

# Short Term Wind Prediction Using Bayesian Networks in WEKA

<https://github.com/npollard/windprediction>

Dr. Ole Mengshoel  
Nelson Pollard

## DATA SOURCE

### Weather Underground

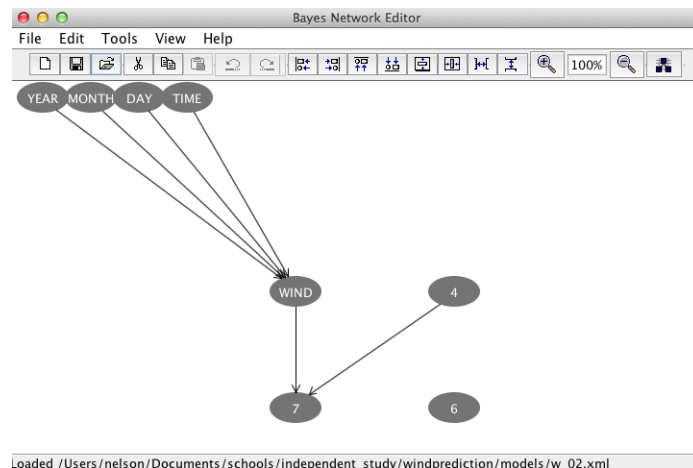
- <http://www.wunderground.com>
- Real-time weather information
- Nationwide
- 35 stations in Santa Cruz county
- used three stations along the Santa Cruz coast



```
Time,TemperatureF,DewpointF,PressureIn,WindDirection,WindI
2012-01-01 00:01:00,57.0,37.0,30.01,West,267,2.0,6.0,47,0.00,,
2012-01-01 00:06:00,56.8,36.9,30.01,West,259,5.0,6.0,47,0.00,,
2012-01-01 00:11:00,56.1,37.2,30.01,WNW,294,4.0,6.0,49,0.00
2012-01-01 00:16:00,54.3,37.9,30.01,NW,326,4.0,8.0,54,0.00,,(
2012-01-01 00:21:00,52.0,38.5,30.01,WNW,290,5.0,8.0,60,0.00
2012-01-01 00:27:00,49.8,38.5,30.01,WNW,300,7.0,8.0,65,0.00
2012-01-01 00:32:00,48.6,38.5,30.01,WNW,300,5.0,7.0,68,0.00
2012-01-01 00:37:00,47.8,38.5,30.01,NW,319,5.0,7.0,70,0.00,,(
2012-01-01 00:42:00,47.3,38.3,30.01,WNW,285,6.0,8.0,71,0.00
2012-01-01 00:47:00,47.1,38.1,30.01,WNW,291,6.0,8.0,71,0.00
2012-01-01 00:52:00,46.9,38.3,30.00,West,278,5.0,8.0,72,0.00,,
2012-01-01 00:57:00,46.9,38.3,30.01,West,277,5.0,8.0,72,0.00,,
2012-01-01 01:02:00,46.8,38.3,30.01,West,269,7.0,8.0,72,0.00,,
2012-01-01 01:07:00,46.8,38.3,30.01,WNW,288,5.0,9.0,72,0.00
2012-01-01 01:12:00,46.6,38.1,30.01,WNW,295,4.0,9.0,72,0.00
2012-01-01 01:17:00,46.6,38.1,30.00,West,278,4.0,9.0,72,0.00,,
2012-01-01 01:23:00,46.9,38.1,30.00,WNW,300,2.0,8.0,71,0.00
2012-01-01 01:28:00,48.0,37.9,30.01,WNW,299,4.0,8.0,68,0.00
2012-01-01 01:33:00,47.5,37.4,30.01,WSW,253,4.0,6.0,68,0.00,
2012-01-01 01:38:00,47.5,37.4,30.01,West,276,2.0,6.0,68,0.00,,
2012-01-01 01:43:00,48.6,37.4,30.01,WNW,283,2.0,6.0,65,0.00
2012-01-01 01:48:00,48.7,37.0,30.01,NNW,327,4.0,5.0,64,0.00,
2012-01-01 01:53:00,47.5,37.0,30.01,NW,319,3.0,5.0,67,0.00,,(
2012-01-01 01:58:00,46.6,37.0,30.01,West,278,3.0,4.0,69,0.00,,
2012-01-01 02:03:00,46.4,36.9,30.01,West,267,2.0,4.0,69,0.00,,
2012-01-01 02:08:00,46.2,36.7,30.01,West,280,2.0,4.0,69,0.00,,
2012-01-01 02:13:00,46.0,36.9,30.02,WNW,285,4.0,4.0,70,0.00
2012-01-01 02:18:00,45.9,36.7,30.02,West,271,2.0,3.0,70,0.00,,
```

## Weka

- <http://www.cs.waikato.ac.nz/ml/weka>
- collection of machine learning algorithms for data mining
- written in Java
- command line and GUI
- generate Bayesian network models

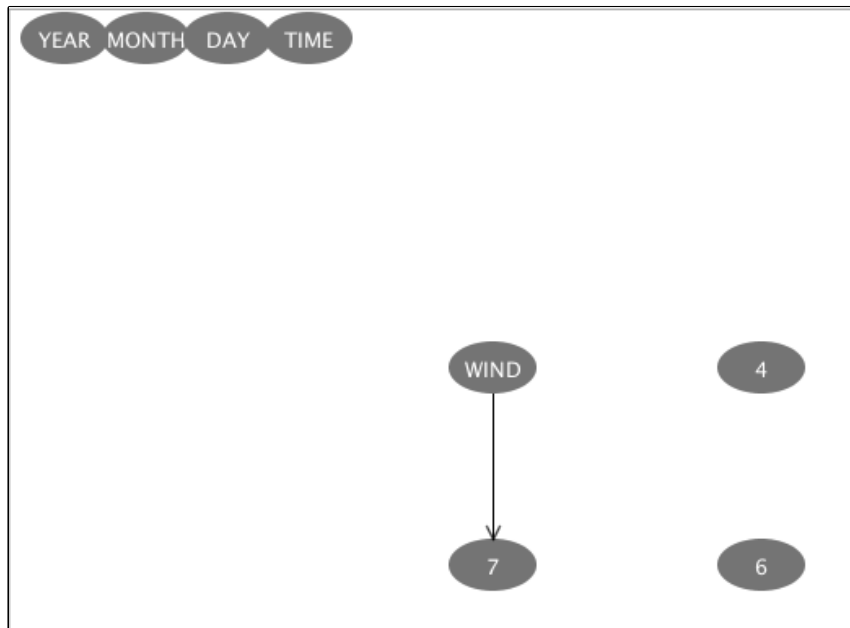


## PYTHON SCRIPTS

- grab CSV from Weather Underground
- clean up and reorder attributes
- convert to WEKA-usable format
- run WEKA BayesNet algorithm for:
  - all datasets
  - all models

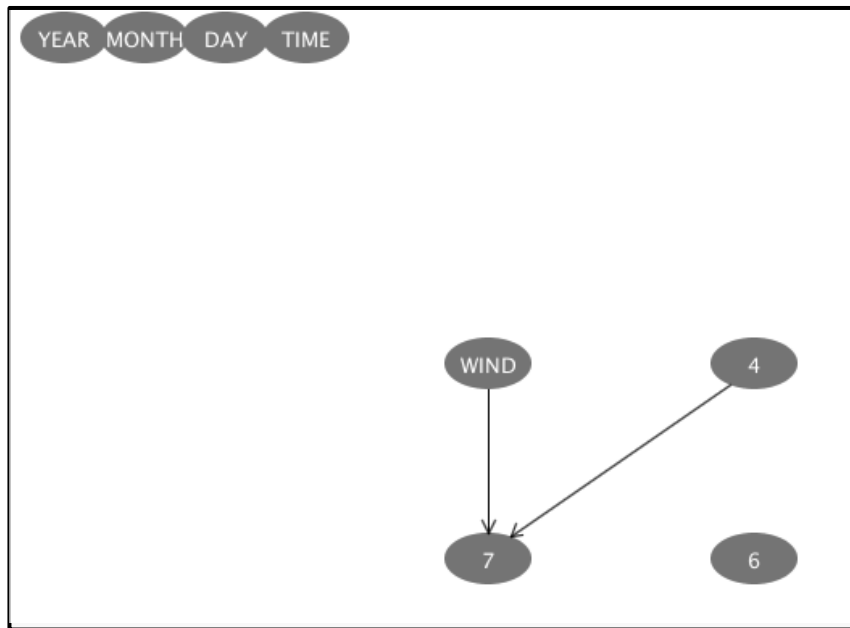


## RESULTS - w\_00



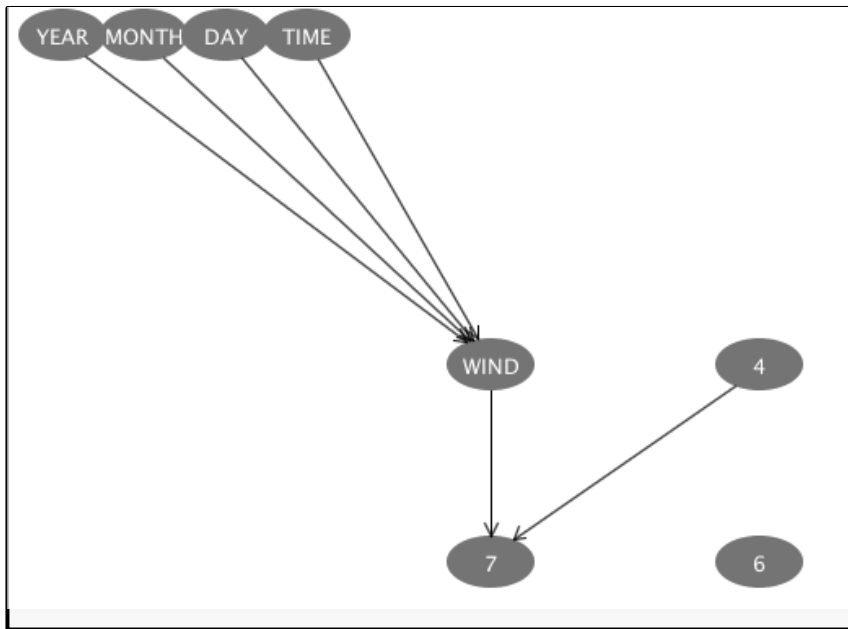
Station	Prediction Accuracy	Cross Validation
KCASANTA132	73.8183 %	73.8156 %
KCASANTA154	72.1022 %	71.3163 %
KCASANTA166	64.6465 %	64.1414 %

## RESULTS - w\_01



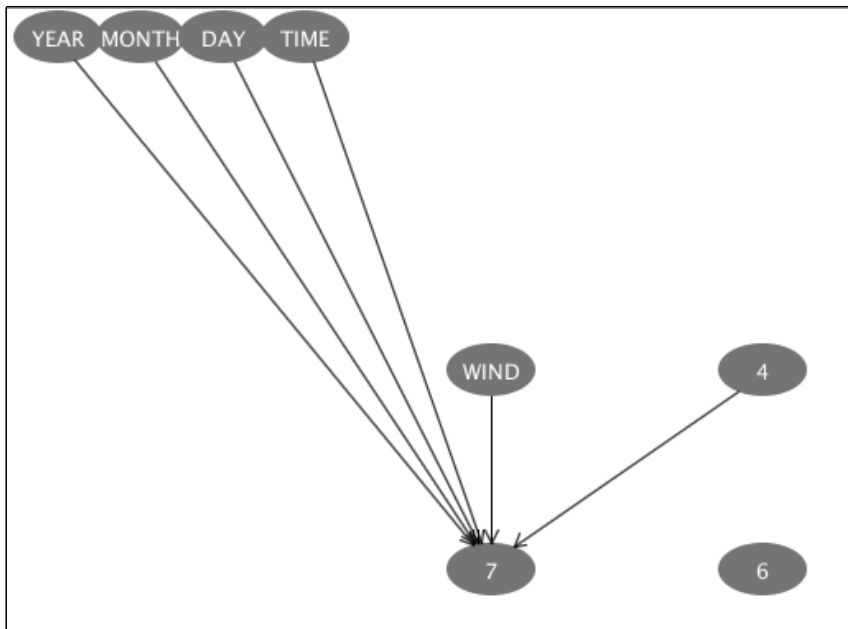
Station	Prediction Accuracy	Cross Validation
KCASANTA132	74.0947 %	74.0477 %
KCASANTA154	74.6562 %	67.387 %
KCASANTA166	71.7172 %	53.5354 %

## RESULTS - w\_02



Station	Prediction Accuracy	Cross Validation
KCASANTA132	74.0947 %	74.0477 %
KCASANTA154	74.6562 %	67.387 %
KCASANTA166	71.7172 %	53.5354 %

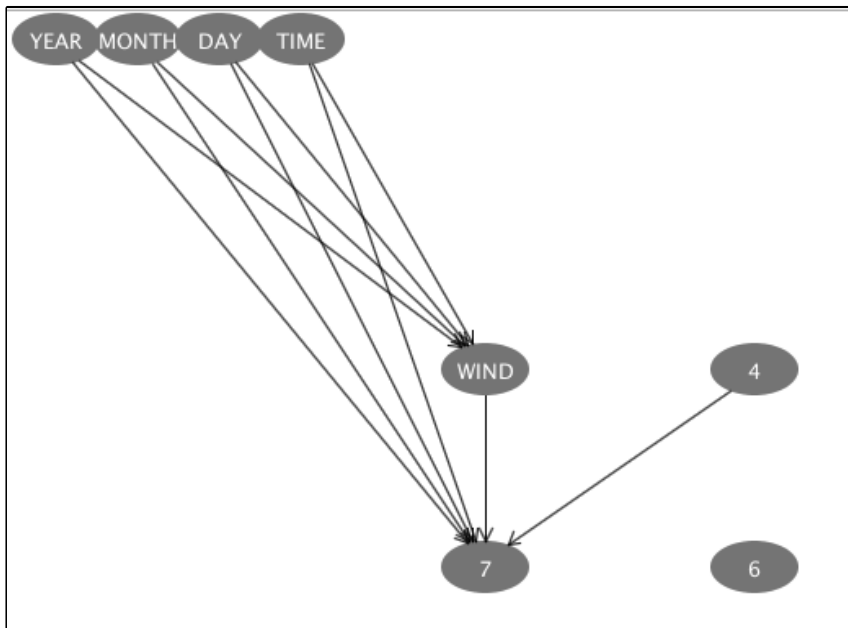
## RESULTS - w\_03



Station	Prediction Accuracy	Cross Validation
KCASANTA132	TODO	TODO
KCASANTA154	81.7289 %	51.6699 %
KCASANTA166	78.7879 %	42.4242 %

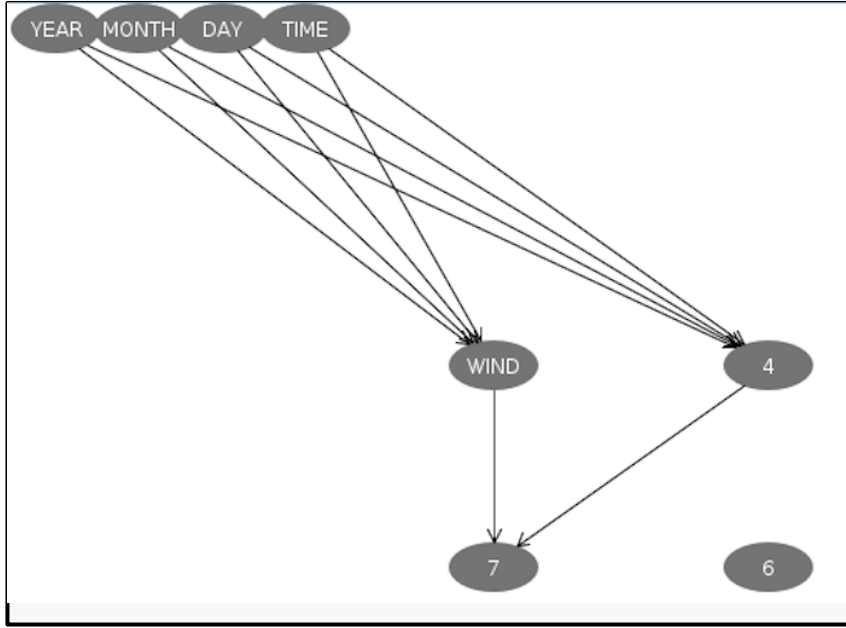


## RESULTS - w\_04



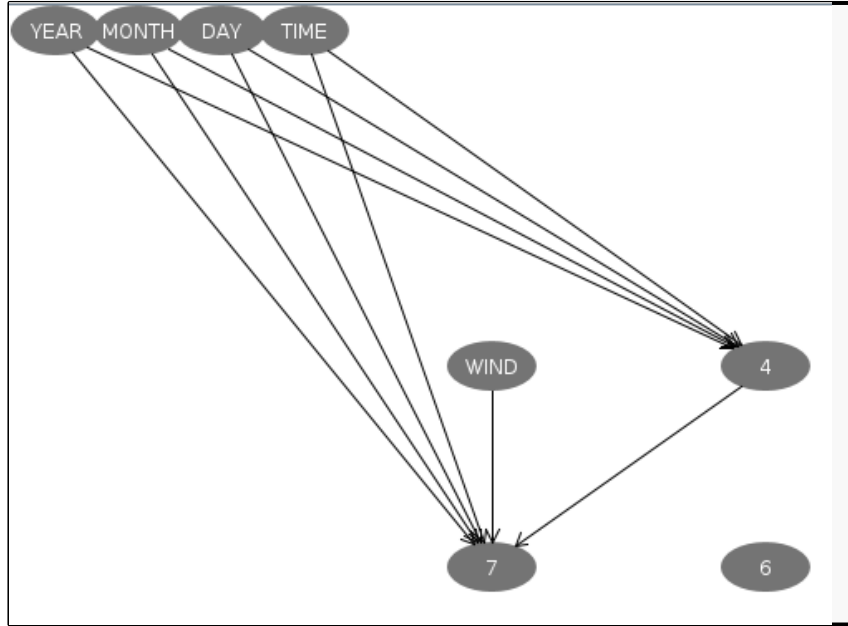
Station	Prediction Accuracy	Cross Validation
KCASANTA132	80.1401 %	72.5416 %
KCASANTA154	81.7289 %	51.6699 %
KCASANTA166	78.7879 %	42.4242 %

## RESULTS - w\_05



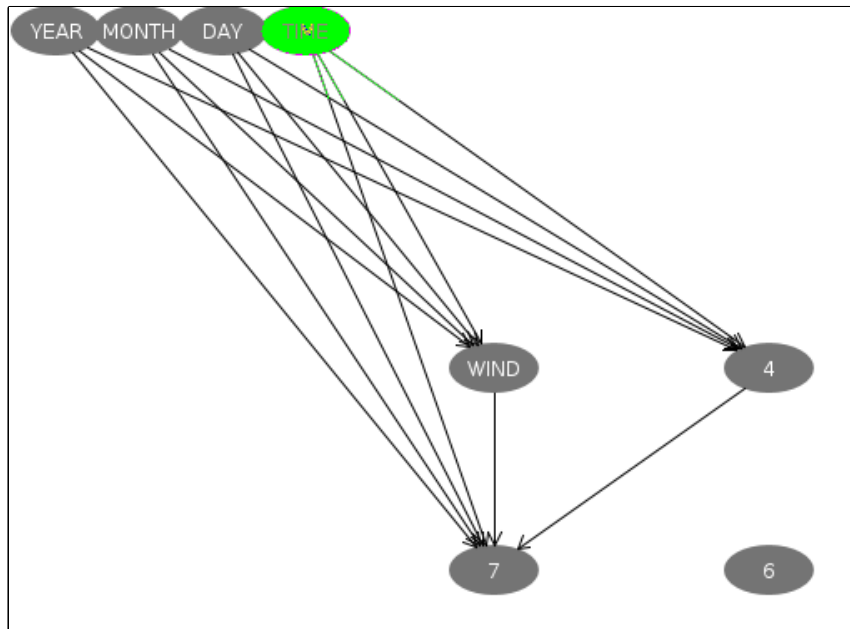
Station	Prediction Accuracy	Cross Validation
KCASANTA132	74.0947 %	74.0477 %
KCASANTA154	74.6562 %	67.387 %
KCASANTA166	71.7172 %	53.5354 %

## RESULTS - w\_06



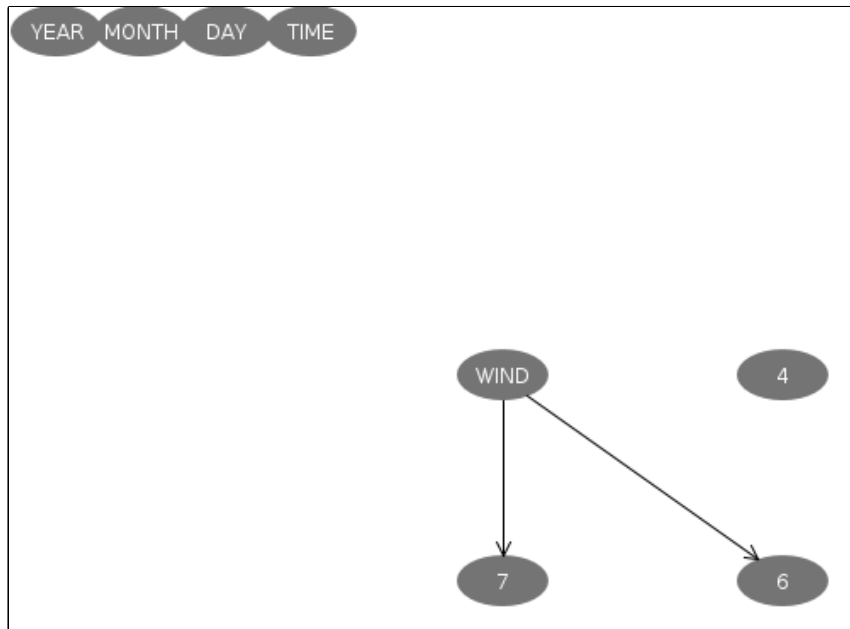
Station	Prediction Accuracy	Cross Validation
KCASANTA132	80.1401 %	72.5416 %
KCASANTA154	81.7289 %	51.6699 %
KCASANTA166	78.7879 %	42.4242 %

## RESULTS - w\_07



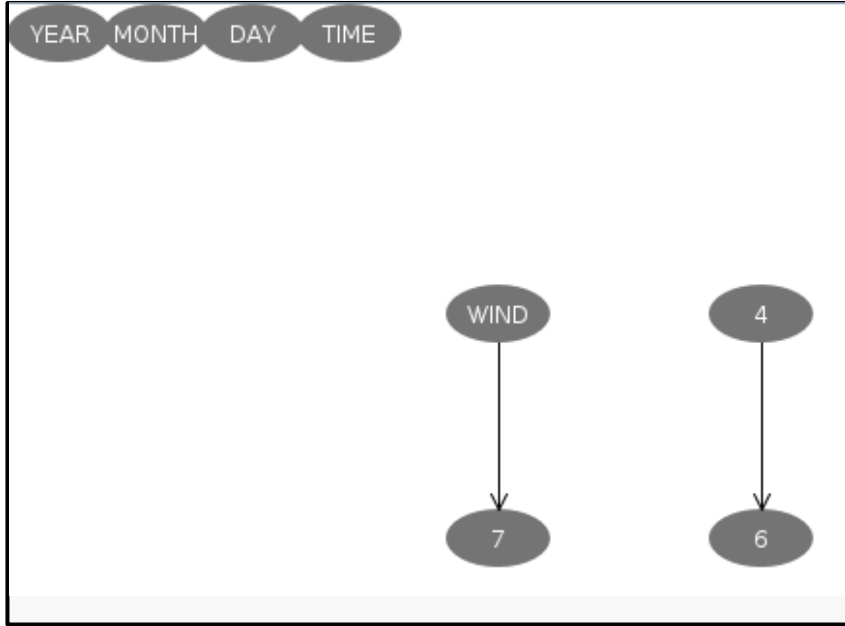
Station	Prediction Accuracy	Cross Validation
KCASANTA132	80.1401 %	72.5416 %
KCASANTA154	81.7289 %	51.6699 %
KCASANTA166	78.7879 %	42.4242 %

## RESULTS - d\_00



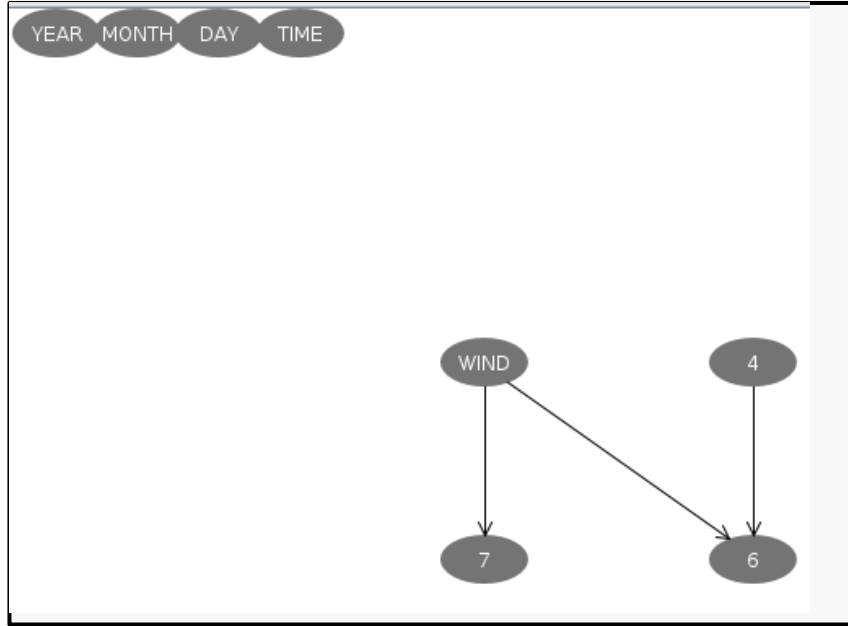
Station	Prediction Accuracy	Cross Validation
KCASANTA132	73.8183 %	73.8156 %
KCASANTA154	72.1022 %	71.3163 %
KCASANTA166	64.6465 %	64.1414 %

## RESULTS - d\_01



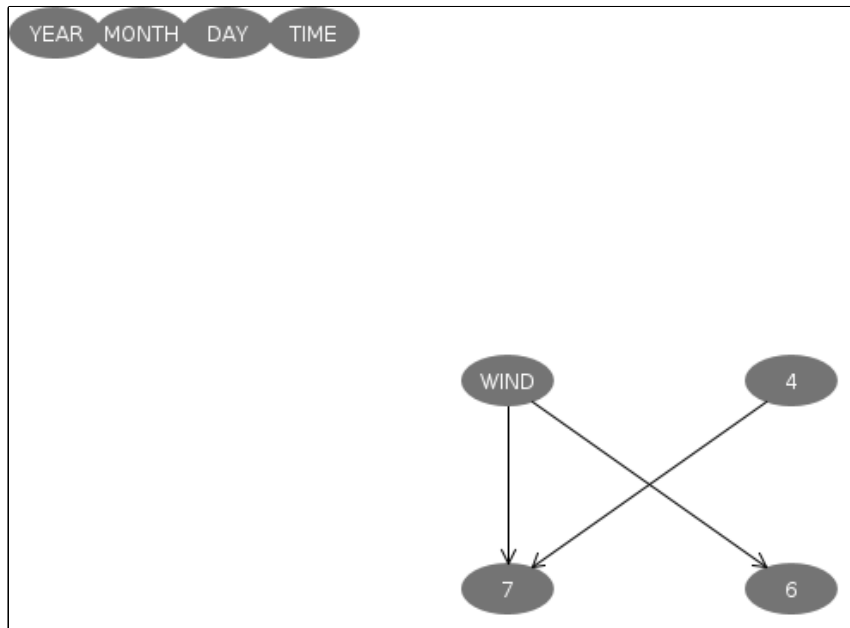
Station	Prediction Accuracy	Cross Validation
KCASANTA132	73.8183 %	73.8156 %
KCASANTA154	72.1022 %	71.3163 %
KCASANTA166	64.6465 %	64.1414 %

## RESULTS - d\_02



Station	Prediction Accuracy	Cross Validation
KCASANTA132	73.8183 %	73.8156 %
KCASANTA154	72.1022 %	71.3163 %
KCASANTA166	64.6465 %	64.1414 %

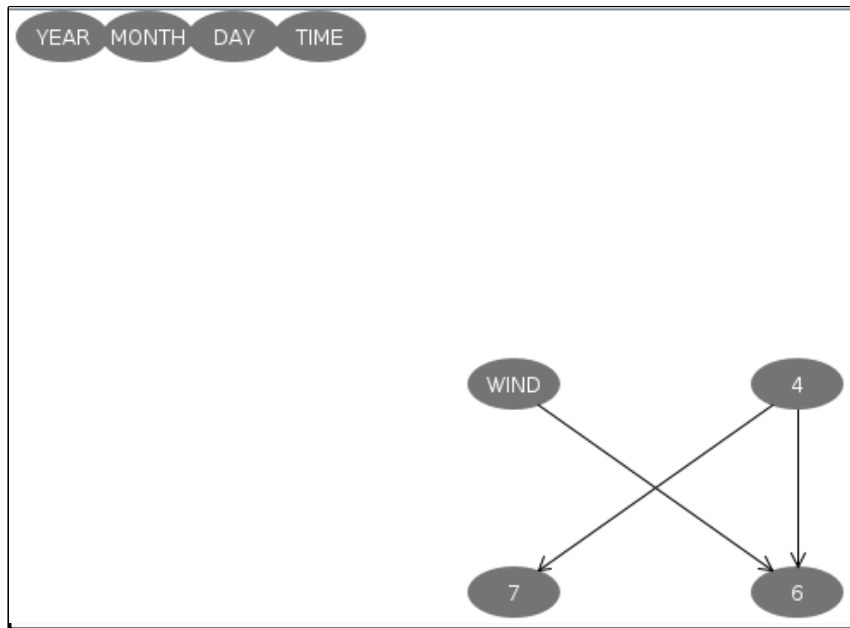
## RESULTS - d\_03



Station	Prediction Accuracy	Cross Validation
KCASANTA132	74.0947 %	74.0477 %
KCASANTA154	74.6562 %	67.387 %
KCASANTA166	71.7172 %	53.5354 %

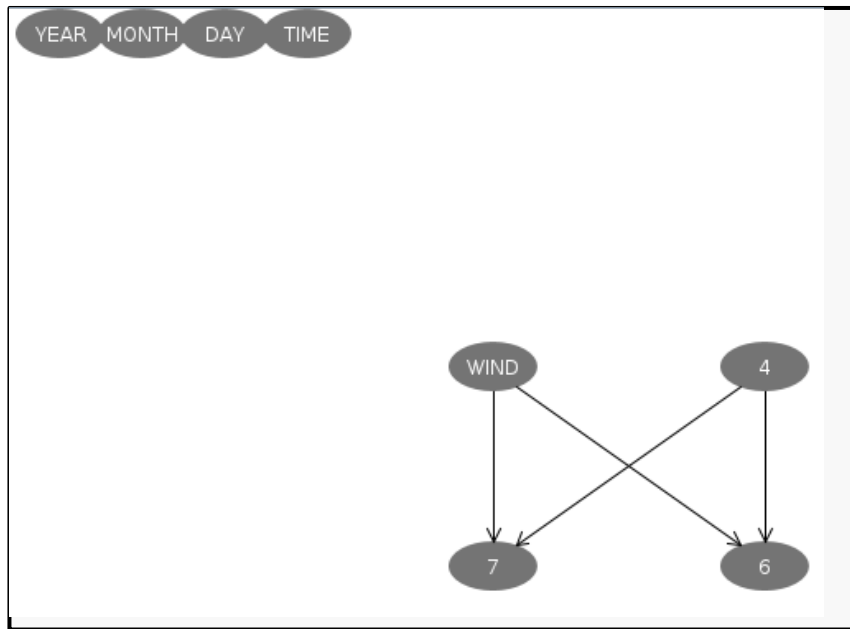


## RESULTS - d\_04



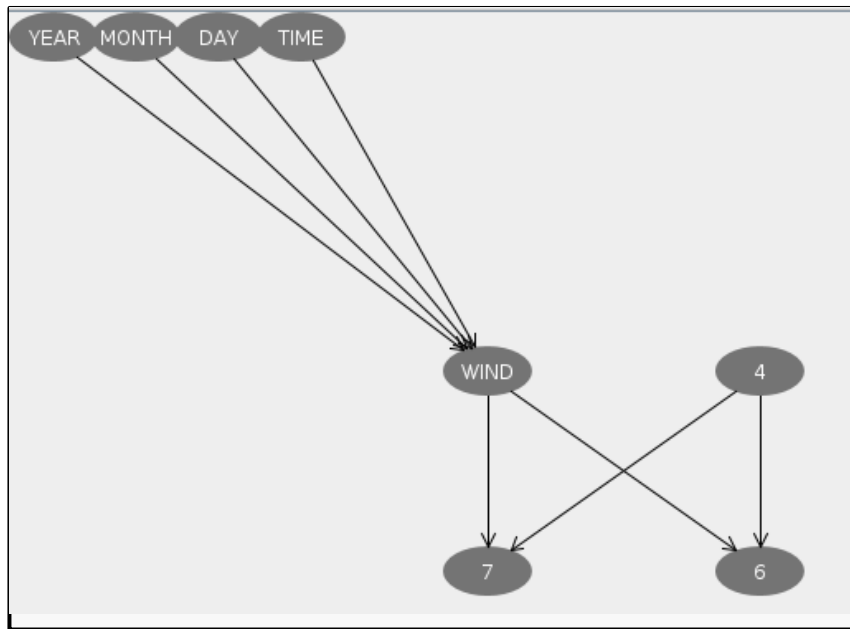
Station	Prediction Accuracy	Cross Validation
KCASANTA132	56.9811 %	56.9811 %
KCASANTA154	40.668 %	35.167 %
KCASANTA166	31.8182 %	26.2626 %

## RESULTS - d\_05



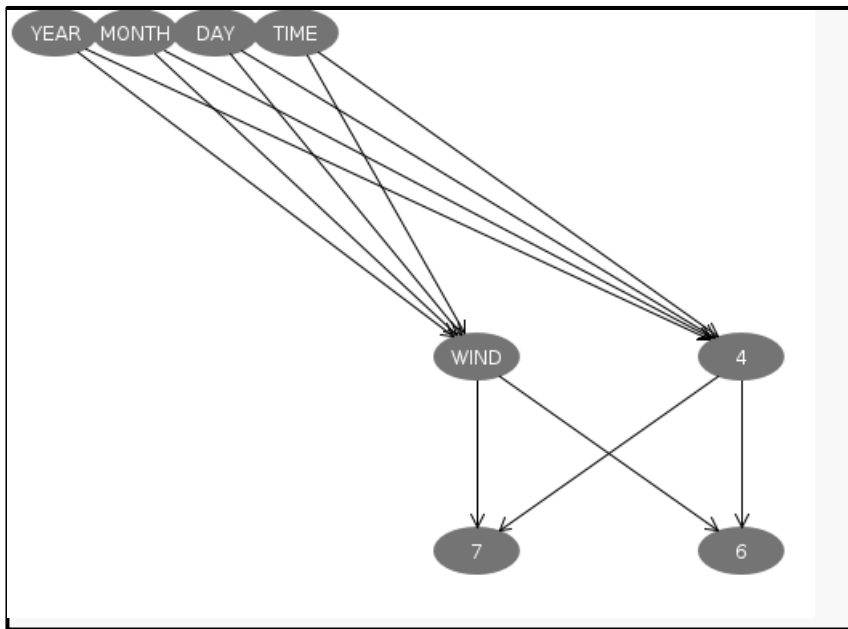
Station	Prediction Accuracy	Cross Validation
KCASANTA132	74.0947 %	74.0477 %
KCASANTA154	74.6562 %	67.387 %
KCASANTA166	71.7172 %	53.5354 %

## RESULTS - d\_06



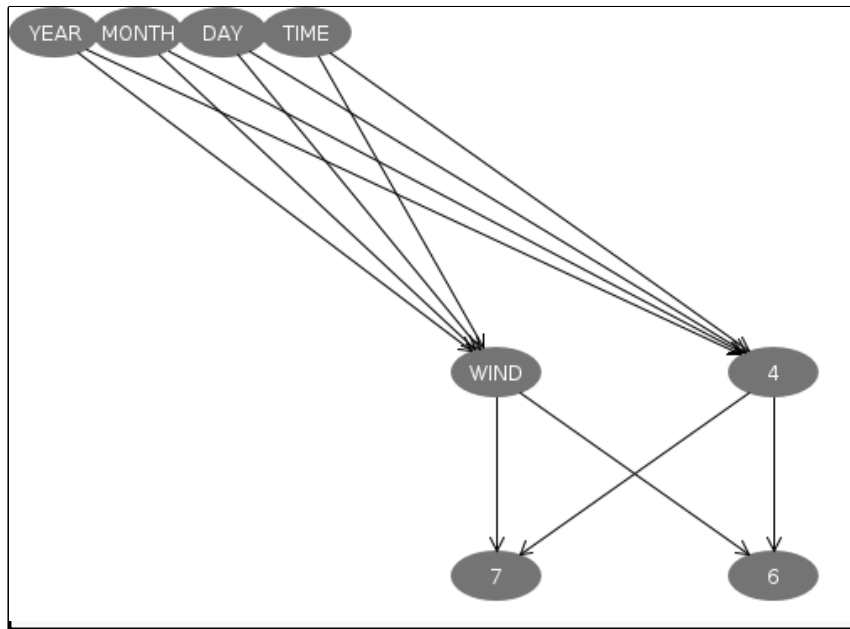
Station	Prediction Accuracy	Cross Validation
KCASANTA132	74.0947 %	74.0477 %
KCASANTA154	74.6562 %	67.387 %
KCASANTA166	71.7172 %	53.5354 %

## RESULTS - d\_07



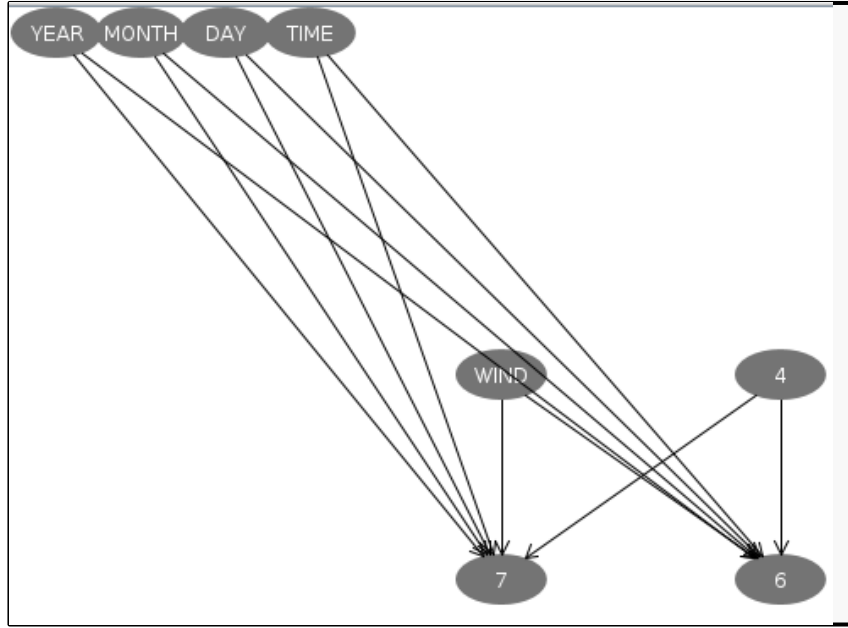
Station	Prediction Accuracy	Cross Validation
KCASANTA132	74.0947 %	74.0477 %
KCASANTA154	74.6562 %	67.387 %
KCASANTA166	71.7172 %	53.5354 %

## RESULTS - d\_08



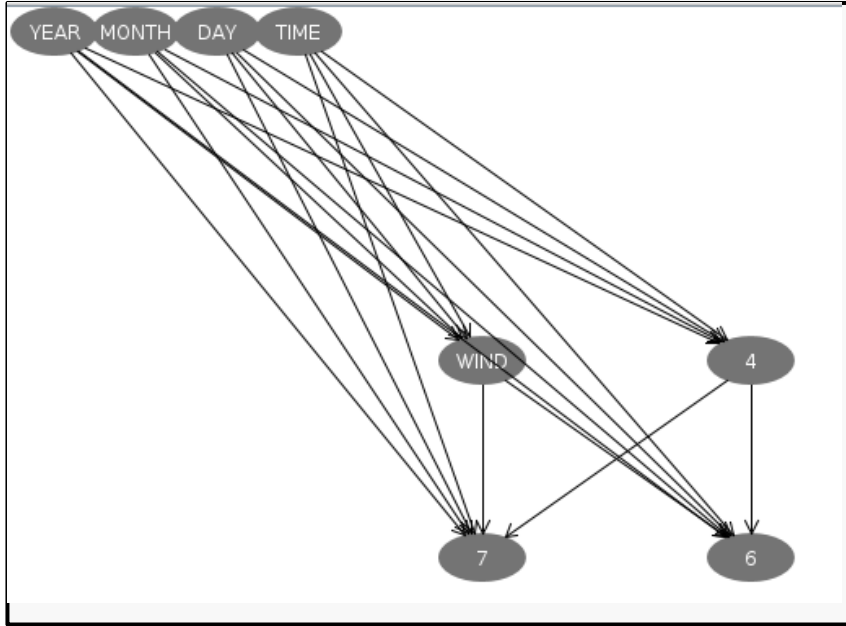
Station	Prediction Accuracy	Cross Validation
KCASANTA132	74.0947 %	74.0477 %
KCASANTA154	74.6562 %	67.387 %
KCASANTA166	71.7172 %	53.5354 %

## RESULTS - d\_09



Station	Prediction Accuracy	Cross Validation
KCASANTA132	80.1401 %	72.5416 %
KCASANTA154	81.7289 %	51.6699 %
KCASANTA166	78.7879 %	42.4242 %

## RESULTS - d\_10



Station	Prediction Accuracy	Cross Validation
KCASANTA132	80.1401 %	72.5416 %
KCASANTA154	81.7289 %	51.6699 %
KCASANTA166	78.7879 %	42.4242 %

## ANALYSIS

- Only immediate parents of wind have an effect
- Considering datetime, previous wind, and previous direction all improve prediction
  - About 80% accurate prediction



## FUTURE WORK

- More datasets
- More models
  - Consider year, month, day, time separately
  - Use other stations' current wind
- Different algorithms
  - CART