**Data Preparation**

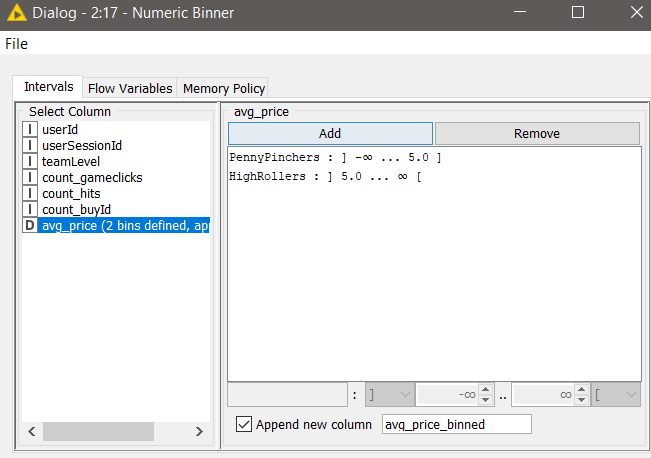
Analysis of combined\_data.csv

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:



*HighRollers (buyers of items that cost more than $5.00) and PennyPinchers (buyers of items that cost $5.00 or less)*

The creation of this new categorical attribute was necessary because it is the target of our prediction.

Attribute Selection

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

|  |  |
| --- | --- |
| **Attribute** | **Rationale for Filtering** |
| Id columns… | They don’t add predictive value, they are a random number |
| avg\_price | Because it would be cheating, it is what derives our target |
|  |  |
|  |  |

**Data Partitioning and Modeling**

The data was partitioned into train and test datasets.

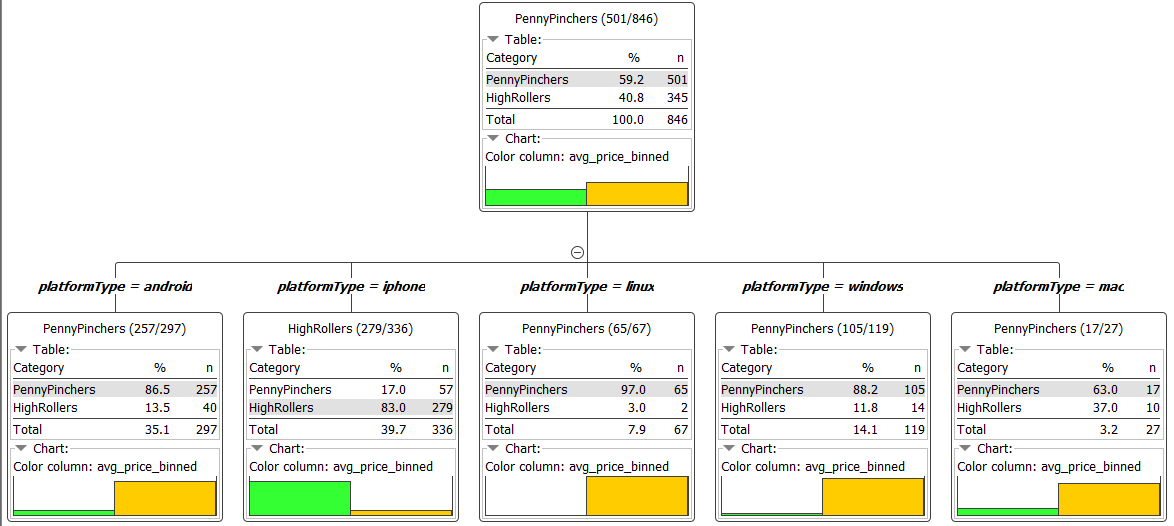
The train data set was used to create the decision tree model.

The trained model was then applied to the test dataset.

This is important because we need results to generalize and not be over fitted.

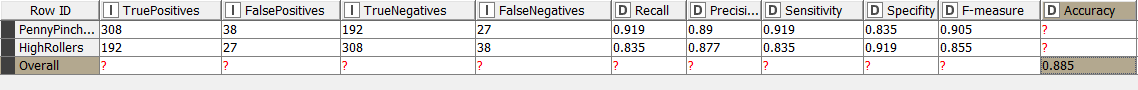
When partitioning the data using sampling, it is important to set the random seed because… this way results are consistent relative to iterating with tweaks

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



**Evaluation**

A screenshot of the confusion matrix can be seen below:



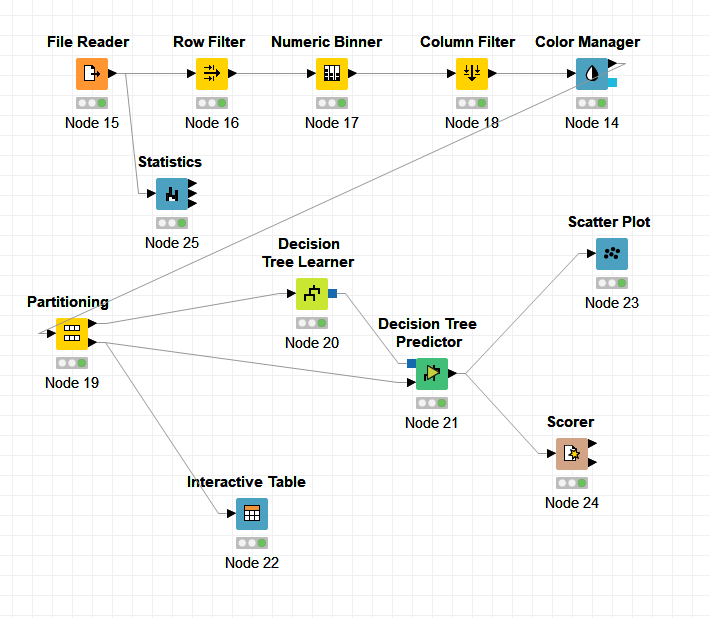


As seen in the screenshot above, the overall accuracy of the model is 88%

TruePositives and TrueNegatives have correctly been predicted as seen in above screenshots.

**Analysis Conclusions**

The final KNIME workflow is shown below:



What makes a HighRoller vs. a PennyPincher?

**People who use iPhone are HighRollers**

|  |
| --- |
| **Specific Recommendations to Increase Revenue** |
| 1. Take marketing and technology measures that will cater for iPhone customers, to increase that user base:  - adds in platform’s app store / facebook / google, if Platform == ‘iPhone’  - prioritize development of features that are more catered towards iPhone OS / Chipset.  - an important decision could be made to force phone App only support, this way all users would either be iPhone or Android, increasing the relative % of iPhone. |
| 2. Run targeted promotions segmented by platform: - if not iPhone then have a promo pop-up with a cheaper price that will increase the buys  - add more expensive features that are targeted to iPhone whale customers |