**Data Preparation**

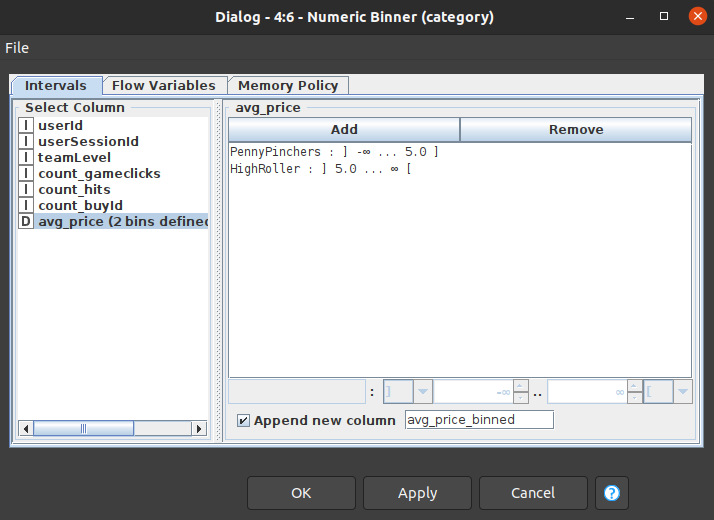
Analysis of combined\_data.csv

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:



**This attribute will categorize spenders at or below 5.0 as PennyPinchers and others as HighRoller.**

The creation of this new categorical attribute was necessary because now **we can use a Decision Tree classification model to predict a users spending type given new data**.

Attribute Selection

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

|  |  |
| --- | --- |
| **Attribute** | **Rationale for Filtering** |
| <avg\_price> | **We have already binned this into category for training and will not have this information available in the test data during prediction.** |
|  |  |

**Data Partitioning and Modeling**

The data was partitioned into train and test datasets.

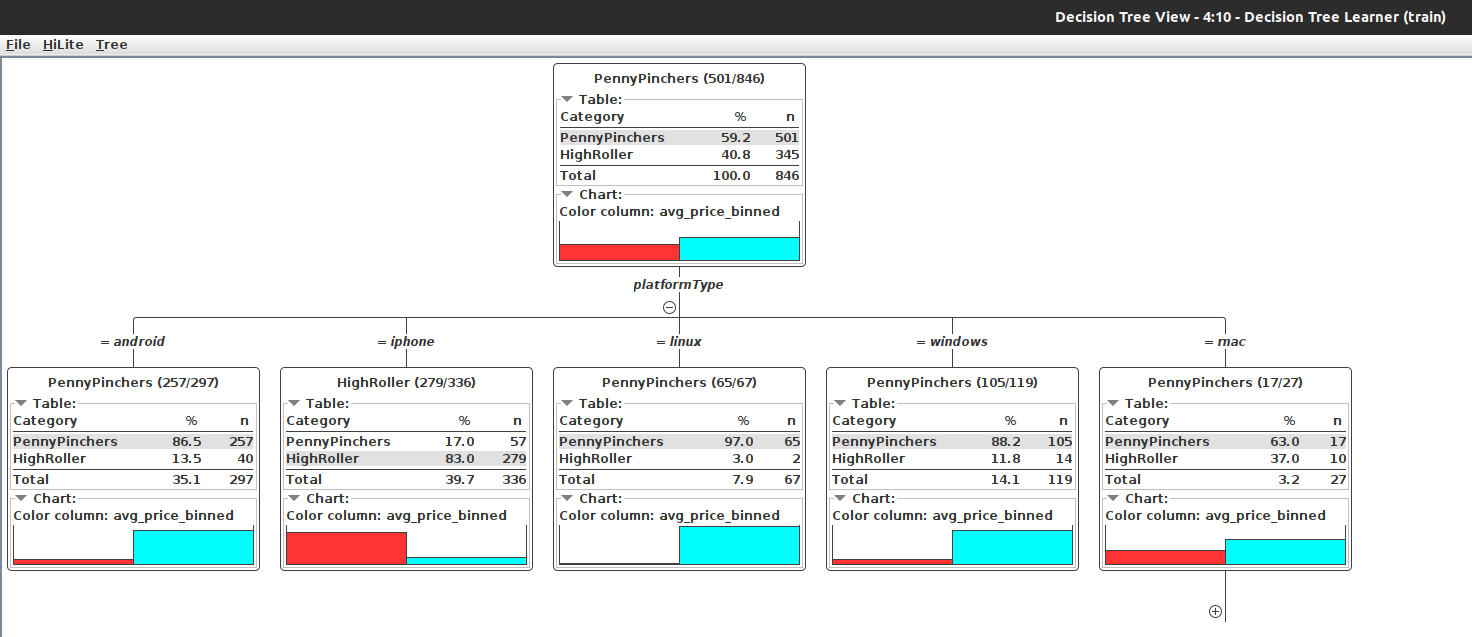
The **train (60%)** data set was used to create the decision tree model.

The trained model was then applied to the **test (40%)**  dataset.

This is important because… **it allows us to evaluate whether the learned model is overfitting, underfitting or generalized.**

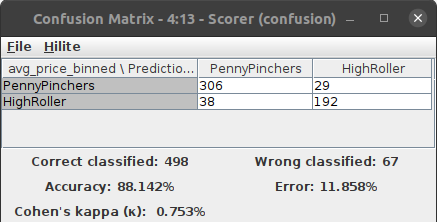
When partitioning the data using sampling, it is important to set the random seed because… **it will help evaluate to similar sampling as done by the evaluator.**

A screenshot of the resulting decision tree can be seen below



**Evaluation**

A screenshot of the confusion matrix can be seen below:



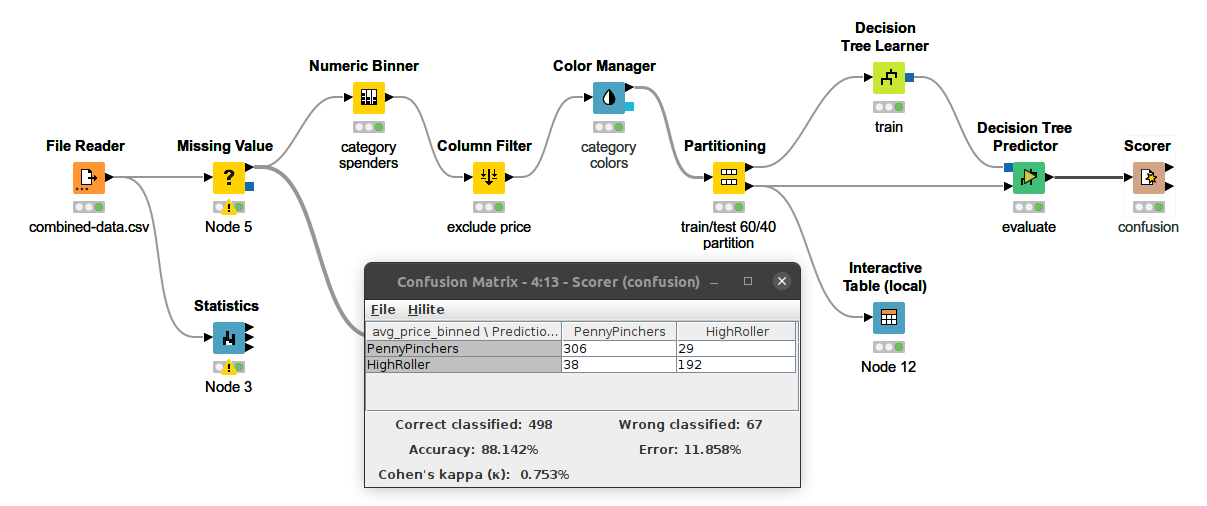
As seen in the screenshot above, the overall accuracy of the model is <**88.142%**>

The **diagonal matrix values are correctly predicted** which means 306 PennyPinchers and 192 HighRoller are accurate.

* **There are 29 PennyPinchers in the original dataset which have been incorrectly predicted as HighRoller**
* **There are 38 HighRoller misclassified as PennyPinchers**

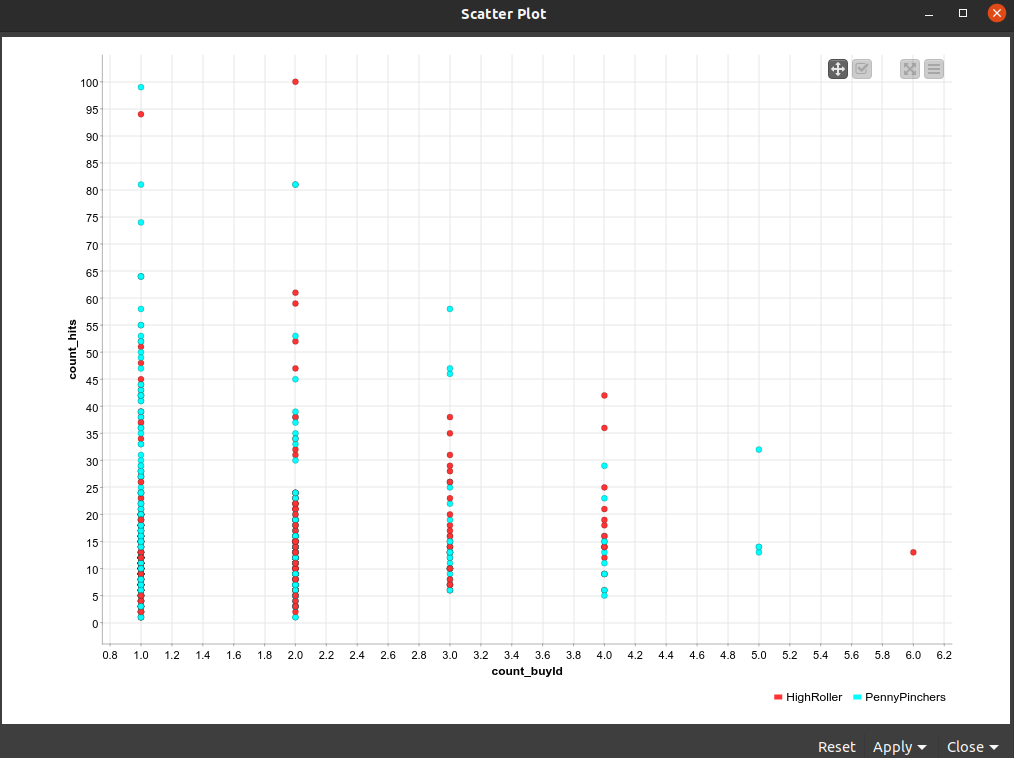
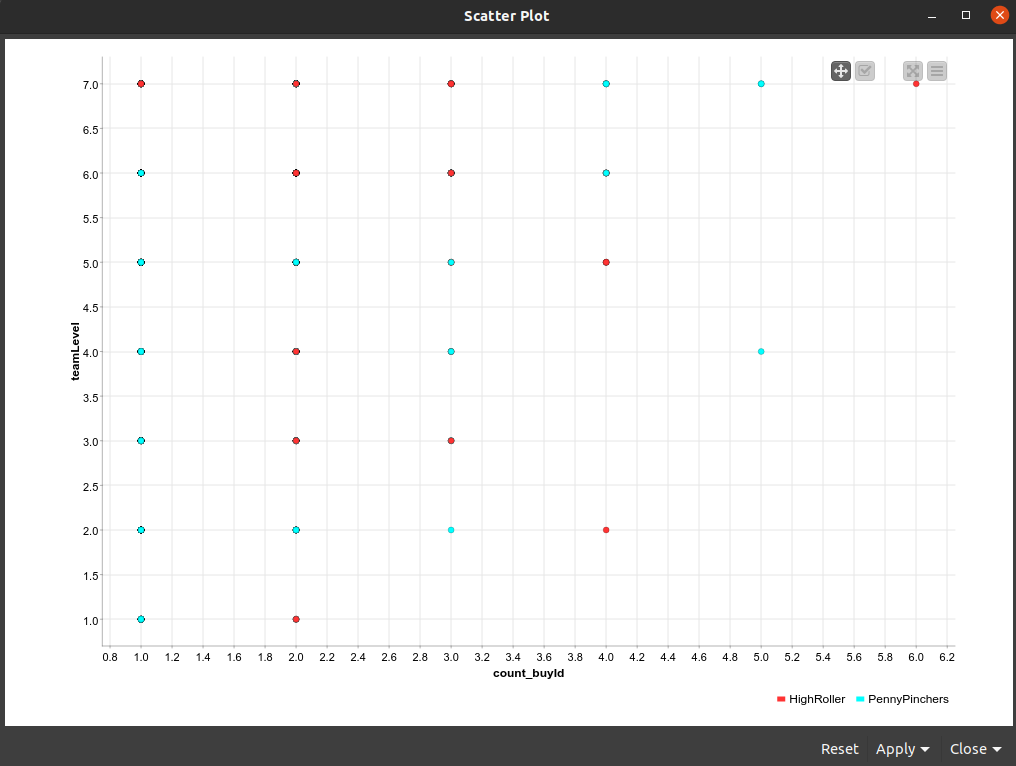
**Analysis Conclusions**

The final KNIME workflow is shown below:



What makes a HighRoller vs. a PennyPincher?

**A HighRoller is the player spending more than 5 in avg\_price and others are PennyPinchers**

**Team Level (left fig) and count\_hits (right fig) vs the product count\_buyId need to be understood here. Looks like until teamLevel 7 is achieved product count\_buyId 1 is not purchased at all.**

**Predictions:**

* The top two game\_clickers are also HighRollers and the classifier makes some mistakes about the them too. It may be due to the fact that classifier may have learnt that clicks may have a positive influence on spending but this seems to be an outlier.
* 8 out of top ten count\_hits are PennyPinchers and the classifier has misclassified 3 of them

|  |
| --- |
| **Specific Recommendations to Increase Revenue** |
| 1. Looks like iPhone users are the target audience and will need more marketing to |
| 2. The **count\_buyId=6** only has HighRoller and looks like we may need to move more users towards that category. There are only PennyPinchers on **count\_buyId=5**. We will need to see in detail what is going on there and learn from 3 or 2 which are doing relatively better. |