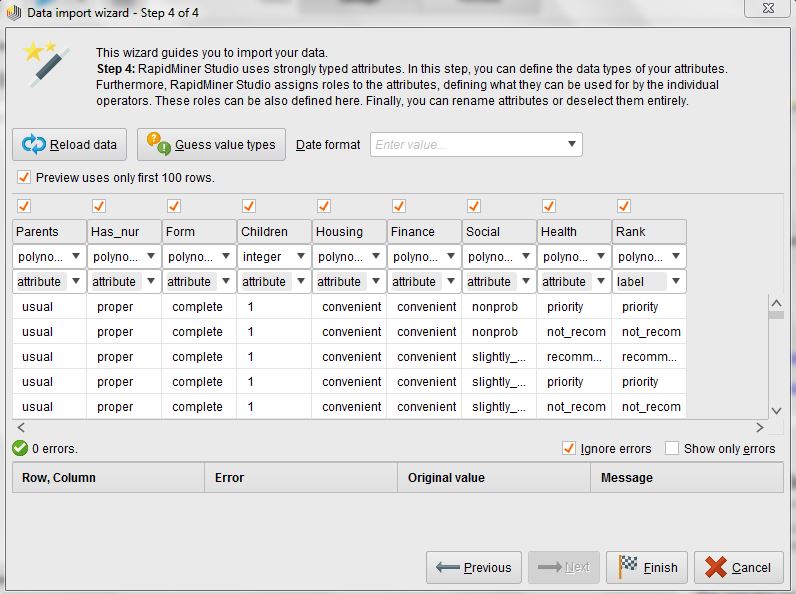
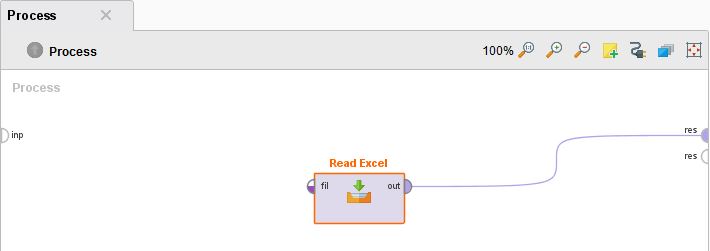
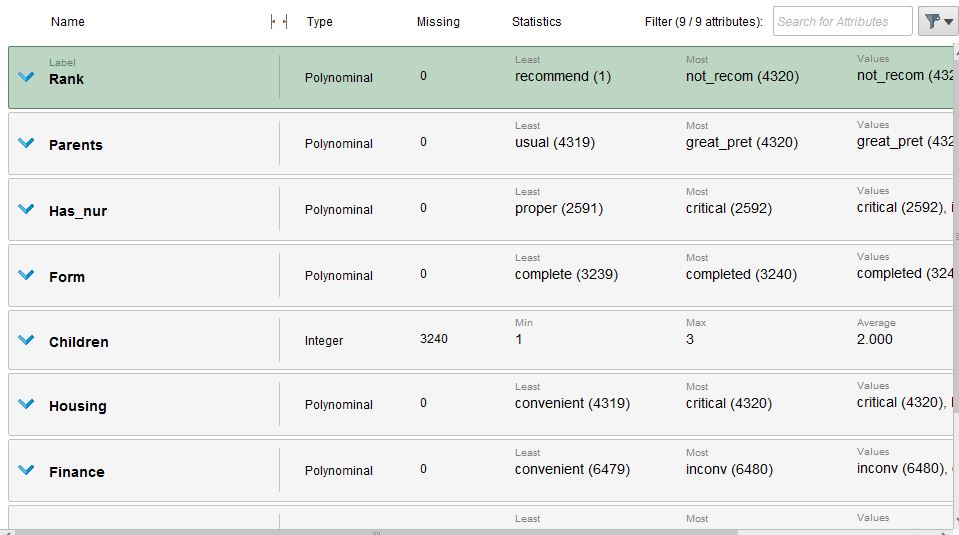
Course Project #2 RapidMiner Nursery by Joshua Troup



Load data screen with data import wizard

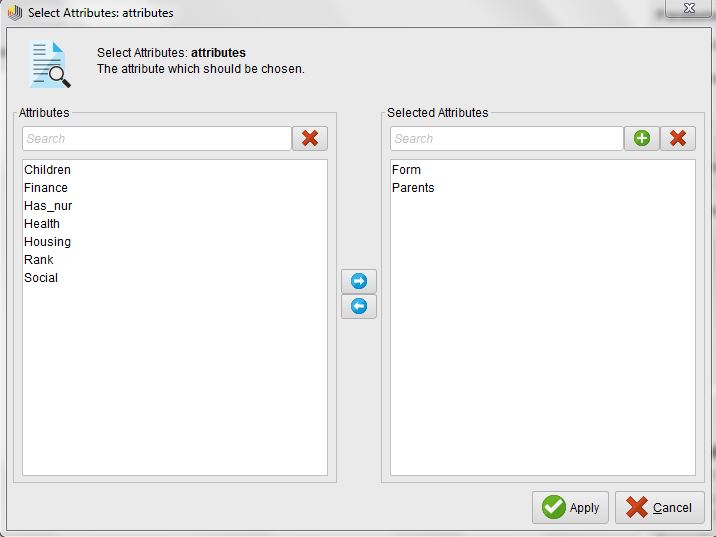


Read Excel output connected to results

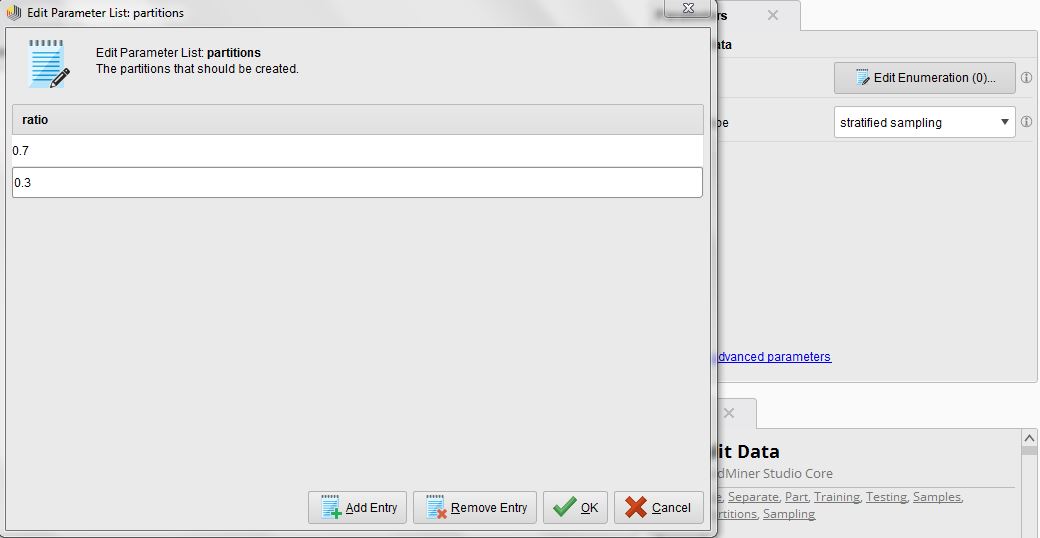


5. Question: Do you see missing data for any attribute. Comment on the missing data.

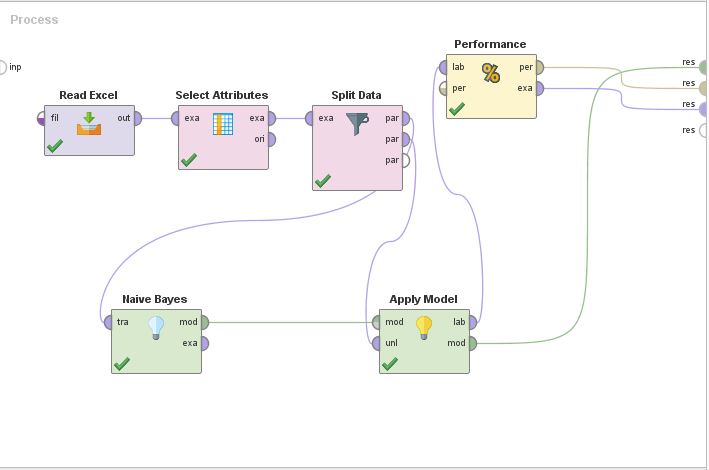
**Yes Children is missing 3,240 records**



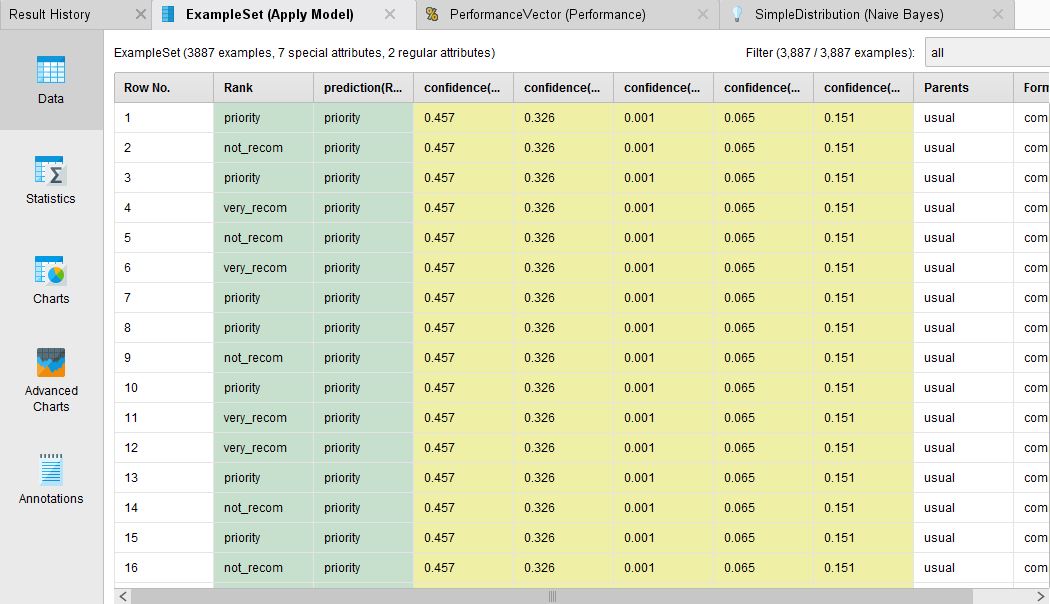
Preprocessing data—Attributes Form and Parents selected



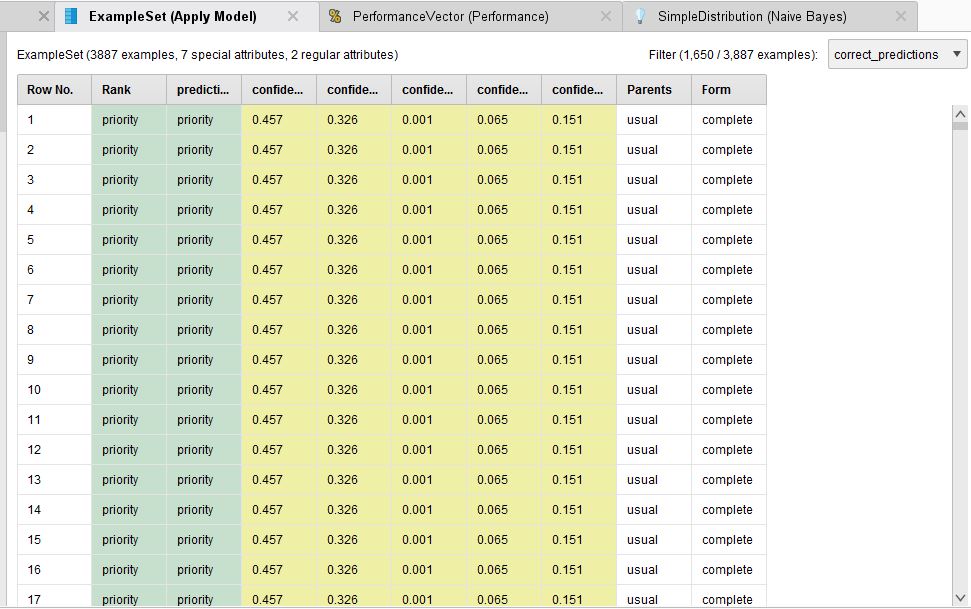
Classification model-Split data operator---70% training/30% testing



Modeling steps---Training---Testing---Performance Classification

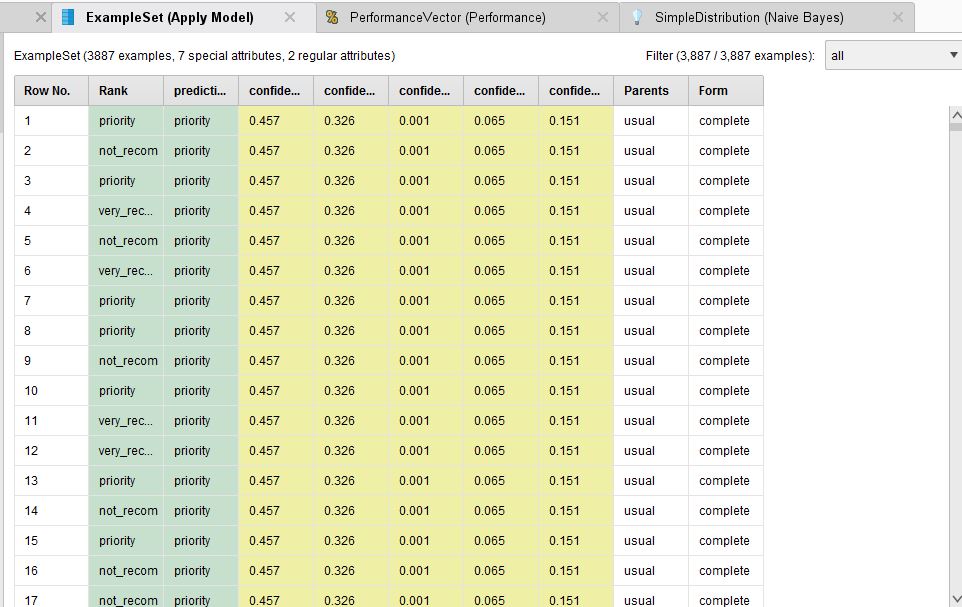


Results of process



Q1. How many correct predictions were made of the 3,887 training data?

**1,650 correct predictions**

****

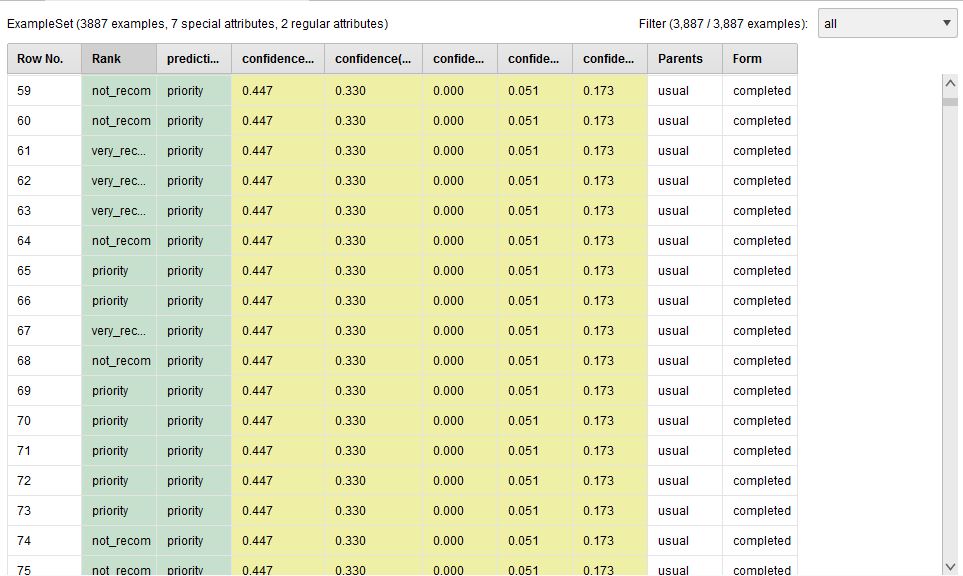
What is the Overall accuracy (%)?

**42.4% overall accuracy (1650/3887=.424491**

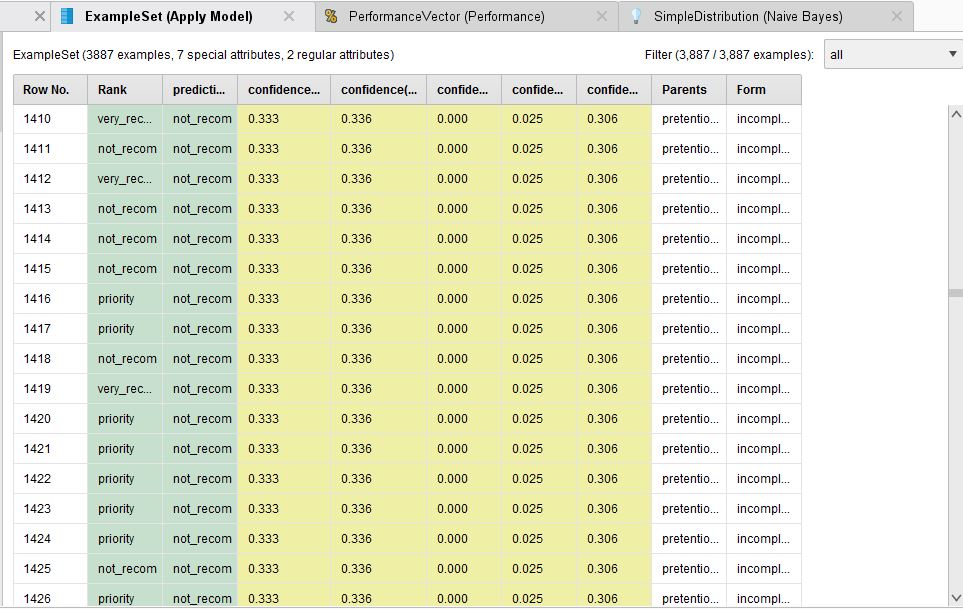
****

Q2: For each classification label, select the row and interpret your prediction model.

**Row 1 indicates Rank and Prediction are both priority. The highest confidence is 0.457 for label value = priority. This row is predicted as priority. The same label prediction is projected to be priority for the 17 examples shown above with a confidence level of 0.457.**



**Row 59 indicates Rank is not recommended and Prediction is priority. The highest confidence is 0.447 for label value = priority. The second highest confidence level is 0.330 for label value = not recommended. This row is predicted as priority.**



**Row 1410 indicates Rank is very recommended and Prediction is not recommended. The highest confidence is 0.336 for label value = not recommended. The second highest confidence level is 0.333 for label value = priority. This row is predicted as not recommended.**