Journal of Computers Vol. 37 No. 2, April 2026, pp. 85-98
https://doi.org/10.63367/199115992026043702006

# Analysis of Teaching Competence Evaluation Based on Bayesian Network

Liangjun Yu[1,2*], Xian Zhou[1,3], and Wanjuan Song[1,2]

[1] College of Computer, Hubei University of Education,
Wuhan 430205, China
yuliangjun@hue.edu.cn

[2] Hubei Provincial Collaborative Innovation Center for Digitalization of Basic Education,
Wuhan 430205, China
key_swj@163.com

[3] College of Physics and Electronic Science, Hubei Normal of University,
Huangshi 435002, China
zhouxian@hue.edu.cn

*Received 16 June 2025; Revised 9 October 2025; Accepted 16 October 2025*

**Abstract.** This paper proposes a hybrid modeling method that employs dual weighting of attributes and instances for the analysis and prediction of the teacher evaluation data. This method extends the Averaged One-Dependence Estimator (AODE) model by incorporating attribute weights and instance weights. Instance weighting integrates instance weights into the calculation of conditional probabilities, while attribute weighting incorporates attribute weights into the classification formula of the classifier. The improved weighted model, attribute and instance weighted one-dependence estimator (AIWODE), combines an attribute weighting method based on information gain with an instance weighting method based on the mode of the instance set and the similarity to other instances. The experimental data is derived from teacher recording data within the Hubei Province AI-Assisted Teacher Project. By comparing various Bayesian approaches and employing ten-fold cross-validation to validate model performance, the experiments assessed four evaluation metrics: accuracy, recall, AUC, and F1 score. Experimental results show that AIWODE, with its dual weighting mechanism, significantly outperforms existing methods in teaching evaluation tasks.

**Keywords:** attribute weighting, instance weighting, information gain, Bayesian network

## 1 Introduction

Teacher evaluation plays a pivotal role in managing education quality, offering systematic insights into teaching effectiveness through scientifically grounded and objective methods. Existing approaches to evaluation often rely on subjective methods, such as student feedback, peer assessments, and administrative observations. Although these methods can partially reflect teaching quality, their inherent subjectivity often raises concerns about the objectivity and fairness of the evaluation outcomes.

With the rapid advancements in information technology and the introduction of the Ministry of Education's "AI + Teacher Development" initiative, artificial intelligence has increasingly permeated the education sector, creating both opportunities and challenges for evaluating teaching performance. Scholars are actively exploring the potential of data mining and machine learning techniques to enhance the scientific rigor and objectivity of teacher evaluations. For example, Bayesian networks [1] have been used to construct directed acyclic graphs and assign weights to various evaluation indicators, enabling a comprehensive assessment of factors influencing teaching performance. Decision tree approaches [2] have been employed to analyze the characteristics of teachers and students, offering deeper insights into the relationships between teaching evaluation metrics and their underlying attributes. Similarly, BP neural networks [3] have demonstrated their capacity to process diverse data inputs and generate composite scores for teaching performance, facilitating precise and objective assessments of teaching effectiveness. Furthermore, numerous machine learning approaches have been extensively applied in teacher evaluation tasks. Among them, naive Bayes stands out for its stable classification performance, simple

* Corresponding Author

structure, and robustness against missing data. This approach has seen widespread application in areas such as text classification [4], fault diagnosis [5], and spam detection [6, 7].

This study proposed the improved weighted model, attribute and instance weighted one-dependence estimator (AIWODE), which analyzes an educational dataset comprising multiple attribute dimensions, offering a more comprehensive and objective evaluation of teachers' instructional capabilities. The dataset encompasses eight key attributes: teachers' gaze distribution ratio, classroom movement trajectories, attention coverage of students, students' classroom behaviors, questioning strategies, speech rate, verbal emotional tone, and others. By analyzing these diverse attributes, this approach not only introduces a more scientifically grounded method for evaluating teaching performance but also addresses the limitations of existing evaluation methods, providing actionable insights and support for improving future teaching practices.

## 2 Attribute and Instance Weighted One-dependence Estimator

### 2.1 The Averaged One-Dependence Estimator Model

The Averaged One-Dependence Estimator (AODE) [8] is an improved Bayesian network model enhanced through structural extensions. The core concept of this method is that each attribute node learns a "special" tree-augmented Naive Bayes classifier [9], which is then treated as the parent node of other attribute nodes and subsequently averaged. Assuming there are four attribute nodes, the structure of the Averaged One-Dependence Estimator model is illustrated in Fig. 1:

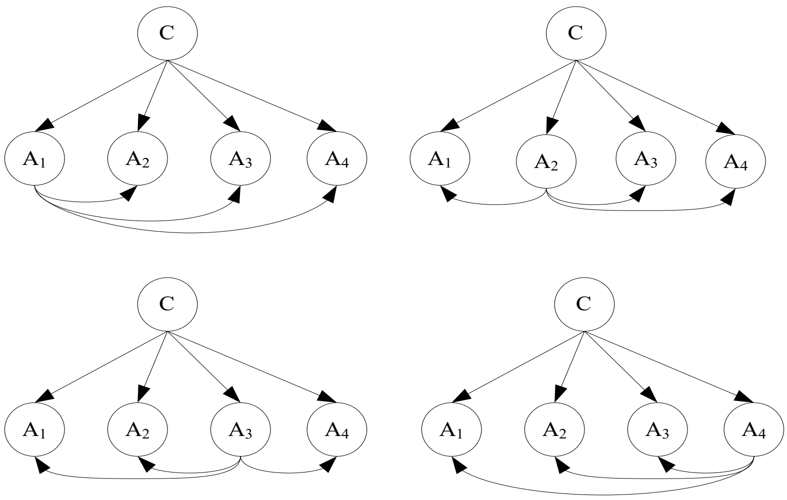


**Fig. 1.** Structure diagram of the averaged one-dependence estimator model

As shown in the Fig. 1, the model represents four attribute variables $A_1$, $A_2$, $A_3$, $A_4$, and a class variable $C$. In all models, one of the attribute nodes acts as the parent node for all other attribute nodes. For a given test instance $x$, its classification formula is expressed as formula (1):

$$c(x) = \arg\max_{c \in C} \frac{\sum_{i=1}^{m} P(a_i, c) \prod_{s=1 \wedge j \neq i}^{m} P(a_s \mid a_i, c)}{numParent} \quad (1)$$

In formula (1), *numParent* represents the number of tree-augmented Naive Bayes classifiers that meet the given conditions. Additionally, Laplace smoothing [10] is applied when calculating the prior probability $P(a_i, c)$ and conditional probability $P(a_s | a_i, c)$, which are expressed in formula (2) and formula (3), respectively:

$$P\left(a_i, c\right) = \frac{\sum_{i=1}^{n} \delta\left(a_{ti}, a_i\right) \delta\left(c_t, c\right) + 1}{n + n_i * q} \tag{2}$$

$$P(a_s \mid a_i, c) = \frac{\sum_{t=1}^{n} \delta\left(a_{ti}, a_i\right) \delta\left(a_{ts}, a_s\right) \delta\left(c_t, c\right) + 1}{\sum_{t=1}^{n} \delta\left(a_{ti}, a_i\right) \delta\left(c_t, c\right) + n_s} \tag{3}$$

Where $m$ is the number of attributes, $n$ represents the number of instances, $c_t$ denotes the class label of the t-th training instance, $q$ is the number of classes, and $\delta(\bullet)$ is a binary function. $a_{ti}$ represents the i-th attribute value of the t-th training instance, $n_i$ denotes the number of possible values for the root attribute $A_i$, and $n_s$ represents the number of possible values for its child attribute $A_s$.

### 2.2 The Attribute and Instance Weighted One-Dependence Estimator Model

Although attribute weighting augments attribute relevance and instance weighting tackles data imbalance, previous models have predominantly concentrated on a single aspect. AIWODE integrates both to capture both attribute importance and instance representativeness, leading to superior performance.

Building upon the AODE model, this paper proposes the attribute and instance weighted one-dependence estimator (AIWODE), which combines attribute weighting based on information gain [11] with instance weighting based on the mode of the instance set and the similarity between each training instance. As shown in Fig. 2. This model integrates these two methods to more effectively handle the weighting of different attributes and instances. By optimizing two weighting approaches, AIWODE outperforms models that rely solely on either attribute weighting [12-14] or instance weighting [15, 16], demonstrating superior performance in classification tasks.

Assuming there are four attribute nodes, denoted as $A_1$, $A_2$, $A_3$, $A_4$, the structure of the AIWODE model is proposed in Fig. 2:

In Fig. 2, the weight of each instance is represented as $w_1^{ins}$, $w_2^{ins}$, ..., $w_3^{ins}$, and the weight of each attribute is also indicated as $w_i^{att}$. Given a test instance set $x$, its classification formula is as follows:

$$c\left(x\right) = \arg\max_{c \in C} \frac{\sum_{i=1}^{m} w_i^{att} P\left(a_i, c\right) \prod_{s=1 \wedge s \neq i}^{m} P(a_s \mid a_i, c)}{\sum_{i}^{m} w_i^{att}} \tag{4}$$

The prior probability $P(a_i, c)$ and the conditional probability $P(a_s | a_i, c)$ are redefined in formulas (5) and (6), respectively:

$$P\left(a_i, c\right) = \frac{\sum_{i=1}^{n} w_t^{ins} \delta\left(a_{ti}, a_i\right) \delta\left(c_t, c\right) + 1}{\sum_{t=1}^{n} w_t^{ins} + n_i * q} \tag{5}$$

$$P(a_i|a_i, c) = \frac{\sum_{t=1}^{n} w_t^{ins} \delta\left(a_{ti}, a_i\right) \delta\left(a_{ts}, a_s\right) \delta\left(c_t, c\right) + 1}{\sum_{t=1}^{n} w_t^{ins} \delta\left(a_{ti}, a_i\right) \delta\left(c_t, c\right) + n_s} \tag{6}$$

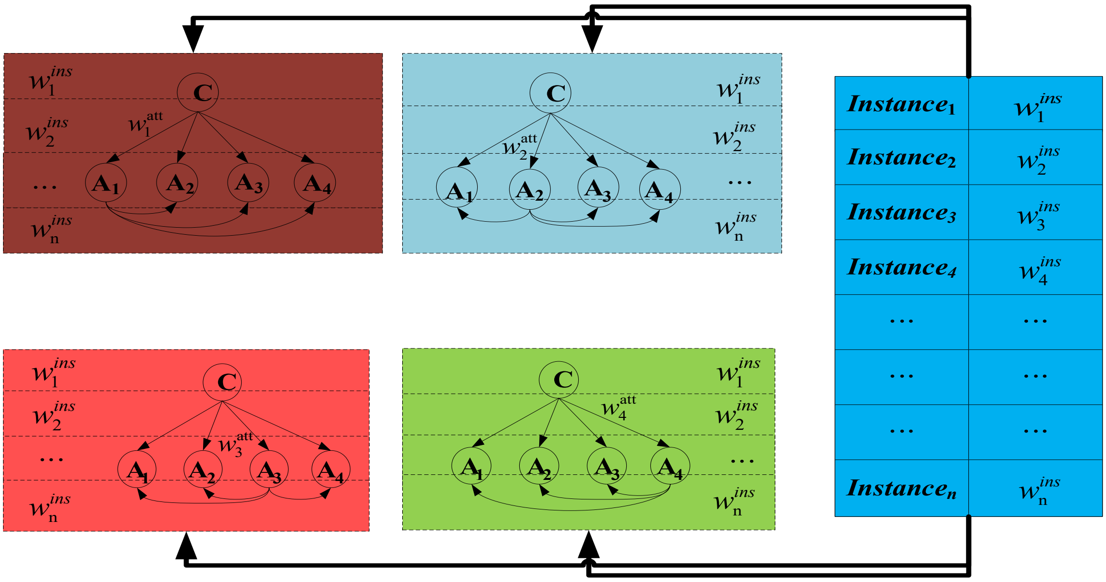


**Fig. 2.** Structure diagram of the AIWODE model

**A. Calculation of Attribute Weights**

In our AIWODE approach, information gain is adopted to calculate attribute weights. Information gain is widely used in attribute selection to evaluate the contribution of attributes to classification tasks [17]. The formula for information gain is as follows:

$$H(C)-H(C\mid A_i)=\sum_{a_i}P(a_i)\sum_{c}P(c\mid a_i)\log P(c\mid a_i)-\sum_{c}P(c)\log P(c) \tag{7}$$

In this paper, the weight is calculated for a certain attribute value, so the calculation formula is adjusted as follows:

$$IG(C|a_i)=\sum_{c}P(c|a_i)\log P(c|a_i)-\sum_{c}P(c)\log P(c) \tag{8}$$

$IG(C|a_i)$ is used to denote the weight $w_i^{att}$ of the corresponding ODE model when the attribute parent node $A_i$ takes the value $a_i$:

$$w_i^{att}=IG(C|a_i)=\sum_{c}P(c|a_i)\log P(c|a_i)-\sum_{c}P(c)\log P(c) \tag{9}$$

**B. Calculation of Instance Weights**

This paper employs an active learning approach that assigns weights to each training instance based on the mode of the instance set and the similarity between each training instance. The similarity between instance $x$ and instance $y_t$ is defined as:

$$s(x,y_t)=\sum_{j=1}^{m}\delta\left(a_j(x),a_j(y_t)\right) \tag{10}$$

Where $m$ represents the number of attributes, $\delta(\bullet)$ is a binary function, $a_j(x)$ is the $j$-th attribute value of the instance $x$, $a_j(y_t)$ is the $j$-th attribute value of the instance $y_t$. Then the weight of the $t$-th training instance is:

$$w_t = 1 + s\left(x, y_t\right) \tag{11}$$

In the process of calculating instance weights, it is necessary to first determine the mode $z$ of the instance set, subsequently compute the similarity $s(z, y_t)$ between the mode $z$ and each instance, and finally calculate the weight for each instance.

### 2.3 Algorithm Implementation

As shown in Table 1, the AIWODE algorithm first incorporates instance weights into conditional probabilities, and then employs an attribute weighting approach to embed the attribute weights into the classifier's classification formula.

**Table 1.** Steps of the AIWODE algorithm

| Algorithm. |
|---|
| Input: Training dataset D, test instance $x$<br>Output: Class label of the test instance $x$<br>Step 1: Given a test instance $x$ and an instance $y_t$ ($t$ = 1, 2, …, $n$), $y_t$ denotes the t-th training instance. Calculate the mode z of the training instance set.<br>1) Based on formula (10), calculate the mode z of the training instance set and the similarity $s(z, y_t)$ between each instance $y_t$ ($t$ = 1, 2, …, $n$) .<br>Calculate the weight $w_t$ ($t$ = 1, 2, …, $n$) of each training instance $y_t$ ($t$ = 1, 2, …, $n$) based on formula (11).<br>**Step 2:** For $m$ attribute variables, establish $m$ ODE (One-Dependence Estimator) topological structures.<br>1) Embed the weights corresponding to each instance into the prior probability $P(a_i, c)$ and conditional probability $P(a_s \mid a_i, c)$ calculations using formulas (5) and (6).<br>2) Based on formula (8), calculate the correlation $IG(C \mid a_i)$ between the attribute parent node $A_i$'s attribute value $a_i$ and the class $C$. Then, use formula (9) to calculate the corresponding weight $w_i^{att}$ for the ODE model.<br>**Step 3:** Use formula (4) to calculate the class label of the test instance $x$.<br>**Step 4:** Return the class label corresponding to the test instance $x$. |

## 3 Experimental Data and Results

We conducted experiments on 36 UCI datasets available on the official website of Waikato Environment for Knowledge Analysis (WEKA) [19]. The AIWODE approach was implemented on the WEKA, and its performance was compared with five representative improved Naive Bayes approaches. These five approaches include the NB [20, 21], AODE [8], AVFWNB [22], CFWNB [23], and AIWNB [24]. Below are the approach abbreviations:

AIWODE: Attribute and Instance Weighted One-Dependence Estimator
NB: Naive Bayes
AODE: Averaged One-Dependence Estimator
AVFWNB: Naive Bayes based on Attribute Value Frequency Weighting
CFWNB: Naive Bayes based on Correlation Feature Weighting
AIWNB: Instance and Attribute Weighted Naive Bayes

### 3.1 Experimental Results Analysis

Table 2 presents the comparison of accuracy results for the AIWODE with other existing approaches. The "w/t/l" column in the table indicates the number of victories, ties, and losses for each approach. Among the six approaches, AIWODE exhibits superior performance on the majority of datasets, with the highest average accuracy of 86.7%. The average accuracy of NB, AODE, AVFWNB, CFWNB, and AIWNB are 83.31%, 85.68%, 84.21%, 84.41%, and 84.94%, respectively. On the “anneal” dataset, AIWODE achieves the best performance with an accuracy of 99.34%. On the “mushroom” dataset, AIWODE demonstrates exceptional performance with an accuracy of 99.96%, which is significantly higher than that of other approaches. AIWODE has the highest number of victories and only a few losses across the datasets.

**Table 2.** Comparison of accuracy results on 36 UCI datasets

| Dataset | AIWODE | NB | AODE | AVFWNB | CFWNB | AIWNB |
|---|---|---|---|---|---|---|
| anneal | 99.34 | 96.13 | 98.01 | 98.62 | 98.5 | 98.94 |
| anneal.ORIG | 94.54 | 92.66 | 93.35 | 93.32 | 94.6 | 95.06 |
| audiology | 80.65 | 71.4 | 71.66 | 78.58 | 74.22 | 83.93 |
| autos | 86.88 | 72.3 | 80.74 | 77.27 | 77.95 | 78.04 |
| balance-scale | 69.28 | 71.08 | 69.34 | 71.1 | 73.76 | 73.75 |
| breast-cancer | 71.03 | 72.94 | 72.53 | 71.41 | 72.46 | 71.9 |
| breast-w | 97.28 | 97.25 | 96.97 | 97.48 | 97.14 | 97.17 |
| colic | 81.47 | 81.39 | 82.64 | 81.47 | 83.34 | 83.45 |
| colic.ORIG | 75.88 | 73.62 | 74.62 | 72.91 | 73.7 | 73.87 |
| credit-a | 86.48 | 86.25 | 86.71 | 86.23 | 86.99 | 87.03 |
| credit-g | 75.84 | 75.43 | 76.5 | 75.38 | 75.7 | 75.81 |
| diabetes | 78.44 | 77.85 | 78.07 | 77.89 | 78.01 | 77.87 |
| glass | 78.15 | 74.39 | 76.08 | 76.25 | 73.37 | 74.02 |
| heart-c | 82.81 | 83.6 | 83.2 | 83.04 | 82.94 | 82.71 |
| heart-h | 85.07 | 84.46 | 84.43 | 84.9 | 83.82 | 84.29 |
| heart-statlog | 83.22 | 83.74 | 83.33 | 83.78 | 83.44 | 83.22 |
| hepatitis | 85.71 | 84.22 | 84.98 | 85.38 | 85.95 | 85.75 |
| hypothyroid | 99.17 | 98.48 | 98.76 | 98.98 | 98.56 | 99.07 |
| ionosphere | 93.85 | 90.77 | 92.79 | 91.94 | 91.82 | 92.4 |
| iris | 93.33 | 94.47 | 93.2 | 94.4 | 94.4 | 94.4 |
| kr-vs-kp | 92.75 | 87.79 | 91.01 | 88.18 | 93.58 | 93.73 |
| labor | 94.27 | 93.13 | 94.7 | 94.33 | 92.1 | 94.33 |
| letter | 91.85 | 74 | 88.76 | 75.07 | 75.22 | 75.56 |
| lymphography | 88.19 | 84.97 | 86.98 | 85.49 | 84.81 | 84.68 |
| mushroom | 99.96 | 95.52 | 99.95 | 99.12 | 99.19 | 99.53 |
| primary-tumor | 46.7 | 47.2 | 47.67 | 45.85 | 47.2 | 47.76 |
| segment | 96.94 | 91.71 | 95.77 | 93.69 | 93.47 | 94.16 |
| sick | 97.65 | 97.1 | 97.39 | 97.02 | 97.36 | 97.33 |
| sonar | 85.24 | 85.16 | 86.6 | 84.49 | 82.56 | 82.23 |
| soybean | 94.54 | 92.2 | 93.28 | 94.52 | 93.66 | 94.74 |
| splice | 96.5 | 95.42 | 96.12 | 95.61 | 96.19 | 96.21 |
| vehicle | 72.71 | 62.52 | 72.31 | 63.36 | 62.91 | 63.59 |
| vote | 94.32 | 90.21 | 94.52 | 90.25 | 92.11 | 92.18 |
| vowel | 88.57 | 65.23 | 80.87 | 67.46 | 68.84 | 69.98 |
| waveform-5000 | 86.43 | 80.72 | 86.03 | 80.65 | 83.11 | 82.98 |
| zoo | 96.15 | 93.98 | 94.66 | 96.05 | 95.96 | 96.05 |
| Average | 86.7 | 83.31 | 85.68 | 84.21 | 84.41 | 84.94 |
| w/t/l | v/ /* | 17/18/1 | 11/25/0 | 13/22/1 | 12/23/1 | 10/24/2 |

# 4 Application of the AIWODE Approach

The dataset used in this study includes a variety of attributes aimed at providing a comprehensive assessment of teachers' teaching abilities. These attributes encompass the teacher's gaze distribution, classroom movement patterns, the extent of student attention coverage, facial expression distribution, classroom behaviors, the frequency and types of teacher questioning, speech rate, verbal emotional tone, and expert evaluations based on the teacher's teaching data. Specifically, the teacher's gaze distribution and movement patterns reflect their attention distribution and mobility within the classroom; student attention coverage indicates the teacher's focus on different groups of students; facial expression and behavior characteristics offer insights into students' emotional states and classroom engagement; the quantity and nature of teacher questions highlight the depth and breadth of classroom interaction; speech rate reveals the efficiency of communication; and verbal emotional tone represents the classroom's emotional environment. Lastly, expert evaluations combine these attributes to provide a quantitative measure of the teacher's overall teaching effectiveness. By analyzing these multifaceted attributes, this study presents a more scientifically robust method for teacher evaluation.

## 4.1 Dataset Description

In this section, the dataset used is derived from the AI evaluation system developed as part of the Hubei Province Artificial Intelligence-Enhanced Teacher Development Project. The dataset consists of 480 instances, 8 attribute dimensions, and expert ratings for each instance. To improve the dataset's quality and ensure the effectiveness of the classification model, several data preprocessing techniques were applied, including discretization, missing value imputation, and removal of irrelevant attributes [18]. First, continuous variables were discretized to reduce noise and decrease the complexity of the data. Next, missing values were addressed using suitable imputation methods to minimize information loss. Additionally, attributes that contributed little to the classification task or were irrelevant were removed after evaluating the correlations between the attributes. This process helped streamline the model and enhance its classification performance. The parameters of the teacher teaching ability AI evaluation dataset are shown in Table 3.

**Table 3.** Parameters of the teacher teaching ability AI evaluation dataset

| Index | Date name | Date type | Child node | Parent node |
|---|---|---|---|---|
| 1 | Teacher's gaze distribution ratio | Numerical | Yes | No |
| 2 | Teacher's classroom movement trajectory | Categorical | Yes | No |
| 3 | Teacher's student attention coverage | Numerical | Yes | No |
| 4 | Student classroom expression ratio | Numerical | Yes | No |
| 5 | Student classroom behavior | Numerical | Yes | No |
| 6 | Teacher's classroom questioning | Numerical | Yes | No |
| 7 | Teacher's classroom speech rate | Numerical | Yes | No |
| 8 | Teacher's verbal emotional tone | Numerical | Yes | No |
| 9 | Expert rating | Numerical | No | Yes |

The teacher teaching dataset used in the experiment is shown in Table 4. The dataset is stored in CSV format and initially contains 480 instances with 34 attributes, organized into 34 columns. The "class" represents the final expert rating, which serves as the label column. Some attributes in the dataset, such as "mention students" and "scare", have a majority of values equal to 0. To simplify the dataset and improve the accuracy of the expert scoring, these attributes will be categorized and preprocessed accordingly. This preprocessing step aims to streamline the data and enhance the precision of the expert evaluation results.

**Table 4.** Teacher evaluation dataset

| Serial number | Front right sight | Back right sight | Position left front | Nature | Listen | Number questions | … | Average speech speed | Class |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 32.3 | Yes | 81.1 | 75.3 | 17 | … | 159 | Good |
| 2 | 0 | 54.7 | Yes | 60 | 0 | 0 | … | 0 | Average |
| 3 | 7.2 | 70.8 | Yes | 76 | 0 | 24 | … | 190 | Average |
| 4 | 3 | 47.5 | Yes | 94.5 | 17 | 20 | … | 187 | Average |
| 5 | 83.7 | 0.8 | Yes | 96.9 | 84.3 | 49 | … | 164 | Good |
| 6 | 0 | 100 | Yes | 94.2 | 97 | 0 | … | 0 | Average |
| 7 | 35.9 | 7.6 | Yes | 89.8 | 68.1 | 0 | … | 154 | Good |
| 8 | 18.4 | 21 | Yes | 85.5 | 59.6 | 11 | … | 80 | Good |
| 9 | 2.9 | 54.6 | Yes | 86.7 | 14.8 | 11 | … | 139 | Good |
| 10 | 0 | 32.3 | Yes | 80 | 35.1 | 4 | … | 100 | Excellent |
| … | … | … | … | … | … | … | … | … | … |

### 4.2 Data Preprocessing

The dataset used in this study is sourced from the AI evaluation system developed under the Hubei Province Teacher Development Project, which includes 480 instances and 8 attribute dimensions. Before applying the new approach to teaching assessment, data preprocessing is first performed, which includes attribute extraction, handling missing values, discretization, and other necessary steps.

Attribute Extraction: The dataset contains information on the distribution of teacher gaze, including four directions: front-left, back-left, front-right, and back-right. However, the majority of the gaze data is concentrated in the front and back directions, so the left-right directions are not considered. Similarly, in the student classroom expression data, natural expressions have the highest proportion, followed by happy expressions, with the other types of expressions being negligible. Therefore, the seven types of expressions are reduced to two: natural and happy expressions. In terms of student classroom behavior, the behaviors of listening, standing, and raising hands have the highest proportions, while the other two behaviors have a zero proportion. Thus, student classroom behavior is simplified to three categories: listening, standing, and raising hands.

Handling Missing Values: The dataset contains some missing values, which were handled using the "Replace Missing Value" filter under the "unsupervised filters" section in Weka. For categorical attributes, the mode substitution method was applied, where missing values were replaced with the most frequent category under the assumption that it best represents the underlying distribution. For numerical attributes, the mean substitution method was used, leveraging the global statistical trend to maintain the attribute's overall central tendency while mitigating distortions caused by missing data. In addition, irrelevant attributes were removed using the Remove filter to prevent noise propagation. This strategy preserved data integrity while avoiding the loss of information that would result from case deletion.

Discretization: Continuous attributes were transformed using Weka's supervised Discretize filter. Specifically, an equal-width binning approach (default 10 intervals) was adopted, where numerical ranges were partitioned into uniformly spaced intervals. This procedure effectively transformed continuous values into categorical ones, thus simplifying probability estimation in Bayesian classifiers. Additionally, it mitigated data sparsity and bolstered computational stability. By incorporating class information during discretization, the method improved inter-class separability and strengthened the discriminative power of the transformed attributes.

The teacher evaluation dataset after preprocessing is shown in Table 5. These preprocessing steps improved the dataset's consistency, representativeness, and suitability for probabilistic modeling, thereby providing a solid foundation for the effective training of the proposed AIWODE model.

To evaluate the results of teacher teaching quality, a correspondence between expert scores and evaluation categories was established. The categories are divided into four types, as shown in Table 6:

**Table 5.** Preprocessed teacher evaluation dataset

| Serial Number | Sight behind | Position left front | Nature | Listen | Number questions | … | Average speech speed | Class |
|---|---|---|---|---|---|---|---|---|
| 1 | ‘\’(21.75-inf)\” | Yes | 81.1 | ‘\’(17.7-inf)\” | ‘\’(-inf-22.5]\” | … | ‘\’(-inf-193.5]\” | Good |
| 2 | ‘\’(21.75-inf)\” | Yes | 60 | ‘\’(-inf-17.7]\” | ‘\’(-inf-22.5]\” | … | ‘\’(-inf-193.5]\” | Average |
| 3 | ‘\’(21.75-inf)\” | Yes | 76 | ‘\’(-inf-17.7]\” | ‘\’(-inf-22.5]\” | … | ‘\’(-inf-193.5]\” | Average |
| 4 | ‘\’(21.75-inf)\” | Yes | 94.5 | ‘\’(-inf-17.7]\” | ‘\’(-inf-22.5]\” | … | ‘\’(-inf-193.5]\” | Average |
| 5 | ‘\’(-inf-21.75]\” | Yes | 96.9 | ‘\’(17.7-inf)\” | ‘\’(22.5-inf)\” | … | ‘\’(-inf-193.5]\” | Good |
| 6 | ‘\’(21.75-inf)\” | Yes | 94.2 | ‘\’(17.7-inf)\” | ‘\’(-inf-22.5]\” | … | ‘\’(-inf-193.5]\” | Average |
| 7 | ‘\’(21.75-inf)\” | Yes | 89.8 | ‘\’(17.7-inf)\” | ‘\’(-inf-22.5]\” | … | ‘\’(-inf-193.5]\” | Good |
| 8 | ‘\’(21.75-inf)\” | Yes | 85.5 | ‘\’(17.7-inf)\” | ‘\’(-inf-22.5]\” | … | ‘\’(-inf-193.5]\” | Good |
| 9 | ‘\’(-inf-21.75]\” | Yes | 86.7 | ‘\’(-inf-17.7]\” | ‘\’(-inf-22.5]\” | … | ‘\’(-inf-193.5]\” | Good |
| 10 | ‘\’(21.75-inf)\” | Yes | 80 | ‘\’(17.7-inf)\” | ‘\’(-inf-22.5]\” | … | ‘\’(-inf-193.5]\” | Excellent |
| … | … | … | … | … | … | … | … | … |

**Table 6.** Correspondence between scores and evaluation categories

| Expert rating score | Evaluation category |
|---|---|
| 85-100 | Excellent |
| 70-85 | Good |
| 60-70 | Average |
| 0-60 | Fail |

## 5 Experimental Results and Analysis

In this section, we validate the effectiveness of the AIWODE approach through experimental comparative analysis. The performance of the approaches used in this study was evaluated based on four metrics: accuracy, recall, AUC (Area Under the Curve), and F-score. Accuracy is an indicator that measures the proportion of correctly classified instances in the entire test set, calculated as the ratio of correctly classified samples to the total number of samples. Recall measures the effectiveness of the model in predicting positive class instances, defined as the ratio of correctly predicted positive class samples to the total number of actual positive class samples. AUC (Area Under Curve) is used to assess the model’s ability to distinguish between positive and negative classes. The higher the AUC value, the stronger the model’s discriminatory power. F-score combines accuracy and recall, serving as a comprehensive indicator of the model’s overall performance. Additionally, when comparing the accuracy and AUC indicators of the algorithms [25], a t-test statistical test with a 95% confidence level was used [26, 27]. In this study, ten-fold cross-validation [28] was performed for each dataset. In each cross-validation iteration, the dataset is divided into ten subsets, nine of which are used for training the classifier and one for testing. Each classifier is trained and predicted on the same cross-validation set, and the final accuracy is obtained by calculating the average value. The experimental results include the outcomes after only discretization and the results after replacing missing values following discretization. These are presented in Table 7 and Table 8, respectively.

**Table 7.** Experimental results after discretization

| Approach | Accuracy | Recall | AUC | F-score |
|---|---|---|---|---|
| AIWODE | 77.52 | 88.17 | 77.02 | 89.6 |
| NB | 79.66 | 92.76 | 80.10 | 87.36 |
| AODE | 80.72 | 95.73 | 81.53 | 88.28 |
| AVFWNB | 77.08 | 85.94 | 74.08 | 85.01 |
| CFWNB | 79.48 | 92.85 | 77.08 | 87.27 |
| AIWNB | 77.89 | 91.08 | 73.38 | 86.1 |

**Table 8.** Experimental results after discretization and missing value replacement

| Approach | Accuracy | Recall | AUC | F-score |
|---|---|---|---|---|
| AIWODE | 83.77 | 93.43 | 90.62 | 85.55 |
| NB | 83.29 | 90.26 | 88.94 | 88.97 |
| AODE | 83.71 | 90.96 | 89.76 | 89.29 |
| AVFWNB | 83.37 | 89.76 | 89.32 | 88.96 |
| CFWNB | 81.98 | 92.49 | 88.64 | 88.48 |
| AIWNB | 82.06 | 92.38 | 88.72 | 88.52 |

From Table 7 and Table 8, it can be observed that after discretization and missing value replacement, all the metrics such as accuracy, recall, AUC, and F-score show improvements. Notably, the improvements in our proposed AIWODE approach are particularly significant.

To facilitate a clear understanding, we use Roman numerals to represent the six subscripts in the chart: I - AIWODE; II - NB; III - AODE; IV - AVFWNB; V - CFWNB; VI-AIWNB. Based on the charts above, the following conclusions can be drawn:

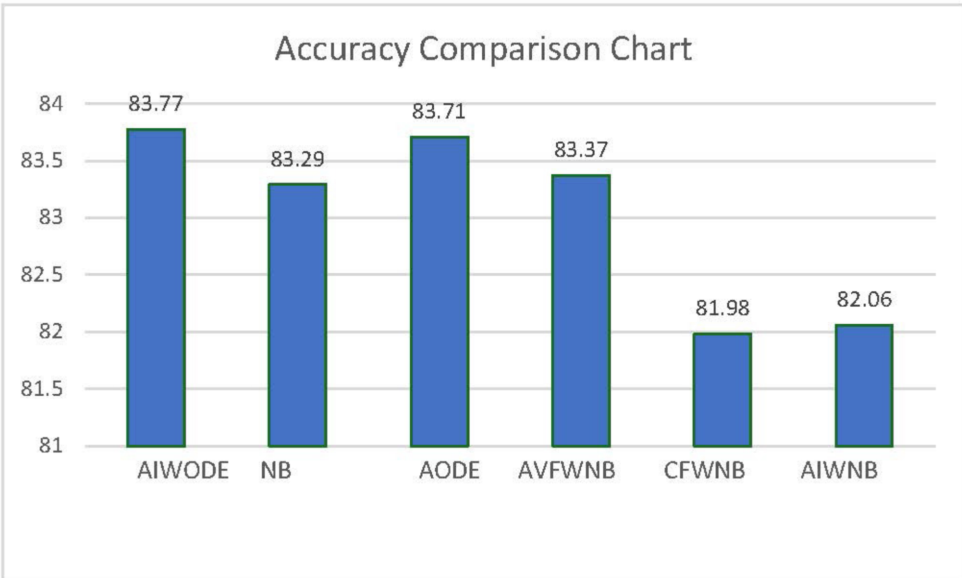


**Fig. 3.** Accuracy comparison chart

As shown in Fig. 3, the accuracy comparison demonstrates that the modified AIWODE approach achieved the highest accuracy of 83.77%. Compared to CFWNB, it shows an improvement of nearly 2%, and also slightly outperforms other classical Bayesian approaches. This indicates that the modified approach has a higher ability to correctly classify positive samples.

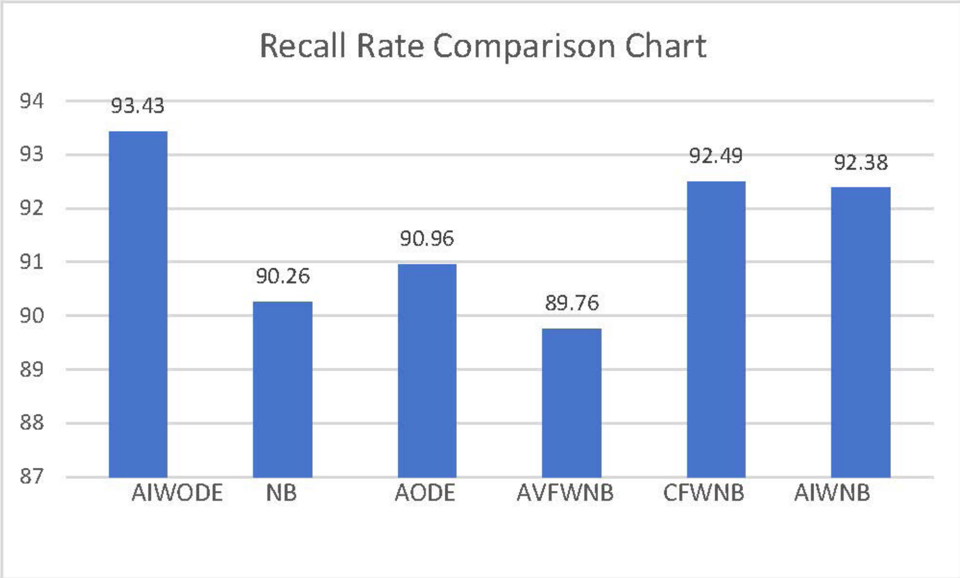


**Fig. 4.** Recall rate comparison chart

As shown in Fig. 4, the recall comparison demonstrates that AIWODE also leads in recall rate, reaching 93.43%, which is more than 3% higher than the NB and AVFWNB approaches, and nearly 3% better than the AODE approach. This indicates that the improved approach performs well in accurately predicting actual positive class samples.

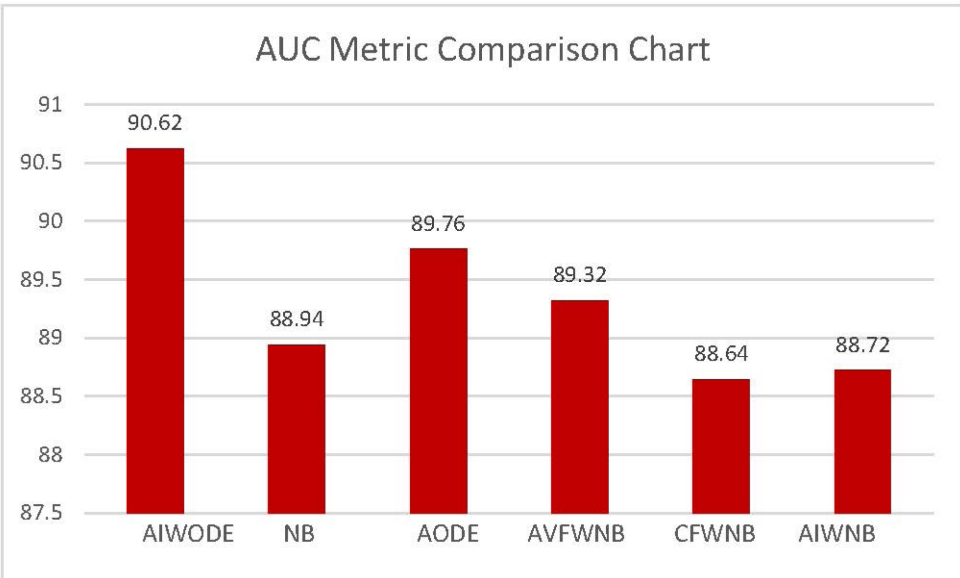


**Fig. 5.** AUC metric comparison chart

As shown in Fig. 5, the AUC comparison demonstrates that the AIWODE approach again performs best, with an AUC of 90.62%. The other approaches are slightly lower.

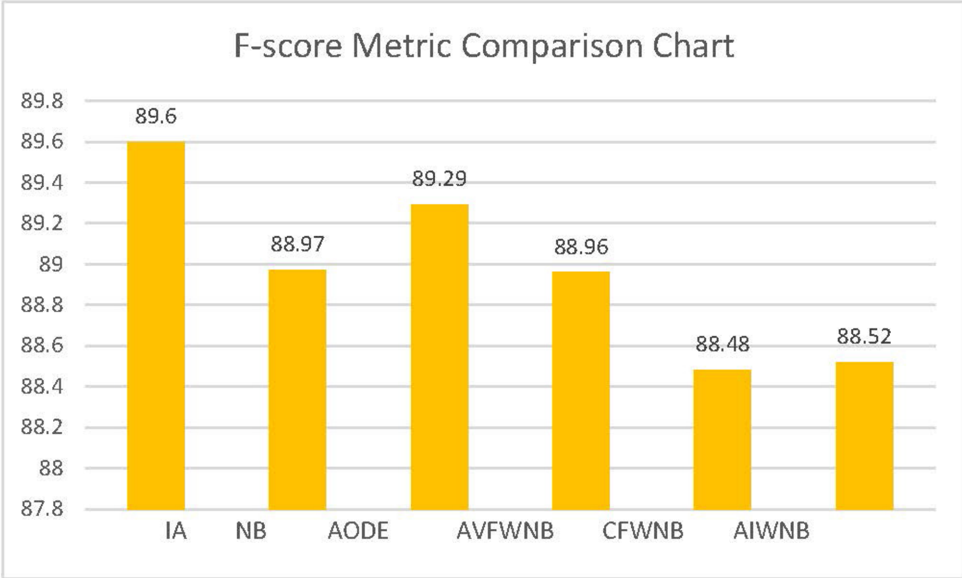


**Fig. 6.** F-score metric comparison chart

As shown in Fig. 6, the F-score comparison demonstrates that AIWODE achieves the highest F-score of 89.6%, maintaining its superior performance in overall classification accuracy and recall balance.

**Table 9.** Summary of AIWODE improvements over best baselines

| Metric | Best baseline (Value) | AIWODE (Value) | Improvement |
|---|---|---|---|
| Accuracy | AODE (83.71%) | 83.77% | +0.06% |
| Recall | CFWNB (92.49%) | 93.43% | +0.94% |
| AUC | AODE (89.76%) | 90.62% | +0.86% |
| F-score | AODE (89.29%) | 89.60% | +0.31% |

AIWODE shows improvements over the best baseline models across key metrics like Accuracy, Recall, AUC, and F - score. In each line, we compare the AIWODE algorithm to the Best Baselines. The results are summarized in Table 9. This demonstrates that by incorporating a dual weighting mechanism for both attributes and instances, the AIWODE model is able to more effectively capture underlying patterns in the data, significantly enhancing the analysis and predictive accuracy of teaching evaluation data.

## 6 Conclusion and Future Outlook

In this paper, an improved approach, AIWODE (attribute and instance weighted one-dependence estimators), is proposed to address the limitations of existing approaches. Building upon the AODE model, the AIWODE approach further explores the method of attribute and instance weighting. First, instance weighting is applied to incorporate the weights of instances into the calculation of conditional probabilities. Second, an attribute weighting method is used to embed the weights of each attribute into the classifier's formula. This approach comprehensively considers the impact of different attributes and instances on the experimental results, thereby significantly improving the classification performance.

In addition, the approach has potential for further optimization. Future improvements could involve introducing more sophisticated weight adjustment strategies to enhance the approach's flexibility and generalization performance, enabling it to better adapt to more complex data and application scenarios. As teacher evaluation datasets continue to expand, more precise evaluation methods could be employed to further validate and enhance the approach's performance.

## 7 Acknowledgement

The research work was financially supported by Hubei Provincial-level Teaching Reform Research Project for Undergraduate Universities (Grant No. 2024488) and Science and technology research projects of Hubei Provincial Department Of education (Grant No. D20213002)

## References

[1] M.J. Culbertson, Bayesian networks in educational assessment: The state of the field, Applied psychological measurement 40(1)(2015) 3-21.
https://doi.org/10.1177/0146621615590401

[2] J. Hou, Online teaching quality evaluation model based on support vector machine and decision tree, Journal of Intelligent & Fuzzy Systems 40(2)(2021) 2193-2203.
https://doi.org/10.3233/JIFS-189218

[3] Z. Zhang, Application of neural network algorithm based on sensor networks in performance evaluation simulation of rural teachers. Measurement: Sensors 32(2024) 101049.
https://doi.org/10.1016/j.measen.2024.101049

[4] W. Zhang, F. Gao, An improvement to naive bayes for text classification, Procedia Engineering 15(2011) 2160-2164.
https://doi.org/10.1016/j.proeng.2011.08.404

[5] X. Wu, W. Cao, D. Wang, M. Ding, L. Yu, Y. Nakanishi, Demand response model based on improved Pareto optimum considering seasonal electricity prices for Dongfushan Island, Renewable Energy 164(2021) 926-936.
https://doi.org/10.1016/j.renene.2020.08.003

[6] L. Yu, S. Gan, Y. Chen, M. He, Correlation-based weight adjusted naive Bayes, IEEE Access 8(2020) 51377-51387.
https://doi.org/10.1109/ACCESS.2020.2973331

[7] X. Liu, Anomaly detection system based on K-nearest neighbors and naive Bayes, Journal of Computational Methods in Sciences and Engineering 24(6)(2024) 3545-3556.
https://doi.org/10.1177/14727978241293274

[8] G.I. Webb, J.R. Boughton, Z. Wang, Not so naive Bayes: aggregating one-dependence estimators, Machine Learning 58(1)(2005) 5-24.
https://doi.org/10.1007/s10994-005-4258-6

[9] S. Chen, Z. Zhang, L. Liu, Attribute selecting in tree-augmented naive Bayes by cross validation risk minimization, Mathematics 9(20)(2021) 2564.
https://doi.org/10.3390/math9202564

[10] T. Sakai, The probability smoothing problem: Characterizations of the Laplace method, Mathematical Social Sciences 135(2025) 102409.
https://doi.org/10.1016/j.mathsocsci.2025.102409

[11] X. Xiao, B. Qi, S. Liu, S. Liu, P. Chen, J. Liang, J. Tong, H. Wang, A dynamic risk-informed framework for emergency human error prevention in high-risk industries: a nuclear power plant case study, Reliability Engineering & System Safety 261(2025) 111080.
https://doi.org/10.1016/j.ress.2025.111080

[12] L. Jiang, C. Li, Two improved attribute weighting schemes for value difference metric, Knowledge and Information Systems 60(2)(2019) 949-970.
https://doi.org/10.1007/s10115-018-1229-3

[13] X. Zhou, Y. Wang, L. Zhang, A. Huang, Z. Wang, An innovative multi-view collaborative optimization framework for weighted naive Bayes, Knowledge-Based Systems 317(2025) 113378.
https://doi.org/10.1016/j.knosys.2025.113378

[14] L. Jiang, L. Zhang, L. Yu, D. Wang, Class-specific attribute weighted naive Bayes, Pattern Recognition 88(2019) 321-330.
https://doi.org/10.1016/j.patcog.2018.11.032

[15] Y. Zhang, J. Wu, C. Zhou, Z. Cai, Instance cloned extreme learning machine, Pattern Recognition 68(2017) 52-65.
https://doi.org/10.1016/j.patcog.2017.02.036

[16] L. Yu, S. Gan, Y. Chen, D. Luo, A novel hybrid approach: instance weighted hidden naive Bayes, Mathematics 9(22)(2021) 2982.
https://doi.org/10.3390/math9222982

[17] C. Fang, Y. Li, Y. Shi, Enhancing undrained shear strength prediction: a robust hybrid machine learning approach with naive Bayes modeling, Journal of Engineering and Applied Science 72(1)(2025) 16.
https://doi.org/10.1186/s44147-025-00586-z

[18] G. Wei, L. Yi, Predicting the impact of yoga on chronic venous insufficiency: a machine learning approach using naive Bayes classifier and optimization systems, Journal of Ambient Intelligence and Humanized Computing 16(1)(2025)

109-131.
https://doi.org/10.1007/s12652-024-04889-6
[19] D. Merlini, M. Rossini, Text categorization with WEKA: a survey, Machine Learning with Applications 4(2021) 100033.
https://doi.org/10.1016/j.mlwa.2021.100033
[20] L. Yu, D. Wang, X. Zhou, X. Wu, Hybrid modeling of structure extension and instance weighting for naive Bayes, Journal of Intelligent Systems 34(1)(2025) 20240400.
https://doi.org/10.1515/jisys-2024-0400
[21] S. Wang, L. Jiang, C. Li, Adapting naive Bayes tree for text classification, Knowledge and Information Systems 44(1)(2015) 77-89.
https://doi.org/10.1007/s10115-014-0746-y
[22] W. Xu, L. Jiang, L. Yu, An attribute value frequency-based instance weighting filter for naive Bayes, Journal of Experimental & Theoretical Artificial Intelligence 31(2)(2019) 225-236.
https://doi.org/10.1080/0952813X.2018.1544284
[23] L. Jiang, L. Zhang, C. Li, J. Wu, A correlation-based feature weighting filter for naive Bayes, IEEE Transactions on Knowledge and Data Engineering 31(2)(2018) 201-213.
https://doi.org/10.1109/TKDE.2018.2836440
[24] H. Zhang, L. Jiang, L. Yu, Attribute and instance weighted naive Bayes, Pattern Recognition 111(2021) 107674.
https://doi.org/10.1016/j.patcog.2020.107674
[25] J. Muschelli, ROC and AUC with a binary predictor: a potentially misleading metric, Journal of Classification 37(3)(2020) 696-708.
https://doi.org/10.1007/s00357-019-09345-1
[26] R.E.W. Robert, Mann-Whitney U test and t-test, Journal of Visual Impairment Blindness 117(1)(2023) 99-100.
https://doi.org/10.1177/0145482X221150592
[27] B.G. Marcot, A.M. Hanea, What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis? Computational Statistics 36(3)(2021) 2009-2031.
https://doi.org/10.1007/s00180-020-00999-9
[28] N.K. Rastogi, B. Rajagopalan, Á. Ossandón, Bayesian hierarchical network model for forecasting daily river stage for rainfed river networks, Journal of Hydrology 654(2025) 132894.
https://doi.org/10.1016/j.jhydrol.2025.132894