**Data Preparation**

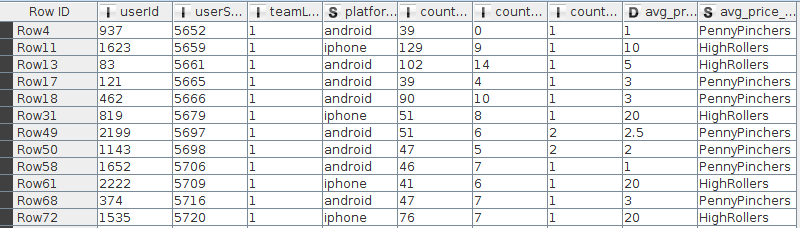
Analysis of combined\_data.csv

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 new attribute avg\_price\_binned uses the avg\_price attribute to classify the instances. When the value of avg\_prive is less than 5 it classify the instance as “PennyPinchers”. And when it is grater than 5 as “HighRollers”

The creation of this new categorical attribute was necessary because it will be the target attribute that we are going to use in the next steps to train the decision tree.

Attribute Selection

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

|  |  |
| --- | --- |
| **Attribute** | **Rationale for Filtering** |
| usserSesionId | Its excluded because its not a significance value, its only the Id to identify the session |
| avg\_price | Its excluded because the target attribute its created from this attribute |

**Data Partitioning and Modeling**

The data was partitioned into train and test datasets.

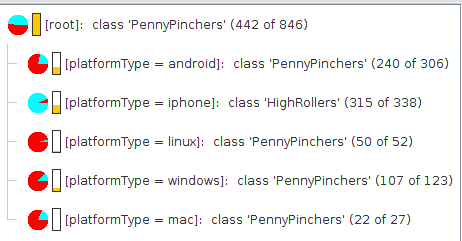
The first partition data set was used to create the decision tree model.

The trained model was then applied to the second partition dataset.

This is important because we need to test our model in data set different from the training data set to see how is it behaving in different data with the same distribution.

When partitioning the data using sampling, it is important to set the random seed because we want to be able to reproduce the same results in each execution.

A screen-shot of the resulting decision tree can be seen below:

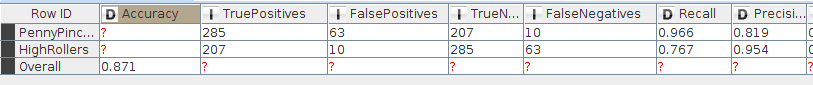


**Evaluation**

A screenshot of the confusion matrix can be seen below:



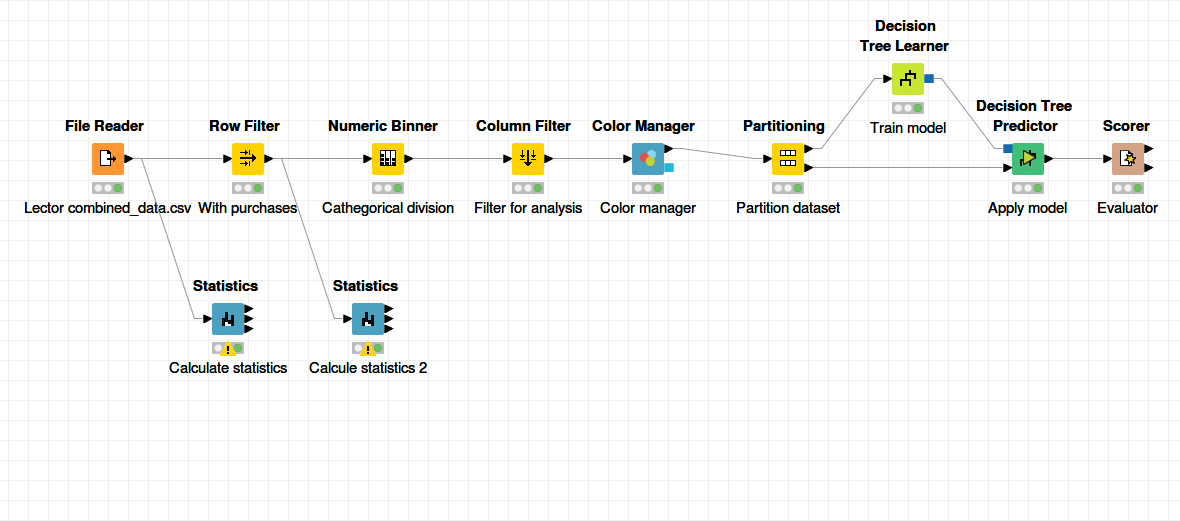
As seen in the screenshot above, the overall accuracy of the model is 0.871



* 285 “PennyPinchers” was correctly predicted
* 207 “HighRollers” was correctly predicted.
* 63 instances was predicted as “PennyPinchers” and they are “HighRollers”
* 10 instances was predicted as “HighRollers” and they are “PennyPinchers”

**Analysis Conclusions**

The final KNIME workflow is shown below:



What makes a HighRoller vs. a PennyPincher?

Based in the decision results the principal attribute is the platform type. The iphone user trend to be high rollers and the users of the oder platmorms trend to be Penny Pinchers.

|  |
| --- |
| **Specific Recommendations to Increase Revenue** |
| 1. We need to focus our efforts to increase iphone users. We could make publicity of our game oriented in this platform. |
| 2. We want to the no iphone platforms users spent more money in our game. We could personalize some characteristics of our game, like an andoid flamingo to increase the interests of this platform users. |