principle | 80% |
| computer interface | 75.6% |
| database principle | 68.9% |
| embedded systems | 63.3% |
| communication principle | 75.6% |
| optical fiber communication | 68.9% |
| digital circuit and logic design | 82.2% |
| communication electronics | 71.1% |

## 4. CONCLUSION

This paper collects the college entrance examination scores of college students and the course score information after admission as research objects, and uses the support vector machine to construct a prediction model. By comparing the actual scores and predicted scores of multiple courses of different majors, the average forecast accuracy is 73.6%. It can be seen that the method adopted in this paper is effective.

## ACKNOWLEDGMENT

Subject of Liaoning province's 13th five-year plan for education science 2017 (JG17DB303).

## REFERENCES

[1] SF Ding, BJ Qi, HY Tan. Review of Support Vector Machine Theory and Algorithms[J]. Journal of University of Electronic Science and Technology of China, 2011, 40(1):2-10.

[2] SQ Li, DH Sun. Review of Crossover Operators in Genetic Algorithms[J]. Computer Engineering and Applications, 2012, 48(1):40-43.

[3] J Li, LL Xu. Comparison and analysis of research hotspot trend prediction models based on machine learning algorithms[J]. Modern Intelligence, 2019, 39(04):24-34.

[4] YQ Zhang. Evaluation Model of Teaching Quality Based on Active Learning Support Vector Machine[J]. Modern Electronic Technology, 2019, 42(07):120-122.

[5] Vapnik V, Vapnik V N. Statistical Learning Theory[J]. 2003, 55(2):371-389.

[6] ZH Zhou. Machine learning. Tsinghua University Press, Beijing, 2016, pp.121-139, 298-300.

[7] H Li. Statistical learning methods. Tsinghua University Press, Beijing, 2012, pp.95-135.

[8] Hsieh, W.W.. Machine learning methods in the environmental sciences: Neural networks and kernels: Cambridge university press, 2009: Chapter 7, pp.157-169.

[9] Sang-Ki Kim, Youn Jung Park, Kar-Ann Toh. SVM-based feature extraction for face recognition[J]. Pattern Recognition, 43(8):2871-2881.

[10] Shan C. Research of Support Vector Machine in Text Classification[M]. Future Computer, Communication, Control and Automation. 2012.