

CISC 333

Assignment 3: Titanic Normalization & Discretization Analysis

Name: Sam Ruscica

Student Number: 10148585

Course: CISC 333

Date: October 6, 2017

Procedure & Results

This analysis was conducted to examine the effects of normalization and discretization techniques on a decision tree prediction model. These models would then be compared to past models that were made. Below are the results.

Confusion Matrix - 0:9 - Scorer		
File Hilite		
survived \ Prediction (survived)	0.0	1.0
0.0	145	23
1.0	29	71
Correct classified: 216		Wrong classified: 52
Accuracy: 80.597 %		Error: 19.403 %
Cohen's kappa (κ) 0.58		

Figure 1: Normalization Matrix

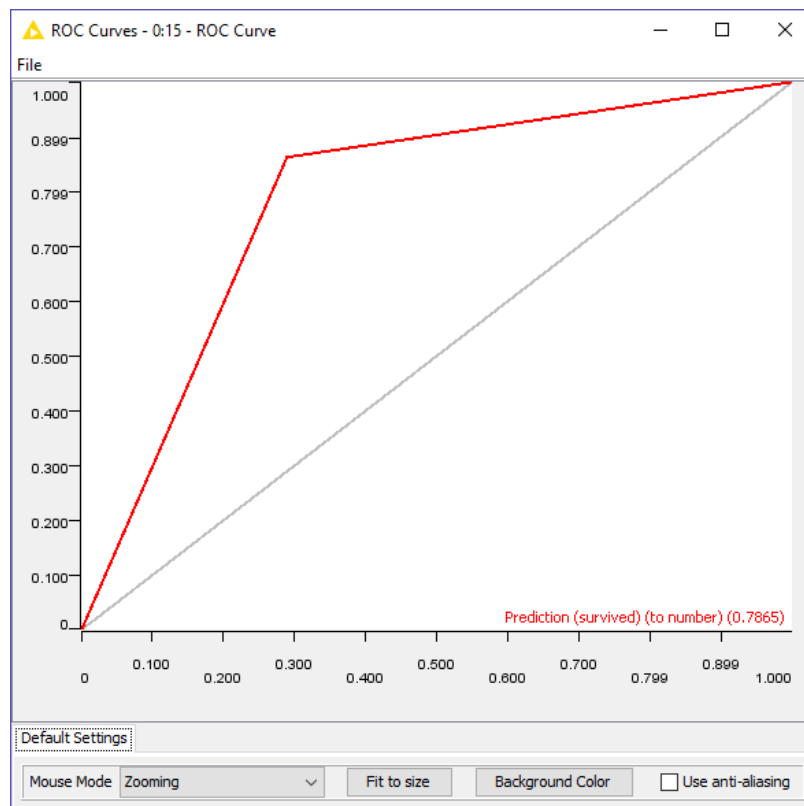


Figure 2: Normalization ROC Curve

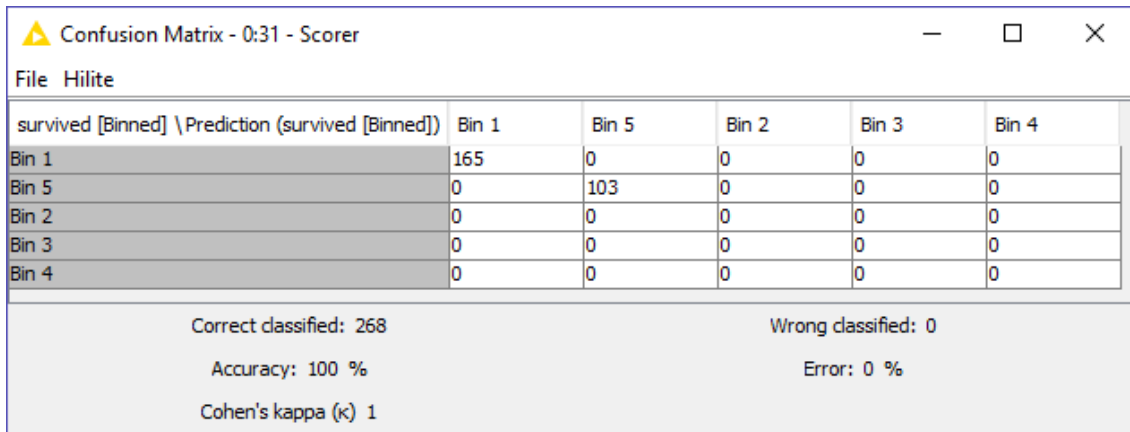


Figure 3: Discretization Matrix

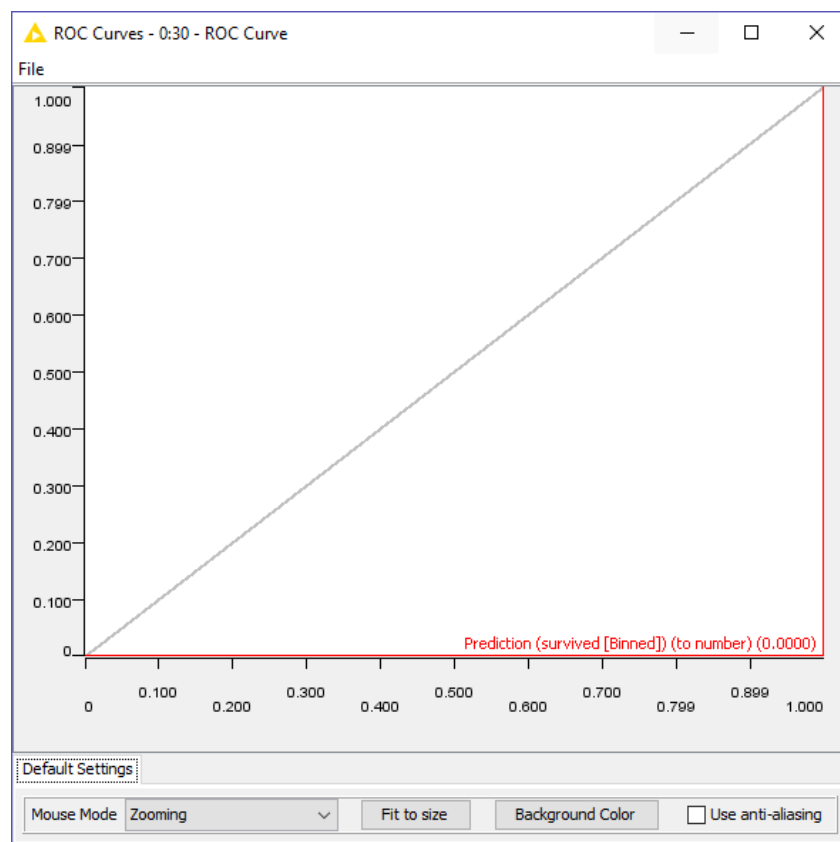


Figure 4: Discretization ROC Curve

Analysis

The results from the new models are similar to older models. The normalized model is actually slightly worse than prior ones. This may be caused due to rounding limitations when converting to normalized values. Additionally, it is just representing the data in a different way so the results should be similar. The discretized model seems to have issues with the scorer and ROC curve. This might be caused by the fact that it is comparing the predicted bins to the original bins.

KNIME Nodes

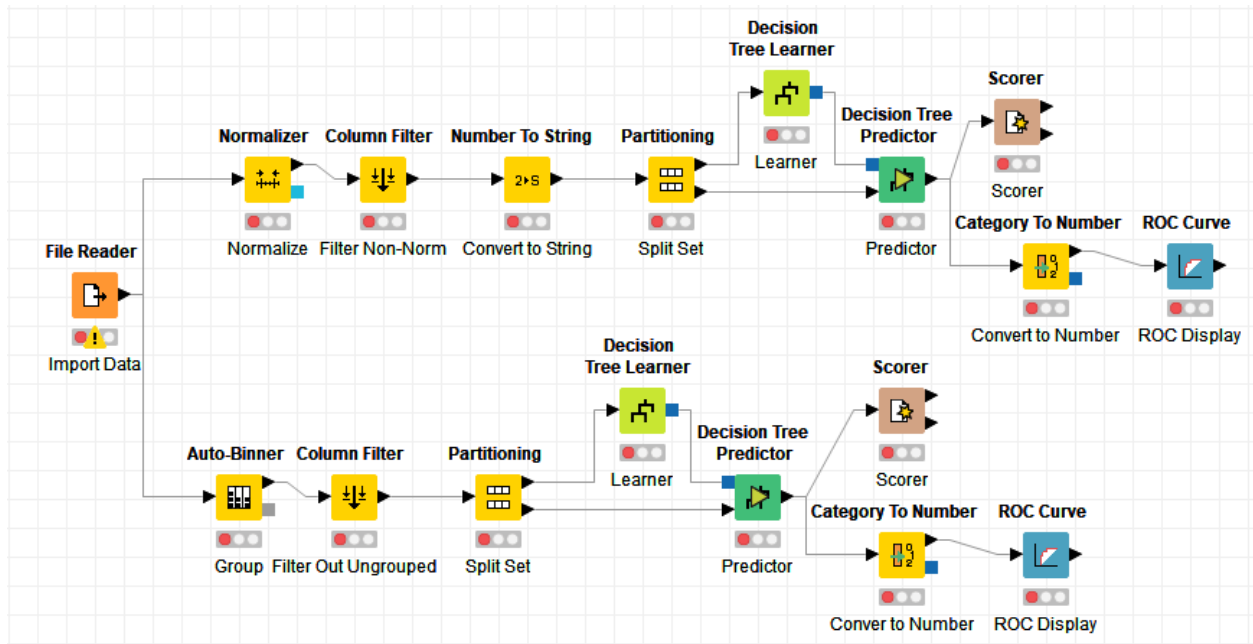


Figure 5: KNIME Node Structure for Analysis