The purpose of the exercise is to find which game players made in game purchases and then further classify the buyers as either “HighRollers” or “PennyPinchers.” Highrollers are defined as users spending more than $5.00 per purchase and PennyPinchers are defined as users spending less than $5.00 per purchase. The main drawback of this exercise was that the data set was less than 5000 lines so the analysis could have easily been done in Excel with a pivot table.

**Data Preparation**

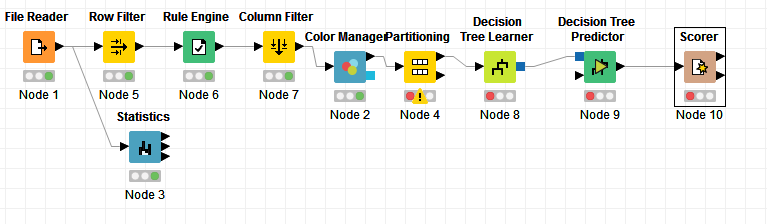
Analysis of combined\_data.csv

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:



Note: A **Rule Engine** node is not the only node that could be inserted after the **Row Filter** node. A **Numerical Binner** node could also have been used.

Locate every user who made a purchase. Based on the amount spent, assign users who made purchases a title of either “PennyPinchers” or “HighRollers.”

The creation of this new categorical attribute was necessary because the purpose of this exercise is to conduct analysis attributed to PennyPinchers and HighRollers.

Attribute Selection

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

|  |  |
| --- | --- |
| **Attribute** | **Rationale for Filtering** |
| userSessionId | Decided not relevant to analysis |

**Data Partitioning and Modeling**

The data was partitioned into train and test datasets.

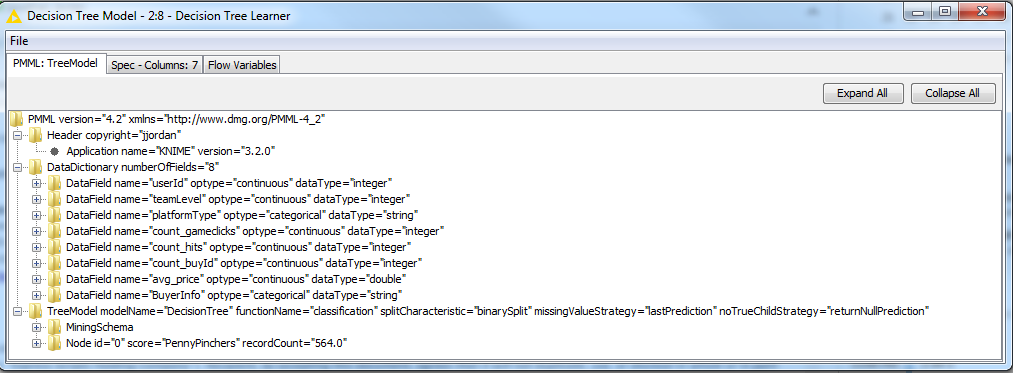
The Buyer Info data set was used to create the decision tree model.

The trained model was then applied to the Buyer Info dataset.

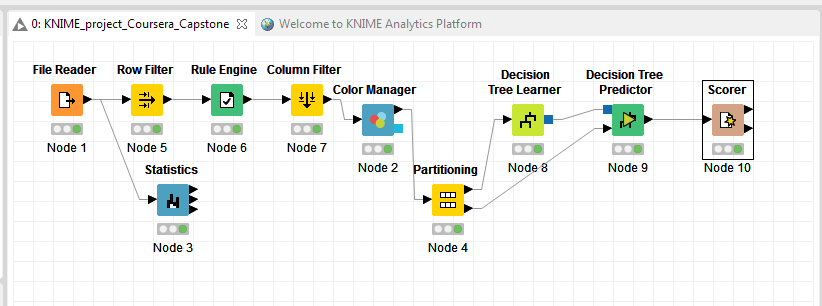
This is important because the Buyer Info is the dataset where we are conducting the analysis.

When partitioning the data using sampling, it is important to set the random seed because you want to select random lines to reduce the chance the rows are related.

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

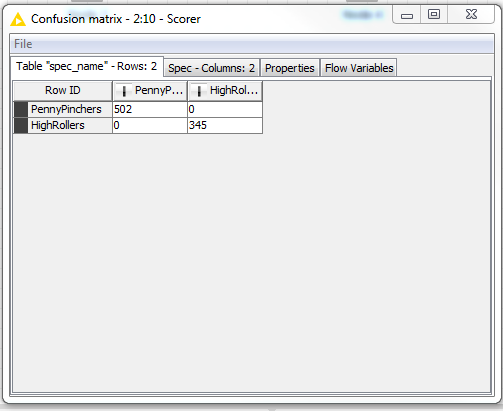


The final KNIME model looks like:



**Evaluation**

A screenshot of the confusion matrix can be seen below:



As seen in the screenshot above, the overall accuracy of the model is 100%

The model identified 502 PennyPinchers.

The model identified 345 HighRollers.

**Analysis Conclusions**

What makes a HighRoller vs. a PennyPincher? A HighRoller spends more than $5 per purchase averaging $10.25 per purchase. A PennyPincher spends less than $5 per purchase averaging $1.97 per purchase.

|  |
| --- |
| **Specific Recommendations to Increase Revenue** |
| 1. iPhone users spend the most followed by Mac users. Increase marketing to Apple users. Develop additional and more expensive in game purchases that can be made by Apple users. |
| 2. Linux and windows users spend the least amount of money. Cease marketing on Linux and Windows platforms. |