

FILTERING A DATASET

ABSTRACT

A computerized method of filtering a dataset for processing is presented. The method starts with receiving a dataset with a plurality of data records. Afterwards, an estimation module determines selection estimation values for the data records, based on which subsequently pass-through probabilities are determined by a pass-through function. The method further comprises generating a subset of data records by discarding at least a portion of the dataset based on the pass-through probabilities. The subset of data records is then processed and one or more data records are selected. Finally, weights and labels are assigned to the data records of the subset of data records for updating the estimation module and the pass-through function.

Fig. 2