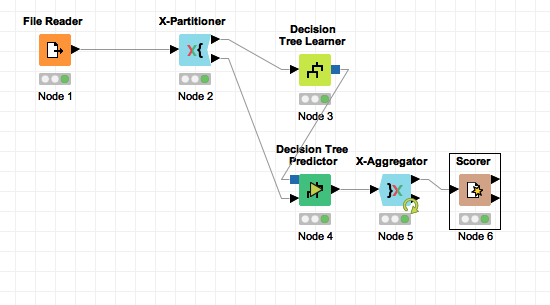
Broc Nickodemus

Machine Learning hw3

Background Information:

The data used in the following models was based on Iris plant data and census income data. The models were created using Knime and use a decision tree learner and a mlp learner.

1. Replicate the general machine learning workflow (on page 2) that uses a decision tree classifier to classify the canonical “Iris” dataset. When you are finished, take a screen shot of your workflow and document the classification results (e.g., accuracy, precision, recall, etc.,).



Classification Results:

The decision tree predictor classified the following:

iris-setosa

accuracy of 100%

precision of 100%

recall of 100%

iris-versicolor

accuracy of 94.6%

precision of 90.4%

recall of 94%

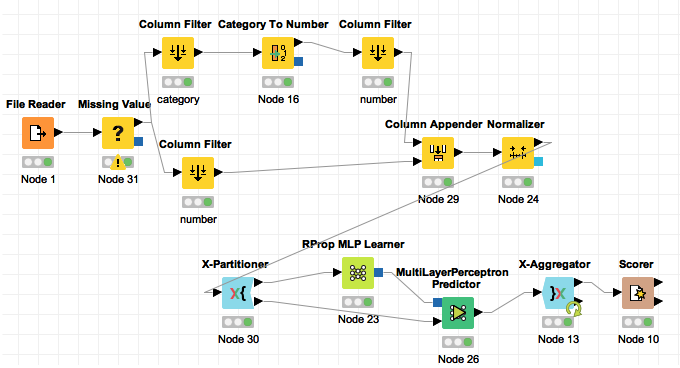
iris-virginica

accuracy of 94.6%

precision of 93.8%

recall of 90%

2. Create a pattern recognition workflow to experiment with the “Census Income” dataset from homework #1. Take a screen shot of your workflow and document the classification results.



a. What type of classifier did you use?

I used a multilayer perception predictor as my classifier. I decided to use a multilayer perception predictor because I wanted to use a neural network.

b. What features did you use? Why?

I ended up using a missing value node to remove missing values in the dataset. I also needed to separate the columns so that I could convert the data formatted as strings into numbers. Since the MLP Learner only accepts data between 0-1, I needed to run the columns through a normalizer. The scorer then showed the accuracy statistics of the multilayer perceptron predictor.

c. What accuracy did you reach? recall? precision?

In the model, I compared column 14 (representing income >50K or <=50K) to the predicted column 14. The model had an accuracy of 74.79%. If I increased the number of cross validations in the X-Partitioner, the accuracy might have increased. Since the model resulted in 0 true positives, the precision and the recall were both 0%. I learned that a multilayer perceptron predictor may not be the best approach to try to categorically organize multidimensional data.