

ASSIGNMENT 3

Chunyao Wang(chw132),
Wenxing Li(wel85),
Jie Rong(jir18)

Table of Contents

1. OVERVIEW	2
2. QUESTIONS AND ANSWERS.....	4
2.1 OVERALL CLASSIFICATION ACCURACY	4
2.1.1 <i>House Votes Manual Model</i>	4
2.1.2 <i>House Votes Naive Model</i>	4
2.1.3 <i>House Votes PC Model</i>	5
2.1.4 <i>Analysis</i>	5
2.2 SENSITIVITY AND SPECIFICITY FOR EACH OF THE TWO PARTIES	5
2.2.1 <i>House Votes Manual Model</i>	5
2.2.2 <i>House Votes Naive Model</i>	6
2.2.3 <i>House Votes PC Model</i>	7
2.2.4 <i>Analysis</i>	7
2.3 POSITIVE AND NEGATIVE PREDICTIVE VALUE FOR EACH OF THE TWO PARTIES	8
2.3.1 <i>House Votes Manual Model</i>	8
2.3.2 <i>House Votes Naive Model</i>	8
2.3.3 <i>House Votes PC Model</i>	8
2.3.4 <i>Analysis</i>	8
2.4 CALIBRATION CURVE FOR A SELECTED BIN COUNT OR WINDOW SIZE	9
2.4.1 <i>House Votes Manual Model</i>	9
2.4.2 <i>House Votes Naive Model</i>	10
2.4.3 <i>House Votes PC Model</i>	11
2.4.4 <i>Analysis</i>	12

This project analyzes voting records of United States representatives on 16 different issues. There are House Votes Manual model, House Votes Naïve model and House Votes PC model. We implement GeNIe to test the three models, in which we use the "leave-one-out" cross-validation method. Here are imported data files and models.

Fig1.1 import data



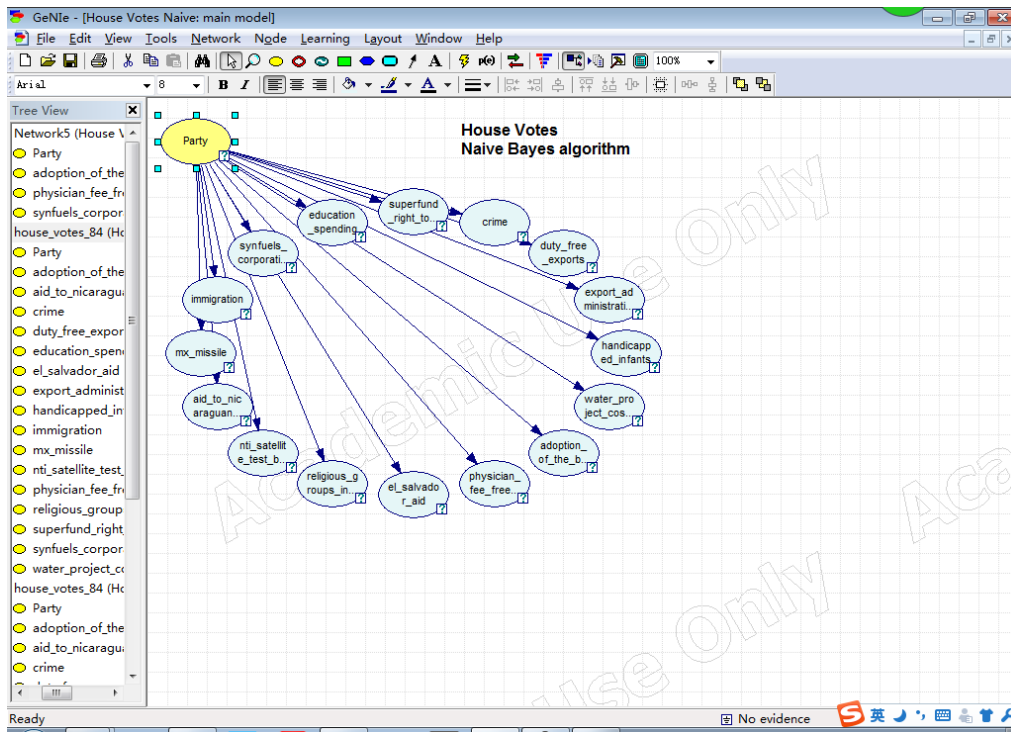


Fig1.3 House Votes Naïve model

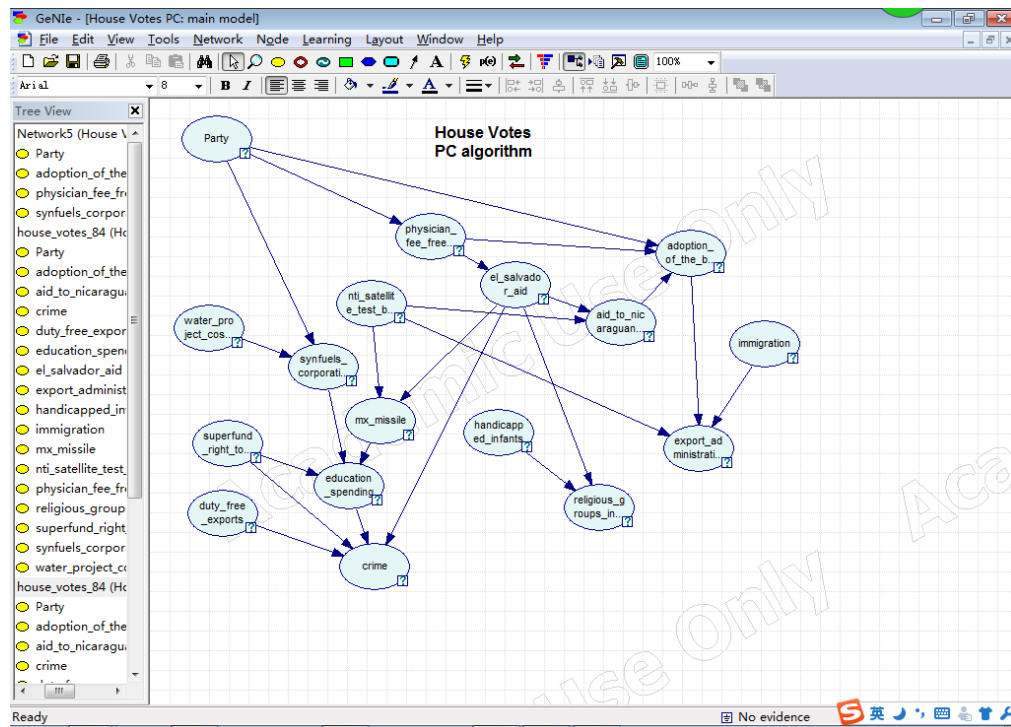


Fig1.4 House Votes PC model

2. Questions and Answers

2.1 Overall classification accuracy

In this part, we use the item “party” to make a validation and compare results of three models to analyze their accuracy. Here are validation results of Manual model, Naïve model and PC model.

2.1.1 House Votes Manual Model

Fig2.1.1 shows the validation result of House Votes Manual model. According to the figure, the classification accuracy of Manual model is $0.96092 \approx 96\%$.

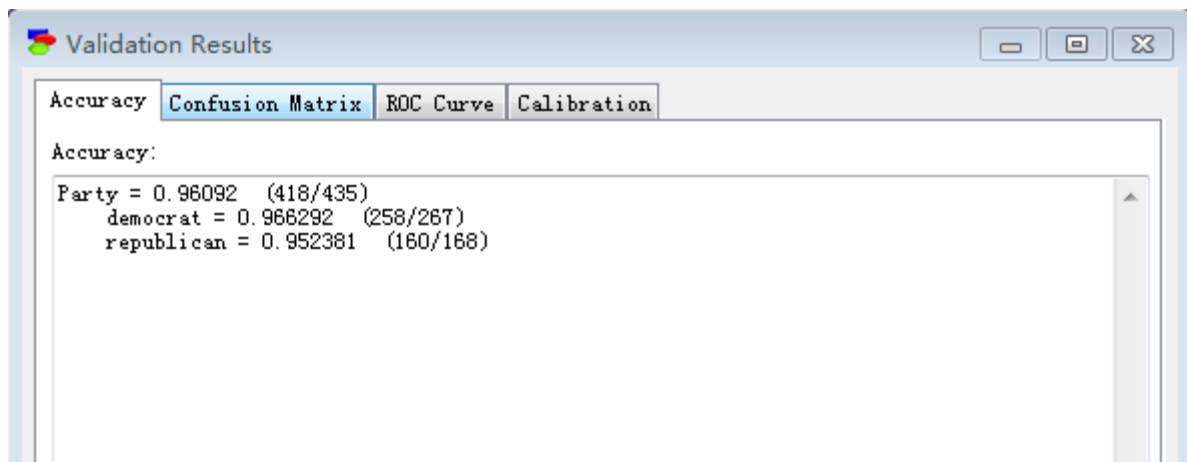


Fig2.1.1 result of House Votes Manual model

2.1.2 House Votes Naive Model

Fig2.1.2 shows the validation result of House Votes Naive model. According to the figure, the classification accuracy of Manual model is $0.901149 \approx 90\%$.

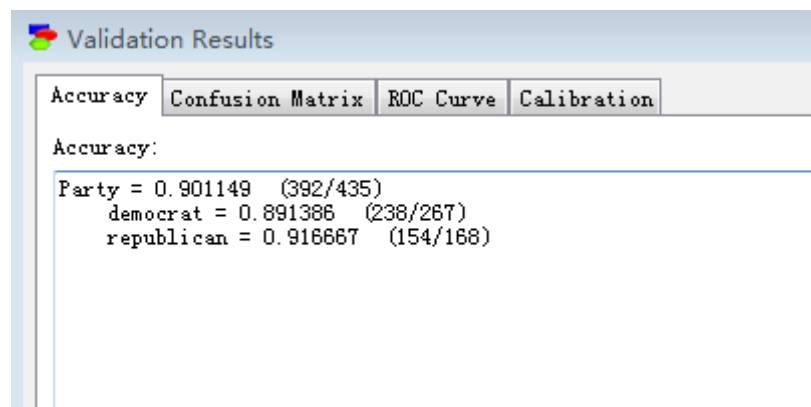


Fig2.1.2 result of House Votes Naïve model

2.1.3 House Votes PC Model

Fig2.1.3 shows the validation result of House Votes PC model. According to the figure, the classification accuracy of Manual model is $0.958621 \approx 96\%$.

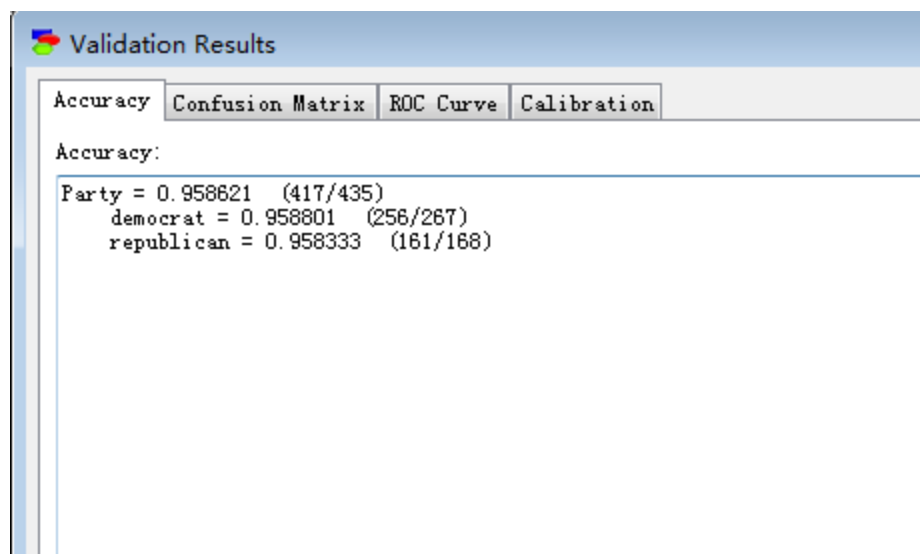


Fig2.1.3 result of House Votes PC model

2.1.4 Analysis

As it can be seen, the Naïve model has the least classification accuracy. The accuracy of Manual model approximates that of PC model. In conclusion, the Manual model and PC model are more accurate than Naïve model.

2.2 Sensitivity and specificity for each of the two parties

In this part, we calculate the sensitivity and specificity of Democrat and Republican. We get the confusion matrix of each model by the formula of sensitivity and specificity calculation.

The formulas we use are:

$$\text{sensitivity(Democrat)} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{specificity(Democrat)} = \text{TN} / (\text{FP} + \text{TN})$$

$$\text{sensitivity(Republican)} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{specificity(Republican)} = \text{TN} / (\text{FP} + \text{TN})$$

2.2.1 House Votes Manual Model

Fig2.2.1 shows the confusion matrix of House Votes Manual model. According to the figure: TP=258; TN=160; FN=8; FP=9. The Chart 2.2.1 below shows the results.

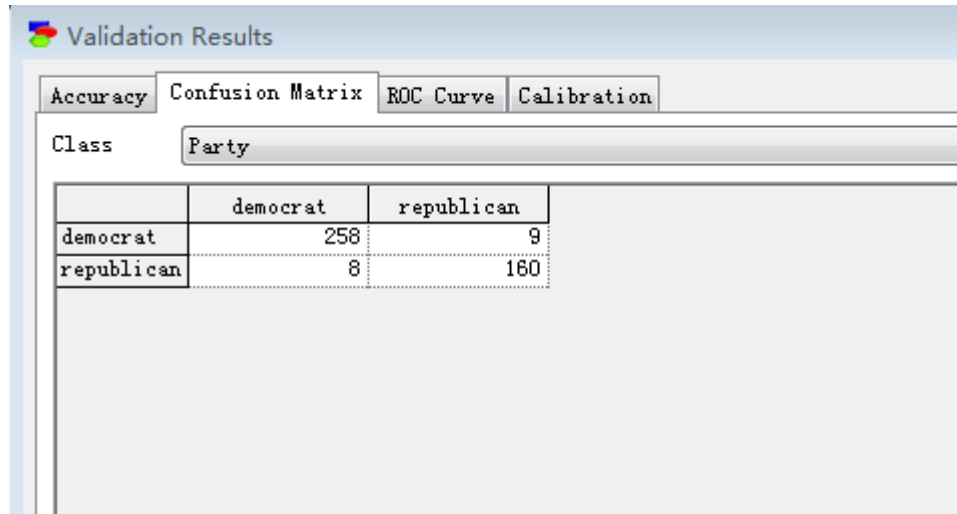


Fig2.2.1 the confusion matrix of Manual model

	Sensitivity(%)	Specificity(%)
Democrat	97	95
Republican	95	97

Chart2.2.1 the sensitivity and specificity of Manual model

2.2.2 House Votes Naive Model

Fig2.2.2 shows the confusion matrix of House Votes Naive model. According to this figure: TP=238; TN=154; FN=14; FP=29. We calculate the sensitivity and specificity of Democrat and Republican in the same way. The chart below shows the results.

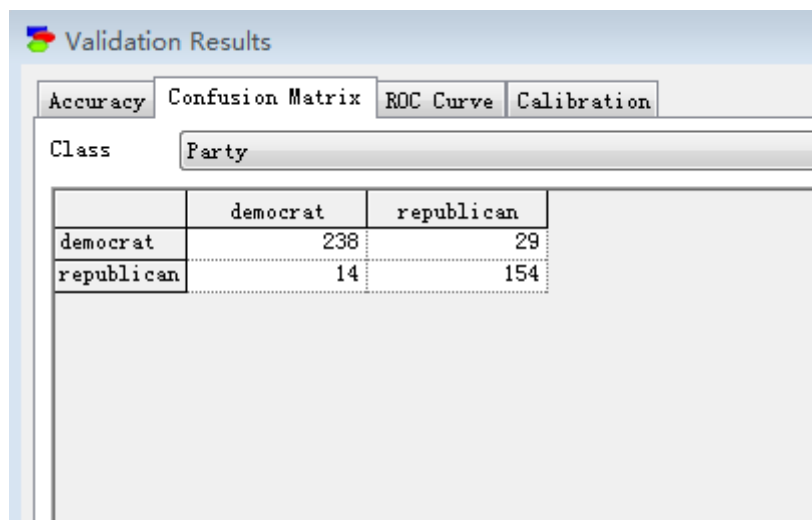


Fig2.2.2 the confusion matrix of Naive model

	Sensitivity(%)	Specificity(%)
Democrat	94	84
Republican	84	94

Chart2.2.2 the sensitivity and specificity of Naive model

2.2.3 House Votes PC Model

Fig2.2.3 shows the confusion matrix of House Votes PC model. According to the figure: TP=256; TN=161; FN=7; FP=11. We calculate the sensitivity and specificity of Democrat and Republican and display the results in Chart2.2.3.

Validation Results		
Confusion Matrix		
Class	Party	
	democrat	republican
democrat	256	11
republican	7	161

Fig2.2.3 the confusion matrix of PC model

	Sensitivity(%)	Specificity(%)
Democrat	97	94
Republican	94	97

Chart2.2.3 the sensitivity and specificity of PC model

2.2.4 Analysis

According to three charts, the results of Manual model and PC model are very closed. What is more, the sensitivity and specificity of Manual or PC model is higher than that of Naïve model. As it can be seen, the sensitivity of positive value equals to the specificity of negative value, from which the conclusion can be drawn that they probably have a certain relationship.

2.3 Positive and negative predictive value for each of the two parties

In this part, we calculate the predictive values of Democrat and Republican in three models.

The formulas we use are:

Positive predictive value(Democrat)= $TP/(TP+FP)$

Negative predictive value(Democrat)= $TN/(FN+TN)$

Positive predictive value(Republican)= $TP/(TP+FP)$

Negative predictive value(Republican)= $TN/(FN+TN)$

2.3.1 House Votes Manual Model

According to Fig2.2.1, $TP=258$; $TN=160$; $FN=8$; $FP=9$.

The chart below shows the results.

	Positive(%)	Negative(%)
Democrat	97	95
Republican	95	97

Chart2.3.1 predictive value of Manual model

2.3.2 House Votes Naive Model

According to Fig2.2.2, $TP=238$; $TN=154$; $FN=14$; $FP=29$. The results are shown in Chart2.3.2.

	Positive(%)	Negative(%)
Democrat	89	92
Republican	92	89

Chart2.3.2 predictive value of Naive model

2.3.3 House Votes PC Model

According to Fig2.2.3, $TP=256$; $TN=161$; $FN=7$; $FP=11$. The results are shown in Chart2.3.3.

	Positive(%)	Negative(%)
Democrat	96	96
Republican	96	96

Chart2.3.3. predictive value of PC model

2.3.4 Analysis

According to three charts above, the positive predictive value of Democrat equals to the negative predictive value of Republican. As it can be seen, predictive values of Manual and PC models are bigger than that of Naive model.

2.4 Calibration curve for a selected bin count or window size

2.4.1 House Votes Manual Model

(a)Democrat

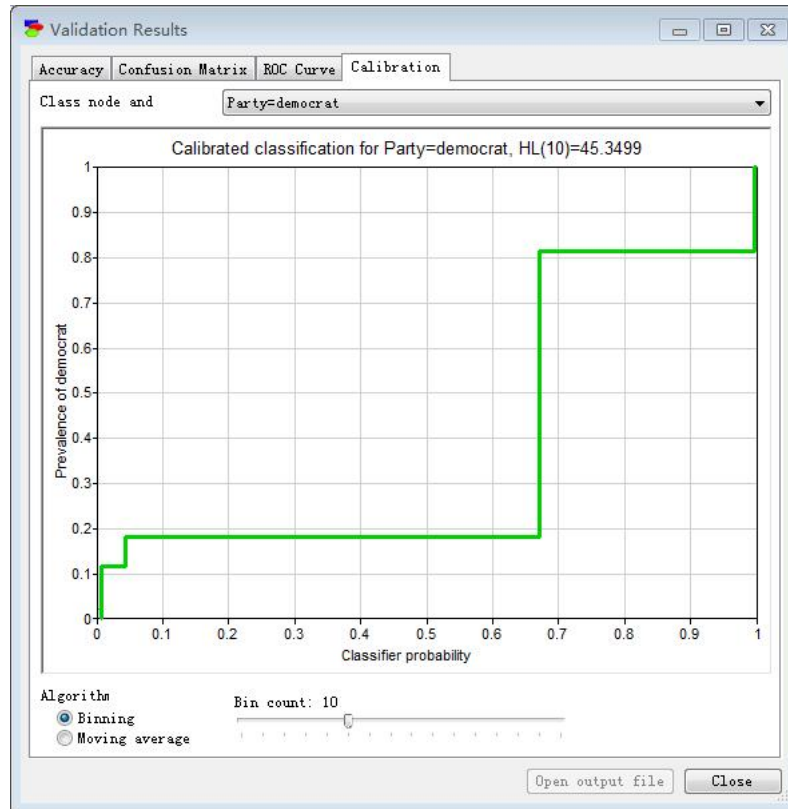


Fig2.4.1(a)

(b)Republican

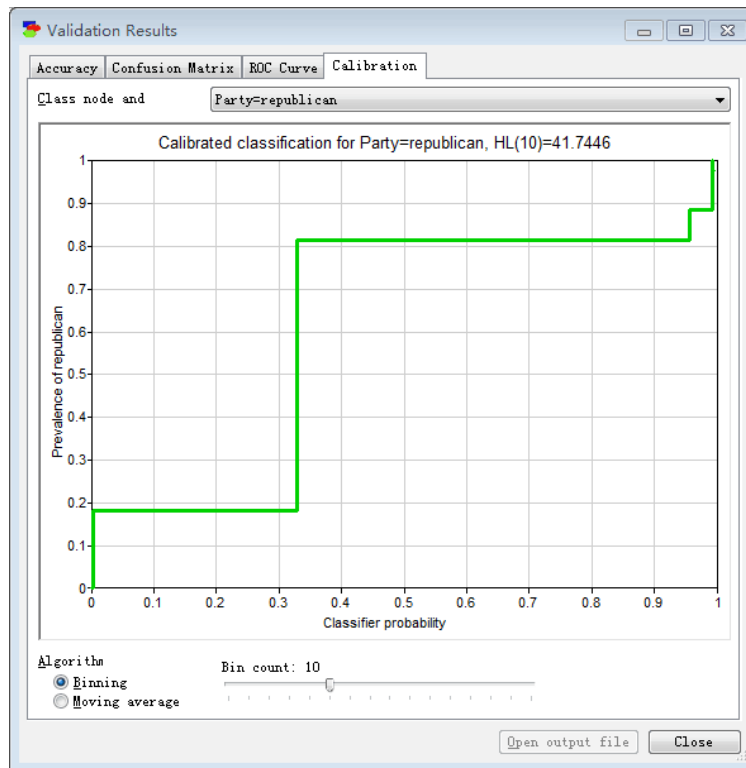


Fig2.4.1(b)

2.4.2 House Votes Naive Model

(a) Democrat

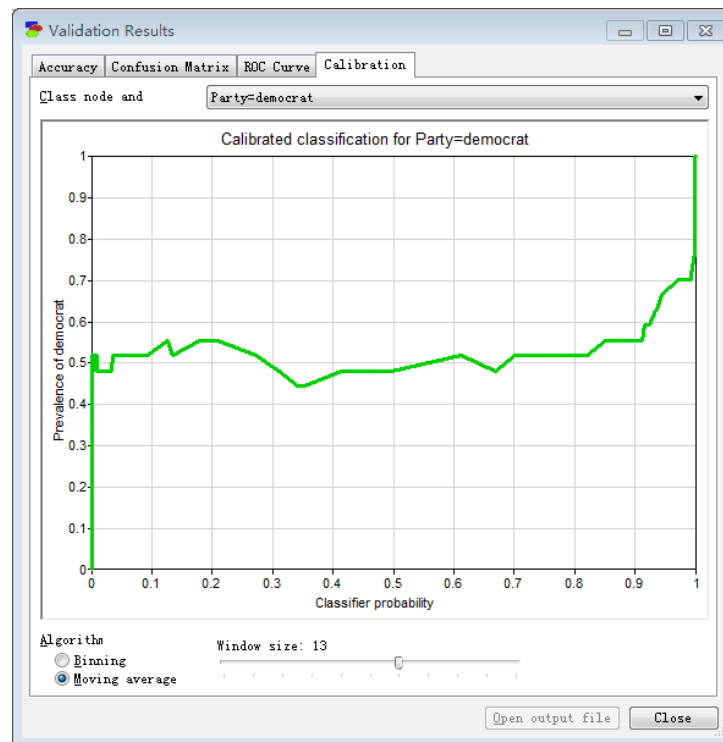


Fig2.4.2(a)

(b) Republican

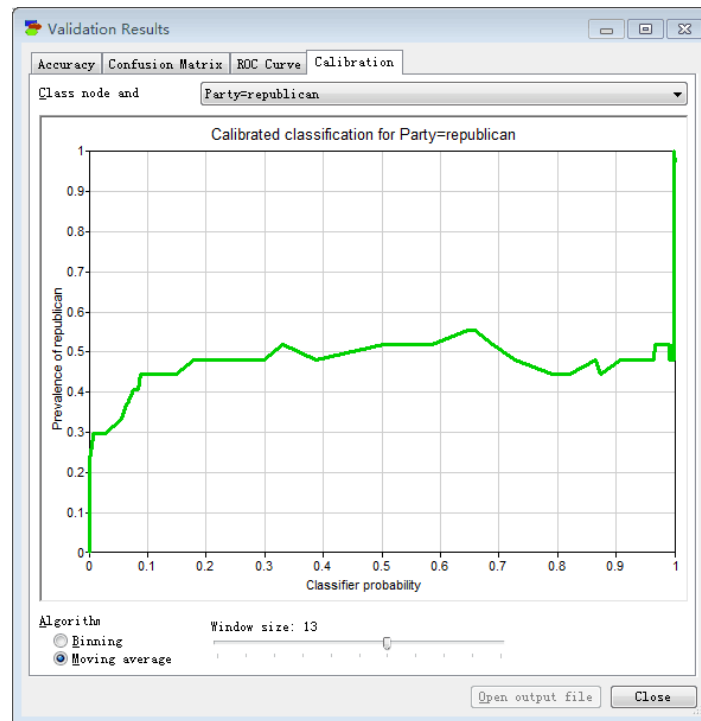


Fig2.4.2(b)

2.4.3 House Votes PC Model

(a) Democrat

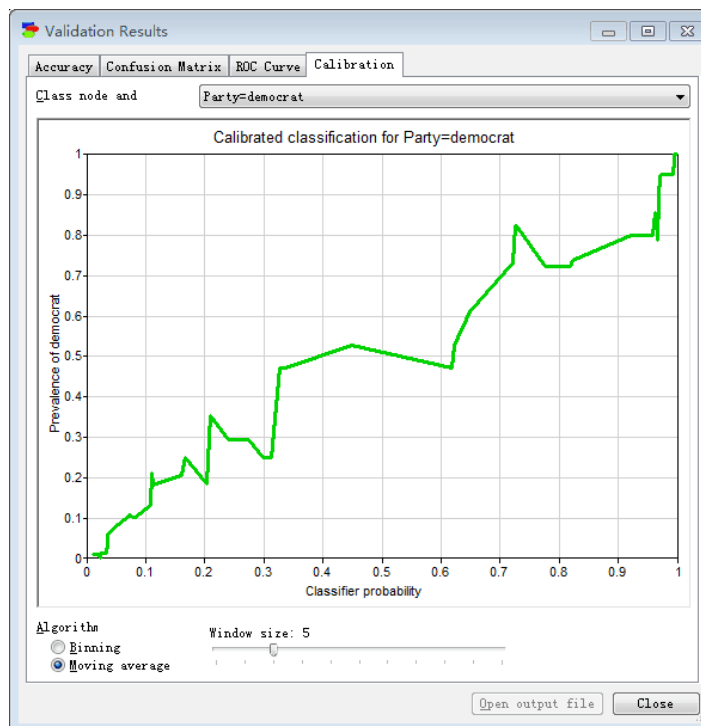


Fig2.4.3(a)

(b) Republican

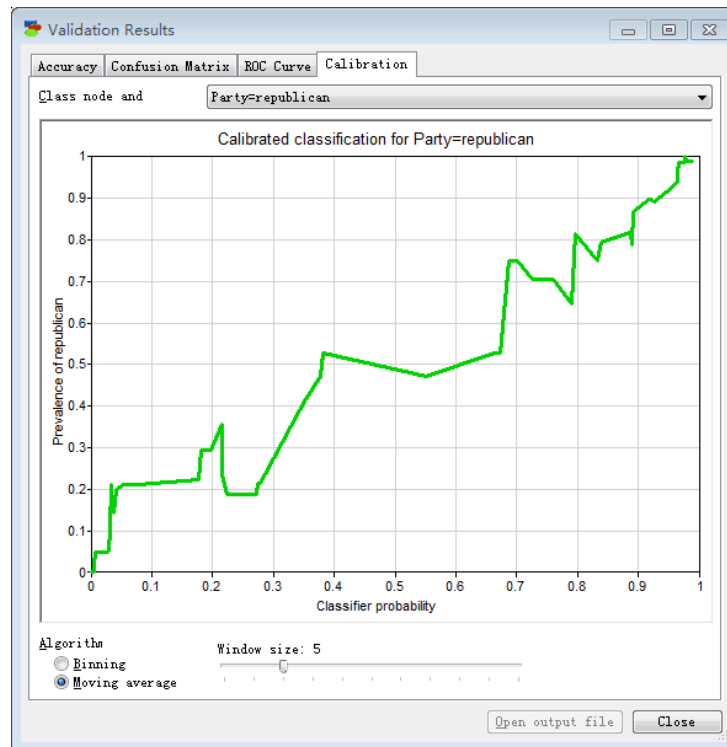


Fig2.4.3(b)

2.4.4 Analysis

As it can be seen, the prevalence changes considerably along with the increment of the classifier probability. Also, the curve alters when the bin count is changed. By comparing figures of Democrat and Republican in each model, trends of two parties' curves are similar to each other. Besides, curves in Fig2.4.1(a) and Fig2.4.1(b), which implement the binning algorithm, are flatter than others.