Evaluation of Decision Tree in KNIME

Learning Objectives

At the end of this activity, you will be able to perform the following operations in KNIME:

1. Create and interpret a confusion matrix for a decision tree
2. Determine the accuracy rate of a decision tree model
3. Use highlighting to analyze classification errors

Problem Description

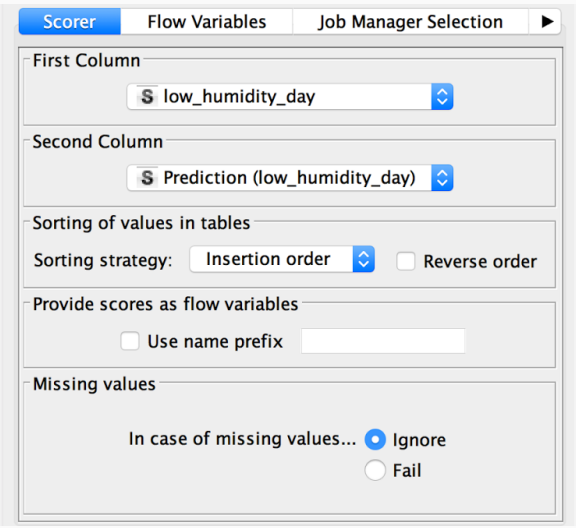
With the decision tree classifier built, we now need to evaluate its performance.

Steps

Generate a Confusion Matrix and Determine Accuracy Rate

A confusion matrix shows the type of errors and correct classifications that a classifier makes. It can be generated using a **Scorer** node.

1. Open the Decision Tree Workflow that you created from the Classification Hands-On reading.
2. Connect a **Scorer** node to the existing **Decision Tree Predictor**.
3. The Scorer Configure Dialog should look like this by default. Click OK.



Execute and view the **Scorer** node. It shows the confusion matrix, along with the accuracy of the prediction. Here you should see an accuracy rate of 80.282% if you followed all the hands-on instructions.

low_humidity_day \ Prediction (low_humidity_day)	humidity_low	humidity_not_low
humidity_low	76	24
humidity_not_low	18	95

Correct classified: 171 Wrong classified: 42

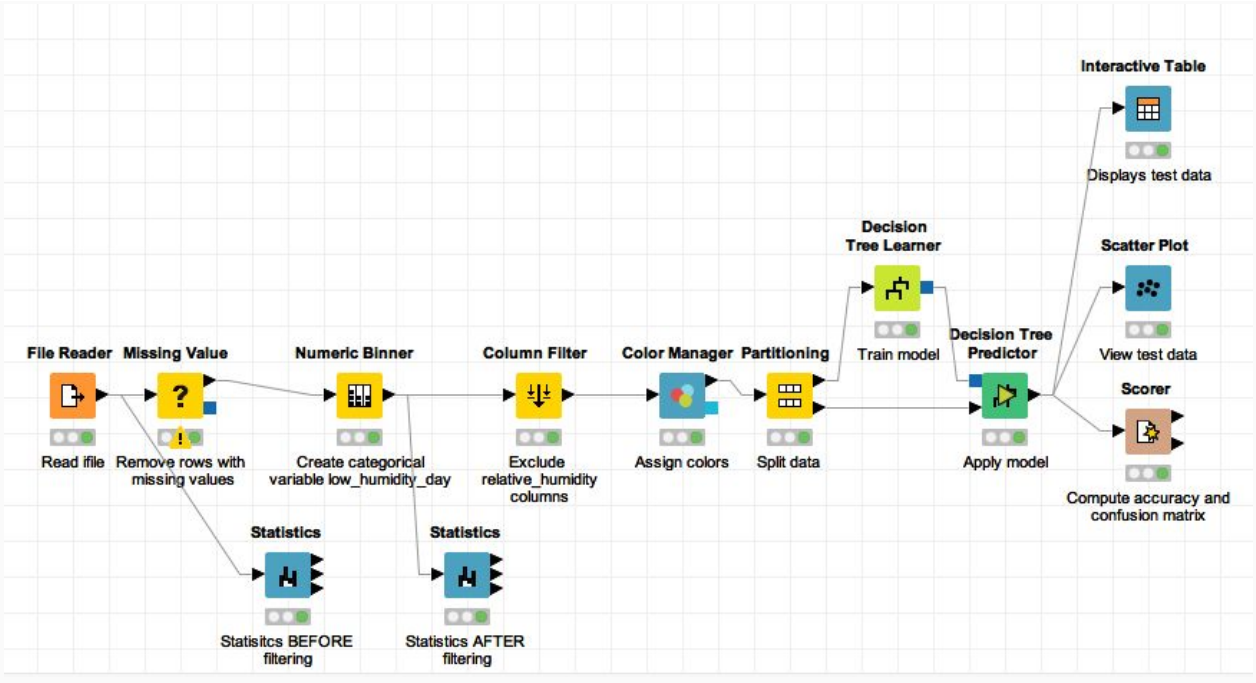
Accuracy: 80.282 % Error: 19.718 %

Cohen's kappa (κ) 0.603

From the confusion matrix, we see the following:

- There are 213 samples in the test data set (the sum of all the values in the confusion matrix)
- 76 humidity_low samples with were correctly classified
- 95 humidity_not_low samples were correctly classified
- The accuracy rate is (76 + 95) / 213 = 171 / 213 = 80.282%
- 24 humidity_low samples were incorrectly classified as humidity_not_low
- 18 humidity_not_low samples were incorrectly classified as humidity_low
- The error rate is (24 + 18) / 213 = 42 / 213 = 19.718%

Use Highlighting and Scatter Plot to Analyze Classification Errors



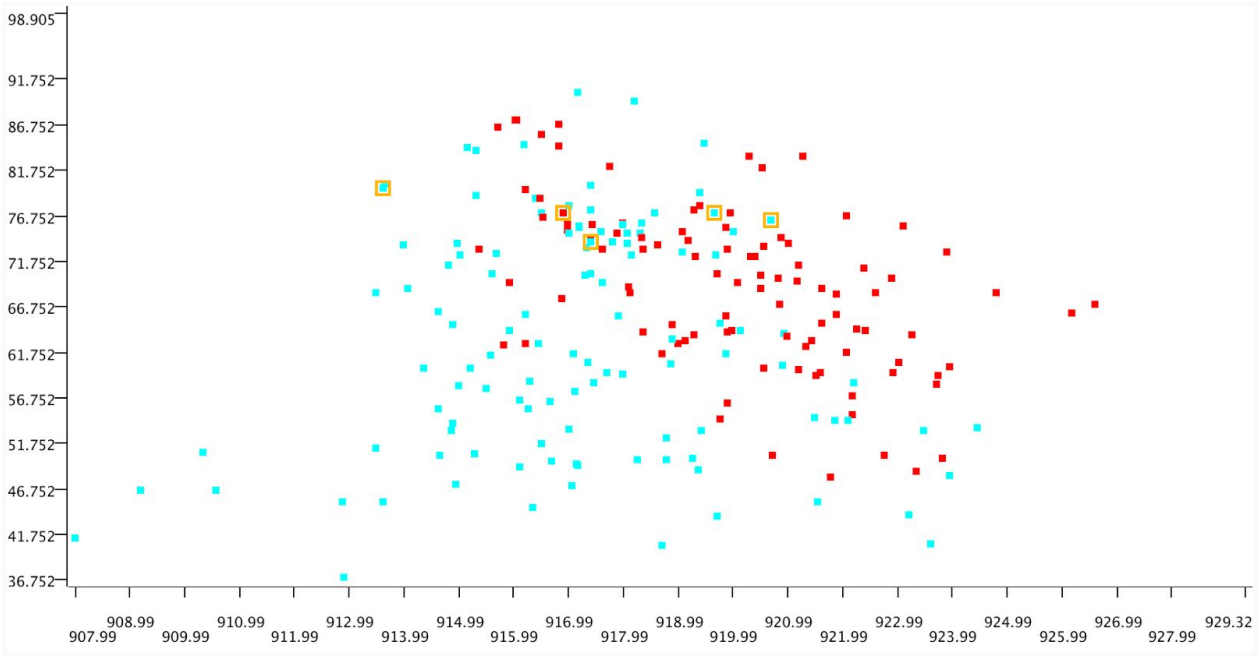
A good way to enhance analysis of incorrect predictions is to visualize them. This can be accomplished using a feature called **hiliting**, and viewing the data in a **Scatter Plot** node.

1. Connect an **Interactive Table** node to the **Decision Tree Predictor**.
2. Execute and view this **Interactive Table** to see the input values for each sample (row), along with the ACTUAL/TRUE low_humidity_day value and the PREDICTED low_humidity_day value. The red and blue squares next to the Row ID color-codes the actual/true label (low or not). You can use this table to analyze samples whose true value differs from the predicted value (incorrect prediction).
3. Connect a **Scatter Plot** node to the **Decision Tree Predictor**.
4. Execute and view the **Scatter Plot** node, and place the window side-by-side with the **Interactive Table** window.
5. Go through the the table looking for rows with predictions that are different from the true value.
6. When you find such a row, click anywhere on that row. At the top of the window click **Hilite > Hilite Selected**. This will make that row yellow in the table and in the Scatter Plot. It may be easiest to use the up and down arrow keys to navigate the rows of the table. In this example, we are just going to highlight the first 5 misclassifications.
7. Do this for any row with a misclassification. This allows you to pinpoint the misclassified samples and analyze them further. Analyzing the misclassified samples can bring insight into how to improve model performance. For example, if many samples with

avg_temperature_9am between 60 and 70 degrees are misclassified, this suggests that more samples with these values for avg_temperature_9am are needed to train the model.

File Hilite Navigation View Output

Row ID	air_pr...	air_te...	avg_w...	avg_w...	max_...	max_...	rain_a...	rain_d...	low_humidi...	Prediction (...
10	919.65	77.036	70.6	3.825	85.5	4.765	0	0	humidity_not_low	humidity_low
15	922.383	70.865	36.174	1.847	58.429	2.529	0	0	humidity_low	humidity_low
17	916.915	77.019	234.539	2.275	229.474	2.907	0	0	humidity_low	humidity_not_low
31	916.28	55.544	174.6	2.595	191.7	3.378	0	0	humidity_not_low	humidity_not_low
33	918.37	63.914	53.7	14.451	72.1	17.045	0	0	humidity_low	humidity_low
38	914.66	50.36	177.6	8.523	186.3	10.021	0	20	humidity_not_low	humidity_not_low
43	914.9	64.688	174.6	5.659	184.5	6.867	0	0	humidity_not_low	humidity_not_low
46	917.9	65.66	183.7	6.062	188.5	6.509	0.021	319	humidity_not_low	humidity_not_low
47	916.05	87.188	210	1.566	109.3	2.595	0	0	humidity_low	humidity_low
50	921.9	65.876	199	4.496	207.2	5.324	0	0	humidity_low	humidity_low
51	918.58	76.982	150.7	1.32	182	1.946	0	0	humidity_not_low	humidity_not_low
63	917.02	53.258	180.2	3.132	216.6	4.452	0	0	humidity_not_low	humidity_not_low
65	921.29	83.228	162.6	2.662	194	3.557	0	0	humidity_low	humidity_low
66	922.06	61.754	56.2	10.424	77	13.422	0	0	humidity_low	humidity_low
70	917.763	82.064	110.258	2.235	129.78	3.206	0	0	humidity_low	humidity_low
76	922.07	76.742	165.736	3.771	205.67	6.004	0	0	humidity_low	humidity_low
79	915.57	61.34	174.3	4.899	180.6	5.503	0	0	humidity_not_low	humidity_not_low
81	917.46	58.298	182.3	5.906	190.3	6.935	0	0	humidity_not_low	humidity_not_low
84	920.93	63.824	196.4	7.65	204	8.567	0	0	humidity_not_low	humidity_not_low
88	915.914	69.389	46.736	10.977	64.645	14.298	0	9.397	humidity_low	humidity_low
91	915.32	83.84	219.9	1.521	245	2.058	0	0	humidity_not_low	humidity_not_low
96	921.874	68.094	70.909	2.147	96.095	2.891	0	0	humidity_low	humidity_low
98	916.52	51.638	190.8	7.404	198	8.389	0.37	2,170	humidity_not_low	humidity_not_low
103	921.5	54.536	220	5.011	234.9	6.442	0	0	humidity_not_low	humidity_not_low
105	917.4	73.868	181.7	4.138	189.8	4.765	0.05	460	humidity_not_low	humidity_low
106	914.95	47.21	183.9	7.628	191.2	8.612	20.02	11,650	humidity_not_low	humidity_not_low
112	916.45	62.654	193.7	3.534	208.8	4.429	0	0	humidity_not_low	humidity_not_low
120	920.7	76.172	190.9	6.263	199.6	7.27	0	0	humidity_not_low	humidity_low
122	918.32	74.732	239.5	2.058	273.4	3.02	0	0	humidity_not_low	humidity_not_low
124	913.61	79.628	180.5	3.266	187.9	3.758	0	0	humidity_not_low	humidity_low



Save Your Workflow

Save your workflow using <control>-s on Windows or <command>-s on Mac, or selecting File>Save or File>Save As.