Article

Research on Model Selection-Based Weighted Averaged One-Dependence Estimators

Chengzhen Zhang [1], Shenglei Chen [2,*] and Huihang Ke [1]

1 School of Computer Science, Nanjing Audit University, Nanjing 211815, China; ma2309019@stu.nau.edu.cn (C.Z.); mp2309011@stu.nau.edu.cn (H.K.)
2 Department of E-Commerce, Nanjing Audit University, Nanjing 211815, China
* Correspondence: shenglei.chen@nau.edu.cn

Abstract: The Averaged One-Dependence Estimators (AODE) is a popular and effective method of Bayesian classification. In AODE, selecting the optimal sub-model based on a cross-validated risk minimization strategy can further enhance classification performance. However, existing cross-validation risk minimization strategies do not consider the differences in attributes in classification decisions. Consequently, this paper introduces an algorithm for Model Selection-based Weighted AODE (SWAODE). To express the differences in attributes in classification decisions, the ODE corresponding to attributes are weighted, with mutual information commonly used in the field of machine learning adopted as weights. Then, these weighted sub-models are evaluated and selected using leave-one-out cross-validation (LOOCV) to determine the best model. The new method can improve the accuracy and robustness of the model and better adapt to different data features, thereby enhancing the performance of the classification algorithm. Experimental results indicate that the algorithm merges the benefits of weighting with model selection, markedly enhancing the classification efficiency of the AODE algorithm.

Keywords: Bayesian network classification; AODE; leave-one-out cross-validation; model selection; mutual information

MSC: 68T01

1. Introduction

Naive Bayes, within the realm of Bayesian network classifiers, has garnered significant interest and ranks among the top ten traditional algorithms in data mining [1–4]. Naive Bayes assumes that the attributes of a given category are independent of each other. This assumption simplifies the computation of the likelihood function and makes it easy to predict the sample category by maximizing the posterior probability. Given a test sample **x** with vector $\langle x_1, \ldots, x_d \rangle$, Naive Bayes predicts the class of the given test sample as follows:

$$y(\mathbf{x}) = arg \max_{y \epsilon Y} P(y) \prod_{j=1}^{d} P(x_j|y) \tag{1}$$

where d is the number of attributes, $x_j(j = 1, 2, \ldots, d)$ is the attribute value of the jth attribute, y is a specific value in random variable Y, and $y(\mathbf{x})$ is the class label of **x** predicted by the Bayesian network classifier.

Despite its popularity, the Naive Bayes algorithm often fails to consider the correlation between features, resulting in inaccurate classification. In response to the limitations of the Naive Bayes algorithm, the AODE (Averaged One-Dependence Estimators) algorithm emerged [5]. The AODE algorithm is built upon the Bayesian network, which considers the relationship between features when constructing the model. Unlike the Naive independence assumption, AODE does not assume independence between features. In order to

consider dependencies between attributes in a limited scope while keeping the network structure simple, AODE allows dependencies between attributes and assumes that they all depend on a common parent attribute, forming a One-Dependence Estimator (ODE) [6]. Then, by rotating all attributes as parent attributes, the average of the posterior probabilities is used to predict the class of the sample, thus it achieves good results in classification tasks.

In AODE, to enhance both the performance and robustness of classification algorithms, some researchers have proposed cross-validation risk minimization strategies, among which the leave-one-out cross-validation (LOOCV) technique [7] is a commonly used method for cross-validation risk minimization. For example, Chen et al. [8] pointed out that the performance of classification algorithms can be evaluated more accurately by a cross-risk minimization strategy to avoid overfitting of training data. The cross-validation risk minimization strategy is a technique used for evaluating and selecting models by estimating the generalization error of the model using cross-validation during model training to select the optimal model. By introducing the cross-validation risk minimization strategy, the AODE algorithm can better adapt to the characteristics of different datasets and improve the generalization ability of the classifier. However, existing cross-validation risk minimization strategies do not consider the differences in attributes in classification decisions, so this paper proposes a Model Selection-based Weighted AODE (SWAODE) algorithm. The SWAODE algorithm adopts the mutual information as the weights for each ODE and evaluates and selects these weighted sub-models using a leave-one-out cross-validation (LOOCV) technique to determine the best model. The AODE algorithm's classification efficiency is greatly enhanced by this technique, which also boasts strong robustness and broad applicability.

The main contributions of this paper are as follows:

1. The variability between ODEs and between sub-models is fully taken into account by weighting each ODE and selecting the sub-models in this paper. In this way, the quality of each ODE can be evaluated more finely, and the optimal set of models can be selected, which provides a new perspective for the optimization of the AODE algorithm.

2. We propose a new Model Selection-based Weighted AODE (SWAODE) algorithm, which effectively combines the advantages of weighting and model selection. The goal of the SWAODE algorithm is to enhance the performance and robustness of the AODE classification algorithm. The SWAODE algorithm is able to classify data more accurately and improve the model's ability to generalize by integrating weighting and model selection strategies.

3. This paper compares the SWAODE algorithm with other advanced algorithms using 70 datasets from the UCI repository [9], along with conducting ablation experiments. Experimental results indicate the superiority of the SWAODE algorithm over other advanced algorithms.

The sections of this paper are structured as follows: In Section 1, we introduce relevant research focused on improving AODE. Section 2 discusses AODE and the process of model selection. The SWAODE algorithm is outlined in Section 3. In Section 4, we provide a detailed description of the experimental setup and its results. Finally, our conclusions are presented in Section 5.

2. Related Work

In recent years, various strategies have been suggested to alleviate the effects of assumptions about attribute independence. Current research can be generally divided into three types: attribute weighting, attribute selection, and structure extension.

2.1. Attribute Weighting

Jiang and Zhang [10] first proposed the idea of assigning different weights to each attribute in AODE. Jiang et al. [11] then argued that it is not reasonable to have the same weight for every One-Dependence Estimator (ODE) in AODE, so in their paper, they proposed the classification model WAODE that assigned different weights to different

ODEs. Wu et al. [12] introduced an adaptive SPODE named SODE, which leveraged the principles of immunity from the artificial immune system to autonomously and flexibly determine the weights of each SPODE.

2.2. Attribute Selection

Zheng et al. [13] introduced attribute selection methods for AODE, including Backward Sequential Elimination (BSE) and Forward Sequential Selection (FSS), but these techniques are not very practical for large datasets. Meanwhile, Yang et al. [14,15] conducted a comparison of attribute selection and weighting techniques in AODE. Chen et al. [16] introduced an innovative method for selecting attributes, suitable for extensive model space searches with just a single extra training dataset. The experimental results indicated that the novel technique markedly diminished the bias of AODE, but the training time was slightly increased. This low bias and efficient computation made it suitable for big data learning, but the article did not mention the effect of model selection.

2.3. Structure Extension to NB

Friedman et al. [17] introduced the Tree-Augmented Naive Bayes (TAN) method as an enhancement to Naive Bayes (NB), incorporating a tree structure to mitigate the independence assumptions of NB. TAN mandates that the class variables lack parent nodes, with each attribute containing the class variable and at most one other attribute as parent nodes. This one-traversal algorithm acquires the necessary probability distributions from the training samples during one-traversal learning to construct the network structure and conditional probability tables.

The K-dependent Bayesian classifier (KDB) [18] is a method to improve Naive Bayes (NB). It relaxes the independence assumption of Naive Bayes by allowing each attribute to possess a maximum of k parent attributes. As a result, NB can be viewed as a zero-dependent Bayesian classifier, whereas KDB can include a higher degree of attribute dependence by increasing the value of k. KDB can construct classifiers at any value of k, retaining most of the computational properties of NB and selecting for each attribute a network structure with up to k parent attributes for each attribute.

Another notable enhancement to NB is AODE [5], which relaxes the independence assumption of Naive Bayes by allowing some degree of dependence between features. AODE constructs multiple One-Dependence Estimators by considering the relationship between each feature and category, and then averages them to obtain the final classification result. This approach can more effectively utilize the correlation between features and ultimately improve classification accuracy.

3. AODE and Model Selection Analysis

We discuss the Averaged One-Dependence Estimators (AODE) algorithm and its model selection process in this section.

In order to make the paper more readable, we summarize all the symbols that are defined in the paper in Table 1 for quick reference and understanding by the reader.

Table 1. Tables of symbols.

Symbols	Definition
$\mathcal{D}$	Set of training data
n	Number of training samples
X, X_i	Variable representing the attribute
x, x_i	Value of attribute variable X or X_i
v_i, v_j	Number of values of attribute X_i or X_j
$\mathbf{x} = \langle x_1, \ldots, x_d \rangle$	Vector representing a sample
$\mathbf{x_i}$	ith sample
$x_{i,j}$	jth value of the ith attribute
Y	Variable representing the category
y	A specific value in random variable Y
y_i	True category of the ith sample
c	Number of categories
d	Number of attributes
m	Smoothing parameter
r	Number of attributes as parent attributes
s	Number of attributes as child attributes
w_j	Weight of the jth ($j = 1, 2, ..., d$) ODE
m'	Threshold

3.1. Constructing the AODE Model

AODE only allows one dependence between attributes; attribute X_i can only depend on some attribute X_j and category Y, where X_j is called the parent attribute of X_i. At the same time, in order to keep the computation simple, it is assumed that all attributes depend on a common parent attribute X_j which constitutes a Bayesian network called One-Dependence Estimator (ODE) [6]. Based on this ODE, the joint probability $p(y, \mathbf{x})$ can be estimated as:

$$P_{ODE}(y, \mathbf{x}) = P(y, x_j) \prod_{i=1}^{d} P(x_i | y, x_j) \tag{2}$$

To eliminate the bias introduced by the selection of the parent attribute, it is allowed that all the attributes can be used as the parent attribute in turn, thus obtaining d ODEs, and finally, the posteriori probabilities estimated from these d ODE are averaged to obtain the posterior probability estimate of the sample. Thus, the AODE algorithm calculates the joint probability as:

$$P_{AODE}(y, \mathbf{x}) = \frac{\sum_{j=1}^{d} P(y, x_j) \prod_{i=1}^{d} P(x_i | y, x_j)}{d} \tag{3}$$

where $P(x_i | y, x_j)$ can be obtained from the ratio of $P(x_i, y, x_j)$ and $P(y, x_j)$, so only the basic probabilities $P(y, x_j)$ and $P(x_i, y, x_j)$ need to be estimated, which can be obtained by an M-estimation:

$$\hat{P}(y, x_j) = \frac{F(y, x_j) + \dfrac{m}{c * v_i}}{n + m} \tag{4}$$

$$\hat{P}(x_i, y, x_j) = \frac{F(x_i, x_j, y) + \dfrac{m}{c * v_i * v_j}}{n + m} \tag{5}$$

where $F(\bullet)$ is the frequency of occurrence of the parameter item in the training dataset. v_i is the number of values of attribute X_i, c is the number of categories, and m is the smoothing parameter in the M-estimation, which is a commonly used parameter estimation method. By introducing the smoothing parameter m, the M-estimation can prevent the probability estimates from being zero and improves the robustness of the estimates.

The frequency F containing class labels and attribute values can be realized in a three-dimensional table in practical implementation, where the first and second dimensions represent the values taken for the first and second attributes. The third dimension represents the values taken for the category, and the values in the table record the frequency values on the values of that dimension. Assuming that there are two attributes X_1, X_2 and two categories $class1, class2$ where X_1 has two attribute values, X_2 has three attribute values, the frequency table is shown in Table 2.

Table 2. Frequency table with two attributes and two class variables.

				Class1	**Class2**
X_2	$x_{2,1}$	X_1	$x_{1,1}$		
		X_1	$x_{1,2}$		
X_2	$x_{2,2}$	X_1	$x_{1,1}$		
		X_1	$x_{1,2}$		
X_2	$x_{2,3}$	X_1	$x_{1,1}$		
		X_1	$x_{1,2}$		

The training process of AODE is described by Algorithm 1.

Algorithm 1 AODE training process

Require: Set of training data $\mathcal{D}$
Ensure: Frequency table F containing a combination of category labels and two attributes
 1: All frequencies in the frequency table F are initialized to zero
 2: **for** each training sample **x** from $\mathcal{D}$ **do**
 3:     Obtain the class label y for sample **x** and set i and j to zero
 4:     **while** $i < d$ **do**
 5:         Read the value x_i of attribute X_i in sample **x**
 6:         **while** $j < i$ **do**
 7:             Read the value x_j of attribute X_j in sample **x**
 8:             Find the frequency at positions x_i, x_j, y in the frequency table F and increase by one
 9:                 $j = j + 1$
 10:         **end while**
 11:         $i = i + 1$
 12:     **end while**
 13: **end for**
 14: **return** the finally calculated frequency table F

Algorithm 1 reveals that the AODE training process's time complexity is influenced by the quantity of samples and attributes, where there are d parent and child attributes, so the total time complexity of the AODE training process is $O(nd^2)$, where d denotes the number of attributes, and n denotes the number of samples for each attribute. AODE typically exhibits a marginally greater time complexity compared to NB, as AODE represents a singular attribute dependency, aligning more closely with the real data. Therefore, the classification performance is significantly improved compared with NB.

3.2. Model Selection-Based AODE

To fully present the Model Selection-based AODE(SAODE) algorithm in this section, the model space is first constructed here. Then, the attributes are ranked based on mutual information. Finally, the best model is selected using the leave-one-out cross-validation error.

3.2.1. Building the Model Space

When constructing the AODE model space, we introduce the threshold m'. When a particular value x_j of the parent attribute in the training data occurs more often than or

as often as a threshold m', the ODE model corresponding to that value is included in the computation of the AODE model. If we choose the former r attributes as parent attributes and the former s attributes as child attributes, where $1 \leq r, s \leq d$, the AODE model is approximated by:

$$P_{AODE}(y, \mathbf{x})_{r,s} = \frac{\sum_{j:1 \leq j \leq r \wedge \mathrm{F}(x_j) \geq m'} \mathrm{P}(y, x_j) \prod_{i=1}^{s} \mathrm{P}(x_i \mid y, x_j)}{|\{j : 1 \leq j \leq r \wedge \mathrm{F}(x_j) \geq m'\}|} \tag{6}$$

where $F(x_j)$ is the frequency of x_j, and m' is the minimum frequency that takes the value x_j. The AODE algorithm with the inclusion of a threshold m' improves the overall performance and prediction accuracy of the model by ensuring that the number of samples for the parent attribute is sufficient, avoiding the problems of high variance and unreliable conditional probability estimation caused by data sparsity. This mechanism enables AODE to maintain high predictive stability and reliability in the face of uneven data distribution. When both r and s are equal to d, it can be seen from the formula that when calculating $P_{AODE}(y, \mathbf{x})_{d,d}$, at most d^2 subsets of attributes are created as sub-models.

All these approximate AODE models are just a small extension of the previous model. For example, $P_{AODE}(y, \mathbf{x})_{1,2}$ is obtained by adding the child attribute x_2 to $P_{AODE}(y, \mathbf{x})_{1,1}$. All of these models can be applied to test instances in a single nested computation. Thus, all models can be evaluated efficiently.

3.2.2. Attribute Sorting

Constructing a model of the later attributes depends on the model of the earlier attributes when constructing the AODE model space. Therefore, this method of nesting models depends on the order of the attributes. Thus, here, the mutual information is used to sort the attributes. The mutual information is calculated as:

$$\begin{aligned} MI(X, Y) &= H(X) - H(X|Y) \\ &= \sum_{y \in Y} \sum_{x \in X} P(x, y) \log_2 \frac{P(x, y)}{P(x)P(y)} \end{aligned} \tag{7}$$

where $H(X)$ is the entropy of X, $H(X|Y)$ is the conditional entropy, $p(x, y)$ is the joint probability of x and y, and $p(x)$ and $p(y)$ are the probabilities of x and y, respectively. Therefore, the MI is used as an indicator of the correlation between attribute X and category Y. The larger the value of the MI, the stronger the correlation between attribute X and category Y is indicated.

An advantage of employing the MI is its ability to efficiently compute the MI between attributes and classes within a single training session. While the MI can identify the discriminative power of individual attributes, it cannot directly assess the discriminative power of combinations of attributes. However, this shortcoming is compensated for the fact that the ranking based on MI can be searched in a wide model space.

3.2.3. Model Selection

For evaluating the distinctiveness of different models and preventing overfitting, leave-one-out cross-validation errors are employed. Through gradual cross-validation, the impact of the absent sample in each fold is deducted from the frequency table to create a model excluding that sample. The technique offers a lower bias estimate of the generalization error and assesses the model using a single training dataset.

In addition, as shown in Equation (6), these models are nested together, with each model being a straightforward extension of another, providing an effective means of evaluating them. That is, these models can be evaluated simultaneously during their construction for the training samples missed in each fold.

Among the more common methods used to evaluate the model selection are the 0–1 loss, Root-Mean-Square Error (RMSE), LogLoss, and AUC value. For example, Chen [19]

proposed the RMSE as a criterion for model evaluation, where a lower RMSE indicated a better model. Therefore, we also use the RMSE as a criterion for model evaluation as a way to select the optimal model in Section 3.

4. Model Selection-Based Weighted AODE

In this section, our focus is on the weighting strategy for AODE and the methodology for model selection on the weighted AODE model, given that we have already described the construction methodology of the AODE model in detail in Section 2.

4.1. Weighting the AODE Model

The contribution of each ODE to the final classification result may be different in the AODE algorithm, and certain sub-models may discriminate more accurately for specific categories while others may perform weakly. Therefore, weighting each sub-model can more accurately reflect its importance in the overall classification process, thus improving the overall model performance [11].

The classification ability of ODEs composed of different parent attributes X_j is different, so different weights can be applied to different ODEs [11]. Thus, the formula transforms into:

$$P_{WAODE}(y, \mathbf{x}) = \frac{\sum_{j=1}^{d} w_j P(y, x_j) \prod_{i=1}^{d} P(x_i | y, x_j)}{d} \tag{8}$$

where w_j is the weight of the jth ($j = 1, 2, ..., d$) ODE, where the weights are obtained by calculating the MI through Equation (8). When attribute X and category Y are completely independent, the MI is 0, indicating that there is no information sharing or dependency between them.

Weighting each ODE can better improve the performance and robustness of the overall model, thus enhancing the reliability and validity of the model in practical applications.

4.2. Model Selection for WAODE

We first construct the model space of WAODE in this subsection. Then, the attributes are ranked according to the MI. Finally, we use the RMSE as the cross-validation error and select the optimal sub-model by minimizing the RMSE.

4.2.1. Building the Model Space

As shown in Equation (6), for the WAODE algorithm, where we also introduce the threshold m', the joint probability is given by:

$$P_{WAODE}(y, \mathbf{x})_{r,s} = \frac{\sum_{j:1 \leq j \leq r \wedge F(x_j) \geq m'} w_j P(y, x_j) \prod_{i=1}^{s} P(x_i \mid y, x_j)}{|\{j : 1 \leq j \leq r \wedge F(x_j) \geq m'\}|} \tag{9}$$

Similar to the AODE model's construction in Section 2, all of these approximate WAODE models constitute the model space depicted in Table 3, since each model is simply a small extension of the previous one. Thus, every model is capable of being assessed with efficiency.

Table 3. Space of approximate models of WAODE with d attributes.

Parent	Children				
	x_1	$\ldots$	x_s	$\ldots$	x_d
x_1	$P_{WAODE}(y, \mathbf{x})_{1,1}$	$\ldots$	$P_{WAODE}(y, \mathbf{x})_{1,s}$	$\ldots$	$P_{WAODE}(y, \mathbf{x})_{1,d}$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
x_r	$P_{WAODE}(y, \mathbf{x})_{r,1}$	$\ldots$	$P_{WAODE}(y, \mathbf{x})_{r,s}$	$\ldots$	$P_{WAODE}(y, \mathbf{x})_{r,d}$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
x_d	$P_{WAODE}(y, \mathbf{x})_{d,1}$	$\ldots$	$P_{WAODE}(y, \mathbf{x})_{d,s}$	$\ldots$	$P_{WAODE}(y, \mathbf{x})_{d,d}$

4.2.2. Attribute Sorting

In constructing the WAODE model space, the model for later attributes is dependent on the model for earlier attributes, as shown in Table 3. This approach of nested models is influenced by the order in which the attributes are considered. To address this, we utilize the MI to rank the attributes. Additionally, we observe that the sorting process also involves selecting attributes, and by sorting first, we can more easily identify the attributes that have a significant impact on categorization. To calculate the MI, we use Equation (8).

4.2.3. Model Selection

We used a 10-fold CV in our experiments to make the results more objective, and we used the LOOCV error as a criterion for model selection. Figure 1 describes the relationship between LOOCV and 10-fold CV. The test set in the 10-fold CV loops through the 10 folds of samples. At the same time, the test instance in LOOCV loops through all the LOOCV instances.
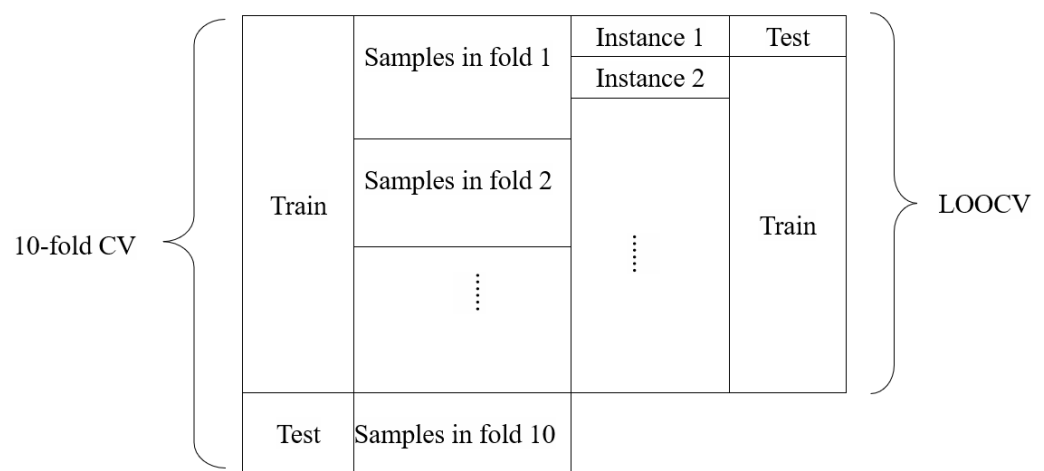

Figure 1. The relationship between LOOCV and 10-fold CV.

LOOCV errors were used to evaluate model distinctiveness and prevent overfitting by excluding one sample at a time from the training data and assessing the model without it. This method provides a lower bias estimate of the generalization error and evaluates the model using nearly all available data for training.

The 0–1 loss (ZOL) and the Root-Mean-Square Error (RMSE) are the most common evaluation criteria used as model selection. The 0–1 loss simply assigns "0" to correct classifications and "1" to misclassifications, considering all misclassifications as equally undesirable. However, the RMSE is more sensitive to the severity of misclassification, so it is able to make more fine-grained probabilistic predictions. The RMSE can be expressed as:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (1 - P(y(\mathbf{x}_i) = y_i \mid \mathbf{x}_i))^2} \tag{10}$$

where y_i is the true class of sample $\mathbf{x}_i$. The smaller the RMSE, the smaller the discrepancy between the model's predictions and the true labels. Compared to the 0–1 loss, the RMSE is able to assess model uncertainty on a continuous basis rather than simply telling us whether the model is classified correctly or not. Meanwhile, the RMSE penalizes model uncertainty more strictly, so it provides a more fine-grained calibration metric for probability estimation. Consequently, the RMSE was employed to assess potential models in our study.

Therefore, the process of choosing the best model can be framed as the following optimization problem:

$$\langle r, s \rangle^* = \underset{\langle r,s \rangle}{\text{argmin}} \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(1 - P_{\text{WAODE}}^{\text{LOO}}(y(\mathbf{x}_i) = y_i \mid \mathbf{x}_i)_{r,s}\right)^2} \tag{11}$$

where $P_{WAODE}^{LOO}(y \mid \mathbf{x}_i)_{r,s}$ can be computed by first estimating $P_{WAODE}^{LOO}(y, \mathbf{x}_i)_{r,s}$ from training set $(\mathcal{D} - \{\langle y_i, x_i \rangle\})$ as in Equation (9), and then normalizing across all possible y's.

4.3. Algorithm Description

Utilizing the aforementioned method, we formulated a training algorithm for the Selection-based Weighted AODE (SWAODE) model, as shown in Algorithm 2.

Algorithm 2 Training algorithm for Model Selection-based Weighted AODE (SWAODE)

1: **First pass:** Form the table of joint frequencies of all combinations of x attribute values and the class label as in Algorithm 1
2: Compute the mutual information
3: Weighting each ODE and rank the attributes
4: **Second pass:** Perform LOOCV on the sample set
5: **for** sample $\mathbf{x} \in \mathcal{D}$ **do**
6:     Remove sample $\mathbf{x}$ from the frequency table
7:     Obtain category for sample $\mathbf{x}$ and set i and j to zero
8:     Build d^2 models for AODE
9:     **for** $j < d$ **do**
10:         **for** $i < d$ **do**
11:             Predict sample $\mathbf{x}$ using all models in Equation (9)
12:             Accumulate the squared error for each model
13:             $i = i + 1$
14:         **end for**
15:         $j = j + 1$
16:     **end for**
17:     Add sample $\mathbf{x}$ back to the frequency table
18: **end for**
19: Compute the root-mean-square error for each model
20: Select the model with the lowest RMSE

The SWAODE algorithm for weighting and model selection needs to consider the time complexity involved in computing the MI and LOOCV, respectively, where the total time complexity of calculating the MI in the WAODE-MI algorithm is $O(cd^2)$, and the time complexity of model selection with LOOCV is $O(cnd^2)$, so the total time complexity of the SWAODE algorithm is $O(cnd^2) + O(cd^2)$, which is almost similar to the time complexity of the SAODE algorithm, where d is the number of attributes, n is the number of samples, and c is the number of categories.

5. Experiments and Discussion

We ran the above algorithm on 70 datasets from the UCI repository [9]. The comprehensive features of the datasets are presented in Table 4, arranged in an increasing sequence based on the count of instances. The experiments were carried out based on the high-performance computing platform of Nanjing Audit University, the computing node CPU was an Intel E5, the amount of memory was 188 G, and the operating system was CentOS7.9-x64. The algorithms were based on the Petal Machine Learning Platform [19] implemented in C++. Compared to the well-known machine learning experimental platform Weka [20], the Petal platform has one significant difference: missing values are viewed as a single value in Petal, whereas the Weka system employs means (numerical attributes) or modes (discrete attributes) instead.

Table 4. Datasets.

No.	Name	Inst	Att	Class
1	contact-lenses	24	4	3
2	lung-cancer	32	56	3
3	labor-negotiations	57	16	2
4	post-operative	90	8	3
5	zoo	101	16	7
6	promoters	106	57	2
7	echocardiogram	131	6	2
8	lymphography	148	18	4
9	iris	150	4	3
10	teaching-ae	151	5	3
11	hepatitis	155	19	2
12	wine	178	13	3
13	autos	205	25	7
14	sonar	208	60	2
15	glass-id	214	9	3
16	new-thyroid	215	5	3
17	audio	226	69	24
18	hungarian	294	13	2
19	heart-disease-c	303	13	2
20	haberman	306	3	2
21	primary-tumor	339	17	22
22	ionosphere	351	34	2
23	dermatology	366	34	6
24	horse-colic	368	21	2
25	house-votes-84	435	16	2
26	cylinder-bands	540	39	2
27	chess	551	39	2
28	syncon	600	60	6
29	balance-scale	625	4	3
30	soybean	683	35	19
31	credit-a	690	15	2
32	breast-cancer-w	699	9	2
33	pima-ind-diabetes	768	8	2
34	vehicle	846	18	4
35	anneal	898	38	6
36	tic-tac-toe	958	9	2
37	vowel	990	13	11
38	german	1000	20	2
39	led	1000	7	10
40	contraceptive-mc	1473	9	3
41	yeast	1484	8	10
42	volcanoes	1520	3	4
43	car	1728	6	4
44	segment	2310	19	7
45	hypothyroid	3163	25	2
46	splice-c4.5	3177	60	3
47	kr-vs-kp	3196	36	2
48	abalone	4177	8	3
49	spambase	4601	57	2
50	phoneme	5438	7	50

Table 4. *Cont.*

No.	Name	Inst	Att	Class
51	wall-following	5456	24	4
52	page-blocks	5473	10	5
53	optdigits	5620	64	10
54	satellite	6435	36	6
55	musk2	6598	166	2
56	mushrooms	8124	22	2
57	thyroid	9169	29	20
58	pendigits	10,992	16	10
59	sign	12,546	8	3
60	nursery	12,960	8	5
61	magic	19,020	10	2
62	letter-recog	20,000	16	26
63	adult	48,842	14	2
64	shuttle	58,000	9	7
65	connect-4	67,557	42	3
66	waveform	100,000	21	3
67	localization	164,860	5	11
68	census-income	299,285	41	2
69	poker-hand	1,025,010	10	10
70	donation	5,749,132	11	2

5.1. Comparison on ZOL

In this experiment, in order to verify the performance of the SWAODE algorithm, we compared it with classical algorithms such as NB [1], KDB (k = 1) [18], AODE [5], WAODE-MI [11], WAODE-KL [21], and so on. We adopted ZOL as the evaluation index, where the loss was one when the sample was misclassified and zero when it was correctly classified, and then we calculated the percentage of total loss in the total number of test samples in order to comprehensively assess the performance of different algorithms in the classification task. The W/D/L metrics tracked the number of wins, draws, and losses for each algorithm across multiple datasets, allowing for a comparison of their performance on the same dataset. For instance, SWAODE demonstrated strong performance with 52 wins, 5 draws, and 13 losses when compared to NB, providing an objective assessment of the algorithms' respective strengths and weaknesses. Through this professional evaluation method, we can more comprehensively and objectively assess the advantages of the SWAODE algorithm over other algorithms and provide a scientific basis for its performance evaluation, as shown in Table 5. Meanwhile, in order to facilitate the observation of SWAODE's experimental data, we bolded the row where SWAODE is located in all subsequent tables and presents the experimental data for each dataset in Appendix Table A1.

Table 5. Win/draw/loss of ZOL for SWAODE.

	NB	KDB	AODE	WAODE-MI	WAODE-KL	SAODE
SWAODE	**52/5/13**	**52/2/16**	**39/8/23**	**34/12/24**	**32/14/24**	**27/22/21**
NB		24/2/44	15/4/51	14/3/53	15/3/52	14/3/53
KDB			21/3/46	21/1/48	20/2/48	15/3/52
AODE				20/15/35	20/13/37	21/8/41
WAODE-MI					9/51/10	25/8/37
WAODE-KL						26/7/37

The analysis in Table 5 reveals that the SWAODE algorithm outperformed other advanced algorithms. Compared to the AODE algorithm, the SWAODE algorithm achieved 39 wins, 8 draws, and 23 losses, representing a significant improvement. Additionally, the weighted AODE classification algorithm also showed improvement when different weights were assigned. The WAODE-KL algorithm, which uses KL divergence as weights,

achieved 37 wins, 13 draws, and 20 losses compared to the AODE algorithm, demonstrating a clear advantage. However, even the excellent WAODE-KL algorithm did not surpass our new algorithm, the SWAODE algorithm. In comparison, the SWAODE algorithm achieved 32 wins, 13 draws, and 24 losses, showing a clear advantage. Overall, our SWAODE algorithm demonstrated strong performance and brought significant improvement to the classification task.

We also represented the scatter plot of SWAODE with respect to WAODE-MI in terms of ZOL in Figure 2. Points above the diagonal represent datasets whose ZOL values are lower than those of WAODE-MI. It can be found that SWAODE consistently provided better predictions than the regular WAODE-MI in a statistically significant way.
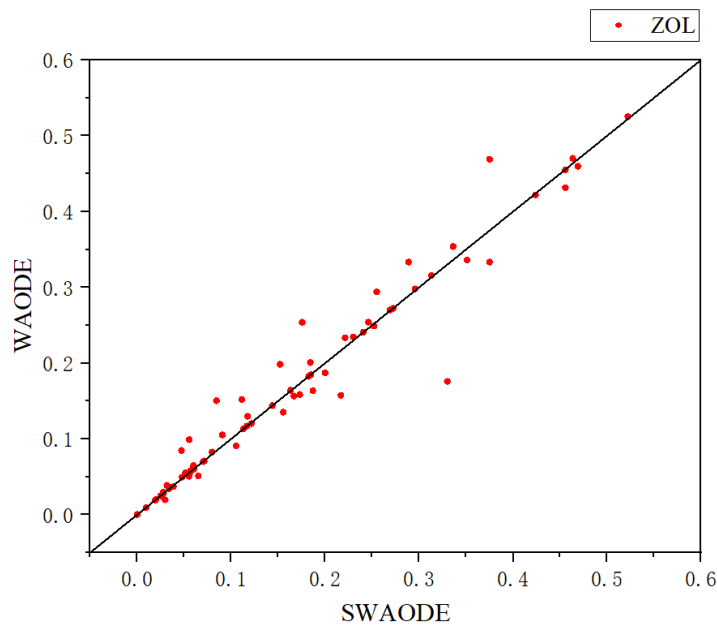

Figure 2. Scatter plot of ZOL.

5.2. Comparison on LogLoss

When assessing the effectiveness of the SWAODE algorithm, it is common to use LogLoss as an evaluation metric. LogLoss is a widely used metric for evaluating the predictive accuracy of a classification model. It measures the deviation between the model's predicted probability for each sample and the actual label. To compare the SWAODE algorithm with other advanced algorithms, their LogLoss values on a test dataset can be calculated and visualized. Additionally, the W/D/L (win/draw/loss) metric can be used to analyze the strengths and weaknesses of different algorithms in the experimental results. By comparing the LogLoss values of the SWAODE algorithm with those of other algorithms, the strengths and weaknesses on different datasets can be determined, as shown in Table 6. Meanwhile, we also show the experimental data in detail in Appendix Table A2.

Table 6. Win/draw/loss of LogLoss for SWAODE.

	NB	KDB	AODE	WAODE-MI	WAODE-KL	SAODE
SWAODE	60/0/10	56/0/14	48/1/21	42/7/21	42/6/22	40/4/26
NB		28/0/42	11/0/59	11/0/59	11/0/59	10/0/60
KDB			18/0/52	18/0/52	18/0/52	12/0/58
AODE				25/1/44	22/1/47	17/2/51
WAODE-MI					19/28/23	24/0/46
WAODE-KL						26/0/44

According to the data analysis in Table 6, the SWAODE algorithm presented excellent performance on LogLoss. In the comparison with AODE, it achieved 48 wins/1 draw/21 losses. In addition, the SWAODE algorithm also performed outstandingly compared to the weighted AODE algorithm. The SWAODE algorithm beat the WAODE-MI algorithm and the WAODE-KL algorithm on 42 of the 70 datasets, respectively. These results show that the SWAODE algorithm is adaptable to various datasets and outperforms other algorithms in most cases. Therefore, the SWAODE algorithm is a very efficient algorithm for AODE improvement.

Meanwhile, we also represented the scatter plot of SWAODE with respect to WAODE-MI in terms of LogLoss in Figure 3. It can be found that SWAODE consistently provided better predictions than the regular WAODE-MI algorithm in a statistically significant way.
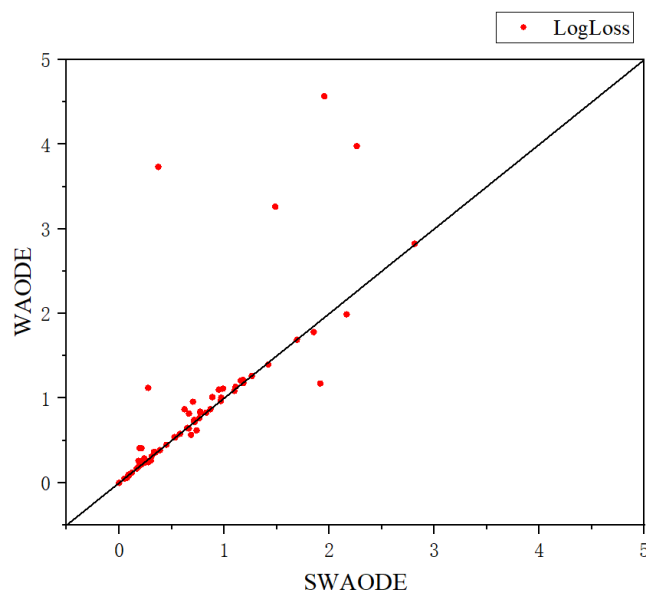

Figure 3. Scatter plot of LogLoss.

5.3. Ablation Studies

To delve deeper into the necessity of weighting and model selection for the AODE classification algorithm, we conducted two ablation study experiments to validate its impact in this section, again using W/D/L (win/draw/loss) as the measure. These experiments aimed to dissect the performance of the SWAODE algorithm in the absence of weighting and model selection, thus highlighting the crucial role of weighting and model selection in improving the classification performance of SWAODE. In our experiments, we implemented the WAODE-MI algorithm, which uses MI as a weight, and the SAODE algorithm, which performs model selection on AODE. The SWAODE algorithm was compared with these two algorithms in terms of ZOL and LogLoss metrics.

According to Table 7, the SWAODE algorithm achieved 34 wins/12 draws/24 losses and 42 wins/7 draws/21 losses in the two comparisons with the WAODE-MI algorithm. It also performed well in the comparison with SAODE, achieving 27 wins/22 draws/21 losses and 40 wins/4 draws/26 losses, respectively. Therefore, we can conclude that both weighting and model selection are necessary and indispensable in the SWAODE algorithm, and the algorithm is able to fully draw on the advantages of weighting and model selection to greatly improve the classification performance of the AODE algorithm.

Table 7. Ablation Studies.

		WAODE-MI	SAODE
ZOL	**SWAODE**	**34/12/24**	**27/22/21**
	WAODE-MI		25/8/37
LogLoss	**SWAODE**	**42/7/21**	**40/4/26**
	WAODE-MI		24/0/46

6. Conclusions

This study proposed a new AODE classification algorithm, the SWAODE algorithm, which aimed to solve the problem of existing cross-validation risk minimization strategies not considering the difference in attributes in classification decisions. The core idea of the algorithm lay in first weighting each ODE in the AODE which used the MI values as the weights. Subsequently, a leave-one-out cross-validation (LOOCV) method was used to perform model selection on these weighted sub-models in order to select the optimal model. Experimental results indicated the SWAODE algorithm markedly surpassed other well-known popular classification algorithms on multiple datasets, exhibiting higher classification efficiency and generalization ability.

However, we recognize that this is only one aspect of model selection and that many potential extensions deserve further exploration. The next step of our work will focus on exploring the extension of attribute-weighted AODE classification models. Overall, further exploration of attribute-weighted AODE classification models is a challenging but promising research direction. By delving into this area, we hope to bring innovative ideas and tools to research related to machine learning and data mining.

Abbreviations

The following abbreviations are used in this manuscript:

AODE	Averaged One-Dependence Estimators
ODE	One-Dependence Estimator
SWAODE	Model Selection-based Weighted AODE
NB	Naive Bayes
LOOCV	Leave-one-out cross-validation
KDB	K-dependent Bayesian
WAODE-MI	Weighted Average of One-Dependence Estimators by Mutual Information
WAODE-KL	Weighted Average of One-Dependence Estimators by Kullback–Leibler
SAODE	AODE under Leave-one-out cross-validation

Appendix A

Detailed results for 0–1 loss and LogLoss ± standard deviation are shown in Tables A1 and A2.

Table A1. ZOL.

Data Set	SWAODE	NB	KDB	AODE	WAODE-MI	WAODE-KL	SAODE
contact-lenses	0.3750+/−0.3425	0.3750+/−0.3425	0.2917+/−0.3543	0.4167+/−0.3574	0.3333+/−0.3581	0.3333+/−0.3581	0.3750+/−0.3425
lung-cancer	0.3750+/−0.3113	0.4375+/−0.2684	0.5938+/−0.3082	0.4688+/−0.2885	0.4688+/−0.2885	0.4688+/−0.2885	0.3750+/−0.3113
labor-negotiations	0.0702+/−0.0966	0.0351+/−0.0422	0.1053+/−0.1146	0.0526+/−0.0675	0.0702+/−0.0966	0.0877+/−0.1269	0.0702+/−0.0966
post-operative	0.2889+/−0.1741	0.3444+/−0.1966	0.3444+/−0.1748	0.3444+/−0.1882	0.3333+/−0.1401	0.3333+/−0.1401	0.2889+/−0.1741
zoo	0.0297+/−0.0600	0.0297+/−0.0477	0.0495+/−0.0614	0.0198+/−0.0384	0.0198+/−0.0384	0.0198+/−0.0384	0.0198+/−0.0384
promoters	0.0472+/−0.0748	0.0755+/−0.0617	0.1321+/−0.0891	0.1038+/−0.0648	0.0849+/−0.0656	0.0849+/−0.0656	0.0660+/−0.0992
echocardiogram	0.3511+/−0.1129	0.2748+/−0.1347	0.3664+/−0.1511	0.3435+/−0.1143	0.3359+/−0.1120	0.3282+/−0.1152	0.3664+/−0.1073
lymphography	0.1554+/−0.1129	0.1486+/−0.0979	0.1757+/−0.0791	0.1486+/−0.0991	0.1351+/−0.1056	0.1419+/−0.1026	0.1554+/−0.1183
iris	0.0600+/−0.0655	0.0733+/−0.0693	0.0733+/−0.0505	0.0600+/−0.0655	0.0600+/−0.0655	0.0600+/−0.0655	0.0600+/−0.0655
teaching-ae	0.4636+/−0.0918	0.5298+/−0.1579	0.4834+/−0.1079	0.4834+/−0.1179	0.4702+/−0.1214	0.4636+/−0.1186	0.4636+/−0.0918
hepatitis	0.2000+/−0.1144	0.1613+/−0.1151	0.2194+/−0.1205	0.1935+/−0.1244	0.1871+/−0.1201	0.1871+/−0.1201	0.2129+/−0.1244
wine	0.0281+/−0.0404	0.0225+/−0.0347	0.0674+/−0.0633	0.0281+/−0.0404	0.0281+/−0.0404	0.0281+/−0.0404	0.0225+/−0.0332
autos	0.1756+/−0.1420	0.3902+/−0.1648	0.2293+/−0.1374	0.2537+/−0.1104	0.2537+/−0.1216	0.2585+/−0.1207	0.1854+/−0.1376
sonar	0.1731+/−0.0978	0.2452+/−0.0889	0.2548+/−0.0914	0.1394+/−0.0888	0.1587+/−0.0849	0.1346+/−0.0918	0.1490+/−0.1027
glass-id	0.1869+/−0.0575	0.2570+/−0.1019	0.2383+/−0.0720	0.1589+/−0.0576	0.1636+/−0.0664	0.1636+/−0.0664	0.1776+/−0.0580
new-thyroid	0.0651+/−0.0410	0.0419+/−0.0487	0.0651+/−0.0454	0.0512+/−0.0544	0.0512+/−0.0468	0.0512+/−0.0468	0.0698+/−0.0492
audio	0.2301+/−0.0817	0.2389+/−0.0548	0.3097+/−0.1054	0.2301+/−0.0649	0.2345+/−0.0701	0.2434+/−0.0671	0.2345+/−0.0805
hungarian	0.1667+/−0.0520	0.1565+/−0.0698	0.2075+/−0.0625	0.1429+/−0.0676	0.1565+/−0.0773	0.1565+/−0.0773	0.1667+/−0.0667
heart-disease-c	0.1848+/−0.1062	0.1683+/−0.0803	0.2178+/−0.1428	0.1848+/−0.1067	0.1848+/−0.1022	0.1848+/−0.1022	0.1848+/−0.1054
haberman	0.2549+/−0.1070	0.2647+/−0.1285	0.2778+/−0.1024	0.2712+/−0.1188	0.2941+/−0.1152	0.2941+/−0.1152	0.2386+/−0.1068
primary-tumor	0.5221+/−0.1028	0.5162+/−0.0883	0.5841+/−0.1119	0.5162+/−0.0984	0.5251+/−0.0914	0.5251+/−0.0914	0.5133+/−0.1031
ionosphere	0.0798+/−0.0399	0.1197+/−0.0854	0.0684+/−0.0441	0.0826+/−0.0405	0.0826+/−0.0405	0.0826+/−0.0405	0.0798+/−0.0497
dermatology	0.0191+/−0.0310	0.0191+/−0.0242	0.0301+/−0.0258	0.0219+/−0.0275	0.0191+/−0.0282	0.0191+/−0.0282	0.0246+/−0.0318
horse-colic	0.1522+/−0.0627	0.2065+/−0.0928	0.2120+/−0.0615	0.2038+/−0.0590	0.1984+/−0.0591	0.1984+/−0.0591	0.1603+/−0.0596
house-votes-84	0.0552+/−0.0435	0.0943+/−0.0256	0.0690+/−0.0353	0.0529+/−0.0346	0.0506+/−0.0358	0.0506+/−0.0358	0.0552+/−0.0435
cylinder-bands	0.2167+/−0.0355	0.2093+/−0.0326	0.2074+/−0.0575	0.1611+/−0.0421	0.1574+/−0.0409	0.1574+/−0.0429	0.2167+/−0.0355
chess	0.0907+/−0.0500	0.1125+/−0.0551	0.0998+/−0.0354	0.1053+/−0.0631	0.1053+/−0.0598	0.0998+/−0.0613	0.0889+/−0.0515
syncon	0.0200+/−0.0136	0.0483+/−0.0398	0.0200+/−0.0156	0.0200+/−0.0163	0.0200+/−0.0163	0.0200+/−0.0163	0.0200+/−0.0136
balance-scale	0.1168+/−0.0119	0.0832+/−0.0207	0.1424+/−0.0307	0.1120+/−0.0159	0.1168+/−0.0119	0.1168+/−0.0119	0.1184+/−0.0174
soybean	0.0556+/−0.0191	0.0893+/−0.0244	0.0644+/−0.0205	0.0542+/−0.0184	0.0542+/−0.0184	0.0542+/−0.0184	0.0556+/−0.0191
credit-a	0.1217+/−0.0309	0.1449+/−0.0303	0.1696+/−0.0417	0.1261+/−0.0210	0.1203+/−0.0251	0.1203+/−0.0251	0.1261+/−0.0292
breast-cancer-w	0.0386+/−0.0275	0.0258+/−0.0223	0.0486+/−0.0181	0.0386+/−0.0248	0.0372+/−0.0235	0.0372+/−0.0235	0.0401+/−0.0274

Table A1. *Cont.*

Data Set	SWAODE	NB	KDB	AODE	WAODE-MI	WAODE-KL	SAODE
pima-ind-diabetes	0.2461+/−0.0655	0.2591+/−0.0707	0.2578+/−0.0583	0.2513+/−0.0636	0.2539+/−0.0663	0.2539+/−0.0663	0.2409+/−0.0584
vehicle	0.3132+/−0.0533	0.4090+/−0.0477	0.3026+/−0.0627	0.3132+/−0.0563	0.3156+/−0.0577	0.3156+/−0.0577	0.3109+/−0.0565
anneal	0.0601+/−0.0262	0.0891+/−0.0261	0.0445+/−0.0156	0.0735+/−0.0232	0.0646+/−0.0242	0.0646+/−0.0242	0.0512+/−0.0250
tic-tac-toe	0.2724+/−0.0406	0.3069+/−0.0427	0.2463+/−0.0382	0.2683+/−0.0432	0.2724+/−0.0406	0.2724+/−0.0406	0.2683+/−0.0432
vowel	0.1131+/−0.0274	0.4061+/−0.0557	0.2162+/−0.0272	0.0808+/−0.0296	0.1131+/−0.0274	0.1131+/−0.0274	0.0778+/−0.0283
german	0.2520+/−0.0451	0.2520+/−0.0325	0.2660+/−0.0634	0.2410+/−0.0535	0.2490+/−0.0474	0.2490+/−0.0474	0.2450+/−0.0515
led	0.2690+/−0.0621	0.2670+/−0.0622	0.2640+/−0.0603	0.2700+/−0.0604	0.2700+/−0.0604	0.2700+/−0.0604	0.2700+/−0.0630
contraceptive-mc	0.4691+/−0.0453	0.4949+/−0.0534	0.4684+/−0.0276	0.4671+/−0.0455	0.4596+/−0.0394	0.4582+/−0.0404	0.4684+/−0.0439
yeast	0.4239+/−0.0370	0.4245+/−0.0504	0.4394+/−0.0326	0.4205+/−0.0402	0.4218+/−0.0385	0.4225+/−0.0378	0.4245+/−0.0400
volcanoes	0.3362+/−0.0287	0.3421+/−0.0278	0.3520+/−0.0258	0.3539+/−0.0331	0.3539+/−0.0340	0.3539+/−0.0340	0.3467+/−0.0292
car	0.1053+/−0.0244	0.1400+/−0.0255	0.0567+/−0.0182	0.0845+/−0.0193	0.0909+/−0.0183	0.0920+/−0.0173	0.0793+/−0.0181
segment	0.0515+/−0.0084	0.1476+/−0.0245	0.0567+/−0.0158	0.0563+/−0.0091	0.0550+/−0.0078	0.0550+/−0.0078	0.0519+/−0.0079
hypothyroid	0.0278+/−0.0105	0.0360+/−0.0112	0.0338+/−0.0137	0.0348+/−0.0118	0.0294+/−0.0104	0.0297+/−0.0102	0.0278+/−0.0105
splice-c4.5	0.0318+/−0.0072	0.0444+/−0.0112	0.0482+/−0.0152	0.0375+/−0.0087	0.0387+/−0.0101	0.0387+/−0.0101	0.0334+/−0.0102
kr-vs-kp	0.0569+/−0.0125	0.1214+/−0.0217	0.0544+/−0.0171	0.0854+/−0.0187	0.0582+/−0.0115	0.0582+/−0.0115	0.0573+/−0.0109
abalone	0.4556+/−0.0206	0.4893+/−0.0249	0.4656+/−0.0237	0.4551+/−0.0214	0.4549+/−0.0212	0.4549+/−0.0212	0.4558+/−0.0208
spambase	0.0602+/−0.0115	0.1050+/−0.0149	0.0702+/−0.0121	0.0635+/−0.0114	0.0606+/−0.0112	0.0602+/−0.0115	0.0646+/−0.0138
phoneme	0.1843+/−0.0177	0.2615+/−0.0129	0.2120+/−0.0123	0.2100+/−0.0144	0.2008+/−0.0139	0.2010+/−0.0145	0.1863+/−0.0155
wall-following	0.0843+/−0.0099	0.1743+/−0.0149	0.1043+/−0.0094	0.1514+/−0.0101	0.1503+/−0.0099	0.1503+/−0.0099	0.0845+/−0.0097
page-blocks	0.0479+/−0.0075	0.1376+/−0.0126	0.0590+/−0.0102	0.0502+/−0.0066	0.0495+/−0.0062	0.0495+/−0.0062	0.0477+/−0.0077
optdigits	0.0274+/−0.0083	0.0861+/−0.0124	0.0454+/−0.0070	0.0283+/−0.0095	0.0285+/−0.0093	0.0286+/−0.0093	0.0281+/−0.0087
satellite	0.1175+/−0.0104	0.2022+/−0.0168	0.1392+/−0.0135	0.1301+/−0.0131	0.1298+/−0.0125	0.1298+/−0.0125	0.1175+/−0.0106
musk2	0.1115+/−0.0138	0.2496+/−0.0101	0.0867+/−0.0097	0.1511+/−0.0101	0.1520+/−0.0095	0.1514+/−0.0098	0.1097+/−0.0138
mushrooms	0.0000+/−0.0000	0.0196+/−0.0036	0.0006+/−0.0009	0.0002+/−0.0005	0.0000+/−0.0000	0.0000+/−0.0000	0.0001+/−0.0004
thyroid	0.2211+/−0.0126	0.2754+/−0.0152	0.2319+/−0.0146	0.2421+/−0.0136	0.2333+/−0.0129	0.2332+/−0.0128	0.2213+/−0.0104
pendigits	0.0252+/−0.0029	0.1447+/−0.0112	0.0529+/−0.0066	0.0254+/−0.0029	0.0251+/−0.0029	0.0251+/−0.0029	0.0253+/−0.0029
sign	0.2957+/−0.0083	0.3851+/−0.0114	0.3055+/−0.0140	0.2960+/−0.0119	0.2977+/−0.0090	0.2977+/−0.0090	0.2936+/−0.0110
nursery	0.0713+/−0.0063	0.0973+/−0.0066	0.0654+/−0.0061	0.0733+/−0.0059	0.0708+/−0.0065	0.0708+/−0.0065	0.0707+/−0.0058
magic	0.1825+/−0.0081	0.2478+/−0.0118	0.1759+/−0.0107	0.1726+/−0.0084	0.1825+/−0.0081	0.1825+/−0.0081	0.1721+/−0.0082
letter-recog	0.1439+/−0.0107	0.3226+/−0.0110	0.1920+/−0.0112	0.1514+/−0.0089	0.1440+/−0.0105	0.1440+/−0.0105	0.1452+/−0.0089
adult	0.1631+/−0.0047	0.1809+/−0.0050	0.1638+/−0.0044	0.1679+/−0.0032	0.1640+/−0.0048	0.1640+/−0.0047	0.1631+/−0.0050
shuttle	0.0095+/−0.0012	0.0311+/−0.0022	0.0163+/−0.0012	0.0101+/−0.0010	0.0093+/−0.0010	0.0093+/−0.0010	0.0095+/−0.0012
connect-4	0.2407+/−0.0039	0.2783+/−0.0059	0.2406+/−0.0030	0.2422+/−0.0047	0.2408+/−0.0039	0.2407+/−0.0039	0.2421+/−0.0048
waveform	0.0339+/−0.0009	0.0432+/−0.0018	0.0396+/−0.0021	0.0343+/−0.0008	0.0343+/−0.0009	0.0343+/−0.0009	0.0338+/−0.0009
localization	0.4556+/−0.0033	0.5449+/−0.0026	0.4642+/−0.0040	0.4333+/−0.0027	0.4314+/−0.0036	0.4314+/−0.0036	0.4556+/−0.0033
census-income	0.0555+/−0.0010	0.2410+/−0.0017	0.0667+/−0.0014	0.1106+/−0.0015	0.0990+/−0.0018	0.0990+/−0.0018	0.0555+/−0.0009
poker-hand	0.3302+/−0.0022	0.4988+/−0.0018	0.3291+/−0.0012	0.4812+/−0.0028	0.1758+/−0.0079	0.1757+/−0.0078	0.3302+/−0.0022
donation	0.0002+/−0.0000	0.0002+/−0.0000	0.0001+/−0.0000	0.0002+/−0.0000	0.0002+/−0.0000	0.0002+/−0.0000	0.0002+/−0.0000

Table A2. LogLoss.

Data Set	SWAODE	NB	KDB	AODE	WAODE-MI	WAODE-KL	SAODE
contact−lenses	0.8874+/−0.8460	1.0171+/−0.8353	1.0277+/−0.7003	1.1270+/−0.8317	1.0118+/−0.8291	1.0015+/−0.8196	0.9293+/−0.8631
lung−cancer	1.9531+/−1.6732	4.6187+/−7.0330	6.7035+/−4.9708	4.5050+/−6.5417	4.5673+/−6.4765	4.5719+/−6.4657	1.9683+/−1.6907
labor−negotiations	0.2764+/−0.3196	0.1463+/−0.1563	0.5502+/−0.4565	0.2172+/−0.2491	0.2435+/−0.2799	0.2528+/−0.2913	0.2402+/−0.2765
post−operative	1.1787+/−0.5878	1.2723+/−0.8020	1.2896+/−0.6286	1.2278+/−0.6653	1.2174+/−0.6698	1.2142+/−0.6689	1.1865+/−0.5906
zoo	0.0801+/−0.0913	0.1111+/−0.0854	0.1624+/−0.1633	0.0803+/−0.0823	0.0746+/−0.0781	0.0753+/−0.0785	0.0803+/−0.0922
promoters	0.1944+/−0.2149	0.3347+/−0.3033	0.9880+/−1.3047	0.3969+/−0.2083	0.4091+/−0.2738	0.4097+/−0.2736	0.1970+/−0.2263
echocardiogram	0.9884+/−0.1735	0.9687+/−0.4870	1.5034+/−1.0267	1.0943+/−0.6294	1.1142+/−0.6790	1.1137+/−0.6767	0.9764+/−0.1816
lymphography	0.6838+/−0.5628	0.6465+/−0.6171	0.8154+/−0.4996	0.5657+/−0.5303	0.5665+/−0.5147	0.5651+/−0.5117	0.6847+/−0.5765
iris	0.2284+/−0.1885	0.3460+/−0.3011	0.2454+/−0.2043	0.2319+/−0.1996	0.2296+/−0.1926	0.2297+/−0.1927	0.2306+/−0.1897
teaching−ae	2.1672+/−0.6669	2.1000+/−0.6756	2.1076+/−0.6395	1.9223+/−0.5151	1.9909+/−0.5181	1.9754+/−0.5172	2.1666+/−0.6675
hepatitis	0.7173+/−0.5595	0.9701+/−0.8161	0.9867+/−0.6371	0.7285+/−0.5726	0.7432+/−0.6007	0.7414+/−0.6012	0.7980+/−0.6568
wine	0.1567+/−0.1901	0.1304+/−0.1976	0.2670+/−0.2300	0.1314+/−0.1795	0.1325+/−0.1774	0.1327+/−0.1776	0.1204+/−0.1321
autos	1.4860+/−1.9171	4.2030+/−2.7437	4.8262+/−4.2331	3.2524+/−3.2344	3.2625+/−3.2916	3.2552+/−3.2922	1.5943+/−1.8917
sonar	1.1577+/−0.7019	1.6809+/−1.1193	1.8069+/−0.7765	1.0254+/−0.7477	1.2091+/−0.8276	1.0368+/−0.7248	1.1754+/−0.7230
glass−id	0.7369+/−0.3984	1.0000+/−0.3915	0.9401+/−0.3718	0.6229+/−0.2004	0.6192+/−0.1978	0.6193+/−0.1975	0.7352+/−0.4003
new−thyroid	0.3004+/−0.2133	0.2465+/−0.2526	0.3084+/−0.2155	0.2648+/−0.1762	0.2620+/−0.1801	0.2619+/−0.1799	0.3019+/−0.2100
audio	2.2635+/−1.4149	3.9563+/−2.6628	5.3522+/−2.3274	3.9528+/−2.6879	3.9795+/−2.6823	3.9806+/−2.6828	2.2886+/−1.4082
hungarian	0.5854+/−0.2865	0.8202+/−0.4467	0.7913+/−0.4361	0.6276+/−0.3111	0.5994+/−0.2900	0.5995+/−0.2902	0.6182+/−0.2790
heart−disease−c	0.6624+/−0.2799	0.7119+/−0.3646	0.9289+/−0.4819	0.6468+/−0.3014	0.6434+/−0.2982	0.6433+/−0.2981	0.6548+/−0.2812
haberman	0.7724+/−0.2210	0.7815+/−0.2614	0.8572+/−0.2611	0.8325+/−0.2585	0.8348+/−0.2658	0.8349+/−0.2659	0.7700+/−0.2192
primary−tumor	2.8134+/−0.5805	2.9163+/−0.6153	3.3812+/−0.7186	2.8284+/−0.5753	2.8250+/−0.5777	2.8249+/−0.5776	2.8192+/−0.5803
ionosphere	0.7014+/−0.4000	1.5528+/−0.9964	0.7280+/−0.6498	0.9810+/−0.5568	0.9590+/−0.5437	0.9591+/−0.5439	0.6841+/−0.3761
dermatology	0.0762+/−0.0778	0.0588+/−0.0654	0.1170+/−0.0991	0.0624+/−0.0689	0.0615+/−0.0694	0.0616+/−0.0694	0.0890+/−0.0782
horse−colic	0.6230+/−0.1680	1.2551+/−0.4164	1.2111+/−0.4258	0.8826+/−0.3060	0.8699+/−0.2724	0.8696+/−0.2718	0.6366+/−0.1638
house−votes−84	0.2481+/−0.2320	0.9110+/−0.4323	0.2866+/−0.2091	0.2513+/−0.2617	0.2402+/−0.2647	0.2402+/−0.2648	0.2500+/−0.2268
cylinder−bands	1.9149+/−0.8050	1.6171+/−0.2745	2.9088+/−0.8703	1.1335+/−0.3156	1.1736+/−0.3168	1.1321+/−0.3167	1.9137+/−0.8063
chess	0.3455+/−0.0948	0.4057+/−0.1043	0.3380+/−0.0931	0.3843+/−0.0956	0.3612+/−0.0857	0.3581+/−0.0841	0.3397+/−0.0791
syncon	0.0911+/−0.0780	0.4910+/−0.4111	0.1593+/−0.1696	0.0907+/−0.0663	0.0888+/−0.0657	0.0888+/−0.0656	0.0908+/−0.0803
balance−scale	0.8296+/−0.0975	0.7287+/−0.0691	0.8618+/−0.0978	0.8271+/−0.0987	0.8296+/−0.0975	0.8296+/−0.0975	0.8321+/−0.0948
soybean	0.1860+/−0.0681	1.0345+/−0.5277	0.2515+/−0.1666	0.2741+/−0.0997	0.2596+/−0.0907	0.2596+/−0.0907	0.1860+/−0.0681
credit−a	0.5354+/−0.1861	0.6433+/−0.2210	0.7901+/−0.2521	0.5482+/−0.1860	0.5379+/−0.1793	0.5377+/−0.1795	0.5231+/−0.1862
breast−cancer−w	0.2096+/−0.1981	0.4577+/−0.4431	0.2955+/−0.2811	0.2209+/−0.2063	0.2183+/−0.2007	0.2181+/−0.2005	0.2141+/−0.1975
pima−ind−diabetes	0.7112+/−0.1365	0.7868+/−0.1729	0.7983+/−0.2034	0.7293+/−0.1559	0.7312+/−0.1482	0.7311+/−0.1482	0.7065+/−0.1371
vehicle	0.9724+/−0.1347	3.1607+/−0.6142	0.9929+/−0.1886	1.0031+/−0.1559	1.0077+/−0.1587	1.0076+/−0.1586	0.9761+/−0.1338
anneal	0.2316+/−0.1124	0.5108+/−0.1970	0.1882+/−0.0953	0.2794+/−0.1146	0.2450+/−0.1127	0.2446+/−0.1126	0.2183+/−0.1098

Table A2. *Cont.*

Data Set	SWAODE	NB	KDB	AODE	WAODE-MI	WAODE-KL	SAODE
tic−tac−toe	0.7191+/−0.0543	0.7854+/−0.0616	0.7077+/−0.0680	0.6953+/−0.0542	0.7191+/−0.0543	0.7191+/−0.0543	0.6953+/−0.0542
vowel	0.4498+/−0.1249	1.5849+/−0.1954	1.0296+/−0.1684	0.3227+/−0.1028	0.4498+/−0.1247	0.4504+/−0.1247	0.3176+/−0.1226
german	0.7635+/−0.1002	0.7690+/−0.1040	0.8958+/−0.1954	0.7613+/−0.0983	0.7632+/−0.0980	0.7632+/−0.0981	0.7509+/−0.0999
led	1.1813+/−0.1834	1.1759+/−0.1870	1.2015+/−0.1877	1.1806+/−0.1839	1.1805+/−0.1832	1.1805+/−0.1832	1.1816+/−0.1841
contraceptive−mc	1.4203+/−0.0854	1.5016+/−0.1295	1.4185+/−0.0813	1.4044+/−0.0890	1.3988+/−0.0860	1.3988+/−0.0860	1.4233+/−0.0885
yeast	1.6929+/−0.1452	1.7185+/−0.1370	1.8312+/−0.1735	1.6864+/−0.1362	1.6899+/−0.1411	1.6901+/−0.1412	1.6889+/−0.1430
volcanoes	1.1081+/−0.0618	1.1167+/−0.0756	1.1341+/−0.0726	1.1177+/−0.0731	1.1353+/−0.0822	1.1353+/−0.0821	1.1170+/−0.0623
car	0.3879+/−0.0277	0.4640+/−0.0340	0.2661+/−0.0321	0.3988+/−0.0323	0.3854+/−0.0299	0.3857+/−0.0299	0.3720+/−0.0310
segment	0.2568+/−0.0566	1.0099+/−0.2586	0.2876+/−0.0707	0.2620+/−0.0599	0.2630+/−0.0554	0.2630+/−0.0554	0.2577+/−0.0570
hypothyroid	0.0901+/−0.0263	0.1892+/−0.0525	0.1110+/−0.0393	0.1297+/−0.0377	0.0975+/−0.0302	0.0976+/−0.0302	0.0901+/−0.0263
splice−c4.5	0.1661+/−0.0350	0.2111+/−0.0613	0.2206+/−0.0575	0.1687+/−0.0395	0.1684+/−0.0385	0.1684+/−0.0385	0.1676+/−0.0367
kr−vs−kp	0.2394+/−0.0229	0.4199+/−0.0339	0.2386+/−0.0457	0.3463+/−0.0291	0.2899+/−0.0225	0.2897+/−0.0225	0.2400+/−0.0225
abalone	1.2628+/−0.0378	2.6815+/−0.2753	1.2791+/−0.0392	1.2643+/−0.0381	1.2629+/−0.0377	1.2629+/−0.0377	1.2642+/−0.0382
spambase	0.3326+/−0.0927	0.8490+/−0.1867	0.3938+/−0.1151	0.3535+/−0.0958	0.3663+/−0.1143	0.3328+/−0.0927	0.3527+/−0.1057
phoneme	0.9483+/−0.1088	1.4351+/−0.0936	1.3346+/−0.1252	1.1686+/−0.0663	1.1014+/−0.0690	1.1008+/−0.0691	0.9509+/−0.1096
wall−following	0.2769+/−0.0238	1.6069+/−0.1649	0.5949+/−0.0691	1.1436+/−0.1329	1.1227+/−0.1329	1.1228+/−0.1329	0.2782+/−0.0238
page−blocks	0.1968+/−0.0417	0.7670+/−0.0913	0.2991+/−0.0834	0.2219+/−0.0471	0.2179+/−0.0462	0.2179+/−0.0462	0.1967+/−0.0417
optdigits	0.1853+/−0.0759	0.9326+/−0.1575	0.3560+/−0.1081	0.1942+/−0.0772	0.1917+/−0.0779	0.1917+/−0.0779	0.1865+/−0.0763
satellite	0.6644+/−0.0841	5.3687+/−0.5379	1.0206+/−0.1639	0.8222+/−0.1142	0.8188+/−0.1139	0.8189+/−0.1139	0.6674+/−0.0853
musk2	0.3730+/−0.0301	6.9568+/−0.4979	1.5495+/−0.2119	3.9347+/−0.4082	3.7331+/−0.3837	3.9152+/−0.4121	0.3723+/−0.0298
mushrooms	0.0003+/−0.0004	0.0913+/−0.0229	0.0019+/−0.0036	0.0005+/−0.0009	0.0003+/−0.0004	0.0003+/−0.0004	0.0004+/−0.0007
thyroid	0.7717+/−0.0436	1.7390+/−0.1826	0.8803+/−0.0753	0.8960+/−0.0608	0.8424+/−0.0583	0.8423+/−0.0583	0.7733+/−0.0435
pendigits	0.1204+/−0.0152	1.1452+/−0.0962	0.2674+/−0.0439	0.1204+/−0.0152	0.1203+/−0.0152	0.1203+/−0.0152	0.1205+/−0.0152
sign	0.9674+/−0.0183	1.2576+/−0.0242	1.0335+/−0.0342	0.9621+/−0.0185	0.9674+/−0.0184	0.9674+/−0.0184	0.9560+/−0.0193
nursery	0.3096+/−0.0111	0.3766+/−0.0121	0.2274+/−0.0120	0.3136+/−0.0096	0.3104+/−0.0109	0.3104+/−0.0109	0.2765+/−0.0108
magic	0.5786+/−0.0248	0.7345+/−0.0296	0.5755+/−0.0201	0.5624+/−0.0244	0.5786+/−0.0248	0.5786+/−0.0248	0.5609+/−0.0234
letter−recog	0.6486+/−0.0327	1.9090+/−0.0682	1.0277+/−0.0508	0.6935+/−0.0358	0.6486+/−0.0328	0.6486+/−0.0328	0.6521+/−0.0342
adult	0.5264+/−0.0144	0.6728+/−0.0200	0.5035+/−0.0125	0.5614+/−0.0123	0.5407+/−0.0115	0.5407+/−0.0115	0.5281+/−0.0136
shuttle	0.0506+/−0.0036	0.1404+/−0.0051	0.0592+/−0.0051	0.0540+/−0.0037	0.0496+/−0.0035	0.0496+/−0.0035	0.0512+/−0.0036
connect−4	0.8693+/−0.0059	0.9840+/−0.0102	0.8600+/−0.0081	0.8766+/−0.0056	0.8694+/−0.0059	0.8694+/−0.0059	0.8753+/−0.0056
waveform	0.0993+/−0.0023	0.5733+/−0.0223	0.1312+/−0.0111	0.1015+/−0.0027	0.1012+/−0.0027	0.1012+/−0.0027	0.0992+/−0.0022
localization	1.8528+/−0.0098	2.1440+/−0.0054	1.8267+/−0.0107	1.7891+/−0.0083	1.7824+/−0.0094	1.7824+/−0.0094	1.8528+/−0.0098
census−income	0.2131+/−0.0027	1.9789+/−0.0172	0.2467+/−0.0058	0.4898+/−0.0062	0.4086+/−0.0050	0.4086+/−0.0050	0.2132+/−0.0027
poker−hand	1.0977+/−0.0048	1.4158+/−0.0048	1.0821+/−0.0027	1.2089+/−0.0034	1.0865+/−0.0031	1.0865+/−0.0030	1.0977+/−0.0048
donation	0.0006+/−0.0001	0.0009+/−0.0001	0.0004+/−0.0001	0.0007+/−0.0001	0.0007+/−0.0001	0.0007+/−0.0001	0.0005+/−0.0001

References

1. Wu, X.; Kumar, V.; Quinlan, J.R.; Ghosh, J.; Yang, Q.; Motoda, H.; McLachlan, G.J.; Ng, A.; Liu, B.; Yu, P.S.; et al. Top 10 algorithms in data mining. *Knowl. Inf. Syst.* **2008**, *14*, 1–37. [CrossRef]
2. Halbersberg, D.; Wienreb, M.; Lerner, B. Joint maximization of accuracy and information for learning the structure of a Bayesian network classifier. *Mach. Learn.* **2020**, *109*, 1039–1099. [CrossRef]
3. Zhang, W.; Zhang, Z.; Chao, H.C.; Tseng, F.H. Kernel mixture model for probability density estimation in Bayesian classifiers. *Data Min. Knowl. Discov.* **2018**, *32*, 675–707. [CrossRef]
4. Jiang, L.; Zhang, L.; Li, C.; Wu, J. A correlation-based feature weighting filter for naive Bayes. *IEEE Trans. Knowl. Data Eng.* **2019**, *31*, 201–213. [CrossRef]
5. Webb, G.I.; Boughton, J.R.; Wang, Z. Not so naive Bayes: Aggregating one-dependence estimators. *Mach. Learn.* **2005**, *58*, 5–24. [CrossRef]
6. Webb, G.I.; Boughton, J.R.; Zheng, F.; Ting, K.M.; Salem, H. Learning by extrapolation from marginal to full-multivariate probability distributions: Decreasingly Naive Bayesian classification. *Mach. Learn.* **2012**, *86*, 233–272. [CrossRef]
7. Gelfand, A.E.; Dey, D.K. Bayesian model choice: Asymp-totics and exact calculations. *J. R. Stat. Soc. Ser. B* **1994**, *56*, 501–514. [CrossRef]
8. Chen, S.; Webb, G.I.; Liu, L.; Ma, X. A novel selective naïve Bayes algo-rithm. *Knowl.-Based Syst.* **2020**, *192*, 105361. [CrossRef]
9. Dua, D.; Graff, C. UCI Machine Learning Repository. Available online: http://archive.ics.uci.edu/ml (accessed on 8 June 2024).
10. Jiang, L.; Zhang, H. Weightily averaged one-dependence estimators. In Proceedings of the 9th Pacific Rim International Conference on Artificial Intelligence, Guilin, China, 7–11 August 2006; pp. 970–974.
11. Jiang, L.; Zhang, H.; Cai, Z.; Wang, D. Weighted average of one-dependence estimators†. *J. Exp. Theor. Artif. Intell.* **2012**, *24*, 219-230. [CrossRef]
12. Wu, J.; Pan, S.; Zhu, X.; Zhang, P.; Zhang, C. Sode: Self-adap-tive one-dependence estimators for classification. *Pattern Recognit.* **2016**, *51*, 358–377. [CrossRef]
13. Zheng, F.; Webb, G.I. Finding the right family: Parent and child selection for averaged one-dependence estimators. In Proceedings of the 18th European Conference on Machine Learning, Warsaw, Poland, 17–21 September 2007; pp. 490–501.
14. Yang, Y.; Webb, G.I.; Cerquides, J.; Korb, K.B.; Boughton, J.; Ting, K.M. To select or to weigh: A comparative study of linear combination schemes for superparent-one-dependence estimators. *IEEE Trans. Knowl. Data Eng.* **2007**, *19*, 1652–1665. [CrossRef]
15. Yang, Y.; Korb, K.; Ting, K.-M.; Webb, G. Ensemble selection for su-perparent-one-dependence estimators. In Proceedings of the 18th Australian Joint Conference on Artificial Intelligence, Sydney, Australia, 5–9 December 2005; pp. 102–111.
16. Chen, S.; Martinez, A.M.; Webb, G.I. *Highly Scalable Attribute Selection for Averaged One-Dependence Estimators*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 86–97.
17. Friedman, N.; Geiger, D.; Goldszmidt, M. Bayesian network classifiers. *Mach. Learn.* **1997**, *29*, 131–163. [CrossRef]
18. Sahami, M. Learning limited dependence Bayesian classifiers. In Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, Portland, OR, USA, 2–4 August 1996; ACM: New York, NY, USA, 1996; pp. 335–338.
19. Chen, S.; Martínez, A.M.; Webb, G.I.; Wang, L. Sample-based attribute selective AnDE for large data. *IEEE Trans. Knowl. Data Eng.* **2017**, *29*, 172–185. [CrossRef]
20. Witten, I.H.; Frank, E.; Trigg, L.; Hall, M.A.; Holmes, G.; Cunningham, S.J. Weka: Practical Machine Learning Tools and Techniques with Java Implementations. *Acm. Sigmod. Record.* **1999**, *31*, 76–77. [CrossRef]
21. Chen, S.; Gao, X.; Zhuo, C.; Zhu, C. Research on Averaged One-Dependence Estimators Classification Algorithm Based on Divergence Weighting. *J. Nanjing Univ. Sci. Technol.* **2024**, *48*.