Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 15, No. 1, 2026

Research on Student Performance Prediction based on BKA-BiLSTM Algorithm

Ling Chen

School of Computing Science, Wuhu University, Wuhu Anhui, 241008, China

Abstract: Based on the online resources and online data of Object Procedural Programming (OPP) course, this paper proposed the course dataset of students (Dataset of OPP: DOPP), and conducts a prediction study on students' final exam scores through this dataset. The hybrid algorithm of Black Kite Algorithm (BKA) and BiLSTM: BKA-BiLSTM algorithm model, is used for prediction, which can not only capture the data upstream and downward information, but also optimize the hyperparameters of the model through BKA. The results show that BKA-BiLSTM has better results in evaluation index than other regression analysis algorithms for the dataset DOPP of this course, and has good practical application value.

Keywords: Performance Prediction; BKA; BiLSTM; Deep Learning.

1. Introduction

In the process of students' online systematic learning, the gradually accumulated teaching process data has attracted widespread attention from scholars, especially the process performance of students on teaching platforms such as MOOC, Rain Classroom, and Learning Pass, which has become the basis for scholars to predict students' performance, learning performance, and learning behavior. There are many methods for predicting student performance, performance, and behavior based on online platform data, which can be mainly divided into: traditional regression prediction methods, such as linear regression (LR); machine learning algorithms such as Random Forest (RF); Deep learning algorithms, such as recurrent neural networks (RNNs), long short-term memory networks (LSTM), and bidirectional long short-term memory networks (BiLSTM) [1]. This study is mainly based on the procedural online data of students in the process programming course, and uses the BiLSTM algorithm to predict students' performance, and compares the prediction results with other algorithm models, which can achieve better evaluation performance. Specifically, the main work of this paper is as follows:

(1) Dataset: The student data in the process programming course is exported, sorted, and preprocessed, and the experimental student dataset is obtained, which is expressed as DOPP.

(2) BKA-BiLSTM prediction algorithm: Combined with different prediction algorithms, experiments are carried out on the dataset DOPP, and according to the results, it is concluded that BKA-BiLSTM can get the best evaluation effect in the prediction method used, so BKA-BiLSTM is selected as the final implementation algorithm of this paper. Combined with the implementation mechanism of BKA-BiLSTM, the block diagram and principle of the overall implementation of this paper are drawn, and the theoretical introduction of BKA-BiLSTM is given in detail.

2. Related Works

A custom dataset composed of eight data, including course audio and video, chapter tests, discussions, assignments, exams, check-ins, classroom interactions, and comprehensive scores, was used to predict final grades using the Transformer model, and hyperparameters were optimized to pay attention to and supervise students with a prediction grade of D[2]. A decision-level fusion performance prediction method based on Gaussian regression and partial least squares is proposed on the datasets of students in seven majors of a school, which provides support for the improvement of teaching methods [3]. Similarly, the partial least squares method is used for performance prediction, which is based on the performance dataset of students in a certain major in Literature [4], considering the correlation of different courses. BP neural network is used for performance prediction based on the dataset composed of existing student grades and other student information [5]. The academic performance of students in a certain school is used as the experimental dataset, and the model is optimized by using the Stacking model combined with Bayesian hyperparameter optimization [6]. Based on the OULAD public dataset of the Open University in the United Kingdom, the DA-GRU-CNN algorithm model with dual attention mechanism was used to predict the results[7]. Based on the performance dataset of analytical geometry and advanced algebra courses of a school, the performance prediction method is combined with multiple linear regression for feature selection[8]. Uses the dataset of a middle school student provided by an education bureau to predict student performance based on the multi-layer feature fusion method of LSTM and attention mechanism, extracts information from the two dimensions of time and course, and realizes the complementary mechanism based on the information of similar students to solve the problem of insufficient early training data [9]. Used the online programming system performance dataset to predict student performance based on graph convolution and similarity graph[10]. Different research methods have achieved certain results in their respective research topics, and this paper combines the existing BiLSTM algorithm with the recently proposed BKA algorithm to achieve the research of student performance prediction.

3. Basic Principles

3.1. BiLSTM Model Framework

The network building block of BiLSTM is shown in Figure

1. The network unit is composed of forward and reverse LSTM neural networks, which can realize forward and reverse LSTM training of time series, which effectively improves the understanding ability of the network in feature selection. The formula for calculating the mechanism of BiLSTM:

$$\boldsymbol{h}_t^{(1)} = f(\boldsymbol{W}_x^{(1)}\boldsymbol{x}_t + \boldsymbol{W}_h^{(1)}\boldsymbol{h}_{t-1}^{(1)} + \boldsymbol{b}^{(1)}) \tag{1}$$

$$\boldsymbol{h}_t^{(2)} = f(\boldsymbol{W}_x^{(2)}\boldsymbol{x}_t + \boldsymbol{W}_h^{(2)}\boldsymbol{h}_{t-1}^{(2)} + \boldsymbol{b}^{(2)}) \tag{2}$$

$$h_t = h_t^{(1)} \oplus h_t^{(2)} \tag{3}$$

$h_t^{(1)}$ represents the output of the hidden layer forward at the time t $h_t^{(2)}$ represents the output of the reverse hidden layer BiLSTM sends the input data twice in the opposite direction□ $h_t^{(1)}$ depends on the hidden state at the previous moment $h_{t-1}^{(1)}$ and the input of the current moment $\boldsymbol{x}_t$, $h_t^{(2)}$ depends on the hidden state at the next moment $h_{t+1}^{(1)}$ and the input for the current moment $\boldsymbol{x}_t$ □ $\boldsymbol{W}_x$ □ $\boldsymbol{W}_h$ are weight matrix represent a bidirectional output connection.

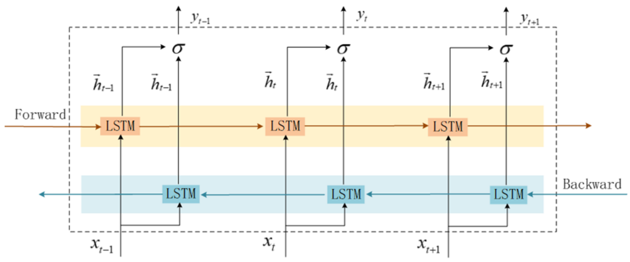


Fig 1. Network building blocks of BiLSTM

3.2. BKA black-winged Kite Algorithm

The mathematical model design of the black-winged kite algorithm BKA [11], a metaheuristic optimization algorithm released in 2024, is reflected in three aspects: initialization stage, predation behavior, and migration behavior. The flow block diagram of the BKA algorithm is shown in Figure 2.

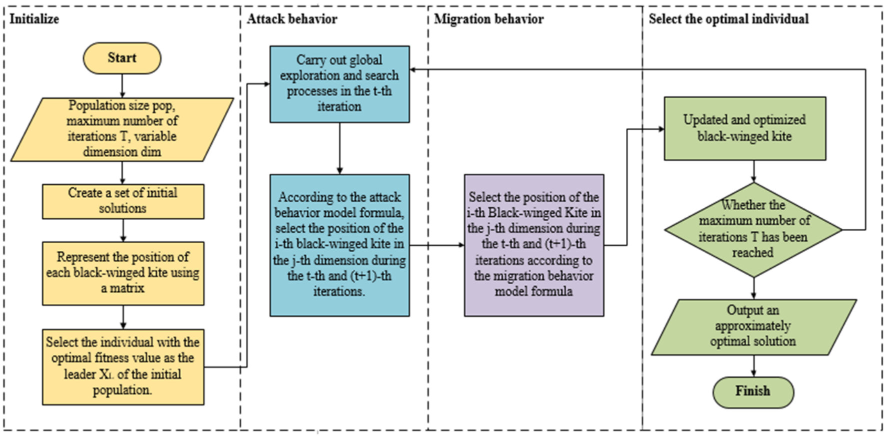


Fig 2. Block diagram of BKA algorithm flow

During the initialization phase, a set of initialization resolutions is first created and a matrix represents the position of each black-winged kite:

$$BK = \begin{bmatrix} BK_{1,1} & BK_{1,2} & \cdots & BK_{1,\dim} \\ BK_{2,1} & BK_{2,2} & \cdots & BK_{2,\dim} \\ \vdots & \vdots & \vdots & \vdots \\ BK_{pop,1} & BK_{pop,2} & \cdots & BK_{pop,\dim} \end{bmatrix} \tag{4}$$

Where pop is the number of potential solutions, dim is the dimension of the given problem, BK_{ij} is the j-dimension of the ith black-winged kite, and the position of each black-winged kite is evenly distributed:

$$X_i = BK_{lb} + rand(BK_{ub} - BK_{lb}) \tag{5}$$

Where ,$i \in [1, pop]$ BK_{lb} and BK_{ub} are the upper and lower limits of the ith black-winged kite in the j-dimension, and rand is the random value between [0,1]. During the initialization process, BKA selects the individual with the best fitness value as the leader in the initial population X_L, which is mathematically represented as:

$$f_{best} = \min(f(X_i)) \tag{6}$$

$$X_L = X(find(f_{best} == f(X_i))) \tag{7}$$

In the attack behavior stage, the black-winged kite adjusts the angle of its wings and tail according to the wind speed in flight, hovers to observe the prey, and then quickly dives to attack, this strategy includes different attack behaviors during global exploration and search, and its mathematical model is expressed as:

$$y_t^{i,j} = \begin{cases} y_t^{i,j} + n(1+\sin(r)) \times y_t^{i,j}, p < r \\ y_t^{i,j} + n(2r-1) \times y_t^{i,j}, \text{others} \end{cases} \tag{8}$$

$$n = 0.05 \times e^{-2\times(\frac{t}{T})^2} \tag{9}$$

Where $y_t^{i,j}$ and $y_{t+1}^{i,j}$ represent the position of the ith black-winged kite in the t and t+1 iterations, respectively, r is the random value between [0,1], p is the constant 0.9, T is the total number of iterations, and t represents the number of iterations that have been completed so far.

In the stage of migration behavior, based on the assumption of bird migration, the mathematical model of migration behavior is abstracted, and its expression is:

$$y_{t+1}^{i,j} = \begin{cases} y_t^{i,j} + C(0,1) \times (y_t^{i,j} - L_t^j), F_i < F_{ri} \\ y_t^{i,j} + C(0,1) \times (L_t^j - m \times y_t^{i,j}), \text{others} \end{cases} \tag{10}$$

$$m = 2 \times \sin(r + \pi / 2) \tag{11}$$

L_t^j is the leading scorer of the Black-Winged Kites until the t-round iteration $y_t^{i,j}$ and $y_{t+1}^{i,j}$ represent the position of the i-th black-winged kite in the t-round iteration and the t+1 iteration, respectively F_i represents the current position of any black-winged kite in the j-dimension in the t-round iteration F_{ri} represents the fitness value of the black-winged kite obtained from the random position of the j-dimension in the t-round iteration $C(0,1)$ represents Cauchy mutation.

Table 1. DOPP Dataset description

Classifications	Features	Characteristic description
Videos	Video task points	8 chapters of knowledge video introduction
Homework	Chapter 1 Homework	Chapter 1 Introduction
	Chapter 2 Homework	Chapter 2 Data Types, Operators, and Expressions
	Chapter 3 Homework	Chapter 3 Keyboard input and screen output
	Chapter 4 Homework	Chapter 4 Control Structure of Procedures
	Chapter 5 Homework	Chapter 5 Functions
	Chapter 6 Homework	Chapter 6 Arrays
	Chapter 7 Homework	Chapter 7 Pointer
	Chapter 8 Homework	Chapter 8 Structures
Experiments	Chapter 2 Experiment (1)	Chapter 2 Data Types, Operators, and Expressions (1)
	Chapter 2 Experiment (2)	Chapter 2 Data Types, Operators, and Expressions (2)
	Chapter 3 Experiment (1)	Chapter 3 Keyboard Input and Screen Output (1) Grades
	Chapter 3 Experiment (2)	Chapter 3 Keyboard Input and Screen Output (2) Grades
	Chapter 4 Experiment (1)	Chapter 4 Control Structure of Procedures (1) Grades
	Chapter 4 Experiment (2)	Chapter 4 Control Structure of Procedures (2) Grades
	Chapter 5 Experiments	Chapter 5 Functions Experimental results
	Chapter 6 Experiments	Chapter 6 Array Experimental results
	Chapter 7 Experiments	Chapter 7 Pointer Experimental results
	Chapter 8 Experiments	Chapter 8 Structure Experimental Results
Midterm	Midterm results	100 points, midterm exam scores
Classroom Questions	Classroom questions	Percentage system, the average score of actual classroom questions
Prediction goal	Final results	Percentage system, final exam scores

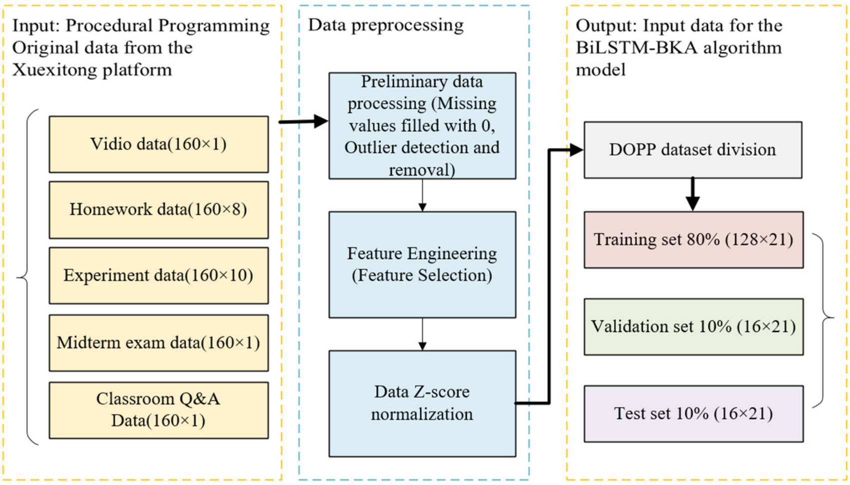


Fig 3. Data preparation and data preprocessing flow chart

4. Dataset DOPP and Model Framework

4.1. Dataset DOPP

In this paper, the data comes from the online data of the learning pass for process programming courses, and the data is preprocessed to obtain the dataset DOPP. The actual performance data of 160 students in a certain grade and major were selected from the process programming course, with a total of 21 grade characteristics and 1 target final score, with a total of 3520 data. The description of the dataset DOPP is shown in Table 1. The dataset DOPP includes classification, special proof, and feature description.

4.2. Data Preparation and Data Preprocessing

Through the Learning Platform, the relevant raw data for the process programming course is exported. Due to the lack of data, data abnormalities, and the completion of video task points need to be converted into fractions to record, some data cannot be directly used in the performance prediction algorithm, so the original data needs to be preprocessed. The data preparation and data preprocessing process is shown in Figure 3.

The main process of data preprocessing involves five aspects: 0 points are counted for missing values (if students do not submit relevant assignments, experiments or exams); For the detected outliers, if the score is not within the range of 0~100 points, the sample is deleted; For the data of video task points, since there are a total of 146 task points in the video, it is calculated according to the number of students who have completed it, for example, if a student completes 73 video task points, then 73/146 = 0.5, and the completion rate is 50%, then the eigenvalue data of the student's video is recorded as 50; feature selection through feature engineering; Standardize the data for ZScore.

The preprocessed data is named the dataset DOPP as the input data for the algorithm for predicting the final exam results. The data dimension of the DOPP dataset is 160×21, and in order to complete the whole process of training, verification and testing of the algorithm, it is divided into three parts: training set, validation set and test set, and the division ratio is 8:1:1, so the training set dimension is 128×21, the validation set dimension is 16×21, and the test set dimension is 16×21.

After data preparation and preprocessing, three data sets are generated, namely training set, validation set and test set, which are combined as input data for the framework process of BKA-BiLSTM algorithm in subsection 3.3.

4.3. BKA-BiLSTM Algorithmic Model Framework

The flow chart of the model framework of the BKA-BiLSTM algorithm is shown in Fig. 4. The framework uses the training set (128×21), validation set (16×21), and test set (16×21) divided into DOPP datasets as inputs, and constructs the model in three processes: The ModelTrain model is trained based on the training set data; The network model ModelValid is obtained by optimizing the network model hyperparameters of ModelTrain through the validation set. The model hyperparameters of ModelValid are optimized by the black-winged kite algorithm BKA to obtain the final model ModelTerminal.

The first process of constructing the model is described in detail as follows: taking the training set as input, using the BiLSTM layer to extract the two-way information of the input data, activating the operation through the ReLu layer, combining the Dropout layer to prevent overfitting of the prediction model, and then integrating the feature information through the FC layer of the fully connected layer to obtain the prediction data, and finally calculating the predicted value and the real value through the regression layer Regression Layer to obtain the loss value.

The second process of constructing the model is described in detail as follows: the ModelTrain model is first optimized by the validation set, and the process implementation idea is consistent with the first process, and the ModelValid model is obtained.

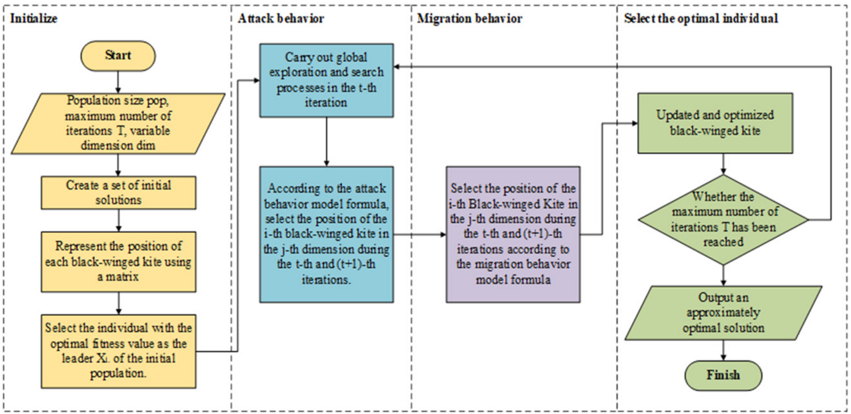


Fig 4. Flow chart of the BKA-BiLSTM algorithm model framework

The second process of constructing the model is described in detail as follows: In the process of being optimized by BKA, the ModelValid model first initializes the parameters of the BKA algorithm, including the number of populations p and the maximum number of iterations T, and then initializes the population matrix BK of BKA and starts iteration. Simulate the attack behavior of black-winged kites and optimize the position values of all individuals in the population. Simulate the migration behavior of black-winged kites and optimize the position values of all individuals in the population. Finally, the optimized black-winged kite is updated, so as to update the hyperparameters of the BKA-BiLSTM algorithm to obtain the final model ModelTerminal.

The final constructed ModelTerminal model is used to predict and process the test set data, compare and analyze the predicted value with the real value of the test set, calculate the evaluation index, visualize the results, and output the data.

5. Experimental Evaluation Indicators, Results and Analysis

5.1. Evaluate the Indicators

In order to verify the performance of different algorithms in predicting student performance tasks, five indicators were used to evaluate them, namely: mean absolute error MAE, mean absolute percentage error MAPE, mean square error MSE, root mean square error RMSE, and coefficient of determination R^2.

MAE is expressed as an equation(12) where y_{ij} represents the true value $\hat{y}_{ij}$ Indicates the predicted value m and n represents the dimension of the data

$$MAE = \frac{1}{mn}\sum_{j=1}^{n}\sum_{i=1}^{m}\left|y_{ij} - \hat{y}_{ij}\right| \tag{12}$$

MAPE is used to measure the relative magnitude of the prediction error, representing the average percentage of error between the predicted and true values.

$$MAPE = \frac{MAE}{\left|\hat{y}_i\right|} \tag{13}$$

MSE represents the average of the squared difference between the predicted and true values, and it is more sensitive to larger errors.

$$MSE = \frac{1}{mn}\sum_{j=1}^{n}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2 \tag{14}$$

RMSE is the square root of MSE, which represents the error in the raw unit data of the data, making it easier to explain the model's prediction error.

$$RMSE = \sqrt{MSE} \tag{15}$$

$R^2$is the proportion of variance predictable from the independent variable in the dependent variable is measured in the range of [0,1], and the closer the value is to 1, the better the data fit of the model.

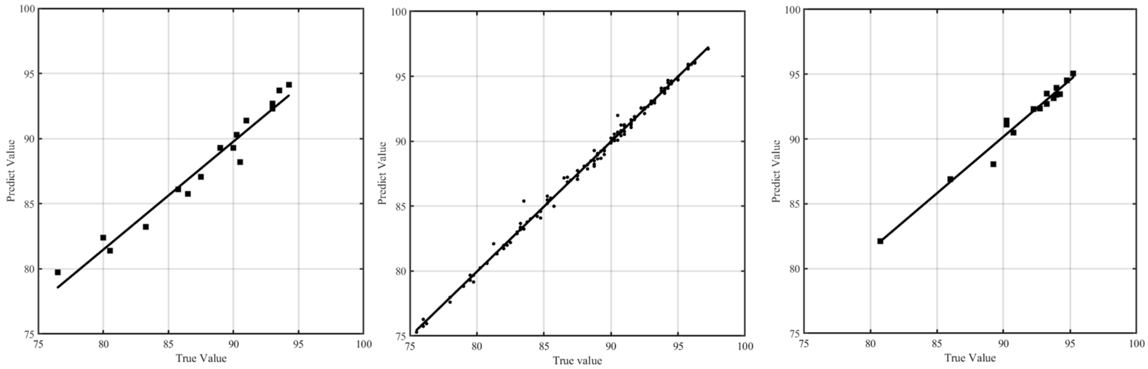
$$R^2 = 1 - \frac{mean(\sum_{i=1}^{m}(y_i - \hat{y}_i)^2)}{mean(\sum_{i=1}^{m}(y_i - \bar{y}_i)^2)} \tag{16}$$

5.2. Experimental Results and Analysis

Based on the DOPP dataset, seven algorithms, including random forest RF, multilayer perceptron MLP, support vector machine SVM, multiple linear (ML), LSTM, BiLSTM, and BKA-BiLSTM, were compared, and their respective evaluation indicators were compared and analyzed.

5.2.1. Fitting Curves

The fitting curves of the BKA-BiLSTM algorithm on the training, validation and test sets are shown in Fig. 5. From the fitting curve results, it can be seen that the prediction effect of the model on the training set, validation set and test set is good, and the predicted value and the true value are relatively close.



(a) Training set fitting curve (b) Validation set fitting curve (c) Test set fitting curve

Fig 5. Fitting curves of the training set, validation set, and test set

5.2.2. Error Contrast Curve

Based on seven different algorithms, namely RF, MLP, SVM, ML, LSTM, BiLSTM, and BKA-BiLSTM, the target final results are predicted on the test set, and the real values are compared with the predicted values. The comparison chart between the real value and the predicted value of the test set is shown in Fig. 6. In Figure 6, Real Value represents the true value of students' final exam scores, and the predicted values of the seven algorithms are reflected in different graphs, and the samples of each student's final performance test set are compared vertically. From the prediction results of the test set, the BKA-BiLSTM algorithm proposed in this paper is more accurate in predicting students' final exam scores, and the position points of the predicted value and the real value are closer. In contrast, the predicted and true values of random forest RF and multilayer perceptron MLP have larger errors and more offsets.

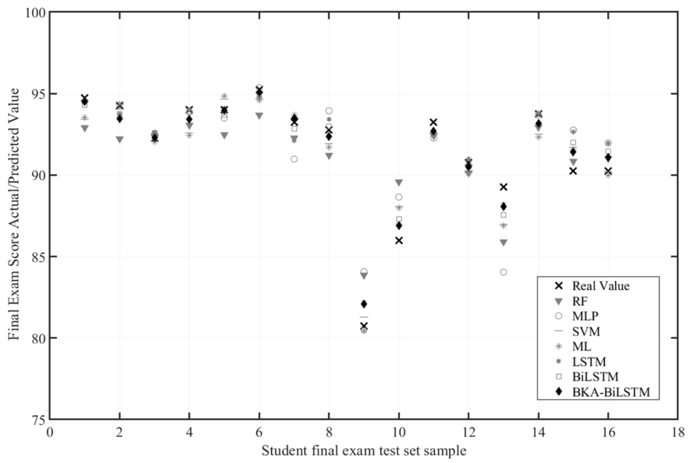


Fig 6. Comparison of the error between the real value and the predicted value of the test set

5.2.3. Comparison of Evaluation Indicator Data

The data results of the evaluation indicators of the seven different algorithms in the target end-of-term performance prediction process are shown in Table 2. Table 2 shows the seven algorithms based on the training set, validation set and test set of the process programming course, and the values of MAE, MAPE, RMSE, MSE and R^2 coefficients are counted. Ours in Table 2 is the BKA-BiLSTM algorithm.

Table 2 shows that in the training set, the MAE value reaches 0.232, the MAPE value reaches 0.003, the RMSE value reaches 0.121, the MSE value reaches 0.348, and the prediction results of the BKA-BiLSTM algorithm reach 0.232, the MAPE value reaches 0.003, the RMSE value reaches 0.121, the MSE value reaches 0.348, The R^2 coefficient value reaches 0.996, and the overall performance is better than that of the comparison algorithm, and the performance of MLP and BiLSTM is close in some indicators. In the validation set, the prediction results of the BKA-BiLSTM algorithm reach 0.821, the MAPE value reaches 0.010, the RMSE value reaches 1.538, and the MSE value reaches 1.240. The R^2 coefficient value reaches 0.942, which shows better performance than the comparison algorithm, and SVM and BiLSTM perform relatively close in some indicators. In the test set, the prediction results of the BKA-BiLSTM algorithm reach 0.583, MAPE value reaches 0.007, RMSE value reaches 0.505, and MSE value reaches 0.711. The R^2 coefficient value reaches 0.962, and the overall performance is better than that of the comparison algorithm, and the effect of LSTM and BiLSTM is relatively close in some indicators, while the random forest algorithm has the worst results compared with other algorithms in the table.

By analyzing and summarizing the data results, it can be concluded that the BKA-BiLSTM algorithm in this paper has better performance in predicting DOPP datasets, and can extract bidirectional information from the data through BiLSTM, and effectively optimize the parameters in combination with BKA.

Table 2. Data of evaluation indicators of different algorithms

Dataset	Algorithm	MAE	MAPE	RMSE	MSE	R^2
Training set	RF	1.079	0.013	2.062	1.436	0.923
	MLP	<u>0.396</u>	<u>0.005</u>	0.797	0.893	0.970
	SVM	0.992	0.012	1.964	1.402	0.927
	ML	1.029	0.012	1.773	1.332	0.934
	LSTM	0.502	0.006	0.512	0.716	0.981
	BiLSTM	0.397	<u>0.005</u>	<u>0.298</u>	<u>0.546</u>	<u>0.989</u>
	Ours	**0.232**	**0.003**	**0.121**	**0.348**	**0.996**
Validation set	RF	2.281	0.027	8.839	2.973	0.668
	MLP	1.911	0.022	8.626	2.937	0.676
	SVM	1.342	<u>0.015</u>	<u>4.187</u>	<u>2.046</u>	<u>0.843</u>
	ML	1.426	0.016	5.317	2.306	0.801
	LSTM	1.457	0.018	6.299	2.510	0.764
	BiLSTM	<u>1.272</u>	<u>0.015</u>	4.816	2.195	0.819
	Ours	**0.821**	**0.010**	**1.538**	**1.240**	**0.942**
Test set	RF	1.514	0.017	3.306	1.818	0.751
	MLP	1.336	0.015	3.896	1.974	0.707
	SVM	0.895	0.010	1.224	1.106	0.908
	ML	0.935	0.010	1.285	1.134	0.903
	LSTM	0.753	<u>0.008</u>	0.978	0.989	0.926
	BiLSTM	<u>0.612</u>	**0.007**	<u>0.678</u>	<u>0.823</u>	<u>0.949</u>
	Ours	**0.583**	**0.007**	**0.505**	**0.711**	**0.962**

Note: The best solution in Table 2 is marked in bold, and the second best solution is underlined

6. Conclusion

This study predicts students' performance based on the online data of the learning platform for process programming courses, and achieves better evaluation performance through the BKA-BiLSTM algorithm. In this paper, the online data of the corresponding courses and semesters on the Learning Tong platform are preprocessed and named as the DOPP dataset as the experimental dataset of this paper. Experimental data show that BKA-BiLSTM achieves better performance in five evaluation index dimensions, and the prediction accuracy and stability are higher. The effectiveness of the prediction of the BKA-BiLSTM algorithm on the DOPP dataset can help teachers understand students' learning in a timelier manner. In future research, we can try to form a large-scale dataset with the data of students in multiple grades and classes for further experimentation and exploration.

Acknowledgments

This work is supported by Research Project of Fundamentals Computing Science and AI Teaching Team for School-level Projects of Wuhu University Project Number: WHSZDW-202502.

References

[1] Tao Lu. Journal of Chizhou University,2020,34(03):27-31.

[2] Lü Hui, Shan Pengfei, Feng Mengmeng. Journal of Liaodong University (Natural Science Edition), 1-8[2025-02-18].

[3] Li Jinsong, Wang Na, Yao Minghai, et al. Research on Performance Prediction Method Based on Decision-Level Integration[J]. Information Technology,2024,(09):78-83.

[4] Gao Yilu, Liu Hongyan, Li Jinsong, et al. Research on student performance prediction based on partial least squares method[J].Information Technology,2024,(06):47-53.

[5] Chen Xixiang, Peng Wei. Prediction and Analysis Method of Student Performance Based on Neural Network[J].Automation Application,2024,65(17):24-26+31.

[6] Li Shipeng, Wei Gaoqi, Zhao Ziyan. Comprehensive Achievement Prediction Method for College Students Based on Multi-Model Fusion[J/OL]. Control Engineering,1-8[2025-02-18].

[7] Zhang Wenjuan, Zhang Bin, Yang Haozhe. Performance prediction based on dual attention mechanism[J].Journal of Nanjing Normal University(Natural Science Edition),2023, 46 (04):103-113.

[8] Liu Xiaoyun, Liu Hongyan, Li Jinsong, et al. Information Technology, 2023, (10):17-22.

[9] Liu Tong, Qi Huiran, Ni Weijian. Computer Engineering and Design,2023,44(10):2973-2978.

[10] Luo Wenjie, Xiao Ziliang. Computer Engineering and Design,2023,44(09):2769-2776.

[11] Wang J, Wang W, Hu X, et al. Black-winged kite algorithm: a nature-inspired meta-heuristic for solving benchmark functions and engineering problems[J]. Artificial Intelligence Review, 2024, 57(4): 98.