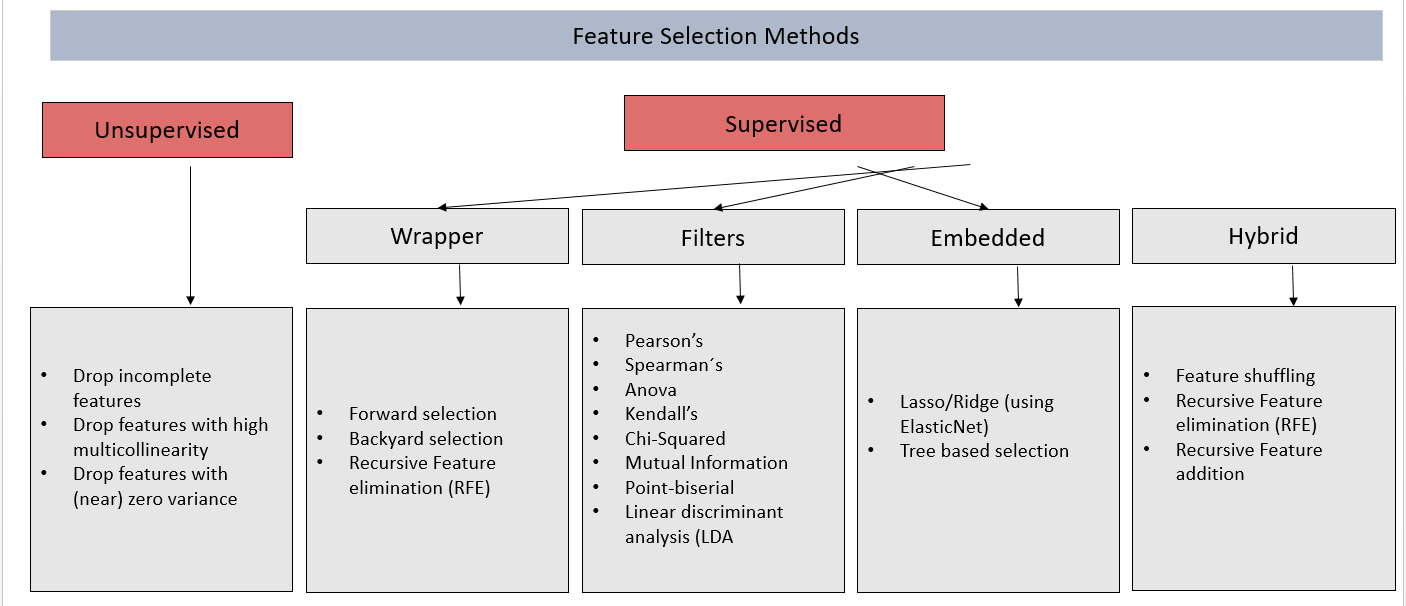
Feature Selection Methods



1. **Filter**: it is based on correlation, uses statistical measures to assign a score to each feature. This method should be used for preliminary screening. It can detect constant, duplicated, and correlated features. Usually not the best performance in terms of reducing features.

Filter method is less accurate. It is great while doing EDA, it can also be used for checking multi co-linearity in data.

Note: Linear discriminant analysis (LDA): Linear discriminant analysis is used to find a linear combination of features that characterizes or separates two or more classes of a categorical variable.

1. **Wrapper Methods**: This approach uses Machine Learning algorithm. Performance of this method depends on model selected and data underlying. Usually can suggests the optimal feature subset. Tries different subset of features to figure out optimal features. Typically very computationally expensive. Can detect interactions between features.

Select a set of features, in which different combinations are prepared, evaluated and compared. A predictive model is used to evaluate the combination of features and assign a score based on model accuracy. Example: Backward Elimination, Forward Selection, Bidirectional Elimination,Exhaustive Search, and RFE (Recursive Feature Elimination). Wrappers are terribly slow when it comes to large datasets.

1. **Embedded Methods:** Performs feature selection as building the model. Generally less computationally expensive than Wrapper methods. Often provides results that are best of both worlds, often more realistic approach.

**Wrapper and Embedded methods** give more accurate results but as they are computationally expensive, these method are suited when you have lesser features (~20).

1. **Hybrid Methods:** Amalgamation of all the techniques above. This approach is less computationally expensive than Wrapper methods and has good performance.

**Qual Método utilizar?**

* Tente usar o RFE caso tenha recursos computacionais para isso.
* Se estiver trabalhando com Classificação e as features forem numéricas utilize f\_classif ou mutual\_info\_classif.
* Se estiver trabalhando com Regressão e as features forem numéricas utilize f\_regression ou mutual\_info\_regression.
* Caso esteja trabalhando com features categóricas utilize chi2
* Automatize essa etapa com Pipelines para evitar erros.

https://scikit-learn.org/stable/tutorial/machine\_learning\_map/index.html