**Features selection in Machine Learning**

**features selection** is one of the important steps while building a machine learning model. Its goal is to find the best possible set of features for building a machine learning model.

When we get any dataset, not necessarily every column (feature) is going to have an impact on the output variable.

One way to think about feature selection methods are in terms of supervised and unsupervised methods. The difference has to do with whether features are selected based on the target variable or not.

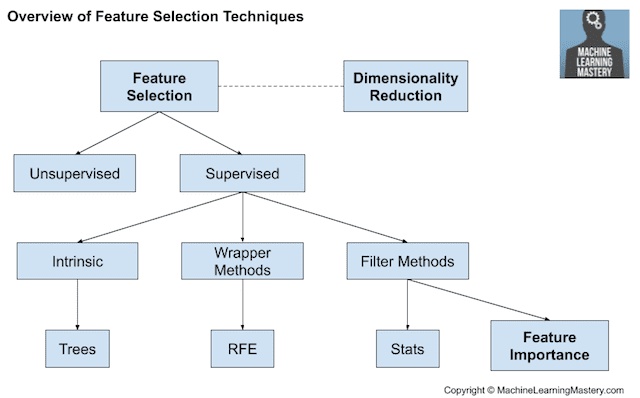
Unsupervised feature selection techniques ignores the target variable, such as methods that remove redundant variables using correlation.

Supervised feature selection techniques use the target variable, such as methods that remove irrelevant variables.

When it comes to implementation of feature selection in Pandas, Numerical and Categorical features are to be treated differently.

Some popular techniques of features selection in machine learning are:

* Filter methods
* Wrapper methods
* Embedded methods (intresinc method)



1. **Filter Methods**

Filter feature selection methods use statistical techniques to evaluate the relationship between each input variable and the target variable, and these scores are used as the basis to choose (filter) those input variables that will be used in the model.

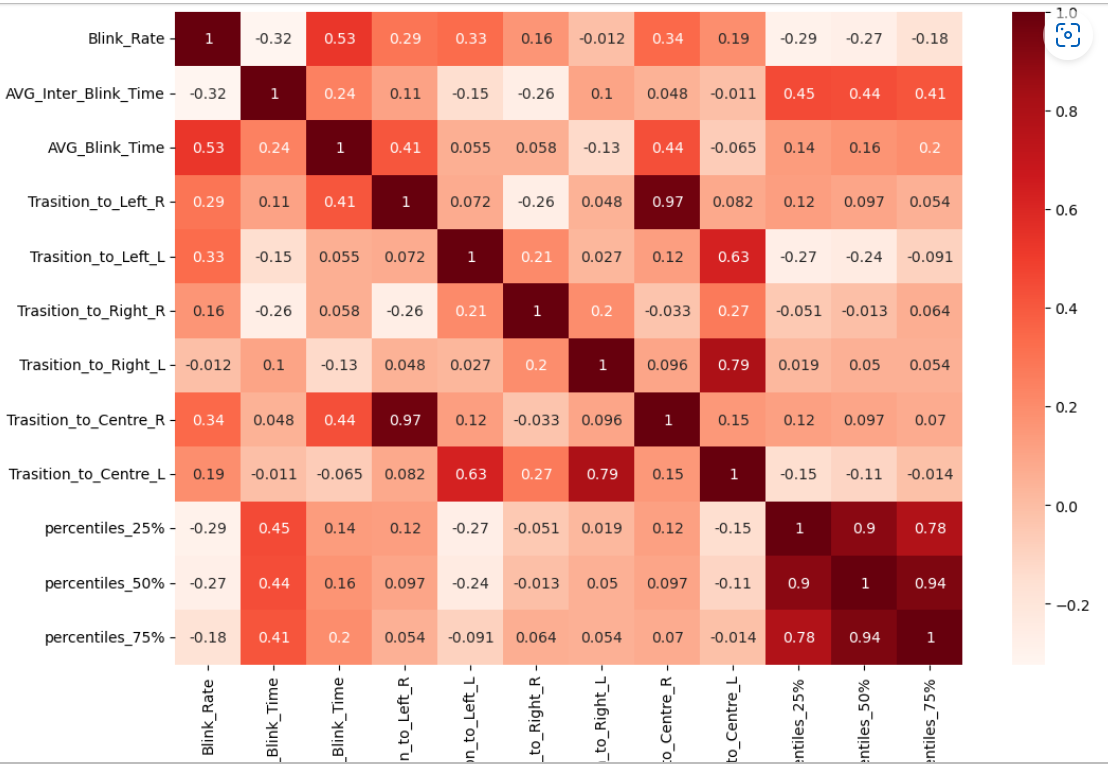
Filter Methods can be:

* Statistical Methods
* Feature Importance Methods
  1. **Selecting using correlation**

It is common to use correlation type statistical measures between input and output variables as the basis for filter feature selection.

We can also measure correlation between feature each other to remove redundant variables:

The correlation coefficient has values between -1 to 1  
— A value closer to 0 implies weaker correlation (exact 0 implying no correlation)  
— A value closer to 1 implies stronger positive correlation  
— A value closer to -1 implies stronger negative correlation

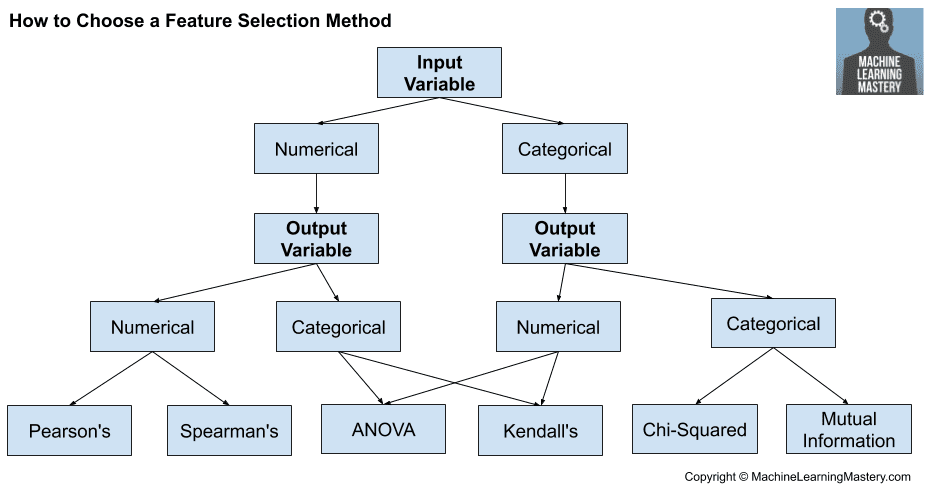


Strong correlation between:

* Transiton\_Center\_Right and Transition\_Left\_Right (0.97)
* Transition\_Center\_ and Transition\_right\_left (0.79)
* Percentiles\_25%, percentiles\_50%, percentiles\_75% (0.78)

**Statistics methods**

The choice of statistical measures is highly dependent upon the variable data types.



### Numerical Input, Categorical Output

This is a classification predictive modeling problem with numerical input variables.

This might be the most common example of a classification problem,

Again, the most common techniques are correlation based, although in this case, they must take the categorical target into account.

* ANOVA correlation coefficient (linear).
* Kendall’s rank coefficient (nonlinear).

Kendall does assume that the categorical variable is ordinal.

**3. Wrapper Method**

A wrapper method needs one machine learning algorithm and uses its performance as evaluation criteria. This means, you feed the features to the selected Machine Learning algorithm and based on the model performance you add/remove the features. This is an iterative and computationally expensive process but it is more accurate than the filter method. There are different wrapper methods such as Backward Elimination, Forward Selection, Bidirectional Elimination and RFE.

Wrapper feature selection methods create many models with different subsets of input features and select those features that result in the best performing model according to a performance metric. These methods are unconcerned with the variable types, although they can be computationally expensive. [RFE](https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.RFE.html) is a good example of a wrapper feature selection method.

1. **Embedded Method**

There are some machine learning algorithms that perform feature selection automatically as part of learning the model. We might refer to these techniques as **intrinsic** feature selection methods.

This includes algorithms such as penalized regression models like Lasso and decision trees, including ensembles of decision trees like random forest.