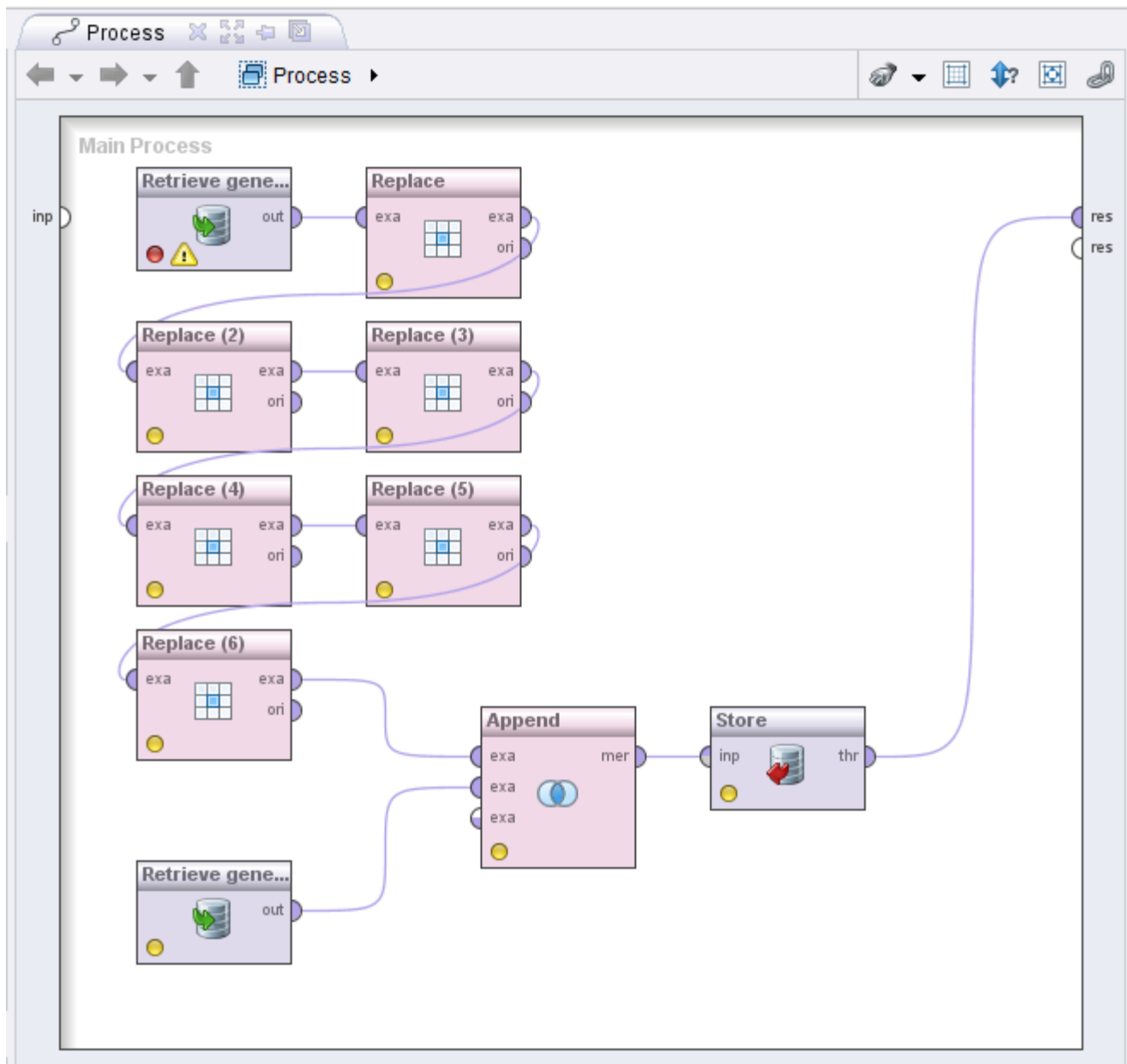


HW2 RapidMiner Extra Credit

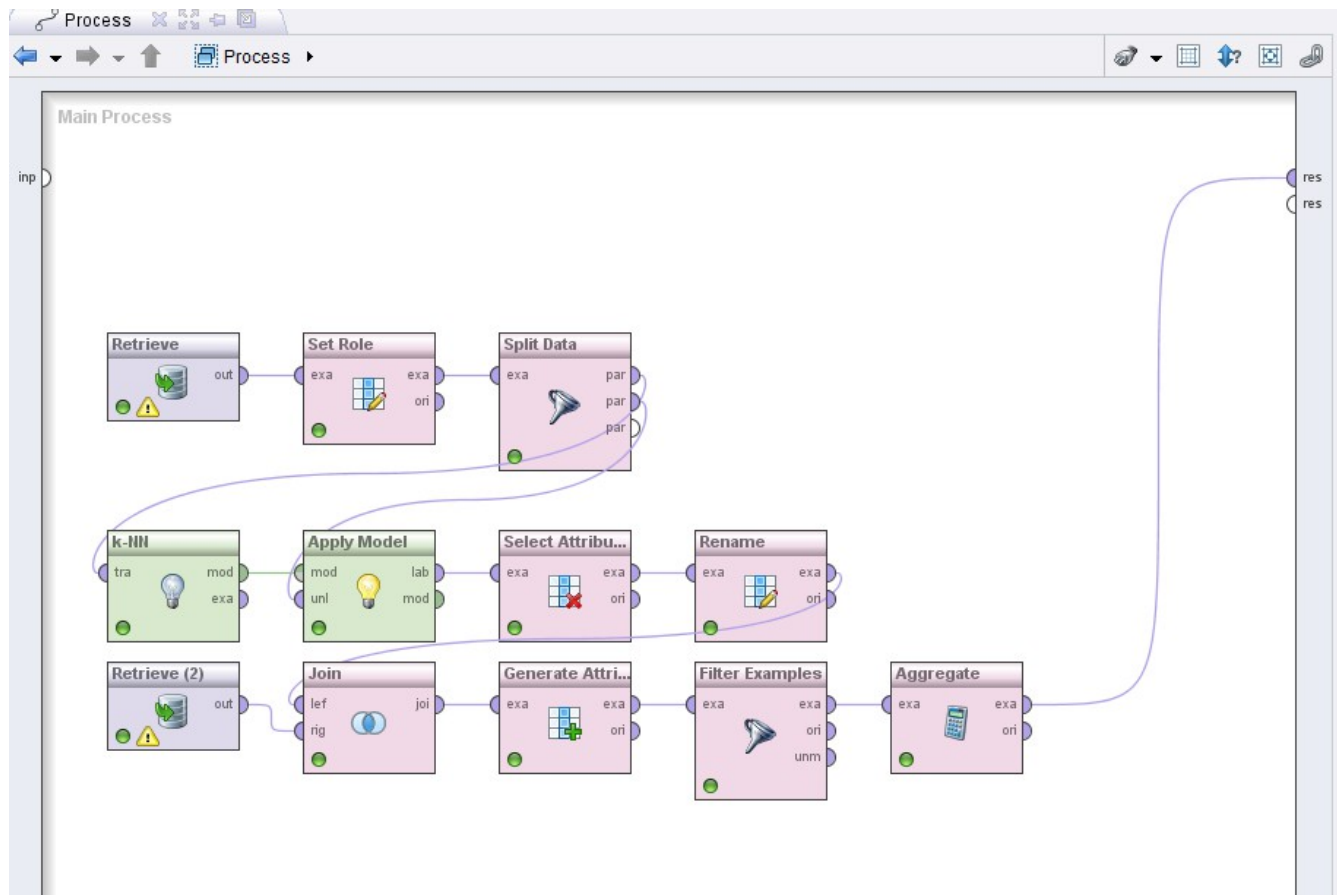
1) Data Cleaning:

I replaced missing elements of the training set with their mode. Though I had found the modes in another process which I seem to have misplaced or failed to save. Then I appended the test set onto the end so that I could later split and feed the set into KNN.



2) Classification

I split the training and test sets before feeding the training set to the knn process and the test set to the apply model process. I then trimmed some new attributes and renamed the attribute that contained the classification so that I could use it in the generate attributes process. I then joined this set with the key values. From there I checked if the predicted localization matched the key set localization and then counted the number that did.



3) Results

Row No.	count(Correct_prediction)
1	908

$$908 / 1929 = .4707$$

This is approximately 10% better than my solution.