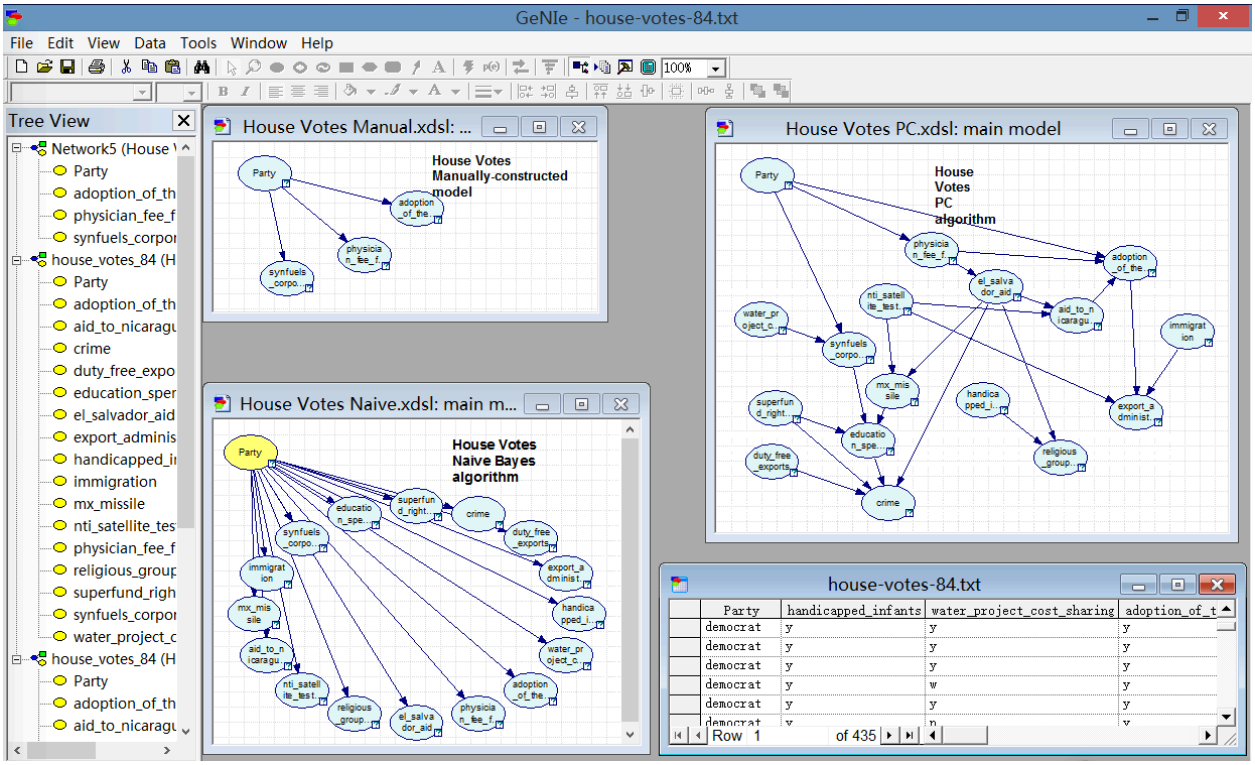
INFSCI 2725: Data Analytics

Assignment 3: Validation and Testing

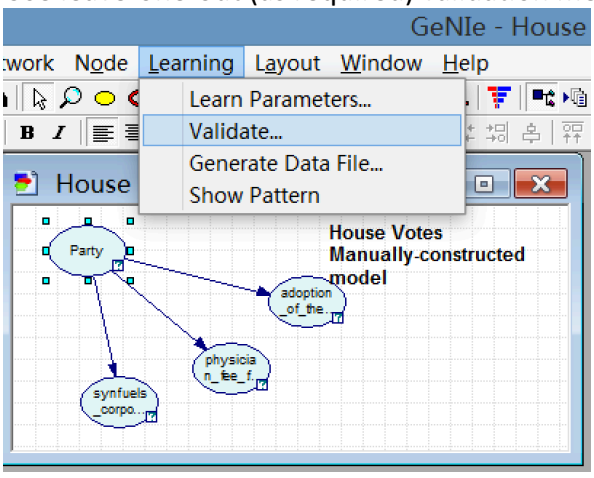
Team members: Zhenyu Peng
Tong Wei

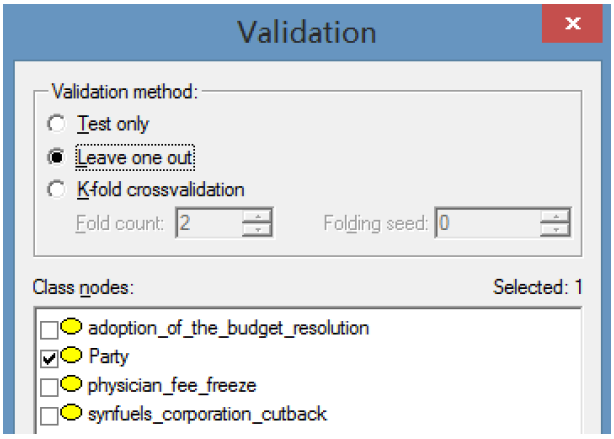
In this assignment, we utilize GeNIe to deal with all these tasks.
First of all, we import three models and load the data file:



1. Overall classification accuracy

Use leave-one-out (as required) validation method to test the model:





1.1 Manually-constructed model:

Accuracy:

```
Party = 0.96092  (418/435)
   democrat = 0.966292  (258/267)
   republican = 0.952381  (160/168)
```

1.2 Naive Bayes algorithm:

Accuracy:

```
Party = 0.901149  (392/435)
   democrat = 0.891386  (238/267)
   republican = 0.916667  (154/168)
```

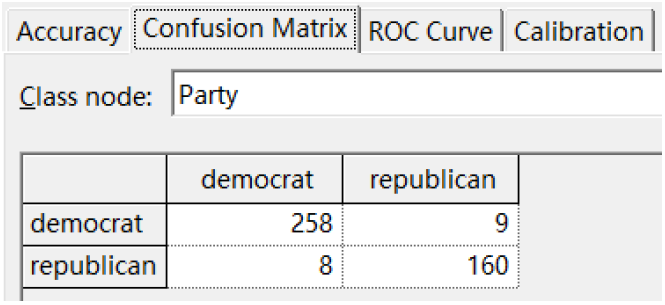
1.3 PC algorithm:

Accuracy:

```
Party = 0.958621  (417/435)
   democrat = 0.958801  (256/267)
   republican = 0.958333  (161/168)
```

2. Sensitivity and specificity for each of the two parties

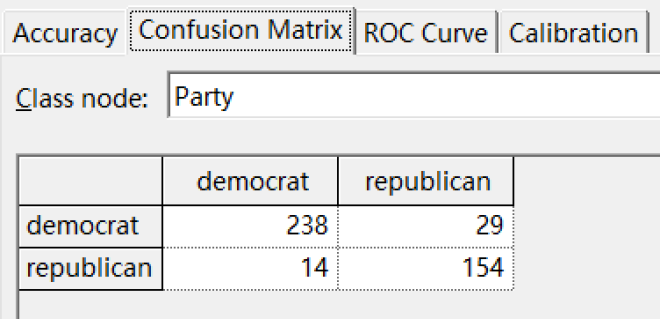
2.1 Manually-constructed model:



	democrat	republican
democrat	258	9
republican	8	160

The sensitivity of democrat: 258 / (258+8) = 0.969925
The sensitivity of republican: 160 / (160+9) = 0.946746
The specificity of democrat: 160 / (160+9) = 0.946746
The specificity of republican: 258 / (258+8) = 0.969925

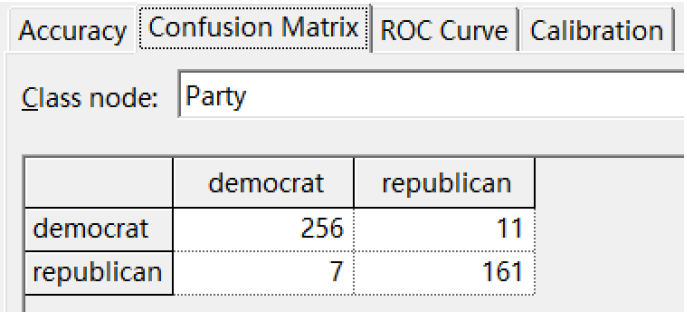
2.2 Naive Bayes algorithm:



	democrat	republican
democrat	238	29
republican	14	154

The sensitivity of democrat: 238 / (238+14) = 0.944444
The sensitivity of republican: 154 / (154+29) = 0.841530
The specificity of democrat: 154 / (154+29) = 0.841530
The specificity of republican: 238 / (238+14) = 0.944444

2.3 PC algorithm:



	democrat	republican
democrat	256	11
republican	7	161

The sensitivity of democrat: 256 / (256+7) = 0.973384
The sensitivity of republican: 161 / (161+11) = 0.936047
The specificity of democrat: 161 / (161+11) = 0.936047
The specificity of republican: 256 / (256+7) = 0.973384

3. Positive and negative predictive value for each of the two parties

3.1 Manually-constructed model:

The positive value of democrat: 9 / (9+258) = 0.033708
The negative value of democrat: 160 / (160+8) = 0.952381
The positive value of republican: 160 / (160+8) = 0.952381
The negative value of republican: 9 / (9+258) = 0.033708

3.2 Naive Bayes algorithm:

The positive value of democrat: 29 / (238+29) = 0.108614
The negative value of democrat: 154 / (154+14) = 0.916667
The positive value of republican: 154 / (154+14) = 0.916667
The negative value of republican: 29 / (238+29) = 0.108614

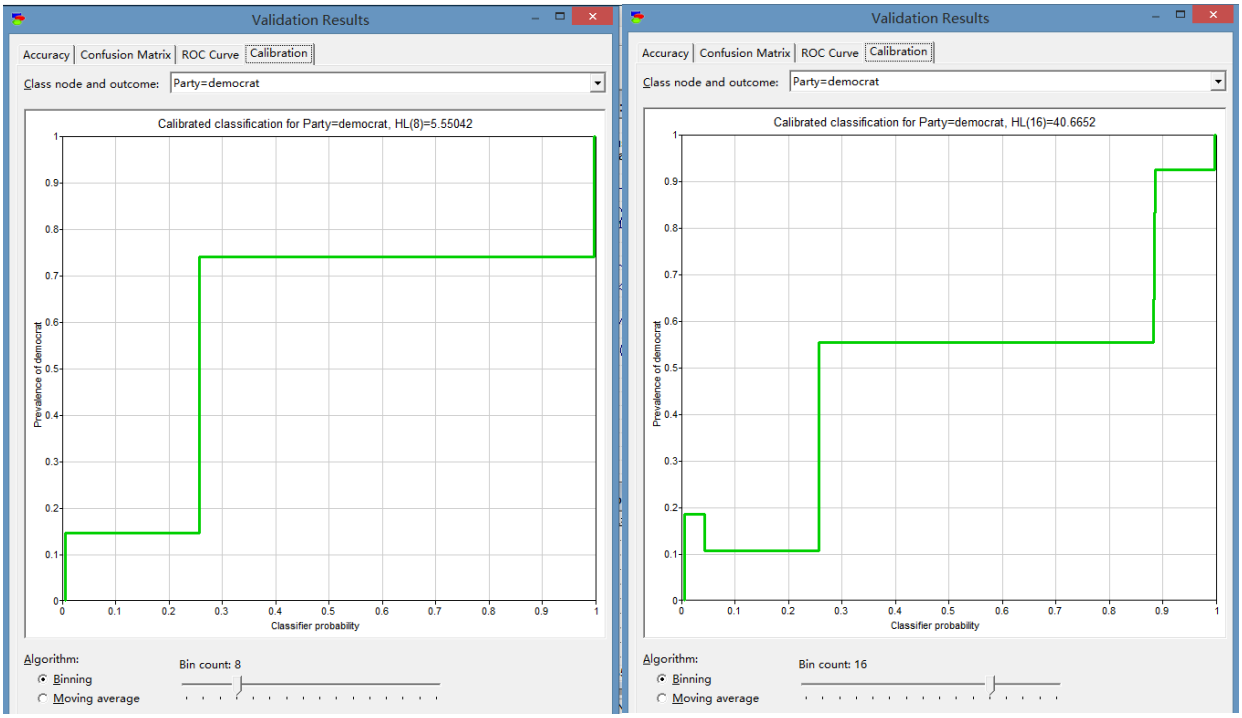
3.3 PC algorithm:

The positive value of democrat: 11 / (256+11) = 0.041199
The negative value of democrat: 161 / (161+7) = 0.958333
The positive value of republican: 161 / (161+7) = 0.958333
The negative value of republican: 11 / (256+11) = 0. 04120

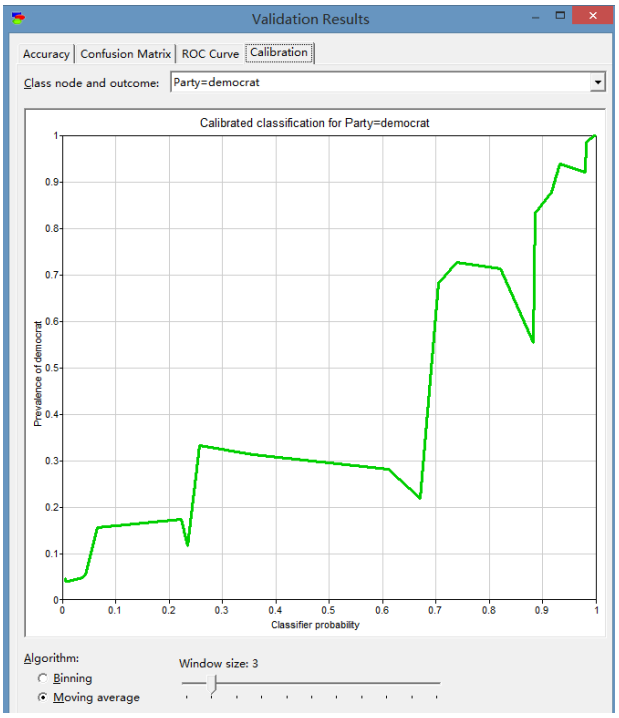
4. Calibration curve for a selected bin count or window size

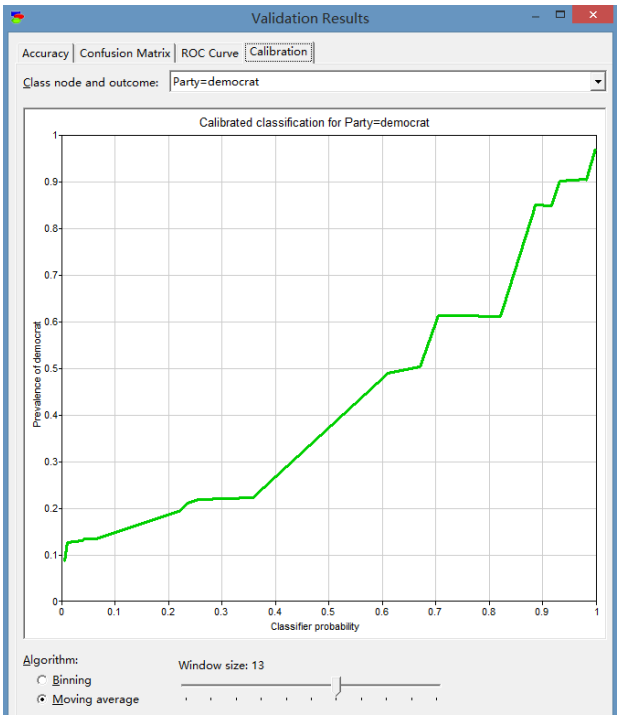
4.1 Manually-constructed model:

Bin count 8 and 16:



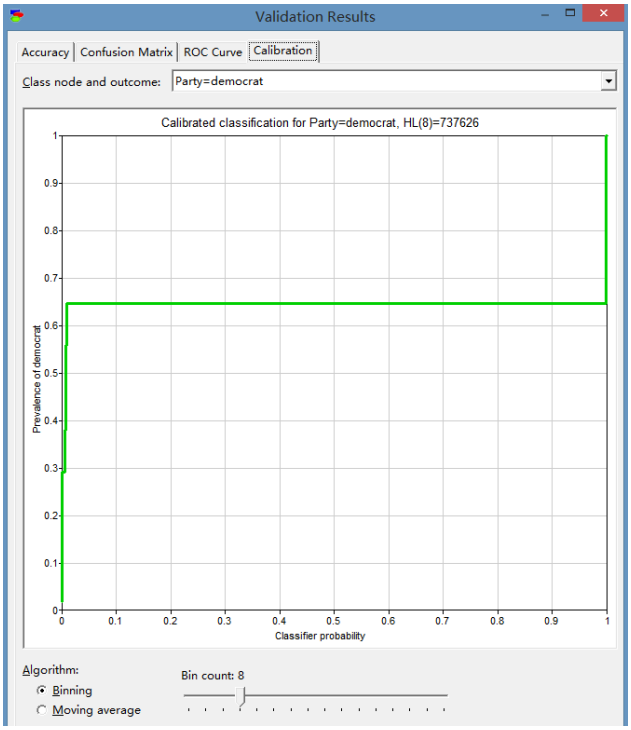
Window size 3 and 13:

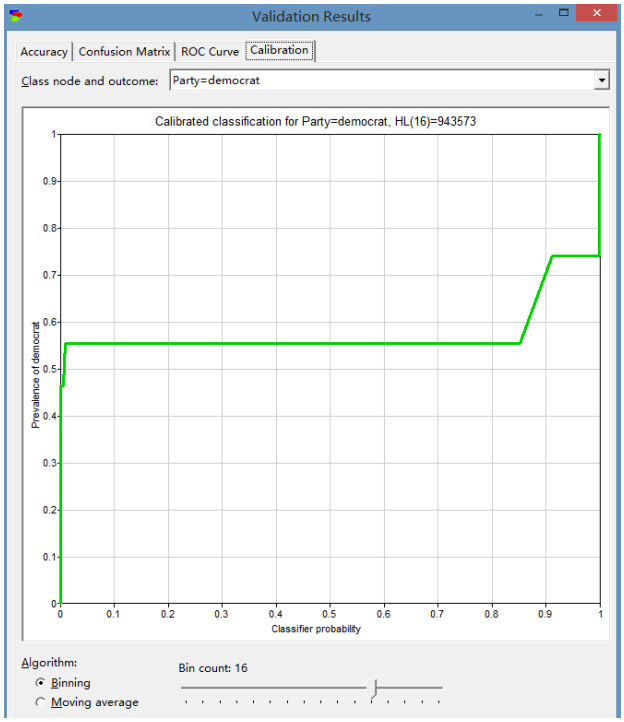




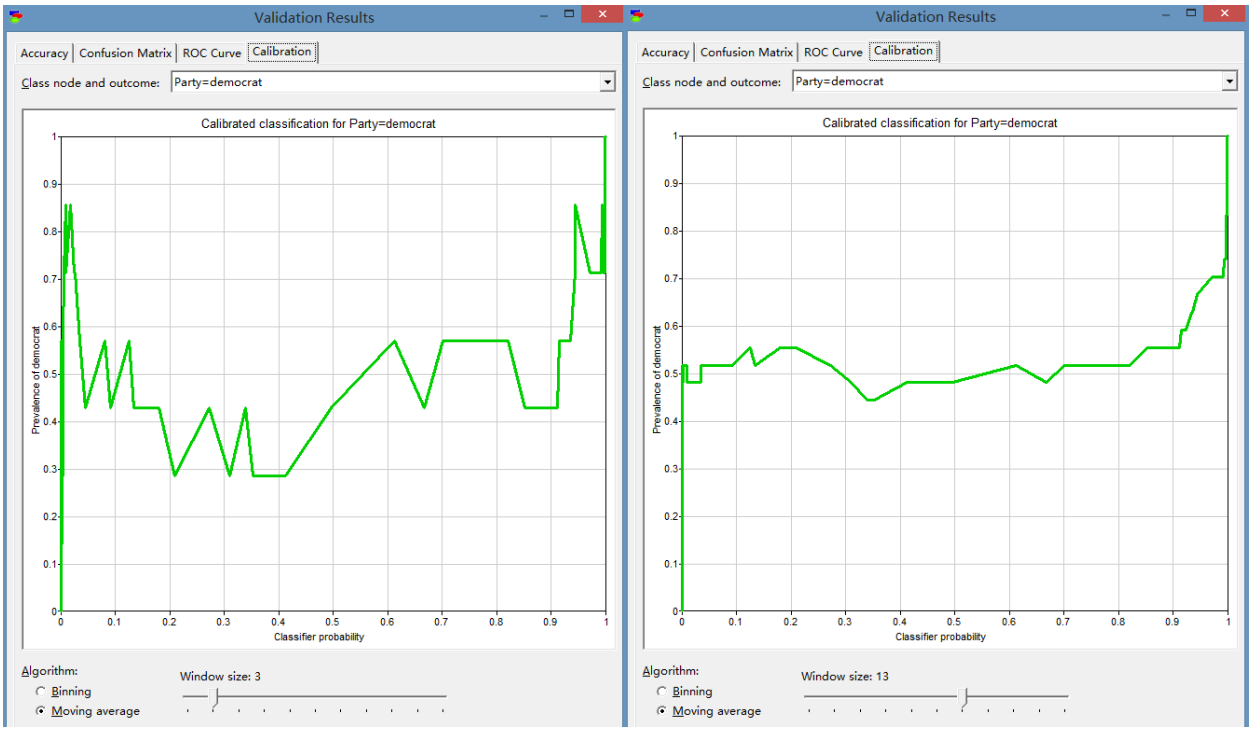
4.2 Naive Bayes algorithm:

Bin count 8 and 16:



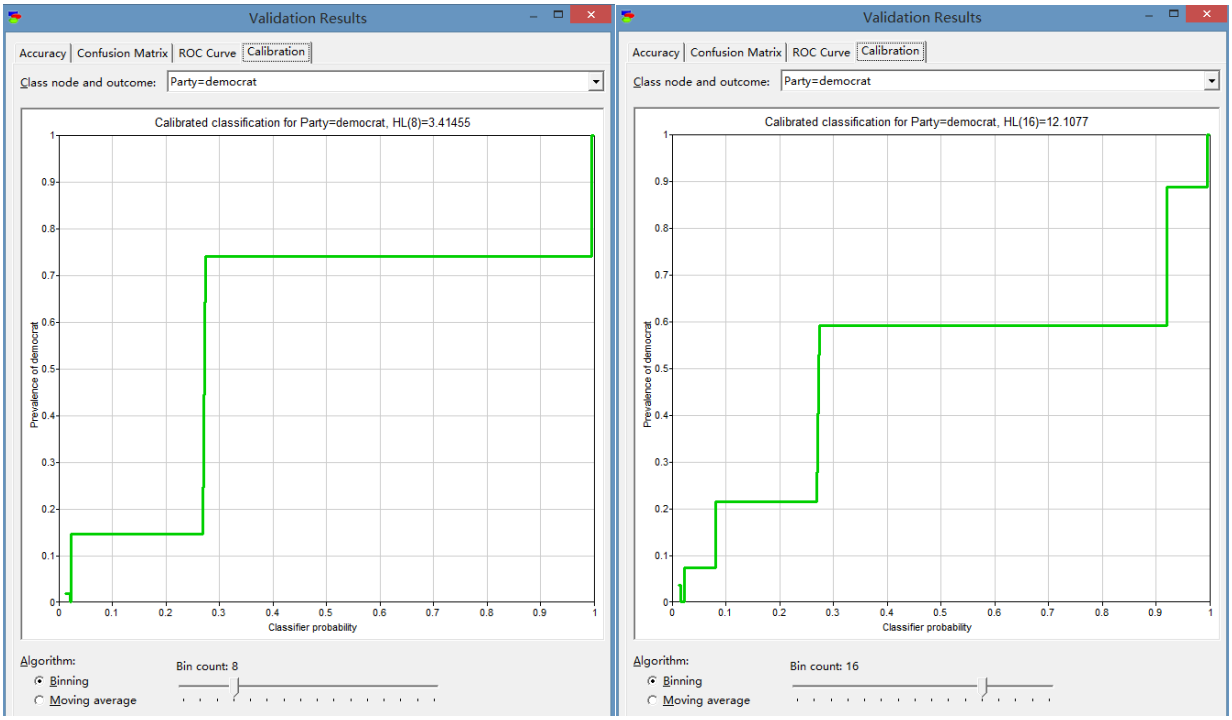


Window size 3 and 13:



4.3 PC algorithm:

Bin count 8 and 16:



Window size 3 and 13: