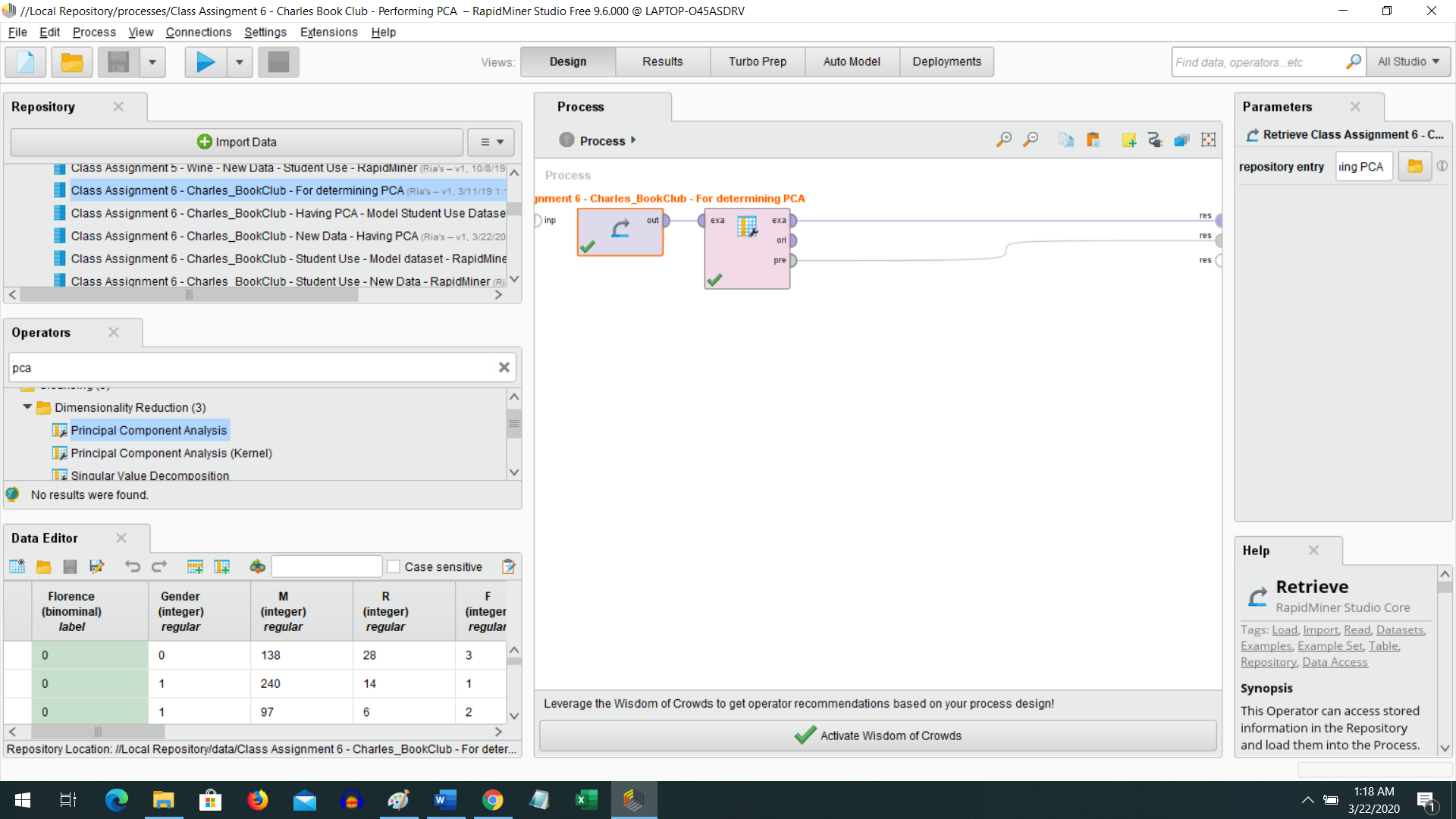
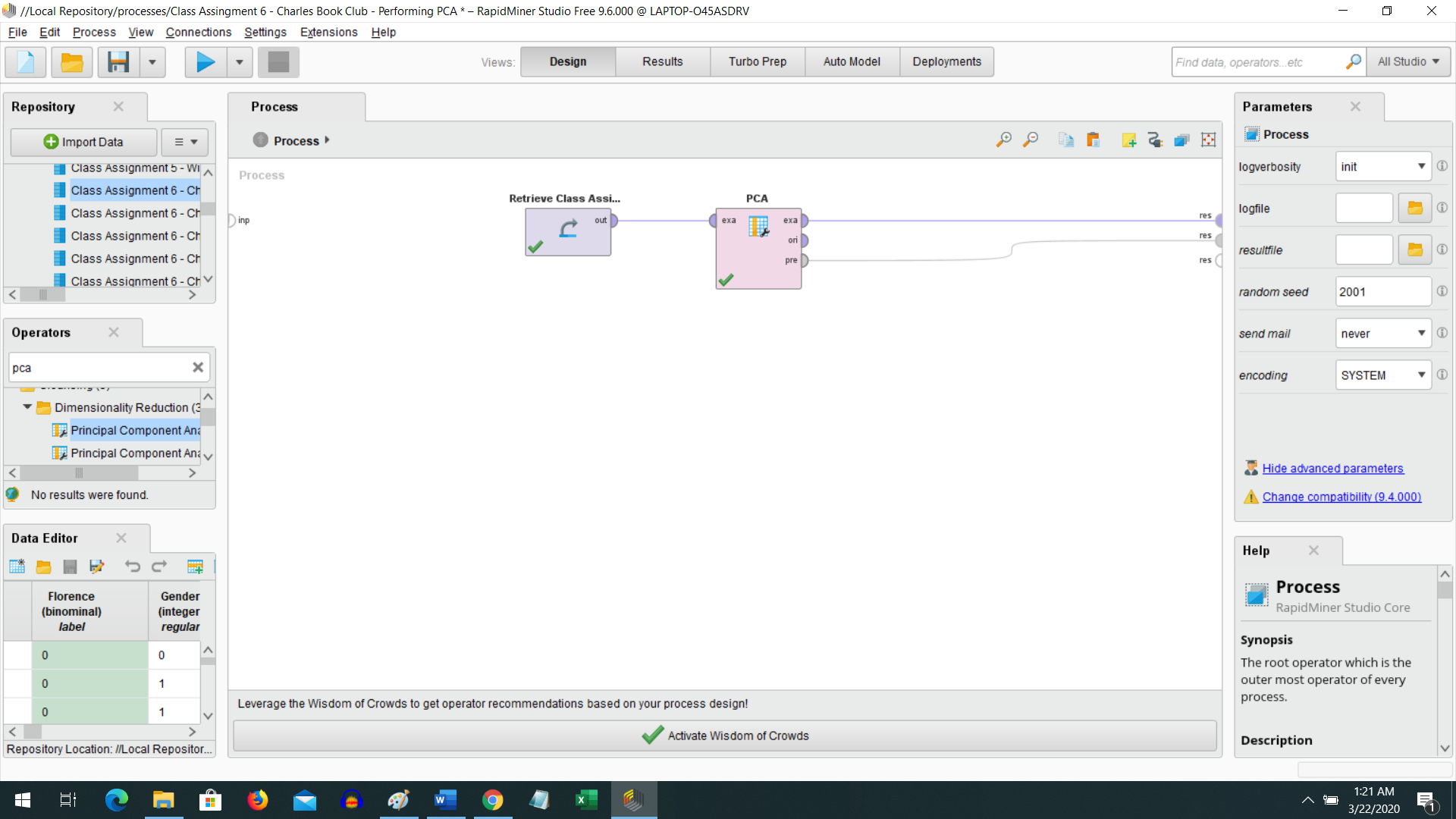
Performing PCA in RapidMiner

1. First of all make sure the input predictors are all in ‘integer’ form otherwise PCA will not work.

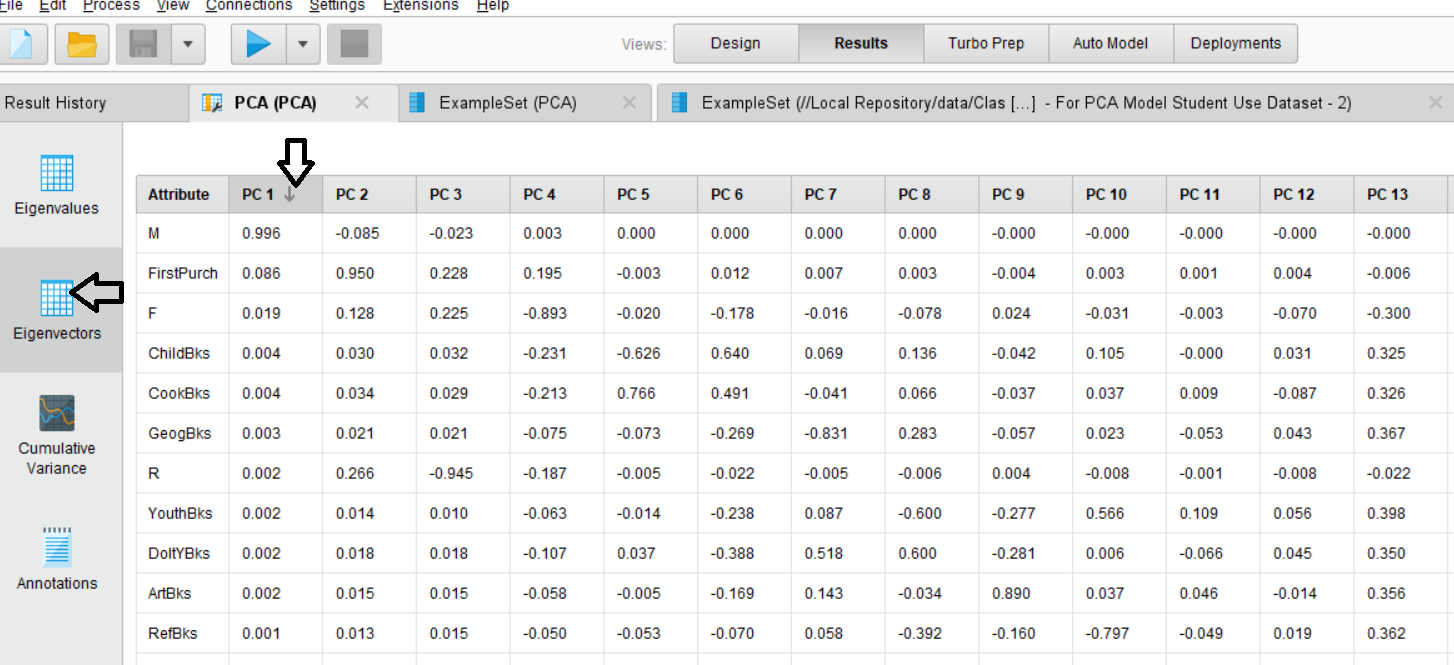
2. Incase there is any input predictor in binomial form then edit the dataset and change it into integer form. If you have already imported it into RapidMiner then right click on it and click ‘Edit’ this opens the Dataset editor and you can edit the Attribute type there.



3. Then create a process to perform PCA. Connect the dataset to PCA operator and run the process.



After you run it then under Results check for the Eigen vectors and determine the top 5 or 7 predictors by sorting individually and selecting the absolute highest value (just like you did in XLminer).



So here we can see the first predictor that contributes the most to Principal component 1 is M. This way look for others…. Then you can select the top 5 or 7 etc and then use only those in your input dataset for for your model (create a separate process). Also modify the new data set to include only those top 5 or 7 (exactly same predictors in the model) and run the scoring process.