**Data Preparation**

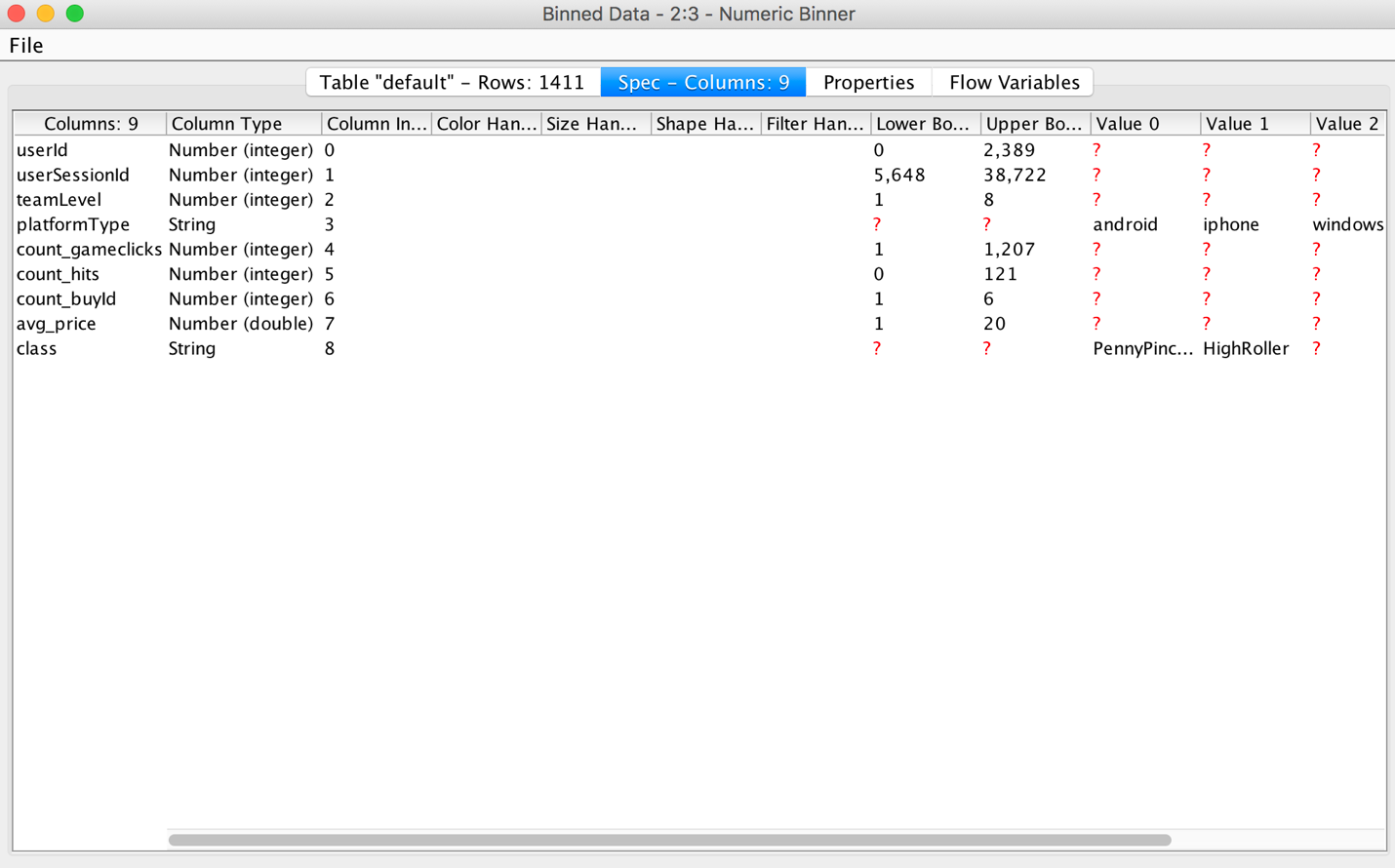
Analysis of combined\_data.csv

Sample Selection

|  |  |
| --- | --- |
| **Item** | **Amount** |
| # of Samples | 4619 |
| # of Samples with Purchases | 1411 |

Attribute Creation

A new categorical attribute was created to enable analysis of players as broken into 2 categories (HighRollers and PennyPinchers). A screenshot of the attribute follows:



The column class is the new categorical attribute. The value is either PennyPincher or HighRoller. Its value depends on avg\_price.

The creation of this new categorical attribute was necessary because the decision tree classifier needs a class label for each data record.

Attribute Selection

The following attributes were filtered from the dataset for the following reasons:

|  |  |
| --- | --- |
| **Attribute** | **Rationale for Filtering** |
| teamLevel | Teams with more people spending money may tend to achieve higher level. |
| platformType | The probability of high roller could be different for users using different devices. |
| count\_buyid | This is purchase-related data. |
| class | This needs to be used as label by decision tree. |

**Data Partitioning and Modeling**

The data was partitioned into train and test datasets.

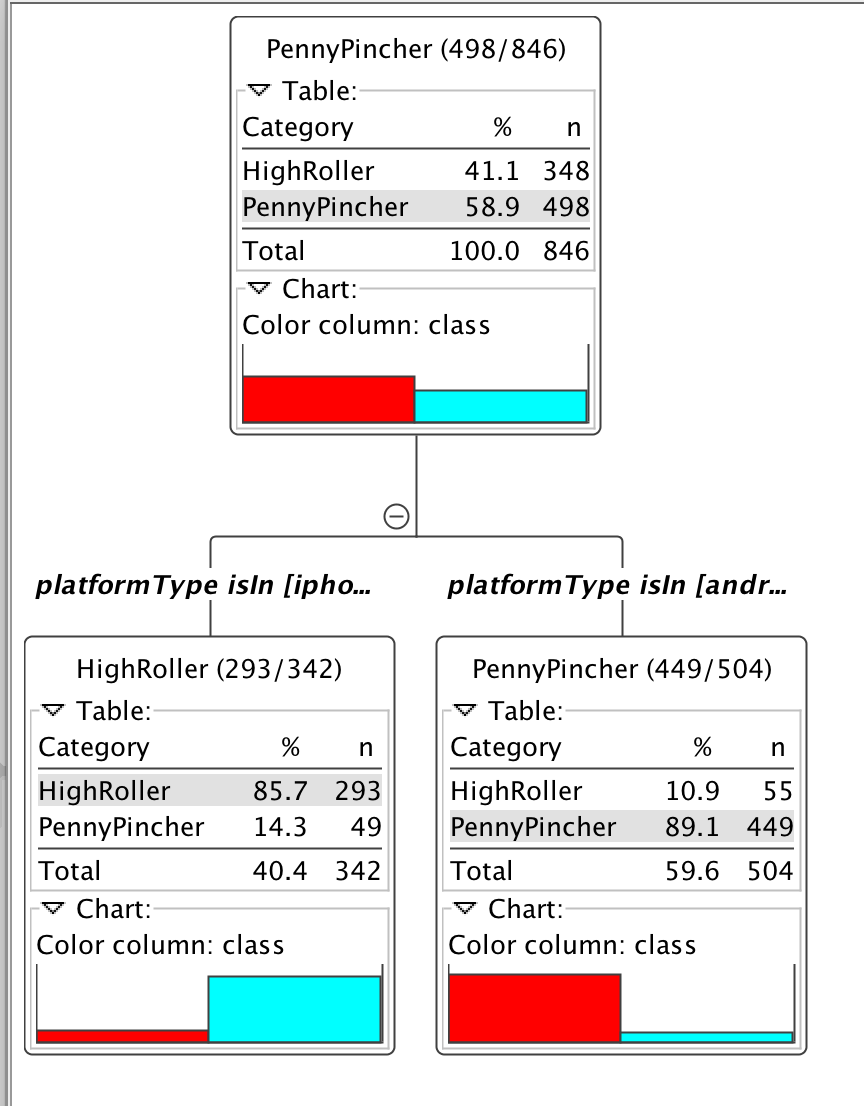
The training data set was used to create the decision tree model.

The trained model was then applied to the testing dataset.

This is important because we need both training and testing dataset. And they normally

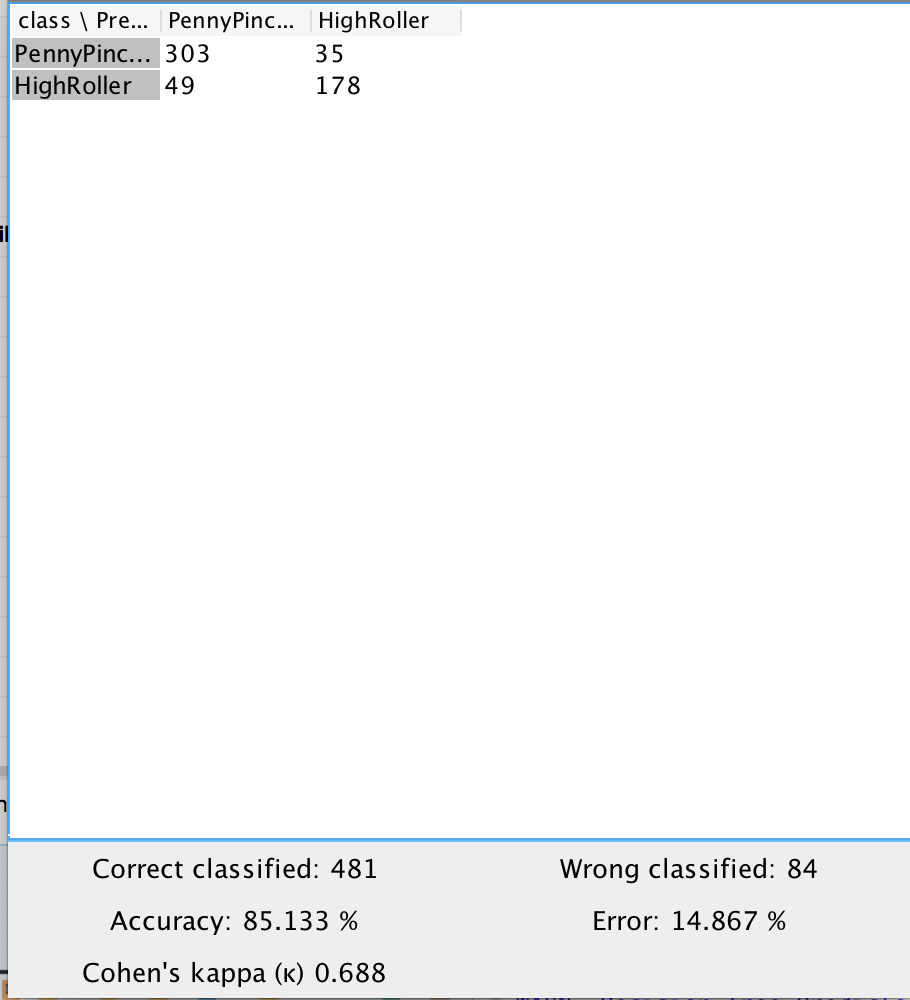
When partitioning the data using sampling, it is important to set the random seed because this will let us get the same partitioning result every time, and hence get the reproducible result.

A screenshot of the resulting decision tree can be seen below:



**Evaluation**

A screenshot of the confusion matrix can be seen below:



As seen in the screenshot above, the overall accuracy of the model is 85.133%

The value 303 at top-left cell means that 303 PennyPickers were correctly classified.

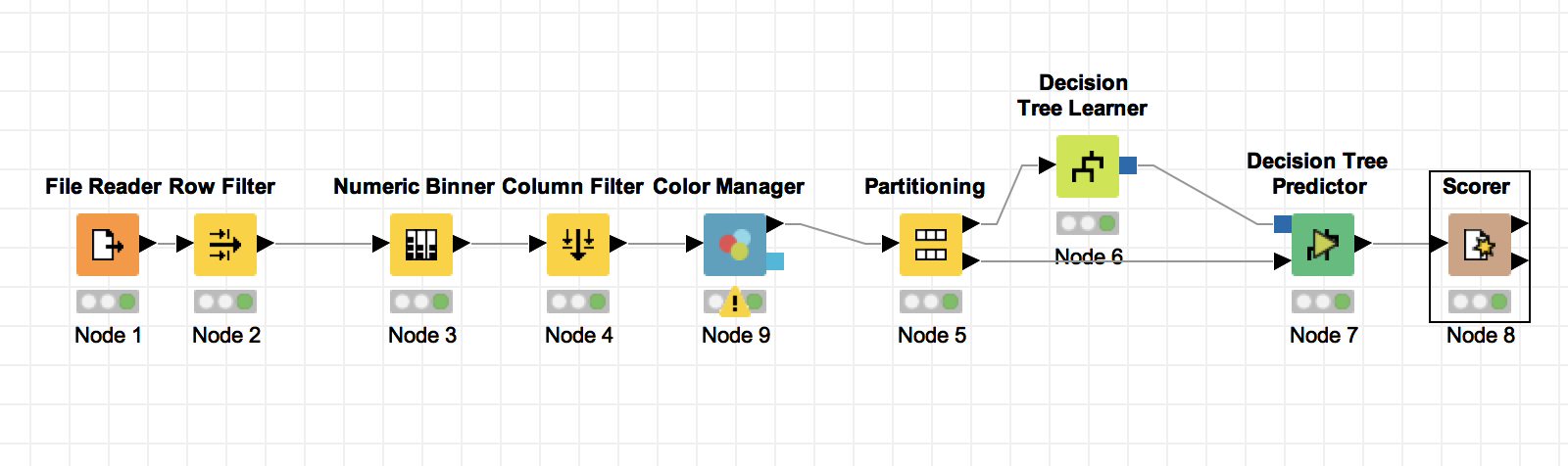
The value 35 at top-right cell means that 35 PennyPickers were incorrectly classified as HighRoller.

The value 49 at bottom-left cell means that 49 HighRollers were incorrectly classified as PennyPickers.

The value 178 at bottom-right cell means that 178 HighRollers were correctly classified.

**Analysis Conclusions**

The final KNIME workflow is shown below:



What makes a HighRoller vs. a PennyPincher?

Based on the structure of the decision tree, the users using iPhone are more likely to be HighRoller, while the users on the other platforms tend to be PennyPincher.

|  |
| --- |
| **Specific Recommendations to Increase Revenue** |
| 1. Compare the differences between the iOS app and the interface on the other platforms, and try to find the possible reasons (i.e. there might be some features only available on iPhone) why iOS users are more likely to spend more. |
| 2. After finding the possible reasons mentioned above, launch A/B testing for some new features accordingly on all the platforms except for iOS, and analyse which features could help to increase the revenue. |