Zachary Khorozian

Homework 4: KNIME

General Approach

The general approach to the machine learning model workflow is to read in data using the File Reader node, feeding - potentially preprocessed using the normalizer in the case of the Neural network input - data into a learner and a predictor, via a partitioning node that does cross-validation or a vanilla train-test split. The Predictor node is connected to a Scorer node which computes the confusion matrix, precision, recall and accuracy.

KNIME vs RapidMiner

Knime and RapidMiner have similar setups and workflows. The available functions are also broad, relatively up-to-date, and similarly easy to use. A large separator of the two technologies is that KNIME is free for a single user. There is no limit on the amount of data that can be processed in KNIME. Only when you’d want to collaborate on a single project would you need to buy a commercial version of KNIME. In RapidMiner, the free version is limited to 10K rows of data, and increases in this limit come at varying, step-wise increasing costs.