Received November 17, 2020, accepted November 30, 2020, date of publication December 3, 2020,
date of current version December 15, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3042180

An Enhanced Naive Bayes Model for Dissolved Oxygen Forecasting in Shellfish Aquaculture

DASHE LI, JIAJUN SUN, HUANHAI YANG, AND XUEYING WANG
School of Computer Science and Technology, Shandong Technology and Business University, Yantai 264005, China
Co-innovation Center of Shandong Colleges and Universities: Future Intelligent Computing, Yantai 264005, China

Corresponding author: Jiajun Sun (2019410040@sdtbu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61772319, Grant 61472227, and Grant 61602277; and in part by the CERNET Innovation Project under Grant NGII20180319.

ABSTRACT It is difficult to predict dissolved oxygen values because they are disordered and nonlinear. Accurate prediction of dissolved oxygen in shellfish aquaculture plays an important role in improving shellfish production, and a reliable model is needed to accurately predict dissolved oxygen values. Therefore, in this paper, an enhanced naive Bayes (NB) model is proposed. Due to the excessive number of different dissolved oxygen values, their direct use as input samples will result in overly few training set categories for each value, which reduces the prediction accuracy. Therefore, the dissolved oxygen differential series dataset is used as the input data to reduce the number of training set categories and improve the training accuracy. To increase the number of samples in the training set, the sliding window concept from network communication protocols is used to partition the differential sequence dataset and generate the features and labels of the training set. The values were predicted as categories, and the dissolved oxygen data were accurately predicted by selecting the labels that correspond to the posterior probability maxima of all training samples. Finally, the algorithm is used to predict the dissolved oxygen data from February 18, 2016, to January 31, 2020, in Yantai, Shandong Province, China. The dissolved oxygen data of a shellfish farm were trained and predicted, and the best values of the feature lengths were optimized by analyzing their effects on the predicted dissolved oxygen values. The proposed algorithm has significantly improved the mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE) compared to the advanced algorithms. The results of the Diebold-Mariano test and 10-fold cross-validation also show that the proposed algorithm has a higher prediction accuracy.

INDEX TERMS Naive Bayes, dissolved oxygen, sliding window, time series, differential sequence.

I. INTRODUCTION

The numerical prediction of dissolved oxygen in water bodies has been extensively studied by scholars. Dissolved oxygen data are nonlinear, cyclical, and nonstationary in nature. Ahmed [1] combined a feedforward neural network (FFNN) and radial basis function neural network (RBFNN) to evaluate and predict the dissolved oxygen parameters in the Surma River. Ji *et al.* [2] designed a model based on a support vector machine (SVM) to predict dissolved oxygen in anoxic river systems. Raheli *et al.* [3] used a multilayer perception integrated with the firefly algorithm (MLP-FFA) model to predict water quality parameters collected at a Malaysian hydrological station. Huan *et al.* [4] combined the ensemble empirical mode decomposition (EEMD) and a least-square support vector machine (LSSVM) to predict the dissolved oxygen sequences. Li *et al.* [5] proposed a hybrid model of multiscale features based on EEMD and used it for dissolved oxygen prediction in aquaculture. Ren *et al.* [6] used a genetic algorithm-optimized fuzzy neural network for the hydroponic system prediction of dissolved oxygen. Although the abovementioned methods can better predict the dissolved oxygen indicator, they lack interpretability. Neural network-based learning algorithms have the problem of overfitting and underfitting, and the final result of the algorithm can easily fall into the local optimum, which cannot accurately predict the dissolved oxygen content changes in the context of practical applications in marine fisheries.

Bayes' equations provide a generative model for data classification from a statistical viewpoint [7]. On this foundation, relying on the assumption of strong independence, a naive Bayes (NB) algorithm is proposed, which shows

The associate editor coordinating the review of this manuscript and approving it for publication was Wentao Fan.

good results and stable classification efficiency in predicting class problems. The algorithm also has the advantage of robustness to missing data and low algorithm complexity. Saritas and Yasar [8] used the algorithm for breast cancer diagnosis. Granik and Mesyura [9] used the algorithm to identify fake news. In the multiclassification problem, Jiang *et al.* [10] classified Chinese texts into five categories and predicted new texts with a combination of NB equations, while Al-Khurayji and Sameh [11] predicted Arabic texts. Xu [12] discussed the differences in text classification performance among different event models compared to NB classifiers. Mubarok *et al.* [13] modeled product evaluations and classified user sentiments into four categories for prediction. Karthick and Harikumar [14] used an NB model to classify oral X-rays to predict different types of oral diseases. In the field of multiclassification, NB algorithms commonly predetermine the number of categories in advance and then make predictions, which leads to improved classification results. However, traditional NB algorithms cannot be used to solve continuous value prediction problems, and no literature has used an NB algorithm for dissolved oxygen prediction.

Based on the above analysis, the problems to solve are the low accuracy of the traditional algorithm for predicting dissolved oxygen and the inability of the NB algorithm to predict continuous values [15]–[21]. Thus, this paper proposes an enhanced NB model to predict dissolved oxygen in shellfish aquaculture based on previous studies. When the difference values of the dissolved oxygen parameter sequence are taken as classification categories and Laplacian correction is performed to correct the observed values in the differential sequence, the predicted dissolved oxygen values are closer to real values than previous models.

Predictions of water quality data, especially dissolved oxygen data, are often made by passing sensor-acquired values directly to predictive models as input parameters. The prediction accuracy is affected by the size of the historical dataset; therefore, choosing a suitable way to expand the training dataset can effectively improve the prediction accuracy. Moreover, the parameters of the algorithmic model are affected by the monitoring points. The selection of different regions will lead to large changes in the model parameters, and the selection of parameters will lead to changes in the prediction accuracy. Problems such as multiple parameters and complex tuning processes have plagued traditional methods. Furthermore, there is substantial randomness in the model building process of many algorithms. Therefore, an enhanced naive Bayesian prediction model is proposed, which simplifies the parameters and no longer necessitates additional parameters, except the sliding window length.

Traditional naive Bayesian algorithms are often used to handle classification problems with a small number of categories; however, predicting continuous values, such as dissolved oxygen values, is often not possible using the naive Bayesian algorithm. This is because the use of naive Bayes requires satisfying the need for a sufficiently large training set size for each predicted category, and a large number of predicted categories exist for continuous values. In this paper, by introducing the method of a sequence of difference values, the continuous values are transformed into a sequence of difference values, thus enabling the use of the naive Bayesian method. The contribution of this paper consists of the following three parts:

1) Continuous dissolved oxygen values are predicted by improving naive Bayesian algorithms. The traditional naive Bayesian algorithm can only classify a finite number of categories. In this paper, the prediction of continuous values by using the dissolved oxygen values as model input is achieved by using the naive Bayesian algorithm.

2) A method of differential series is proposed to enhance the regularity of the data samples. The prediction performance of the naive Bayesian algorithm depends on the selection of historical data samples. The traditional method is limited by the lack of regularity in the distribution of the number of samples per category, resulting in low prediction accuracy. In this paper, the differential series is taken as a categorical category for prediction. It decreases the number of categories and increases the number of studies per category, which enhances the regularity of the data sample distribution.

3) A sliding-window data generation method is proposed to increase the number of training samples. The traditional method directly divides the time series into two parts: the test set and the training set. Using this method will lead to too few training samples, but the data generation method with sliding window proposed can effectively increase the number of training samples and thus improve the prediction accuracy.

The remainder of the paper is organized as follows. In Section II, the derivation process of the NB method and Laplacian correction method are presented. In Section III, the methods in this paper are described in detail. In Section IV, an enhanced NB algorithm is used to predict the dissolved oxygen levels and verify the superiority of the algorithm in the paper in terms of prediction performance by comparing the errors with similar algorithms. We use Diebold-Mariano test and 10-fold cross-validation to verify the advantages of our algorithm. In Section V, the main ideas of the paper are summarized, and the outlook for future work is presented.

II. RELATED WORK

A. NAIVE BAYES

Bayes' theorem [22]–[25] converts the "probability of event Y occurring conditional on event X occurring" to the "probability of event X occurring conditional on event Y occurring", where $P(Y|X)$ is the posterior probability, $P(Y)$ is the prior probability, and $P(Y|X)/P(X)$ is the likelihood function, which can be considered an adjustment factor, as shown in equation (1).

$$\begin{aligned} P(Y|X) &= \frac{P(Y)\,P(X|Y)}{P(X)} \\ &= \frac{P(Y)\,P(X|Y)}{\sum_{c=1}^{N_Y} P(Y=y_c)\,P(X|Y=y_c)} \end{aligned} \tag{1}$$

In reality, many factors influence the events, and the known events are closely related to one another, so it is very difficult to find the conditional probability $P(Y|X)$ for all events in Bayes' theorem.

Based on Bayes' theorem, the interrelationship among known events is no longer considered. A strong independent constraint is added to the set of events X, and each event in the set is considered independent of the other events, which leads to the general form of the NB model (2).

$$\begin{aligned} P\left(X|Y=y_c\right) &= P(X_1=x_1, X_2=x_2, \cdots, X_N=x_N|Y=y_c) \\ &= \prod_{i=1}^{N} P(X_i=x_i|Y=y_c) \end{aligned} \tag{2}$$

Introducing this equation into a Bayes' theorem yields the NB discriminant, as shown in equation (3).

$$\begin{aligned} P\left(Y=y_c|X\right) &= \frac{P\left(Y=y_c\right)P\left(X|Y=y_c\right)}{P\left(X\right)} \\ &= \frac{P\left(Y=y_c\right)\prod_{i=1}^{N} P(X_i=x_i|Y=y_c)}{P\left(X\right)} \end{aligned} \tag{3}$$

For a sequence of features, $Feature = \{f_1, f_2, \cdots, f_L\}$, where f_i is the value of each attribute in the feature. For the same set of feature values, the denominator $P(X)$ is fixed. Therefore, in the actual calculation, the denominator is ignored, and the class with the largest numerator is directly selected as the predicted value based on the size of the numerator, as in equation (4).

$$\begin{aligned} y &= arg\max_{y_c} P(X_i=x_i|Y=y_c) \\ &= \prod_{i=1}^{N} P(X_i=x_i|Y=y_c)P\left(Y=y_c\right) \end{aligned} \tag{4}$$

B. LAPLACIAN CORRECTION

When predicting dissolved oxygen levels, because some values of the sequence of differences in the test set do not exist in the training set, the probability calculation makes $P(X_i = x_i|Y = y_c) = 0$, which makes $P(Y = y_c|X) = 0$, so the probability of the final occurrence of this property y_c affects the final prediction. Laplacian correction (5) was introduced to correct this effect [26]–[30]. In this equation, x_i refers to the observed value of the attribute in the test set. $|D_{c,x_i}|$ refers to the number of observations of the i-th attribute of the observation feature X_i equal to x_i when the predicted value Y is y_c. $|D|$ refers to the number of samples in the training set. N_{X_i} refers to the total number of possible values of the event X.

$$P(X_i=x_i|Y=y_c) = \frac{1+\left|D_{c,x_i}\right|}{|D|+N_{x_i}} \tag{5}$$

In the equation, x_i refers to the observed value of the attribute in the test set. $|D_{c,x_i}|$ refers to the number of observations of the i-th attribute of the observation feature X_i equal to x_i when the predicted value Y is y_c. $|D|$ refers to the number of samples in the training set. N_{X_i} refers to the total number of possible values of event X.

III. CONSTRUCTION OF THE ENHANCED NAIVE BAYES ALGORITHM

A. DIFFERENTIAL SEQUENCE

Dissolved oxygen values are easily affected by many natural factors such as the climate, season, altitude, and time [31]. Therefore, the overall dissolved oxygen data show nonlinear characteristics [32]. There will be large deviations from the dissolved oxygen data collected at different times, so it is necessary to preprocess these data and then make further algorithm predictions.

The most common method of data preprocessing is normalization [33]–[37], i.e., mapping the data to values of 0-1. However, normalization results in values with many decimal places, which cannot be classified into a limited number of categories by the NB algorithm. Therefore, normalization cannot be used with this algorithm; instead, the dissolved oxygen sequence is transformed into a differential sequence, and the differential sequence is predicted.

For a given dissolved oxygen sequence $S = s_1, s_2, \ldots, s_{N^{train}}$, the differential sequence $Diff$ is computed according to equation (6):

$$Diff_k^{train} = s_k^{train} - s_{k-1}^{train}, \quad k = 2, 3, \ldots, N^{train} \tag{6}$$

where N^{train} is the length of the dissolved oxygen sequence.

To further illustrate the effect of the method of differential series introduced in this paper on the distribution of the dissolved oxygen data series, the frequency distribution histograms of the original dissolved oxygen data series and the differential data series are plotted in Figure 1. Among them, the upper part of Figure 1 shows the distribution histogram of dissolved oxygen data, which is the value of dissolved oxygen, so there are only positive values; the lower part of

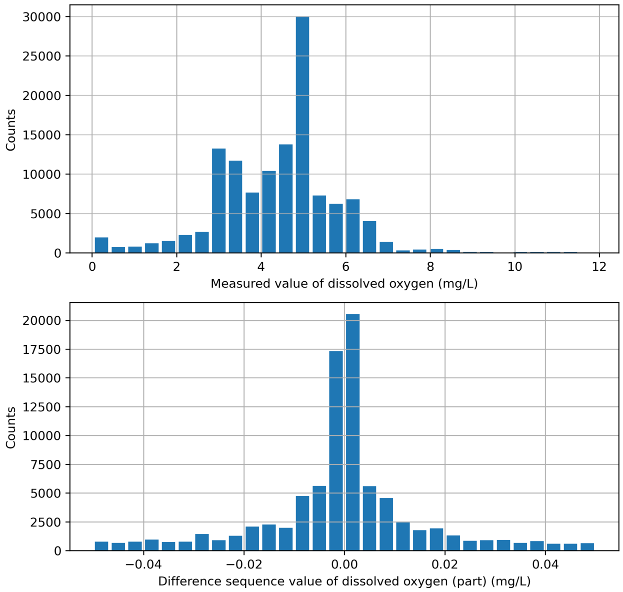


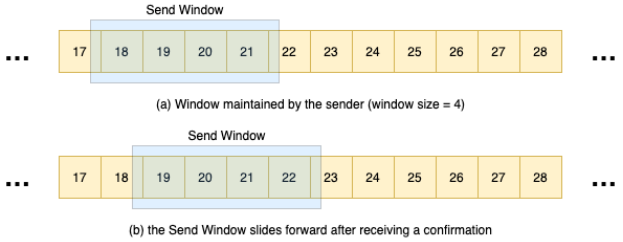
FIGURE 1. Histograms of the frequency distributions of the raw and differential dissolved oxygen data sequences.

Figure 1 is the histogram of dissolved oxygen difference calculated according to Equation (6), with positive and negative values (0 being in the middle).

By comparing the frequency distribution histograms of the raw and differential dissolved oxygen data sequences, we observe that the data after using the differential sequence have a smaller distribution range, which enables the use of the differential sequence as a classification category in the NB classifier to calculate the probabilities. Since the sampling period of the sensor is 10 minutes, the dissolved oxygen values do not significantly change between two adjacent samples, and the difference is concentrated near zero in the frequency distribution histogram. A possible reason for the large difference is that the acquisition is interrupted due to an unexpected power failure of the equipment in some periods, and the dissolved oxygen has greatly changed after restarting. To overcome the error caused by this factor and improve the prediction accuracy, only the data with an absolute value of the difference less than or equal to 0.01 were included in the model.

B. SLIDING WINDOW METHOD TO CONSTRUCT DATASETS

The sliding window technique [38]–[40] was originally a traffic control technique in computer network communication protocols. In the Transmission Control Protocol (TCP), two parties negotiate the size of the sliding window to determine the number of bytes of data sent. As shown in Figure 2, in the data transmission process, the window is constantly sliding backward; eventually, the entire data message is transmitted.



(a) Window maintained by the sender (window size = 4)

(b) the Send Window slides forward after receiving a confirmation

FIGURE 2. Sliding window diagram.

Using the idea of sliding windows, a new dataset can be generated by sliding through the data sequence. This method is commonly used to preprocess time series data. In this paper, the method is used to preprocess a sequence of dissolved oxygen differences. Referring to Figure 3, the specific method is to specify a sliding window size k and then slide backward from the starting position of the difference sequence with length n. The first $k-1$ dissolved oxygen difference records in the sliding window are used as features, the last dissolved oxygen difference record is used as a label, and the window is slid to the end of the difference sequence to generate $n-k$ data segments with features and labels. All data segments form the dissolved oxygen differential sequence dataset. Let $L = k-1$ and let L be the length of the selected feature. Selecting a larger L will cause the algorithm to focus on more features, but an excessively large L will make the model overfit. Selecting a smaller L will cause the algorithm to focus more on the mutated values of the dissolved oxygen difference sequence, but an overly small L will make the algorithm biased toward predicting the mutated values, which reduces the prediction accuracy. To improve the accuracy of the algorithm, an appropriate L must be selected.

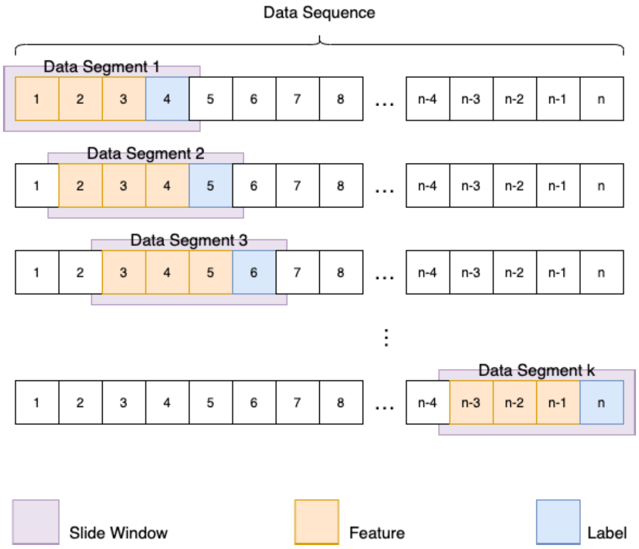


FIGURE 3. Sliding window method to generate data segments.

C. DESCRIPTION OF THE ALGORITHM

The length of the selected feature is specified as L when the dataset is generated using the sliding window method. The percentage of the dataset for use as the training set is $Rate_{train}$ percent. The specific steps of the enhanced NB prediction algorithm are as follows.

Step 1: Dissolved oxygen data preprocessing. First, the dissolved oxygen data sequence is converted into a differential sequence. Then, the first $Rate_{train}$ percent of records of the dissolved oxygen differential sequence dataset is partitioned as the training set, and the last $1 - Rate_{train}$ percent of the records is used as the test set. Finally, the sliding window method is used on the training set to generate data segments with features and labels, and the occurrence probability of each label is calculated.

Step 2: Construction of the enhanced NB prediction model. First, all values Y_i of the labels in the training set with absolute values less than or equal to 0.01 are taken as labels, and a feature space of length L is opened in the memory. The initialization of model space T is completed. Then, with label Y_i, the values of the first L feature elements of Y_i are saved to the feature space corresponding to Y_i in model space T.

Step 3: Iteration through each label in the feature space to find the occurrence probability of the corresponding label. First, a data segment with a feature and a label (i.e., the actual value) is selected from the test set using the sliding window method. Next, to obtain the predicted value of the dissolved oxygen difference sequence from the features, each label on model space T must be traversed. In the process of traversing each label, the number of features x_i of the data segment with an identical position and value in the feature space for the label in model space T is recorded as D_{c,x_i}. The number of species in the feature space that are equal to feature position i is denoted by N_x. The sum of the number of occurrences corresponding to all types of fetches at feature location i is denoted by D. According to equation (7), the conditional probability of that possible predicted value is calculated. Then, the number of occurrences of that label in the training set is recorded as D_c. The total number of labels in model space T is recorded as N, and the total number of data segments generated from the training set is recorded as D. The probability of that label appearing is calculated according to equation (8).

$$P(X_i = x_i|Y = y_c) = \frac{1 + |D_{c,x_i}|}{|D| + N_x} \tag{7}$$

$$P(Y = y_c) = \frac{1 + |D_c|}{|D| + N} \tag{8}$$

Step 4: Completion of the traversal of all labels to make predictions for the given data segment. Step 3 is repeated to calculate the probability of each label on the dissolved oxygen difference data segment in model space T according to equation (9). The label corresponding to the maximum probability is used as the prediction value of the selected data segment. Then, equation (10) is used to reduce the differential sequence to the original dissolved oxygen sequence. In this equation, s_{k-1}^{test} refers to the dissolved oxygen to obtain the predicted value of s_k^{test} for the dissolved oxygen at the next moment. N^{test} refers to the size of the training set.

$$\begin{aligned} P(Y = y_c|X_i = x_i) &= ln\,(P(Y = y_c))) \\ &\quad + \sum_{i=1}^{L} ln\,(P(X_i = x_i|Y = y_c)) \end{aligned} \tag{9}$$

$$s_k^{test} = Diff_k^{test} + s_{k-1}^{test}, \; k = 2, 3, \ldots, N^{test} \tag{10}$$

Step 5: Prediction effect evaluation. Steps 3 and 4 are repeated until the prediction of the test set is completed. The error function is used to evaluate the real and predicted sequences.

In the context of this application of marine pastures, this paper uses historical data combined with an enhanced NB algorithm for modeling. After the sensor reads new water parameters (such as the dissolved oxygen levels), the model can quickly assess the water condition, and abnormal conditions are given as timely feedback to experts for evaluation. The experts then take appropriate treatment measures to achieve risk avoidance and effectively reduce losses.

Algorithm 1 Algorithm of the Enhanced NB Model

```
Input: $DissolvedOxygenSequence$, $L$, $R_{train}$.
Output: $PredictSequence$, $Error$
    Initialization:
    $DO\_S \leftarrow DissolvedOxygenSequence$
    $R \leftarrow R_{train}$
    Dissolved oxygen data preprocessing:
    $DS \leftarrow$ calculate_differential_sequence($DO\_S$)
    $TrainData, TestData \leftarrow$ split_data($DS, R$)
    $P\_train \leftarrow$ calculate_probability($TrainData, L$)
    Construction of the enhanced NB model:
    $T \leftarrow \{\}$
    for $P\_train_i$ in $P\_train$ do
        if abs($P\_train_i$) <= 0.01 and $P\_train_i$ not in $T$ then
            T[$P\_train_i$] $\leftarrow$ [{} for _ in range($L$)]
        end if
    end for
    for $TrainData_i$ in $TrainData$,$T_i$ in $T$ do
        if $TrainData_i = T_i$ then
            $temp_i \leftarrow (TrainData_i).index - L$
            for $i = 0 \rightarrow L$ do
                $value \leftarrow TrainData[i]$
                $T[T_i][i][value] \leftarrow T[T_i][i][value] + 1$
                $temp_i \leftarrow temp_i + 1$
            end for
        end if
    end for
    Prediction by the enhanced NB model:
    $L_{test} \leftarrow$ length($TestData$)
    $L_{train} \leftarrow$ length($TrainData$)
    for $test_i = L \rightarrow L_{test}$ do
        $result = \{\}$
        for $T_i$ in $T$ do
            for $i = 0 \rightarrow L$ do
                $value = TestData[test_i - L + i]$
                $D_{c,xi} = T[T_i][i][value]$
                $D \leftarrow 0$
                for $temp$ in $T[T_i][i]$ do
                    $D \leftarrow D + T[T_i][i][temp]$
                end for
                $N_x \leftarrow$ len($T[T_i][i]$)
                $result[T_i] \leftarrow result[T_i] + \log\left(\frac{D_{c,xi}+1}{D+N_x}\right)$
            end for
            $D_c \leftarrow P\_train[T_i]$
            $N \leftarrow$ length($T$)
            $D \leftarrow L_{train} - L$
            $result[T_i] \leftarrow result[T_i] + \log\left(\frac{D_c+1}{D+N}\right)$
        end for
        $predictValue \leftarrow$max($result_list.values$).index
        $PredictData$.append($predictValue$)
    end for
    $Error \leftarrow$calculate_error($PredictData, TestData$)
    return $Predicts$, $Error$
```

IV. EXPERIMENTS

A. DATA DESCRIPTION

The dissolved oxygen in the water of a shellfish farm in Yantai, Shandong Province (Figure 4), is affected by the

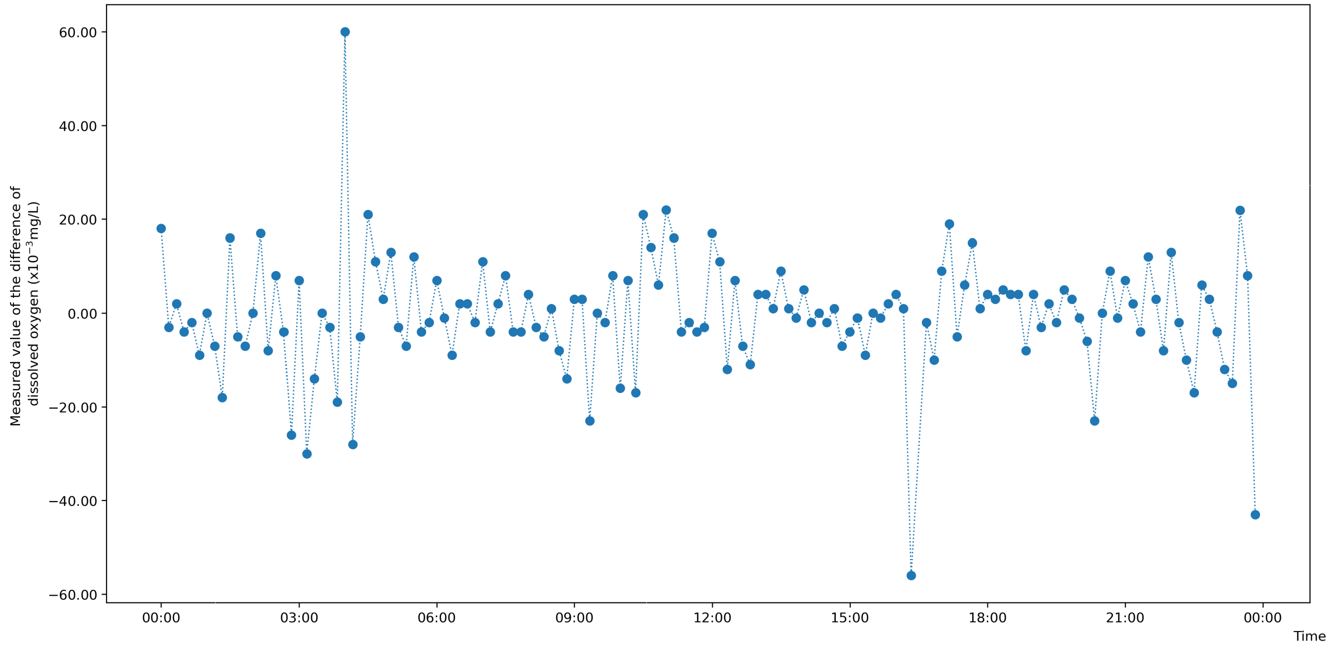


FIGURE 4. Differential sequence plots of the raw dissolved oxygen data.

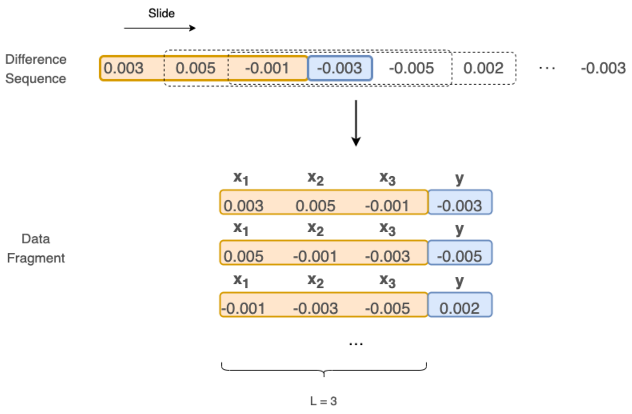


FIGURE 5. Diagram of the differential sequence data generation.

TABLE 1. Dissolved oxygen differential sequence dataset when the length of the differential sequence to be predicted is $L = 3$.

Sequence of differences to be predicted			Predicted value
x_1	x_2	x_3	y
0.003	0.005	-0.001	-0.003
0.005	-0.001	-0.003	-0.005
-0.001	-0.003	-0.005	0.002
…	…	…	…
0.007	0.006	-0.004	-0.002
0.006	-0.004	-0.002	-0.004
-0.004	-0.002	-0.004	-0.003

atmospheric temperature, humidity and other weather factors. Thus, the obtained sequence of dissolved oxygen differential values in the marine pasture significantly changes with nonlinear and nonstationary characteristics, peaks and troughs.

B. ALGORITHM IMPLEMENTATION AND TESTING

Python 3 is used to write a simulation program for the enhanced NB algorithm. First, the differential sequence of dissolved oxygen data is generated; then, the length of the differential sequence to be predicted is set to $L = 3$.

As shown in Table 1, all data slices formed the dissolved oxygen differential sequence dataset; then, the first 99.5% of records in the dataset were selected as the training set, and the last 0.5% of records in the dataset were used as the test set. Finally, the obtained training set and test set were used to model and predict, respectively, the dissolved oxygen differential sequence of shellfish marine pastures.

Figure 6 compares the measured and predicted values of the dissolved oxygen differential sequence. The prediction results for the shellfish marine pastures based on the enhanced NB algorithm are consistent with the dissolved oxygen differential sequence of the actual marine pastures, which can better reflect the nonlinear variation pattern of dissolved oxygen.

Figure 7 shows the relative error plot between measured and predicted values of the dissolved oxygen difference sequence. The smaller relative deviation based on the enhanced NB algorithm enables a more accurate prediction of the dissolved oxygen difference sequence.

Equation (10) is used to convert the dissolved oxygen differential sequence into a dissolved oxygen sequence. The predicted dissolved oxygen levels show that the algorithm generally can predict the trend in dissolved oxygen values with high accuracy.

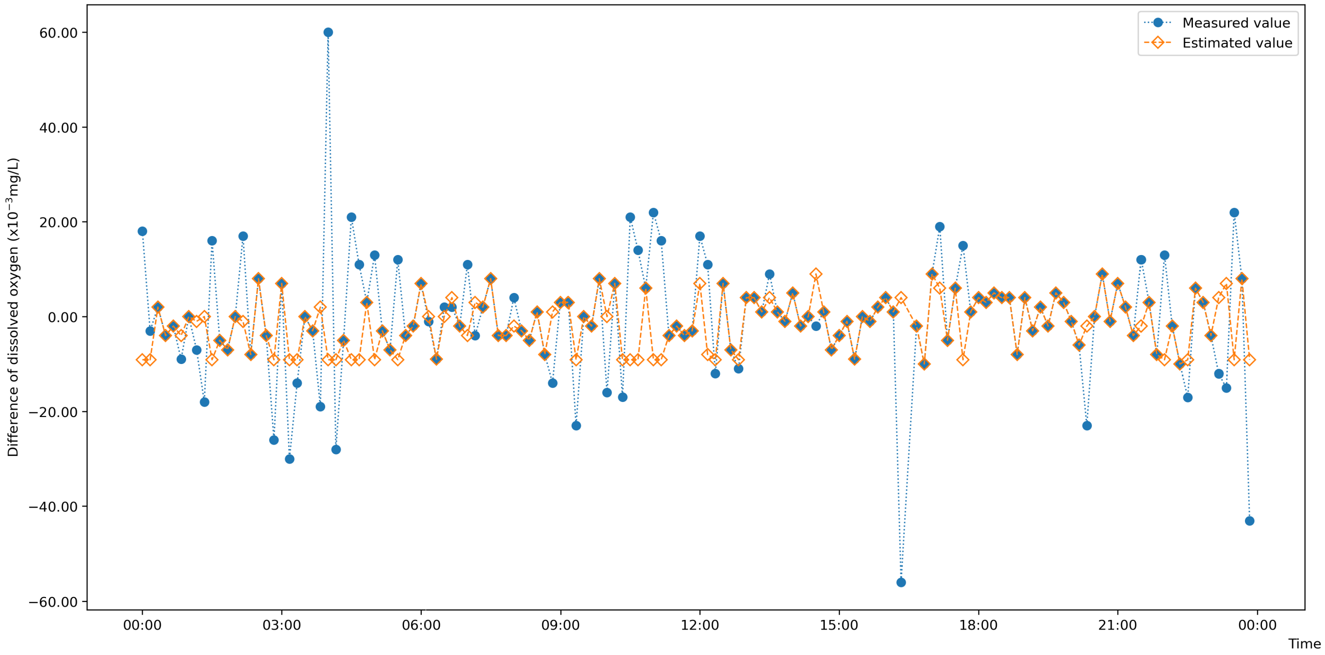


FIGURE 6. Plot of dissolved oxygen differential sequence measurements versus predicted values.

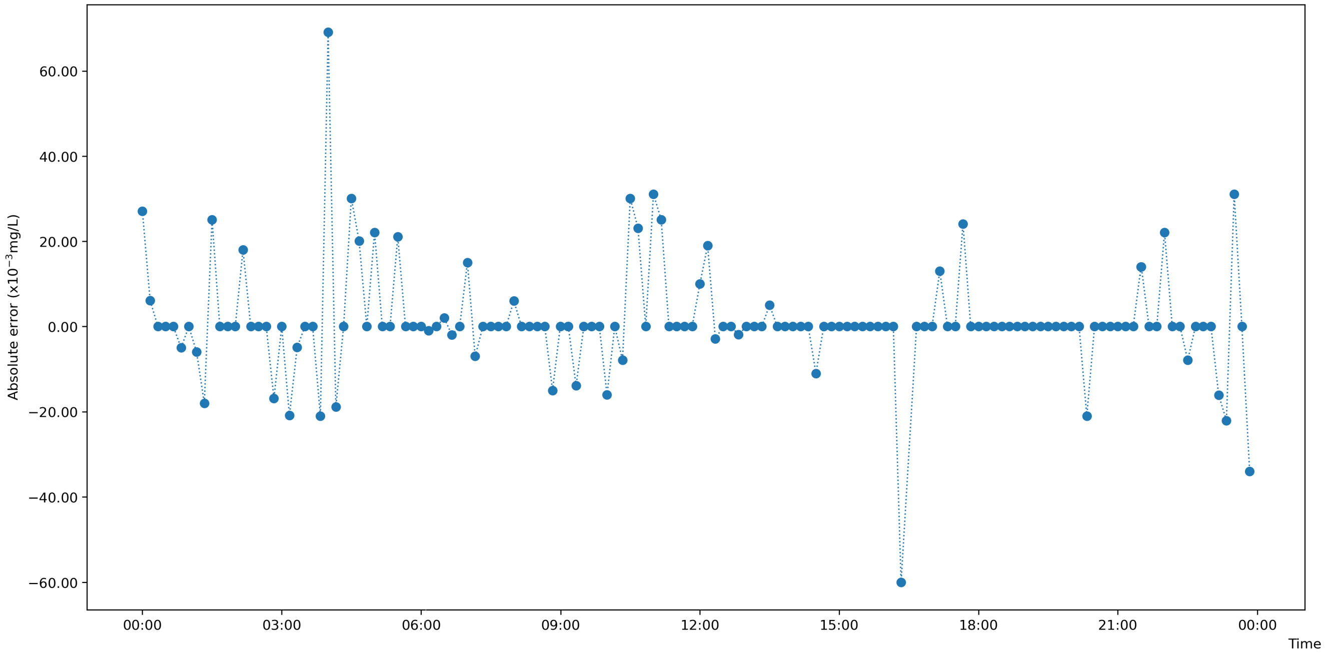


FIGURE 7. Enhanced NB prediction relative error curves.

C. COMPARISON OF THE EFFECTS OF DIFFERENT PREDICTIVE MODELS

Scikit-learn is a Python library that integrates a wide range of machine learning algorithms [41]. In this paper, the multilayer perceptron regressor (MPR) and support vector regression (SVR) algorithms provided by Scikit-learn are chosen to predict the dissolved oxygen time series of the marine pastures. In addition, the same dataset is predicted using the RBFNN, long short-term memory (LSTM) and the autoregressive integrated moving average with exogenous variables (ARIMAX) algorithms [42]. The prediction results are compared with those of the proposed enhanced NB model.

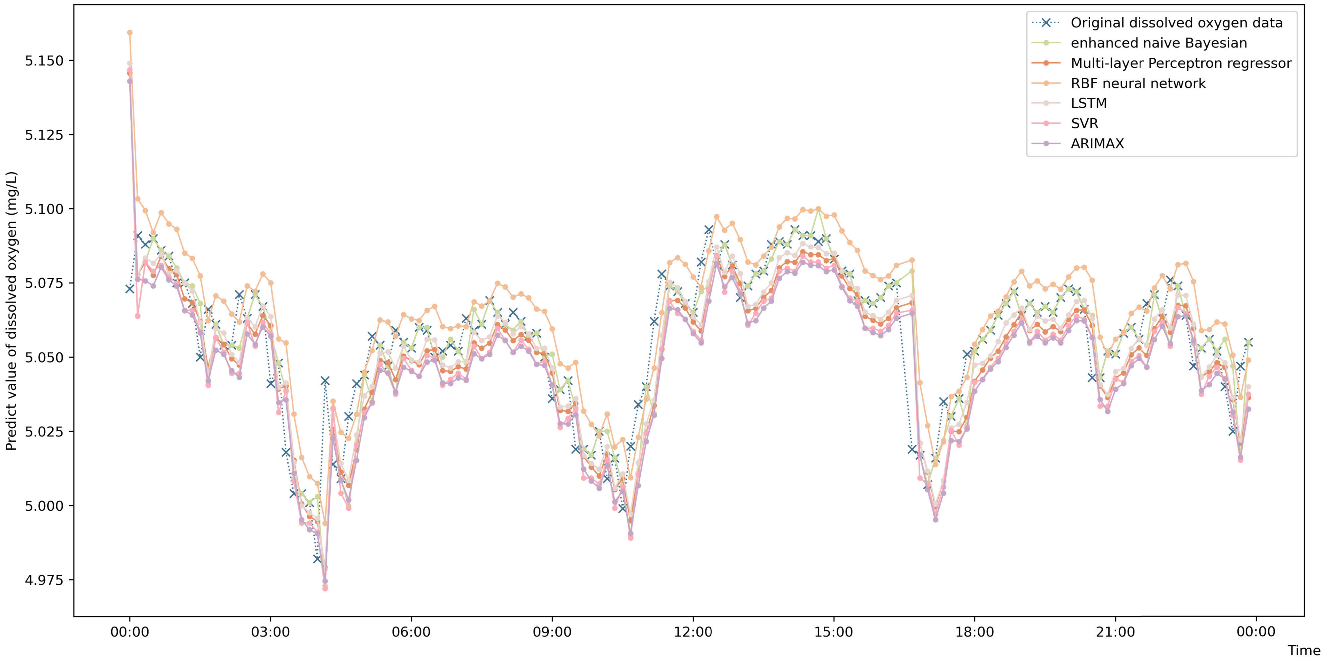


FIGURE 8. Comparison chart of the prediction effect of the enhanced NB, MPR and RBFNN models.

It is concluded that the proposed algorithm obtains relatively good prediction results, as shown in Figure 8.

To further quantify the prediction performance of the algorithm, the algorithm errors were evaluated using the mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE) [43]–[45]. In Table 2, the MAE, RMSE, and MAPE decreased by 0.0141, 0.0043, and 0.002, respectively, when we compared the enhanced NB model with the MPR on the same dataset and test set. Compared with those of the RBFNN model [46]–[50] with 10 hidden layers, the MAE, RMSE, and MAPE of the enhanced NB model decreased by 0.0713, 0.3463, and 0.0076, respectively. The model proposed in this paper decreases the MAE, RMSE and MAPE by 0.0327, 0.0402 and 0.0038, respectively, compared to the ARIMA algorithm. Compared to the SVR algorithm, the MAE, RMSE and MAPE decrease by 0.03, 0.0099 and 0.0034, respectively. Compared to the LSTM algorithm, the MAE, RMSE and MAPE decrease by 0.03, 0.0099 and 0.0034, respectively. The results were 0.0713, 0.1236 and 0.01. Thus, the enhanced NB algorithm can make effective predictions of dissolved oxygen data in marine pastures.

The Diebold-Mariano test is a statistical test method with results that obey a normal distribution [51]–[53], and by comparing the companion p-values from the Diebold-Mariano test, it is possible to determine whether there is a difference between two time series prediction algorithms. The MAE or MAPE is commonly used as an error function, which is combined with the Diebold-Mariano test results to evaluate the predictive ability of the model.

TABLE 2. Comparative analysis of the prediction accuracy of multiple models.

Prediction algorithm	MAE	RMSE	MAPE
MPR	0.0327	0.1945	0.0044
RBFNN	0.0899	0.5365	0.0100
ARIMA	0.0513	0.2304	0.0062
SVR	0.0486	0.2001	0.0058
LSTM	0.0899	0.3138	0.0124
Enhanced NB	**0.0186**	**0.1902**	**0.0024**

In the above Table 3, the results of the DM test are shown in the MAE sense regarding the statistics. The comparison data shows that the companion p-values of each DM-MAE statistic are less than α at the level of $\alpha = 0.05$; i.e., the difference between the predictive ability of each similar algorithm and the model proposed in this paper for the dissolved oxygen data is large in the MAE sense. The confidence level and significant findings indicate that the predictive ability of the enhanced naive Bayesian model is indeed better than that of similar algorithms.

TABLE 3. DM-MAE comparison between prediction models.

Compared algorithm	DM-MAE	P(DM-MAE)
MPR	-7.2838	$3.2446 * 10^{-13}$
RBFNN	-3.6048	$3.0000 * 10^{-4}$
ARIMA	-6.6484	$2.9640 * 10^{-11}$
SVR	-14.8346	$8.7586 * 10^{-50}$
LSTM	-7.5161	$5.6449 * 10^{-14}$

In Table 4, the results of the DM test under MAPE significance indicate that each similar algorithm is significantly different in terms of predictive ability compared to the enhanced naive Bayesian model. At the level of $\alpha = 0.05$, the associated p-values of each DM-MAPE statistic are less than α. The confidence level and significant findings indicate that the predictive ability of the enhanced naive Bayesian model is indeed better than that of the similar algorithms.

TABLE 4. DM-MAPE comparison between prediction models.

Compared algorithm	DM-MAPE	P(DM-MAPE)
MPR	-8.9415	$3.8407 * 10^{-19}$
RBFNN	-3.7462	$2.0000 * 10^{-4}$
ARIMA	-7.5084	$5.9850 * 10^{-14}$
SVR	-17.4518	$3.3348 * 10^{-68}$
LSTM	-5.9560	$2.5847 * 10^{-89}$

D. 10-FOLD CROSS-VALIDATION

In the previous section, we selected the first 99.5% of the dataset as the training set and the last 0.5% of the dataset as the test set and used a sliding window method to expand the training set sample. The prediction results were finally compared with a variety of existing time series prediction algorithms. With the use of this dataset partitioning method, the method proposed in this paper achieved a large improvement in prediction accuracy over other algorithms. To rule out the possibility that this improvement in prediction changes due to the training set partitioning Please ensure that the intended meaning has been maintained in this edit. and, thus, to further illustrate that the enhanced naive Bayesian algorithm proposed in this paper has a higher accuracy in dissolved oxygen time series prediction, the 10-fold cross-validation method was again used to partition the dissolved oxygen dataset used in this paper. According to statistics, the number of valid dissolved oxygen data entries in the dataset implemented in this paper totaled 125,883, which were divided into 10 equal parts; one was selected as the test set, and the remaining data entries were used as the training set. Then, the MAE, RMSE and MAPE errors of the prediction results were calculated using the enhanced naive Bayesian algorithm proposed in this paper. Consequently, the MAE, RMSE and MAPE results of the proposed algorithm are 0.06365, 0.13039 and 0.02173, respectively, after 10-fold cross-validation.

E. EFFECT OF DIFFERENT FEATURE LENGTHS ON THE PREDICTION PERFORMANCE

To further verify the prediction effect of the enhanced NB algorithm for dissolved oxygen values under different feature lengths L, enhanced NB models with different differential sequence lengths L were used to predict the samples obtained from February 18, 2016, to January 31, 2020. Their MAEs, RMSEs and MAPEs are compared in Table 6. The prediction accuracy of the enhanced NB algorithm for dissolved oxygen first increases and then decreases when the differential sequence length L increases, which indicates that the differential sequence length L affects the prediction performance for different water conditions. When there are few data samples in the training set, setting a large L will create insufficient samples to fit the prediction model, which decreases the prediction accuracy. Therefore, when the amount of data is large enough, appropriately increasing L can achieve a better prediction accuracy.

TABLE 5. K-fold cross-validation result.

Test data index		Predicted value		
Start index	End index	MAE	RMSE	MAPE
0	12588	0.0106	0.0377	0.0018
12589	25177	0.0127	0.1209	0.0033
25178	37766	0.0104	0.0447	0.0029
37767	50355	0.0120	0.0384	0.0025
50356	62943	0.0898	0.1768	0.0243
62944	75531	0.1642	0.2546	0.0873
75532	88119	0.1094	0.1879	0.0464
88120	100707	0.1054	0.1618	0.0208
100708	113295	0.1041	0.1612	0.0244
113296	125883	0.0179	0.1199	0.0036
Average		**0.06365**	**0.13039**	**0.02173**

TABLE 6. Comparison of the prediction effectiveness for different difference sequence lengths.

Feature length (L)	MAE	RMSE	MAPE
1	0.019478980	0.189648942	0.002629971
2	0.018752153	0.189872031	0.002478120
3	0.0186317891	0.190196560	0.002419507
4	0.0193596800	0.190364749	0.002554397
5	0.0220724358	0.190602931	0.003086022
6	0.0230393258	0.190790197	0.003274478
7	0.0232382636	0.190949574	0.003311569
8	0.023275523	0.191103255	0.003316870
9	0.0232983870	0.191256960	0.003319380
10	0.0233229402	0.191411109	0.003322211

V. CONCLUSION

To solve the problem of the inability of the traditional NB algorithm to predict continuous-type multicategorical variables, this paper proposes an enhanced NB algorithm and compares it with other models. The results show that compared with the traditional RBFNN algorithm, the proposed algorithm 1) greatly improves the prediction accuracy and 2) can predict continuous attributes. In addition, the prediction model proposed in this paper provides an important reference for dissolved oxygen data in shellfish pastures. The following two factors have led to an improvement in the predictive accuracy of the enhanced NB algorithm:

1) In this paper, we predict the dissolved oxygen (DO) differential series as a classification category, which enables the improved and enhanced naive Bayesian algorithm to predict continuous values. The advantage of naive Bayes is its ability to predict future trends from a probabilistic perspective based on historical experience, but using the naive Bayes

algorithm requires a large number of training samples for each prediction category. In this paper, the dissolved oxygen values are differenced to reduce the number of prediction categories to meet the requirement of using naive Bayes. It also compensates for the fact that, due to the excessive number of different dissolved oxygen values, using them directly as input samples will result in overly few training set categories for each value, which improves the prediction accuracy. At the same time, the distribution of the data after using the differential series tends to be more Gaussian and more regular, which helps the algorithm to learn the regularity of the data itself, thus improving the prediction accuracy.

2) A novel training set sample generation method is used to increase the size of the training samples. To increase the number of training set samples to improve the accuracy of dissolved oxygen prediction, the sliding window concept from network communication protocols is used to partition the dissolved oxygen differential sequence dataset to generate the features and labels of the training set. The values were predicted as categories, and the dissolved oxygen data were accurately predicted by selecting the labels corresponding to the posterior probability maxima of all training samples.

In this paper, the enhanced NB model shows good results for the prediction of dissolved oxygen in shellfish pastures, and the generalization ability of this algorithm will be further discussed and investigated in future work. Considering the strong independence assumptions added to the conditions in the NB equation, which affect the final prediction accuracy, in the next study, application scenarios of the NB equation in this algorithm will be further investigated to obtain better prediction results.

ACKNOWLEDGMENT

The authors would like to thank the editor-in-chief, the associate editor, and the reviewers for their insightful comments and suggestions.

REFERENCES

[1] A. A. M. Ahmed, "Prediction of dissolved oxygen in Surma river by biochemical oxygen demand and chemical oxygen demand using the artificial neural networks (ANNs)," *J. King Saud Univ.-Eng. Sci.*, vol. 29, no. 2, pp. 151–158, Apr. 2017, doi: 10.1016/j.jksues.2014.05.001.

[2] X. Ji, X. Shang, R. A. Dahlgren, and M. Zhang, "Prediction of dissolved oxygen concentration in hypoxic river systems using support vector machine: A case study of Wen-Rui Tang river, China," *Environ. Sci. Pollut. Res.*, vol. 24, no. 19, pp. 16062–16076, Jul. 2017, doi: 10.1007/s11356-017-9243-7.

[3] B. Raheli, M. T. Aalami, A. El-Shafie, M. A. Ghorbani, and R. C. Deo, "Uncertainty assessment of the multilayer perceptron (MLP) neural network model with implementation of the novel hybrid MLP-FFA method for prediction of biochemical oxygen demand and dissolved oxygen: A case study of Langat river," *Environ. Earth Sci.*, vol. 76, no. 14, p. 503, Jul. 2017, doi: 10.1007/s12665-017-6842-z.

[4] J. Huan, W. Cao, and Y. Qin, "Prediction of dissolved oxygen in aquaculture based on EEMD and LSSVM optimized by the Bayesian evidence framework," *Comput. Electron. Agricult.*, vol. 150, pp. 257–265, Jul. 2018, doi: 10.1016/j.compag.2018.04.022.

[5] G. Li, D. Li, Y. Wang, and W. Sun, "Hybrid decoding of finite geometry low-density parity-check codes," *IET Commun.*, vol. 4, no. 10, pp. 1238–1246, 2010, doi: 10.1049/iet-com.2009.0415.

[6] Q. Ren, L. Zhang, Y. Wei, and D. Li, "A method for predicting dissolved oxygen in aquaculture water in an aquaponics system," *Comput. Electron. Agricult.*, vol. 151, pp. 384–391, Aug. 2018, doi: 10.1016/j.compag.2018.06.013.

[7] T. M. Mitchell. *Generative and Discriminative Classifiers: Naïve Bayes and Logistic Regression*. Accessed: Apr. 9, 2020. [Online]. Available: https://www.cs.cmu.edu/~tom/mlbook/NBayesLogReg.pdf

[8] M. M. Saritas and A. Yasar, "Performance analysis of ANN and Naive Bayes classification algorithm for data classification," *Int. J. Intell. Syst. Appl. Eng.*, vol. 7, no. 2, pp. 88–91, Jan. 2019, doi: 10.18201/ijisae.2019252786.

[9] M. Granik and V. Mesyura, "Fake news detection using Naïve Bayes classifier," in *Proc. IEEE 1st Ukraine Conf. Electr. Comput. Eng. (UKRCON)*, Kyiv, Ukraine, May/Jun. 2017, pp. 900–903, doi: 10.1109/UKRCON.2017.8100379.

[10] Q. Jiang, W. Wang, X. Han, S. Zhang, X. Wang, and C. Wang, "Deep feature weighting in Naïve Bayes for chinese text classification," in *Proc. 4th Int. Conf. Cloud Comput. Intell. Syst. (CCIS)*, Aug. 2016, pp. 160–164, doi: 10.1109/CCIS.2016.7790245.

[11] R. Al-khurayji and A. Sameh, "An effective arabic text classification approach based on kernel Naïve Bayes classifier," *Int. J. Artif. Intell. Appl.*, vol. 8, no. 6, pp. 1–10, Nov. 2017, doi: 10.5121/ijaia.2017.8601.

[12] S. Xu, "Bayesian Naïve Bayes classifiers to text classification," *J. Inf. Sci.*, vol. 44, no. 1, pp. 48–59, Feb. 2018, doi: 10.1177/0165551516677946.

[13] M. Mubarok, S. Adiwijaya, and M. D. Aldhi, "Aspect-based sentiment analysis to review products using Naïve Bayes," in *Proc. AIP Conf.*, Budapest, Hungary, 2017, pp. 1–8.

[14] G. Karthick and R. Harikumar, "Comparative performance analysis of Naïve Bayes and SVM classifier for oral X-ray images," in *Proc. 4th Int. Conf. Electron. Commun. Syst. (ICECS)*, Coimbatore, India, Feb. 2017, pp. 88–92.

[15] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. (2020). *Bayesian Data Analysis Third Edition (With Errors Fixed as of 13 February 2020)*. Accessed: May 19, 2020. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.550.6951

[16] F. D. Schönbrodt and E.-J. Wagenmakers, "Bayes factor design analysis: Planning for compelling evidence," *Psychonomic Bull. Rev.*, vol. 25, no. 1, pp. 128–142, Feb. 2018, doi: 10.3758/s13423-017-1230-y.

[17] G. Baele, P. Lemey, and S. Vansteelandt, "Make the most of your samples: Bayes factor estimators for high-dimensional models of sequence evolution," *BMC Bioinf.*, vol. 14, no. 1, p. 85, Mar. 2013, doi: 10.1186/1471-2105-14-85.

[18] M.-G. Xie and K. Singh, "Confidence distribution, the frequentist distribution estimator of a parameter: A review," *Int. Stat. Rev.*, vol. 81, no. 1, pp. 3–39, Apr. 2013, doi: 10.1111/insr.12000.

[19] S. Müller, J. L. Scealy, and A. H. Welsh, "Model selection in linear mixed models," *Stat. Sci.*, vol. 28, no. 2, pp. 135–167, May 2013, doi: 10.1214/12-STS410.

[20] F. D. Schönbrodt, E.-J. Wagenmakers, M. Zehetleitner, and M. Perugini, "Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences," *Psychol. Methods*, vol. 22, no. 2, pp. 322–339, Jun. 2017, doi: 10.1037/met0000061.

[21] L. Zhang and S. Zhang, "Comparison of computational methods for imputing single-cell RNA-sequencing data," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 17, no. 2, pp. 376–389, Apr. 2020, doi: 10.1109/TCBB.2018.2848633.

[22] R. E. Kass and A. E. Raftery, "Bayes factors," *J. Amer. Stat. Assoc.*, vol. 90, no. 430, pp. 773–795, Jun. 1995, doi: 10.1080/01621459.1995.10476572.

[23] B. P. Carlin and T. A. Louis, *Bayes and Empirical Bayes Methods for Data Analysis*. Boca Raton, FL, USA: Chapman & Hall, 2000.

[24] J. S. Maritz and T. Lwin, *Empirical Bayes Methods*. Evanston, IL, USA: Routledge, 2018.

[25] Z. Dienes, "How Bayes factors change scientific practice," *J. Math. Psychol.*, vol. 72, pp. 78–89, Jun. 2016, doi: 10.1016/j.jmp.2015.10.003.

[26] A. Lischke, G. Pang, M. Gulian, F. Song, C. Glusa, X. Zheng, Z. Mao, W. Cai, M. M. Meerschaert, M. Ainsworth, and G. Em Karniadakis, "What is the fractional laplacian?" 2018, *arXiv:1801.09767*. [Online]. Available: http://arxiv.org/abs/1801.09767

[27] L. Brasco and E. Parini, "The second eigenvalue of the fractional p-Laplacian," *Adv. Calculus Variat.*, vol. 9, no. 4, pp. 323–355, Oct. 2016, doi: 10.1515/acv-2015-0007.

[28] B. Mohar, Y. Alavi, G. Chartrand, and O. R. Oellermann, "The Laplacian spectrum of graphs," *Graph Theory, Combinat., Appl.*, vol. 2, nos. 871–898, p. 12, 1991.

[29] O. Sorkine, D. Cohen-Or, Y. Lipman, M. Alexa, C. Rössl, and H.-P. Seidel, ''Laplacian surface editing,'' in *Proc. Eurographics/ACM SIGGRAPH Symp. Geometry Process. (SGP)*, 2004, pp. 175–184, doi: 10.1145/1057432.1057456.

[30] D. A. Field, ''Laplacian smoothing and delaunay triangulations,'' *Commun. Appl. Numer. Methods*, vol. 4, no. 6, pp. 709–712, Nov. 1988, doi: 10.1002/cnm.1630040603.

[31] D. L. Kramer, ''Dissolved oxygen and fish behavior,'' *Environ. Biol. Fishes*, vol. 18, no. 2, pp. 81–92, Feb. 1987, doi: 10.1007/BF00002597.

[32] K. Kaiho, ''Benthic foraminiferal dissolved-oxygen index and dissolved-oxygen levels in the modern ocean,'' *Geology*, vol. 22, no. 8, pp. 719–722, 1994, doi: 10.1130/0091-7613(1994)022<0719:BFDOIA>2.3.CO;2.

[33] J. P. Castro and E. R. Pereira-Filho, ''Twelve different types of data normalization for the proposition of classification, univariate and multivariate regression models for the direct analyses of alloys by laser-induced breakdown spectroscopy (LIBS),'' *J. Anal. At. Spectrometry*, vol. 31, no. 10, pp. 2005–2014, 2016, doi: 10.1039/C6JA00224B.

[34] I. Gibson and C. Amies, ''Data normalization technique,'' Tech. Rep., 2014, vol. 16, pp. 257–269.

[35] N. Golov and L. Rönnbäck, ''Big data normalization for massively parallel processing databases,'' *Comput. Standards Interfaces*, vol. 54, pp. 86–93, Nov. 2017, doi: 10.1016/j.csi.2017.01.009.

[36] J. Quackenbush, ''Microarray data normalization and transformation,'' *Nature Genet.*, vol. 32, no. 4, pp. 496–501, Dec. 2002, doi: 10.1038/ng1032.

[37] X. Shen, X. Gong, Y. Cai, Y. Guo, J. Tu, H. Li, T. Zhang, J. Wang, F. Xue, and Z.-J. Zhu, ''Normalization and integration of large-scale metabolomics data using support vector regression,'' *Metabolomics*, vol. 12, no. 5, p. 89, May 2016, doi: 10.1007/s11306-016-1026-5.

[38] K. Nichols, S. Blake, F. Baker, and D. Black, *Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers*, document RFC 2474, 1998. [Online]. Available: http://www.rfceditor.org/info/rfc2474

[39] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, *An Architecture for Differentiated Service*, document RFC 2475, IETF, Dec. 1998, doi: 10.17487/rfc2475.

[40] W. Fang, N. Seddigh, and B. Nandy, *A Time Sliding Window Three Colour Marker (TSWTCM)*, document Request for Comments 2859, 2000.

[41] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, and J. Vanderplas, ''Scikit-learn: Machine learning in Python,'' *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.

[42] J.-W. Bi, Y. Liu, and H. Li, ''Daily tourism volume forecasting for tourist attractions,'' *Ann. Tourism Res.*, vol. 83, Jul. 2020, Art. no. 102923.

[43] C. Willmott and K. Matsuura, ''Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance,'' *Climate Res.*, vol. 30, pp. 79–82, 2005, doi: 10.3354/cr030079.

[44] R. J. Hyndman and A. B. Koehler, ''Another look at measures of forecast accuracy,'' *Int. J. Forecasting*, vol. 22, no. 4, pp. 679–688, Oct. 2006, doi: 10.1016/j.ijforecast.2006.03.001.

[45] A. de Myttenaere, B. Golden, B. Le Grand, and F. Rossi, ''Mean absolute percentage error for regression models,'' *Neurocomputing*, vol. 192, pp. 38–48, Jun. 2016, doi: 10.1016/j.neucom.2015.12.114.

[46] P. J. De Groot, P. A. Wijnen, and R. B. Janssen, ''Real-time frequency determination of acoustic emission for different fracture mechanisms in carbon/epoxy composites,'' *Compos. Sci. Technol.*, vol. 55, no. 4, pp. 405–412, 1995, doi: 10.1016/0266-3538(95)00121-2.

[47] F. Fatimah, D. Rosadi, R. F. Hakim, and J. C. R. Alcantud, ''Probabilistic soft sets and dual probabilistic soft sets in decision-making,'' *Neural Comput. Appl.*, vol. 31, no. 1, pp. 397–407, Jan. 2019, doi: 10.1007/s00521-017-3011-y.

[48] F. Fatimah, D. Rosadi, R. B. F. Hakim, and J. C. R. Alcantud, ''N-soft sets and their decision making algorithms,'' *Soft Comput.*, vol. 22, no. 12, pp. 3829–3842, Jun. 2018, doi: 10.1007/s00500-017-2838-6.

[49] Q. He, H. Shahabi, A. Shirzadi, S. Li, W. Chen, N. Wang, H. Chai, H. Bian, J. Ma, Y. Chen, X. Wang, K. Chapi, and B. B. Ahmad, ''Landslide spatial modelling using novel bivariate statistical based Naïve Bayes, RBF classifier, and RBF network machine learning algorithms,'' *Sci. Total Environ.*, vol. 663, pp. 1–15, May 2019, doi: 10.1016/j.scitotenv.2019.01.329.

[50] S. K. Satapathy, S. Dehuri, and A. K. Jagadev, ''ABC optimized RBF network for classification of EEG signal for epileptic seizure identification,'' *Egyptian Informat. J.*, vol. 18, no. 1, pp. 55–66, Mar. 2017, doi: 10.1016/j.eij.2016.05.001.

[51] D. Harvey, S. Leybourne, and P. Newbold, ''Testing the equality of prediction mean squared errors,'' *Int. J. Forecasting*, vol. 13, no. 2, pp. 281–291, Jun. 1997, doi: 10.1016/S0169-2070(96)00719-4.

[52] H. Chen, Q. Wan, and Y. Wang, ''Refined diebold-mariano test methods for the evaluation of wind power forecasting models,'' *Energies*, vol. 7, no. 7, pp. 4185–4198, Jul. 2014, doi: 10.3390/en7074185.

[53] F. X. Diebold and R. S. Mariano, ''Comparing predictive accuracy,'' *J. Bus Econ. Statist.*, vol. 13, no. 3, pp. 253–264, 1995.

DASHE LI received the Ph.D. degree from the School of Mechanical and Information Engineering, China University of Mining and Technology-Beijing, in 2011. From 2016 to 2017, he worked as a Visiting Scholar with the University of Washington. He is currently an Associate Professor with the School of Computer Science and Technology, Shandong Technology and Business University, Yantai, Shandong. His research interests include artificial intelligence, software engineering, and computational intelligence.

JIAJUN SUN is currently pursuing the master's degree with the School of Computer Science and Technology, Shandong Technology and Business University, Yantai, Shandong. His current research interests include software engineering and computational intelligence.

HUANHAI YANG received the M.S. degree from Naval Aeronautical Engineering University, Shandong, China, in 2010. He is currently a Lecturer with the School of Computer Science and Technology, Shandong Technology and Business University, Shandong. His research interests include artificial intelligence, data mining, and computational intelligence.

XUEYING WANG is currently pursuing the master's degree with the School of Computer Science and Technology, Shandong Technology and Business University, Yantai, Shandong. Her current research interests include computer applications, artificial intelligence, data mining, and computational intelligence.