The main techniques for the estimation of the model performance are (Vanschoren 2022): holdout, *K-fold cross validation* (CV), *Stratified K-Fold cross-validation, Leave-One-Out cross-validation, Shuffle-Split cross-validation, Bootstrap, Repeated cross-validation, Time series – for details, ref. file cross.docx.*

* holdout (randomly split data into training and test set, train (fit) a model on the training data and score on the test data);
* *K-fold cross validation* (CV) (split data into equal-sized parts, called “folds”; compute evaluation scores each time using a different fold as the test set; examine the score variance to see how sensitive (unstable) models; large gives better estimates (more training data), but is expensive);
* *Stratified K-Fold cross-validation* applicable for unbalance data sets. Startification conserves proportions between classes in each folder (order examples per class; separate the samples of each class in k sets (strata); combine corresponding strata into folds);
* *Leave-One-Out cross-validation* which is the fold cross-validation with equal to the number of samples; it is completely unbiased (in terms of data splits), but computationally expensive. It is recommended only for small data sets. It generalizes less weel towards unseen data;
* *Shuffle-Split cross-validation* (shuffles the data and samples points randomly for the training data set; it is applicable for large data sets);
* *Bootstrap* (it samples (it is the dataset size) data points, with replacement, as training set (the bootstrap) and uses out-of-bootstrap samples as the test set; this is repeated times to obtain scores’);
* *Repeated cross-validation* (it shuffles data randomly and does *k-fold cross-validation* – repeats this times each time obtaining scores; it is unbiased, very robust, but expensive);
* *Time series* for case when data is ordered and other.