Feature Engineering

* Helps to get the best results from the algorithm you are using.
* Getting the most out of your data for your algorithms to work with.

Components:

* Feature selection
* Feature Scaling

**Feature Selection**

* Reducing the number of input variables when developing a predictive model. Primarily focused on reducing non-informative and redundant predictors from models.
* *Many models, especially those based on regression slopes and intercepts, will estimate parameters for every term in the model. Because of this, the presence of non-informative variables can add uncertainty to the predictions and reduce the overall effectiveness of the model.*
* Helps reduce computational cost of model, and sometimes improve model performance.

**Feature Selection Methods:**

1. **Supervised feature selection** uses the target variable, e.g. methods that remove irrelevant variables.

* **Wrapper feature selection** creates 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 is a good example of a wrapper feature selection method.
* **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.
* **Intrinsic feature selection methods** are built into some machine learning algorithms and perform feature selection automatically as part of the learning process

1. **Unsupervised feature selection** ignores the target variable

Diagram

Description automatically generated

<https://www.kaggle.com/c/competitive-data-science-predict-future-sales/data?select=sales_train.csv>

<https://www.kaggle.com/arashnic/hr-analytics-job-change-of-data-scientists>