

Classifying in KNIME to identify big spenders

Predicting which user is likely to purchase big-ticket items while playing Catch the Pink Flamingo is valuable knowledge to have for Egence since in-app purchases are a major source of revenue. In this assignment, you will analyze available data to classify users as buyers of big-ticket items (“HighRollers”) vs. buyers of inexpensive items (“PennyPinchers”). Big-ticket items are those with a price of more than \$5.00, and inexpensive items are those that cost \$5.00 or less.

To predict the user category, Decision tree classification will be used.

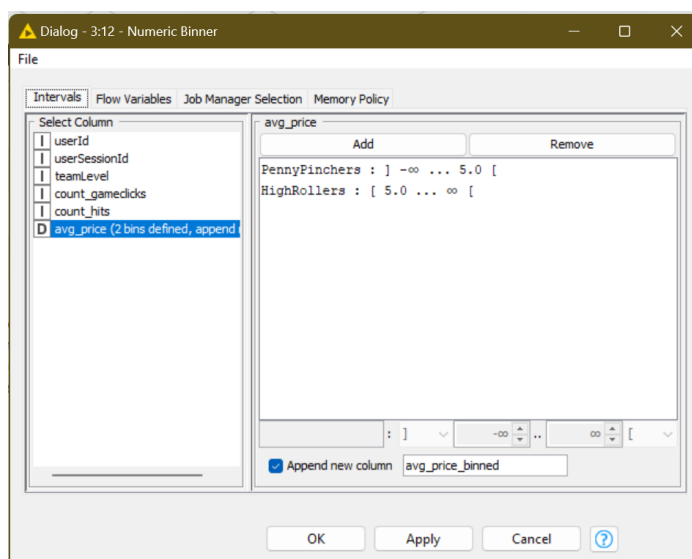
Data Preparation

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 users who have purchased more than \$5 were named as “HighRollers” and the users who have purchased less than \$5 were named as “PennyPinchers”.

The creation of this new categorical attribute was necessary because the objective of the task is to, identify who makes the large purchases.

Attribute Selection

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

Attribute	Rationale for Filtering
userId	Not relevant to the model
usersessionId	Not relevant to the model
avg_price	Categorized to new column

Data Partitioning and Modelling

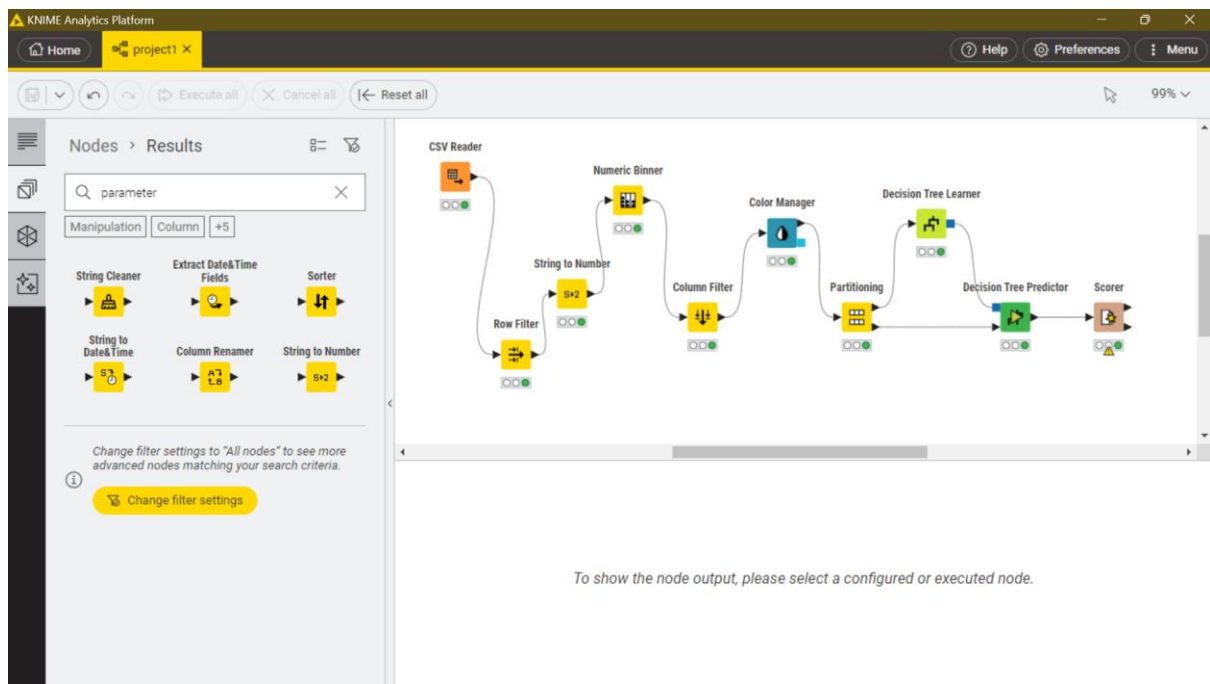
The data set was partitioned into train and test datasets.

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

The trained model was then applied to the test dataset.

This is important because partitioning the data set into train and test data is needed to verify the accuracy of the model.

When partitioning the data using sampling, it is important to set the random seed because it allows us to have reproducible results each time the model is run.



To show the node output, please select a configured or executed node.

The resulting decision tree:-



Evaluation

The confusion matrix:-

avg_price_...	PennyPinc...	HighRollers
PennyPinchers	238	55
HighRollers	67	201

Correct classified: 439	Wrong classified: 122
Accuracy: 78.253%	Error: 21.747%
Cohen's kappa (κ): 0.563%	

As seen in the screenshot above, the overall accuracy of the model is 78.253%

When the buyer type is PennyPincher, the model has classified it correctly 238 times and incorrectly 55 times.

When the buyer type is HighRoller, the model has classified it correctly 201 times, and incorrectly 67 times.

Analysis Conclusions

Users OS makes a HighRoller or a PennyPincher. HighRollers use iOS while PennyPinchers use Mac, Android, Windows, and Linux.

Specific Recommendations to Increase Revenue
1. Focus on improving user experience, performance, and compatibility for these operating systems.
2. Run targeted marketing campaigns to attract more users.
3. Collect feedback from users across all platforms to identify any issues or areas for improvement.