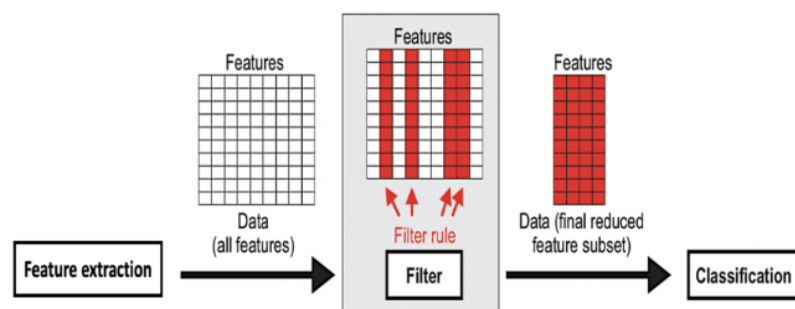


Feature Selection-

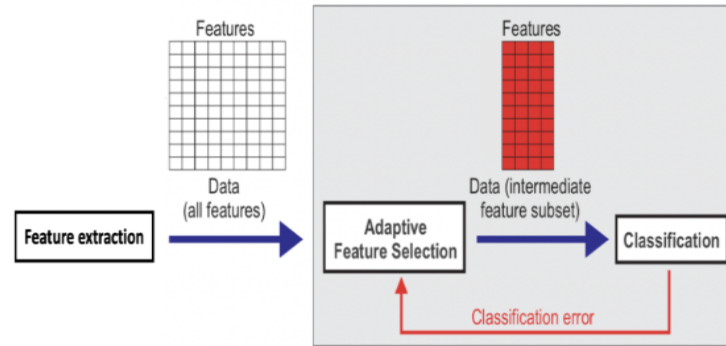
An ML model is trained on features extracted from the dataset. We use Feature Selection to select or drop the features depending on the performance measure. It is used for improving the learning performances of the model, lowering the computational complexity, and decreasing the storage required for the dataset.

There are three methods used for feature selection-

1. **Filter method-** Filter methods choose a feature subset from the large feature set applying some filter rules before training the model. This filter rule may be derived from prior knowledge or statistics of the data. This method needs a strong assumption of the distribution of the data for its optimal score. This method **ranks the features** based on their score individually. Here, the user needs to define the number of features or a threshold score to determine the subset of the features. Different techniques which use this method are :



- a. Correlation coefficient: It calculates the correlation between feature and target variables or between features.
 - b. Chi-square test: the features are selected based on the chi-square value. The chi-square value is the square of the difference between the observed frequency of the target label and the expected frequency of the target label to the expected frequency concerning the individual feature to the target variable.
 - c. Information gain: Measures the reduction in entropy brought by a feature to the target variable.
 - d. Analysis of variance: Also known as ANOVA determines the statistical measure that is significant in separating observed variance data into features concerning the target variable.
2. **The wrapper method-**the process of feature selection takes place in an iterative aspect by tuning the feature subset to a classifier. The **selected subset** is dependent on the error. The feature subset with a very less estimation error is preferred. The wrapper method techniques are very expensive.



- a. Forward Selection: The forward search strategy starts with one feature and builds a large feature set iteratively
 - b. Backward Elimination: starts with a large dataset and eliminates features iteratively.
 - c. Recursive elimination: Being a greedy optimization technique it selects the smaller and smaller set of feature data according to the importance of the features.
3. **Embedded method**- This is a combination of both **filter and wrapping** methods. It avoids the manual specification of the number of features or threshold of a filter. It is faster than the other two methods.
 - a. Regularization technique: LASSO which means Least absolute shrinkage and selection operator is used for the feature selection. It is a kind of penalty that is applied over the features of estimating the wrong target.
 - b. Tree-based technique : Random Forest technique uses performance measures from n number of Random forest Models to rank and select the features.