**INFSCI 2725: Data Analytics**

**Assignment 3: Validation and Testing**

**(1) Overall classification accuracy**

1. **House Votes Manual Model**

Validation Results (party node):

* **Accuracy:**

Party = 0.96092  (418/435)

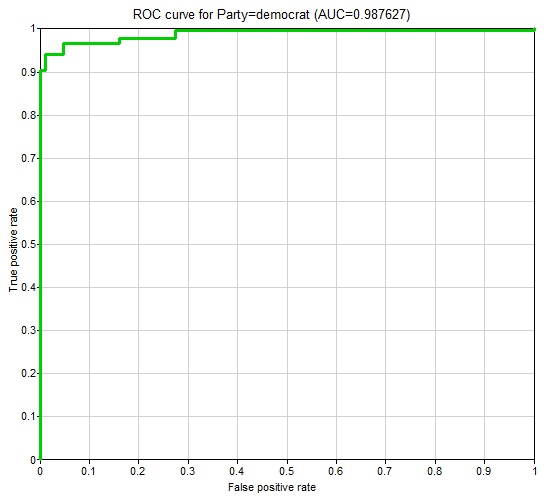
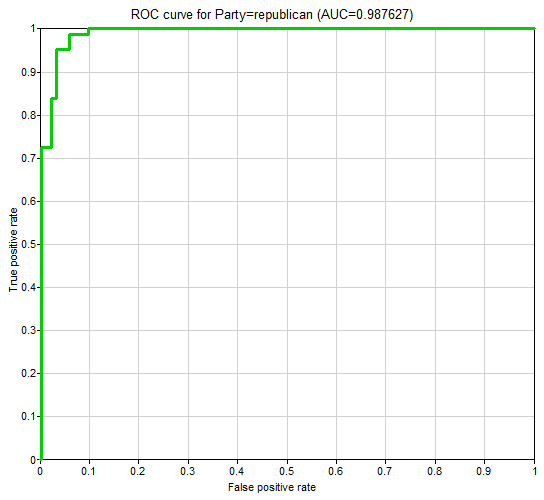
democrat = 0.966292  (258/267)

republican = 0.952381  (160/168)

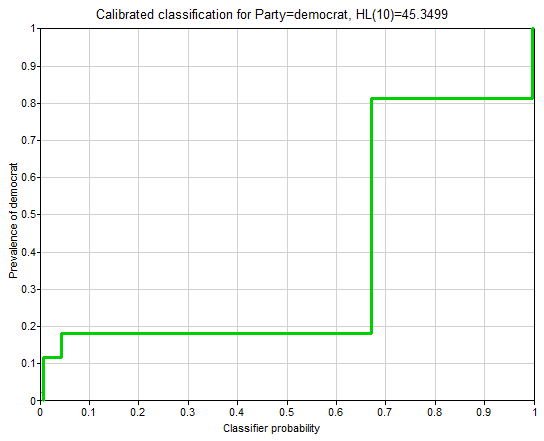
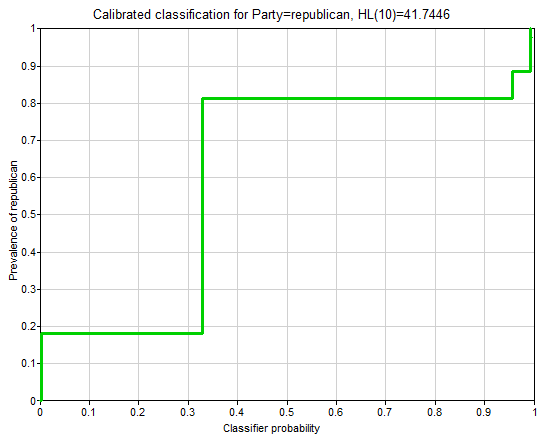
* **Confusion Matrix:**

|  |  |  |
| --- | --- | --- |
|  | democrat | republican |
| democrat | 258 | 9 |
| republican | 8 | 160 |

* **ROC Curve:**

* **Calibration:**

1. **House Votes Naive Model**

Validation Results (party node):

* **Accuracy:**

Party = 0.901149  (392/435)

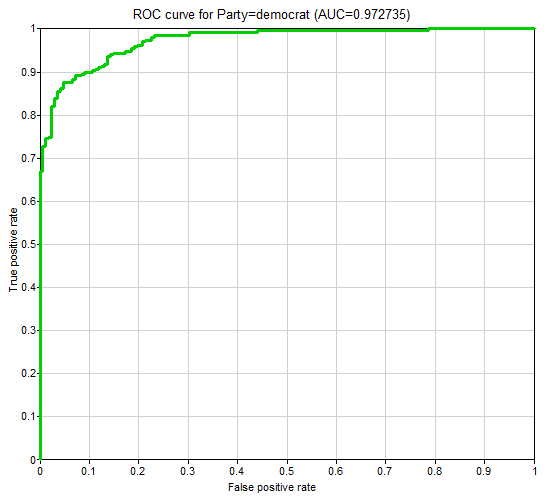
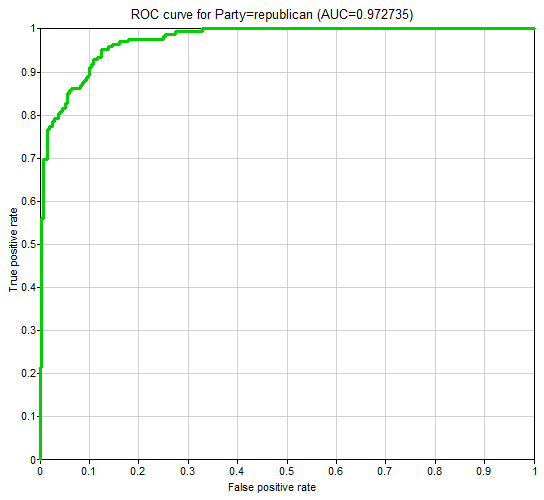
democrat = 0.891386  (238/267)

republican = 0.916667  (154/168)

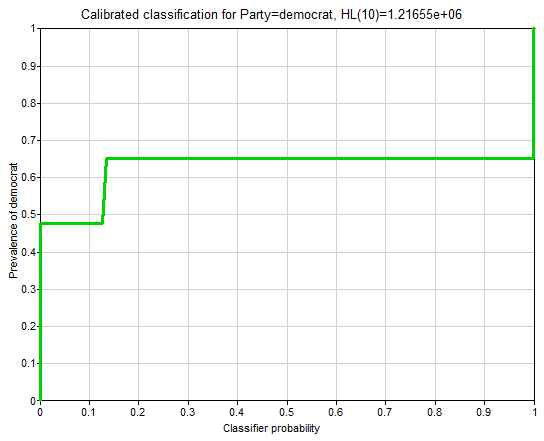
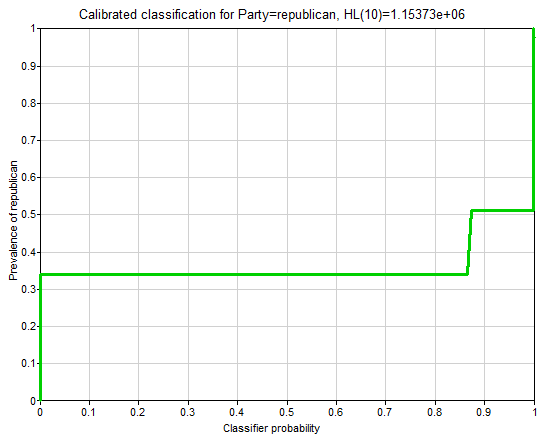
* **Confusion Matrix:**

|  |  |  |
| --- | --- | --- |
|  | democrat | republican |
| democrat | 238 | 29 |
| republican | 14 | 154 |

* **ROC Curve:**

* **Calibration:**

1. **House Votes PC Model**

Validation Results (party node):

* **Accuracy:**

Party = 0.958621  (417/435)

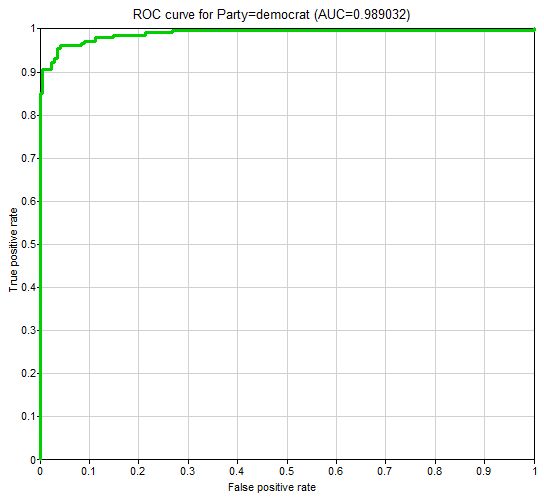
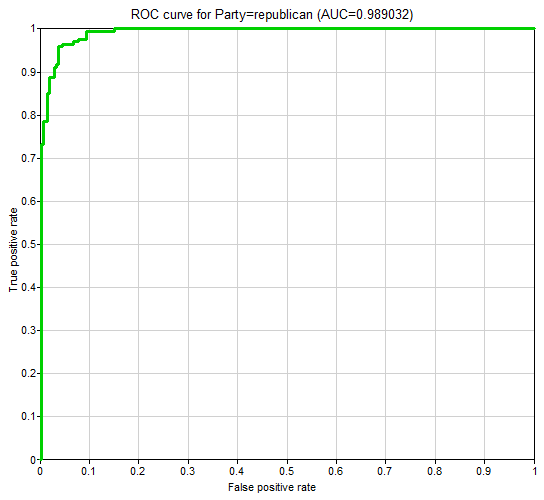
democrat = 0.958801  (256/267)

republican = 0.958333  (161/168)

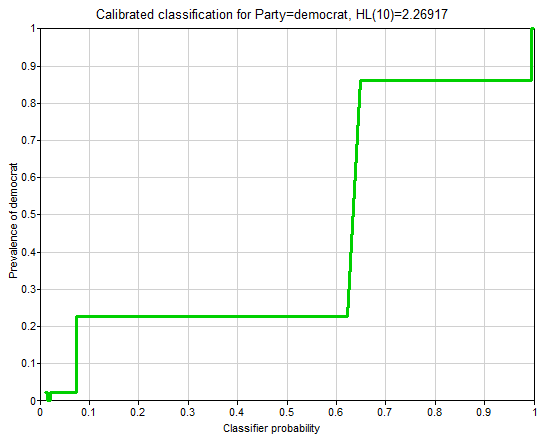
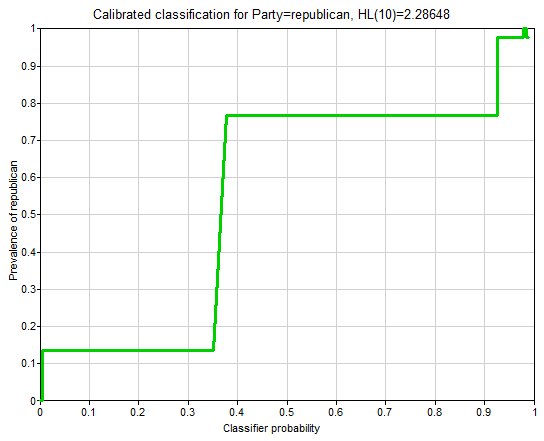
* **Confusion Matrix:**

|  |  |  |
| --- | --- | --- |
| democrat | 256 | 11 |
| republican | 7 | 161 |

* **ROC Curve:**

* **Calibration:**

**(2) Sensitivity and specificity for each of the two parties**

|  |  |  |
| --- | --- | --- |
|  | Sensitivity  TP / (TP+FN) | Specificity  TN / (FP+TN) |
| House Votes Manual | 258/(258+8)  = 96.99% | 160/(9+160)  = 94.67% |
| House Votes Naive | 238/(238+14)  = 94.44% | 154/(29+154)  = 84.15% |
| House Votes PC | 256/(256+7)  = 97.34% | 161/(11+161)  = 93.60% |

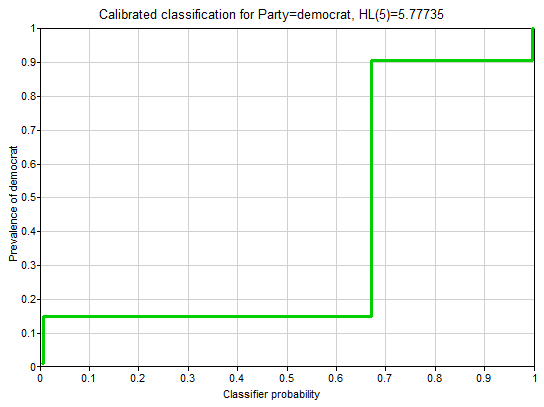
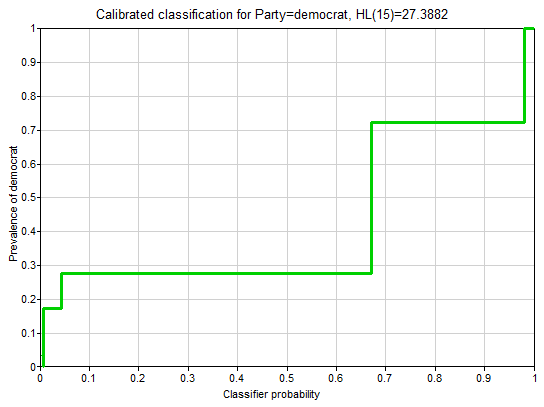
**(3) Positive and negative predictive value for each of the two parties**

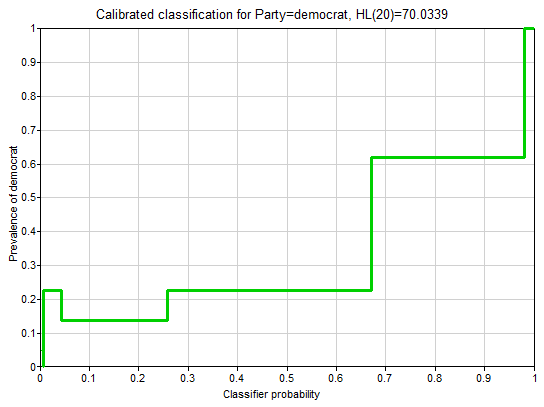
|  |  |  |
| --- | --- | --- |
|  | Sensitivity  TP / (TP+FP) | Specificity  TN / (FN+TN) |
| House Votes Manual | 258/(258+9)  = 96.63% | 160/(8+160)  = 95.24% |
| House Votes Naive | 238/(238+29)  = 89.14% | 154/(14+154)  = 91.67% |
| House Votes PC | 256/(256+11)  = 95.88% | 161/(7+161)  = 95.83% |

**(4 Extra Point!) Calibration curve for a selected bin count or window size**

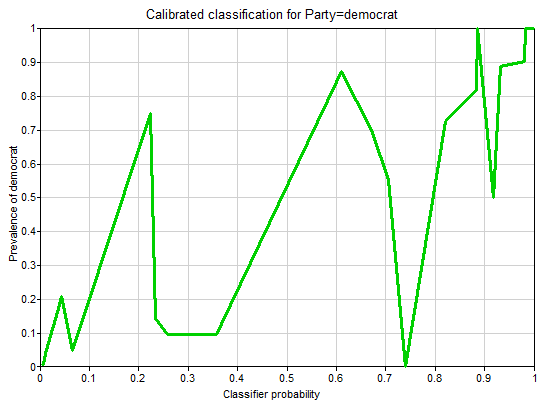
1. **House Votes Manual:**

**Different Bin Count:**

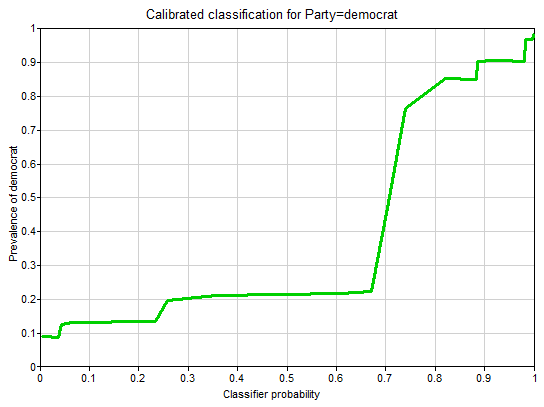
 



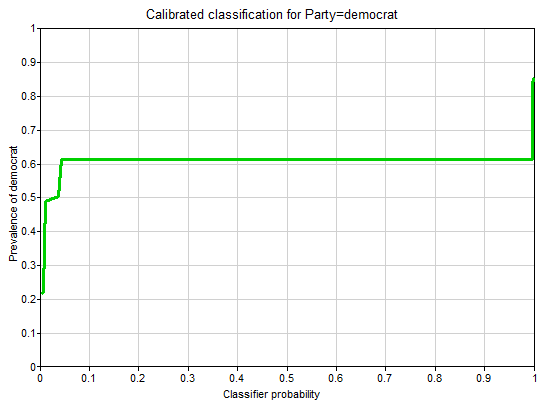
**Window Size: 1**



**Window Size: 11**

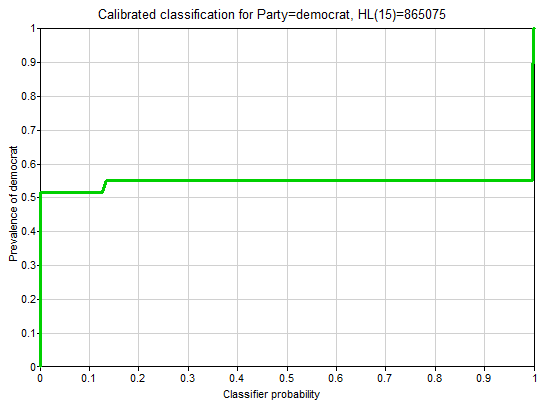
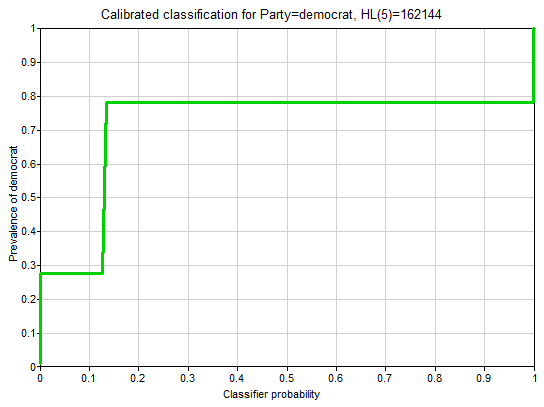


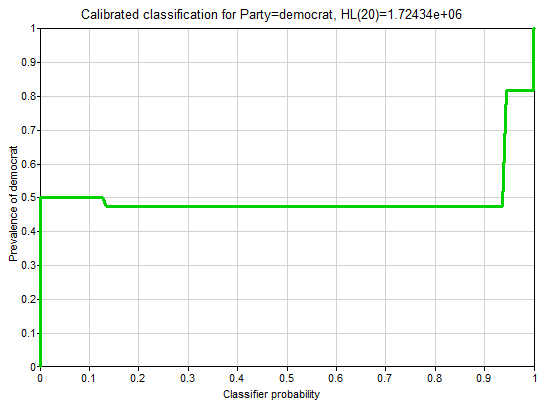
**Window Size: 21**



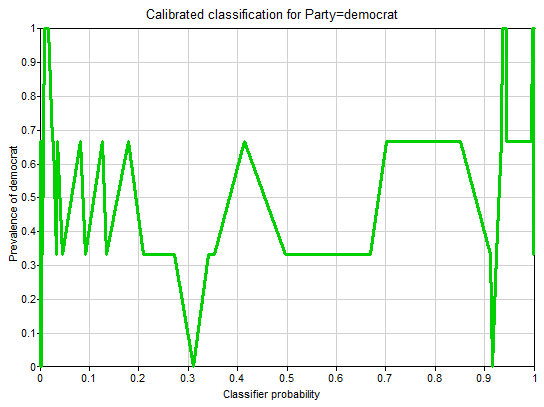
1. **House Votes Naïve:**

**Different Bin Count:**

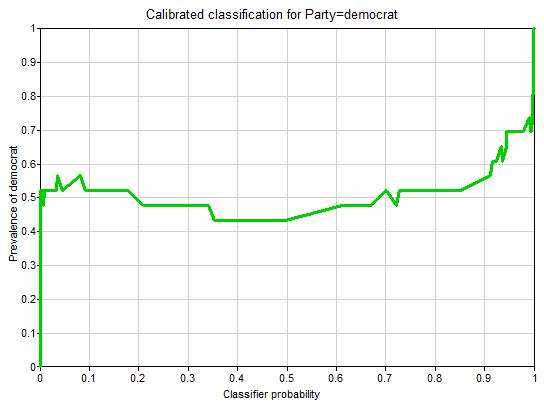




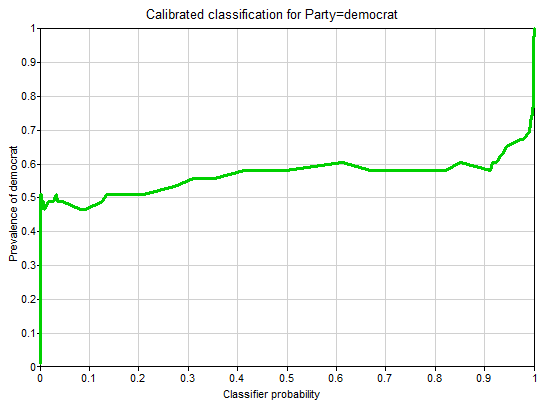
**Window Size: 1**



**Window Size: 11**

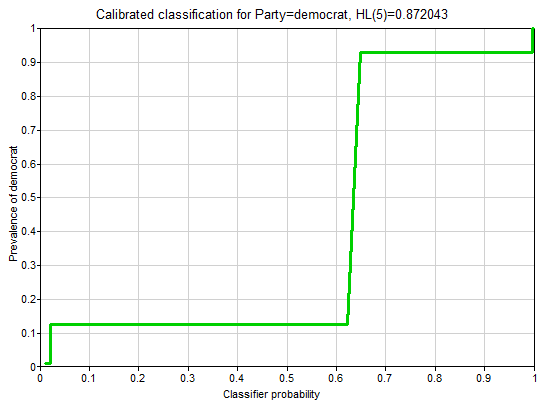
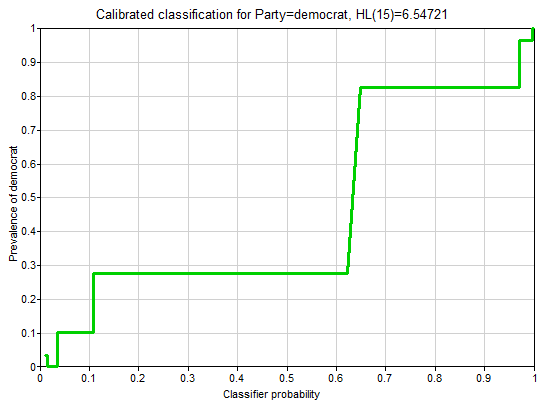


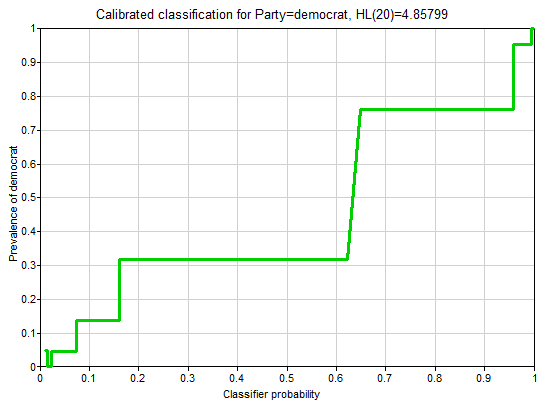
**Window Size: 21**



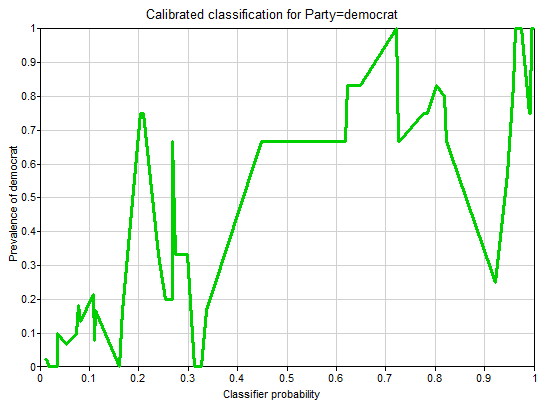
1. **House Votes PC:**

**Different Bin Count:**

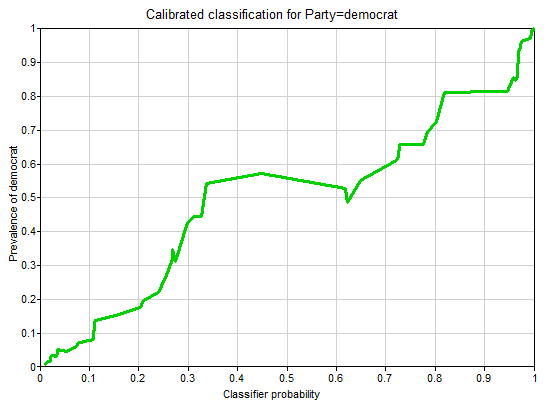
 



**Window Size: 1**



**Window Size: 11**



**Window Size: 21**

