+

CS 522 – Selected Topics in CS

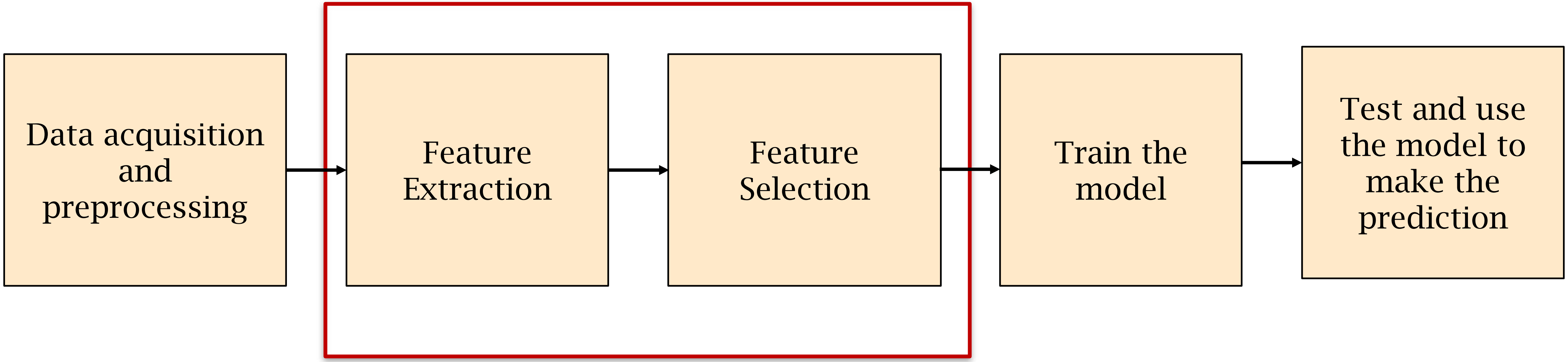
Lecture 03 – Feature Engineering

January 31, 2022

+ Topics to be covered

- What is Feature
- Feature Engineering
- Feature Extraction
- Feature Selection
- Feature Encoding

Feature Engineering



Sometimes feature extraction is a part of preprocessing.
Feature extraction also include feature encoding, standardization and normalization

+ What is Feature (Review)

- The ***initial pick*** of feature is always an ***expression*** of ***prior knowledge.***
- *Images* → pixels, contours, textures, etc.
- *Signal* → samples, spectrograms, etc.
- *Time series* → ticks, trends, reversals, etc. العلامات - اتجاهات- انعكاسات
- *Biological data* → DNA, marker sequences, genes, etc.
- *Text data* → words, grammatical classes and relations, etc.

+ What is Feature Engineering ?

- Feature engineering is the ***science*** (and art) of ***extracting*** information from ***raw data***.
- It is a process that ***transform raw data*** into ***features*** that ***better*** represent the ***underlying problem to the predictive models***, resulting in ***improved*** model ***accuracy*** on ***unseen*** data.

Machine Learning Mastery:
Jason Brownlee

- Feature Engineering is a ***Representation Problem***
  - Machine learning algorithms ***learn a solution to a problem from sample data***.
  - In this context, feature engineering asks: ***what is the best representation of the sample data to learn a solution to your problem?***

+ What is Feature Engineering ?

- ***Some machine learning projects succeed and some fail.***
- ***What makes the difference?***
- ***Easily the most important factor is the features used*****.**

A few useful things to know about Machine Learning: Pedro Domingos

A feature is an *individual measurable property* or *characteristic* of a phenomenon being *observed*.

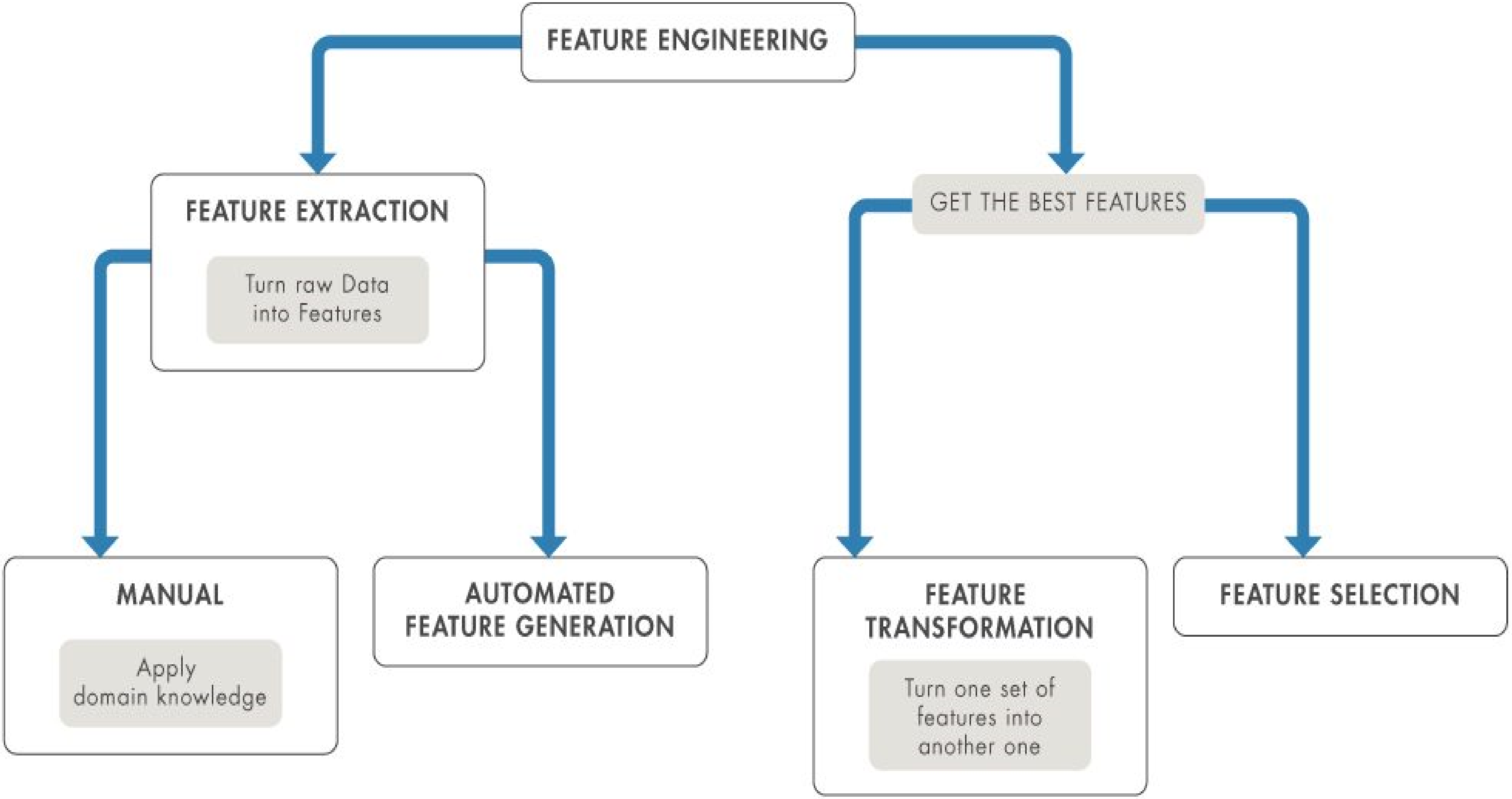
+ Goal of Feature Engineering

- ***Convert unstructured*** data into ***input*** to ***learning*** algorithm.
- ***Expose*** the ***structure*** of the ***concept*** to the ***learning*** algorithm.
- Work ***well*** with the ***structure*** of the model.
- Balance number of ***features***, ***complexity*** of ***concept***, ***complexity*** of ***model***, and the ***amount*** of ***data***.

+ Benefits of Feature Engineering

- ***Improving*** the accuracy of ML model
- ***Solving Overfitting*** problem
- ***Speed up*** your ***computation***
- ***Understandability*** for ML process

+ Feature Engineering



+ Feature Extraction

- Feature extraction refers to the process of ***transforming raw data*** into ***numerical features*** that can be ***processed*** while ***preserving the information*** in the ***original data set***.

- It yields ***better results*** than ***applying*** machine learning ***directly*** to the raw data.

- Feature Extraction can be performed using ***two way***

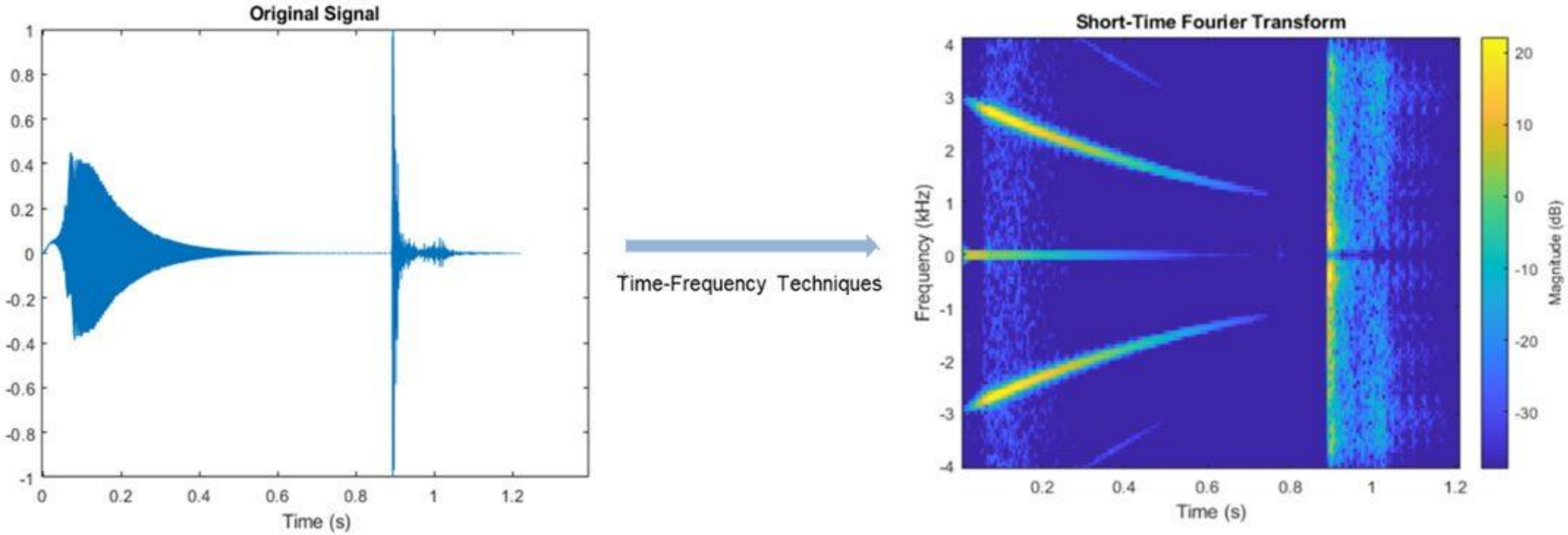
- ***Manual Feature Extraction***
  - Manual feature extraction requires ***identifying*** and ***describing*** the ***features*** that are ***relevant*** for a ***given problem*** and ***implementing*** a way to ***extract*** those features.
  - In many situations, having a ***good understanding*** of the ***background*** or ***domain*** can ***help*** make ***informed decisions*** as to which features could be ***useful***.
  - It is impractical for the huge datasets.

+ Feature Extraction

- **Automated Feature Extraction**
  - Automated feature extraction uses ***specialized algorithms*** or ***deep networks*** to extract features automatically from ***signals*** or ***images*** without the ***need*** for ***human intervention***.
  - This technique can be very ***useful*** when you want to move ***quickly*** from ***raw data*** to developing machine learning algorithms.

+ Feature Extraction in different types of Data

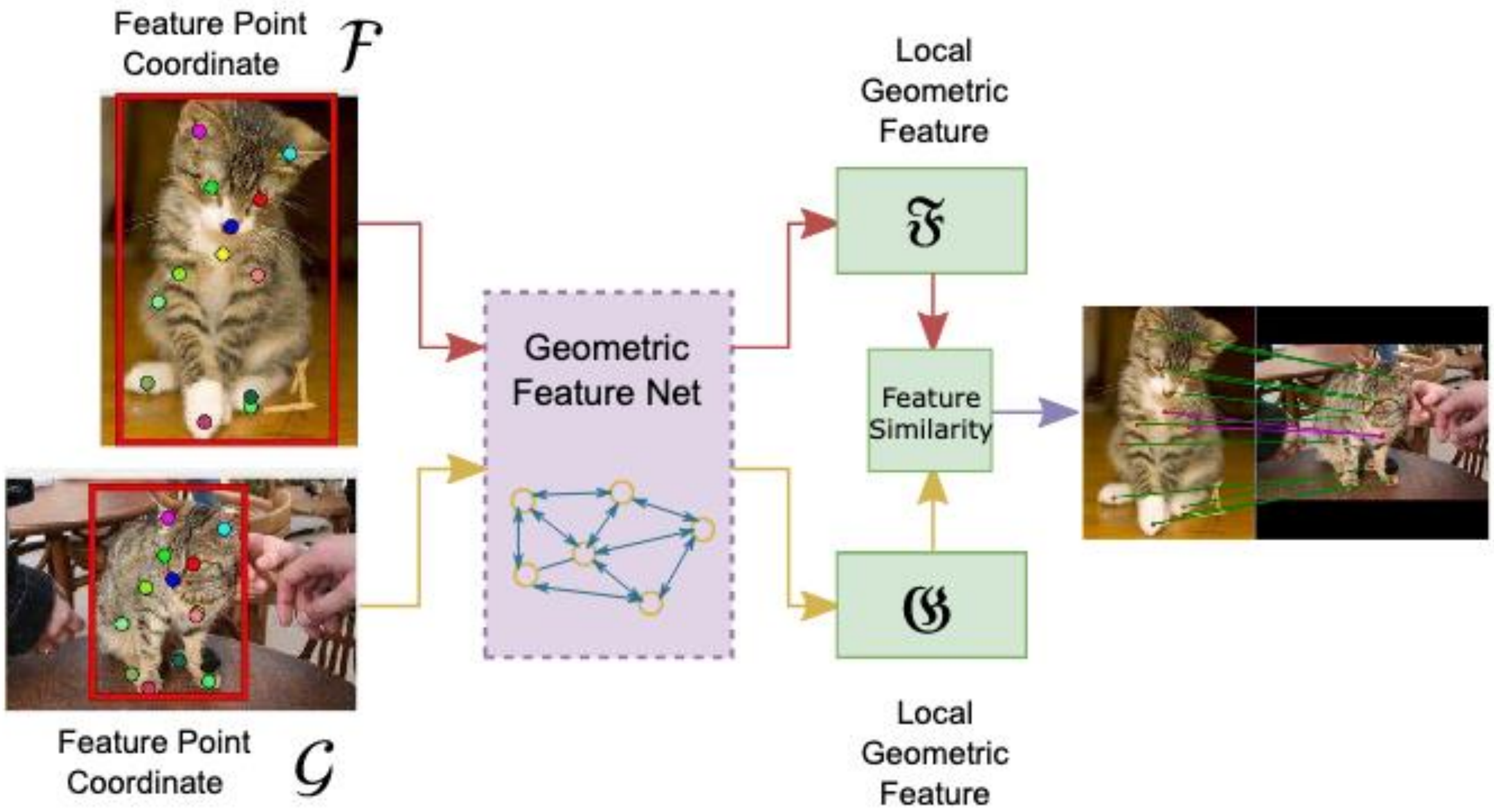
Time Series/Audio Data



Spectrogram of a ***signal*** using short-time ***Fourier transform***.
Spectrogram shows variation of ***frequency*** content over time.

+ Feature Extraction in different types of Data

Image /Visual Data



Feature Extracted from ***Images***

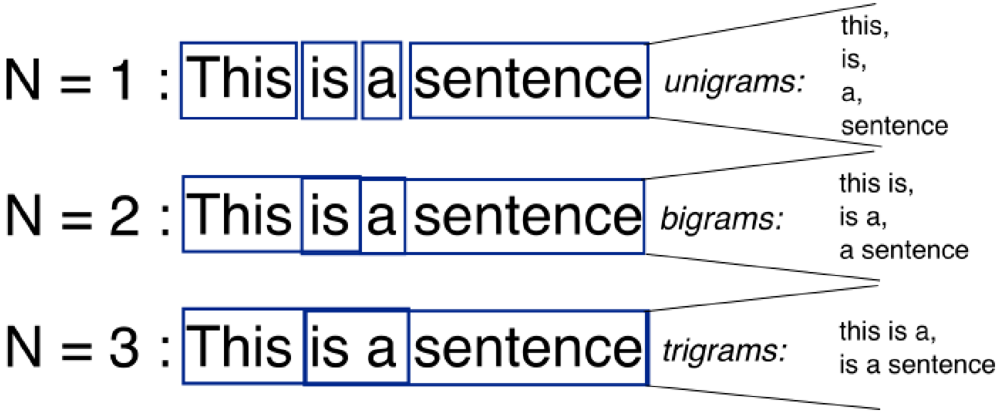
+ Feature Extraction in different types of Data

Textual Data

	1 This	2 movie	3 is	4 very	5 scary	6 and	7 long	8 not	9 slow	10 spooky	11 good	Length of the review(in words)	
Review 1	1	1	1	1	1	1	1	0	0	0	0	7	negative
Review 2	1	1	2	0	0	1	1	0	1	0	0	8	negative
Review 3	1	1	1	0	0	0	1	0	0	1	1	6	positive

- **Vector of Review 1:** [1 1 1 1 1 1 1 0 0 0 0]
- **Vector of Review 2:** [1 1 2 0 0 1 1 0 1 0 0]
- **Vector of Review 3:** [1 1 1 0 0 0 1 0 0 1 1]

Review 1: this movie is very scary and long
Review 2: this movie is very long and is slow
Review 3: this movie is long, spooky, good



Feature Extracted from text

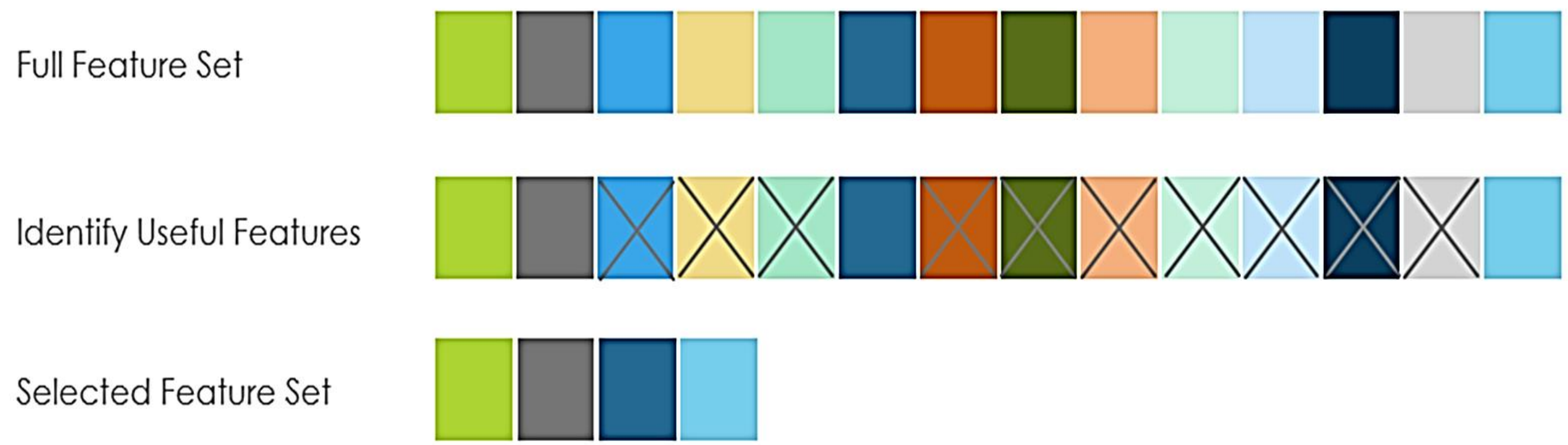
+ Feature Selection

- As a dimensionality reduction technique, feature selection aims to choose a small subset of the relevant features from the original features by removing ***irrelevant, redundant, or noisy features***.

- Feature selection usually can lead to better ***learning performance, higher learning accuracy, lower computational cost, and better model interpretability***.

- ***The aim of feature selection is to maximize relevance and minimize redundancy.***

Feature Selection



- **Given:** a set of predictors ("***features***") ***V*** and a target variable ***T***
- ***Find: minimum set F*** that achieves ***maximum classification performance*** of ***T*** (for a given set of classifiers and classification performance metrics)

+ Feature Selection

- Feature selection is also called ***variable selection*** or ***attribute selection*** or ***dimensionality reduction***

- A process of ***reducing*** the number of features used to ***build*** the model with the ***goal*** of keeping only ***informative, discriminative*** and ***non redundant*** features.
  - It is the ***automatic selection*** of ***attributes*** in your ***data*** (such as columns in tabular data) that are ***most relevant*** to the ***predictive modelling problem*** you are working on.

- To keep ***"relevant features only"***, we will remove the features that are
  - Non informative
  - Non discriminative غير مميز
  - Redundant

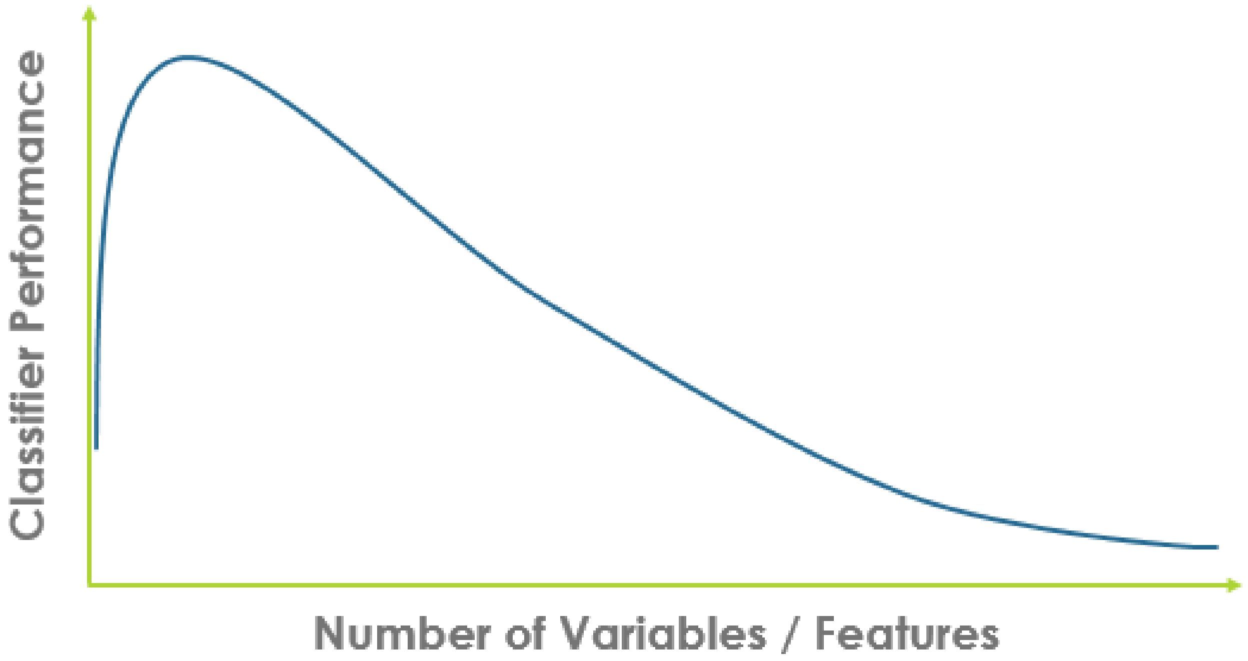
+ Feature Selection

- The ***objective*** of variable selection is three-fold:
  - ***Improving*** the ***prediction performance*** of the predictors,
  - Providing ***faster*** and more ***cost–effective predictors,***
  - Providing a ***better understanding*** of the ***underlying process*** that generated the data.
- ***Removing a redundant variab***le helps to ***improve accuracy***.
- ***Inclusion of a relevant variable*** has a ***positive*** effect on model ***accuracy***.
- Too many variables might result to ***overfitting*** which means model is not able to ***generalize*** pattern
- Too many variables lead to ***slow computation*** which in turns requires ***more memory and hardware***.
- It ***reduces*** the ***complexity*** of a model and makes it ***easier*** to interpret.

+ Curse of dimensionality

- The required number of samples (to achieve the same accuracy) grows ***exponentially*** with the ***number of variables***!

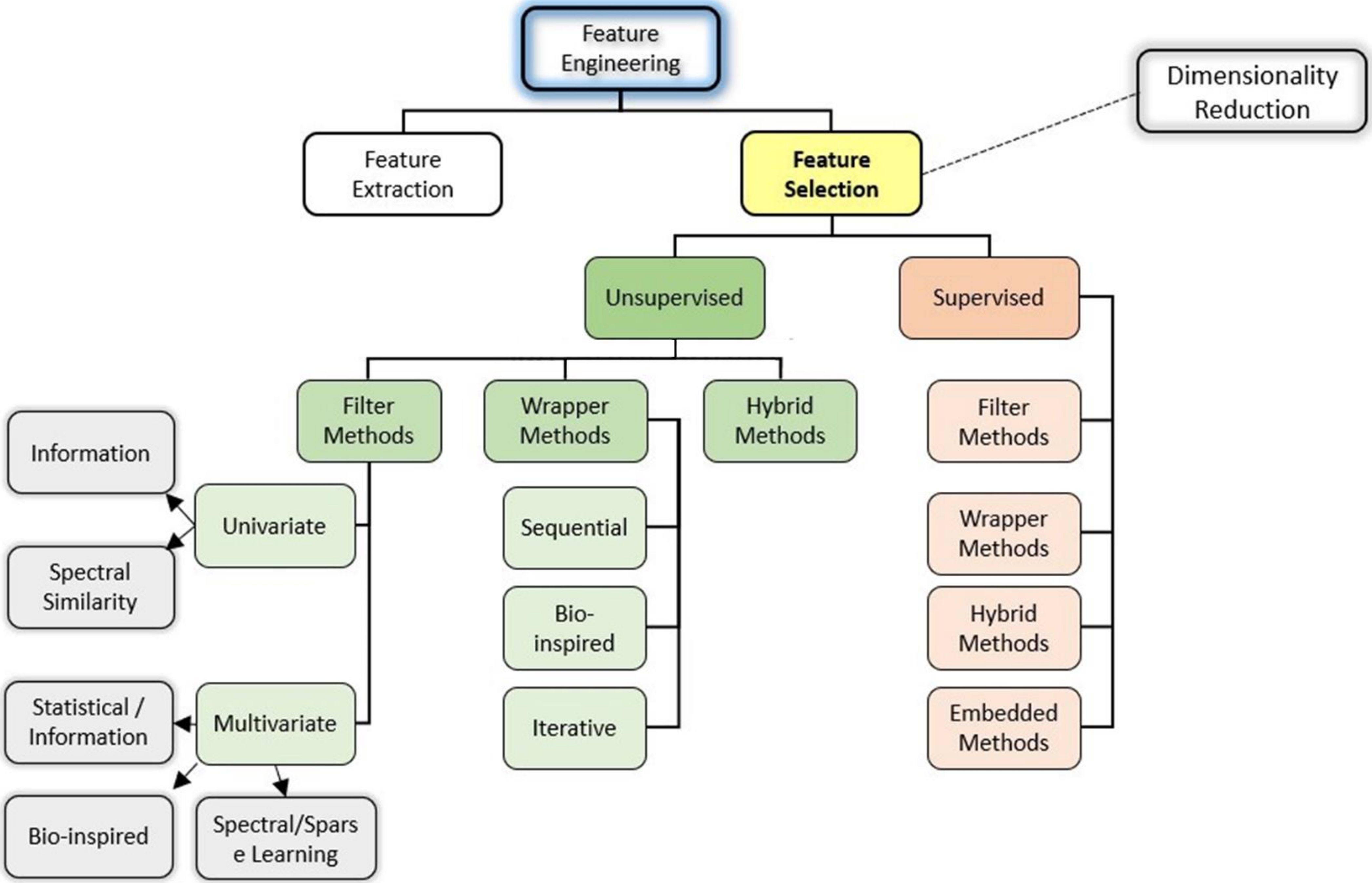
- The ***classifier's performance*** usually will ***degrade*** for a large number of features!



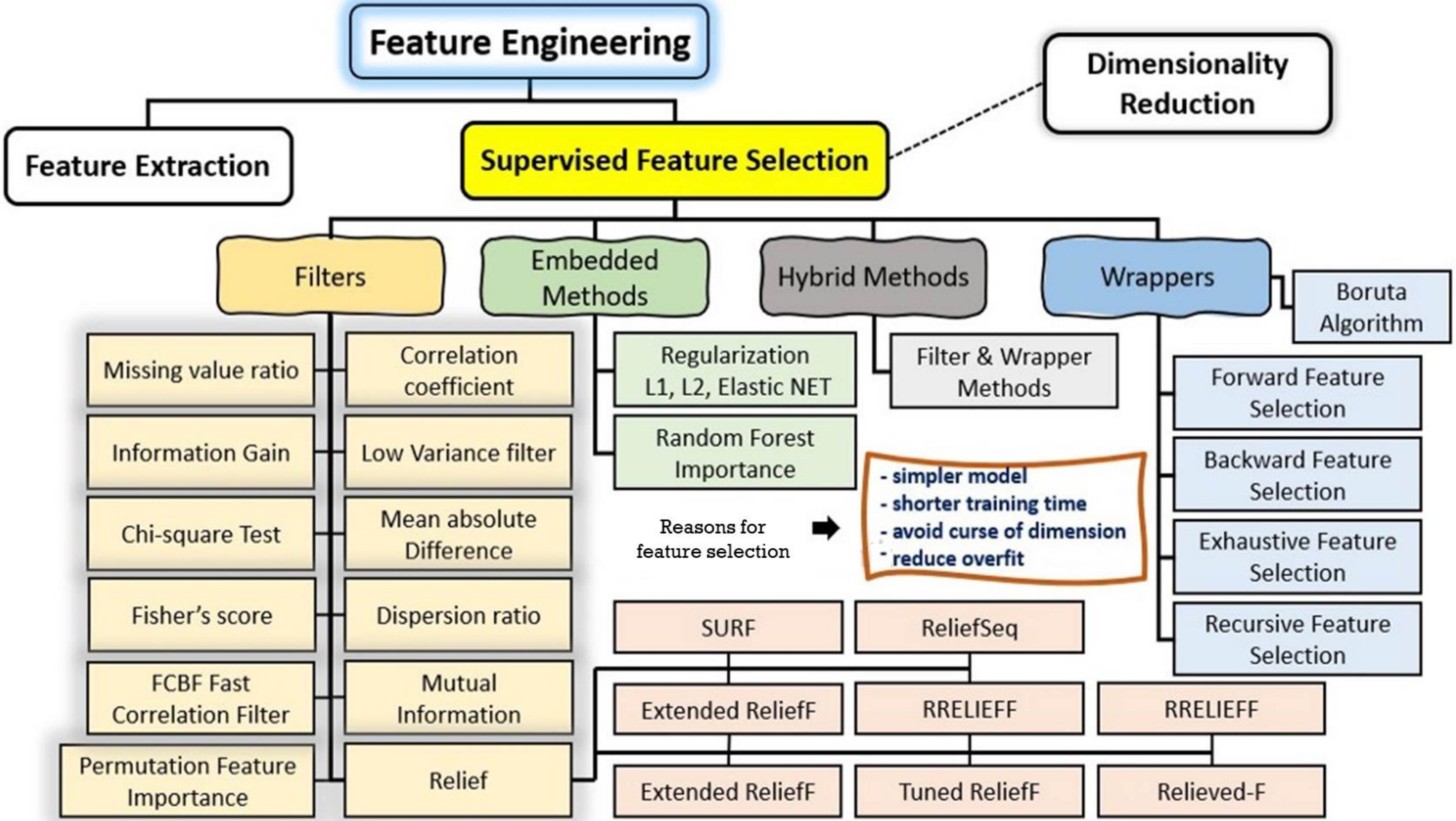
+ Categorization of Feature Selection Techniques

- ***Unsupervised feature selection*** techniques ***ignore*** 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***.


Feature Engineering
Feature Extraction
Feature Selection
Dimensionality Reduction
Unsupervised
Supervised
Filter Methods
Wrapper Methods
Hybrid Methods
Information
Univariate
Spectral Similarity
Sequential
Bio-inspired
Iterative
Statistical / Information
Multivariate
Bio-inspired
Spectral/Sparse Learning
Filter Methods
Wrapper Methods
Hybrid Methods
Embedded Methods

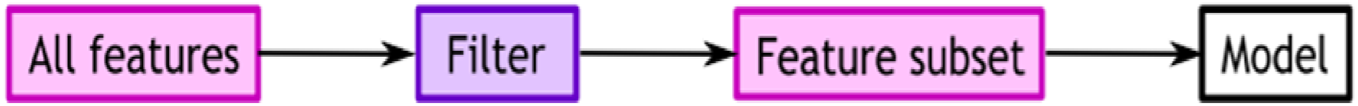
+ Supervised Feature Selection



+ Supervised Feature Selection

- ***Four*** type of supervised Feature Selection
  - Filter
  - Wrapper
  - Embedded
  - Hybrid
- **Before training model**
  - ***Statistical method:*** Removing features with low variance
  - ***Filter method:*** Univariate feature selection
- **While training model**
  - ***Wrapper method:*** Recursive feature elimination
  - ***Embedded method:*** L1-based feature selection

+ 1. Filter Supervised Feature Selection

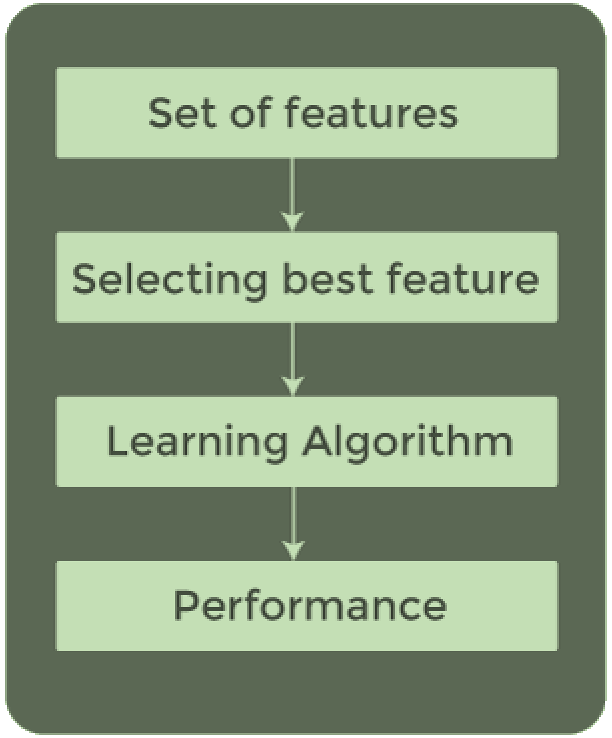


- In Filter Method, features are selected using ***statistics measures.***

- This method ***does not depend*** on the ***learning*** algorithm and ***chooses*** the features as a ***pre-processing step***.

- The filter method ***filters*** out the ***irrelevant*** feature and ***redundant*** columns from the model by using ***different*** metrics through ***ranking***.

- The advantage of using filter methods is that it ***needs low computational time*** and ***does not overfit the*** data.



+ 1. Filter Supervised Feature Selection methodology

1. Creates ***groups*** of the ***features*** as per different criteria.

2. Creates a ***benchmark*** for each group.

3. Tests the ***correlation of features*** inside the ***group***, compared to the predetermined ***group benchmark.***

4. Keeps only the features that are ***less correlated to each other*** than to the group benchmark.

+ 1. Filter Supervised Feature Selection

- Common Filter Feature selection methods are
- Information Gain
- Chi-square Test
- Fisher's Score
- Missing Value Ratio
- Correlation coefficient filter
- High correlation filter
- Low variance filter

+ 1. Filter Supervised Feature Selection

- ***Missing Value Ratio***
  - ***Remove*** those features which have ***high ratio*** of ***missing*** values
  - A ***predefined threshold*** may be defined. In case of ***low missing*** vales, the imputation technique may need to be applied.
  - The variable is having ***more*** than the ***threshold*** value can be ***dropped***.
- ***Information Gain***
  - Information gain determines the ***reduction in entropy*** while ***transforming*** the dataset.
  - It can be used as a ***feature selection*** technique by calculating the ***information gain*** of each variable with respect to the ***target variable***.
- ***Chi–square Test***
  - Chi-square test is a technique to determine the ***relationship*** between the ***categorical*** variables.
  - The chi-square value is calculated between each ***feature*** and the ***target*** variable, and the desired number of features with the ***best chi–square value*** is selected.

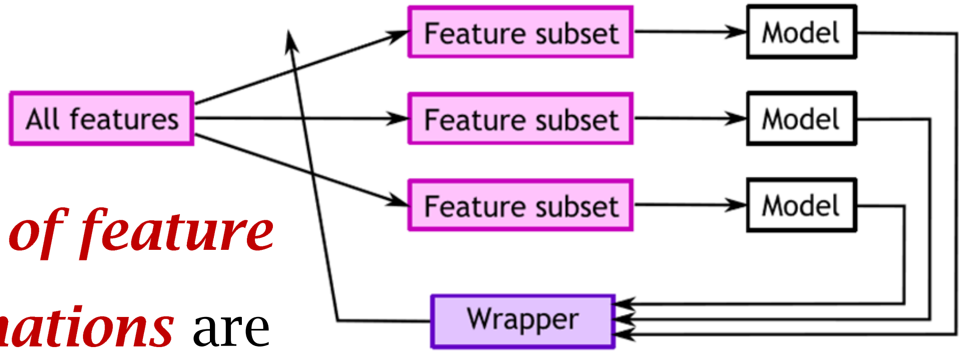
+ 1. Filter Feature Selection

- ***Fisher's Score***
  - Fisher's score is one of the popular ***supervised technique*** of features selection.
  - It returns the ***rank*** of the ***variable*** on the ***fisher's criteria*** in ***descending*** order. Then we can select the variables with a ***large fisher's score***.
- ***Correlation Coefficient***
  - Features should be ***correlated*** with the ***target*** but should be ***uncorrelated*** among ***themselves***.
- ***High Correlation Filter***
  - ***High Correlation*** between ***two features*** means they have ***similar trends*** and are likely to ***carry similar information***.
  - If the correlation coefficient ***crosses*** a certain ***threshold value***, we can ***drop one of*** the features.

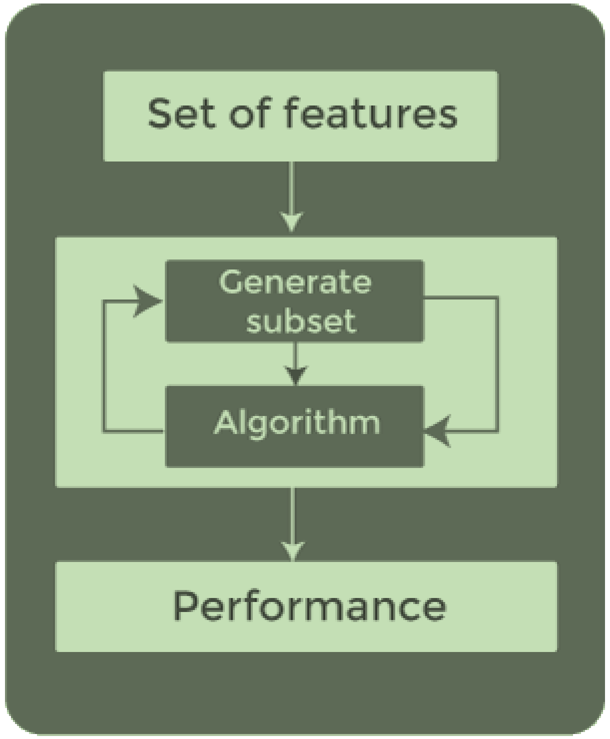
+ 1. Filter Feature Selection

- ***Low variance Filter***
  - Features in a dataset where ***all the observations*** have the ***same*** value***, e.g. 1***
  - Such features have ***0 variance*** and may ***not be significant***.
  - The assumption is that features with ***higher variance*** may contain ***more useful*** information.

+ 2. Wrapper Supervised Feature Selection

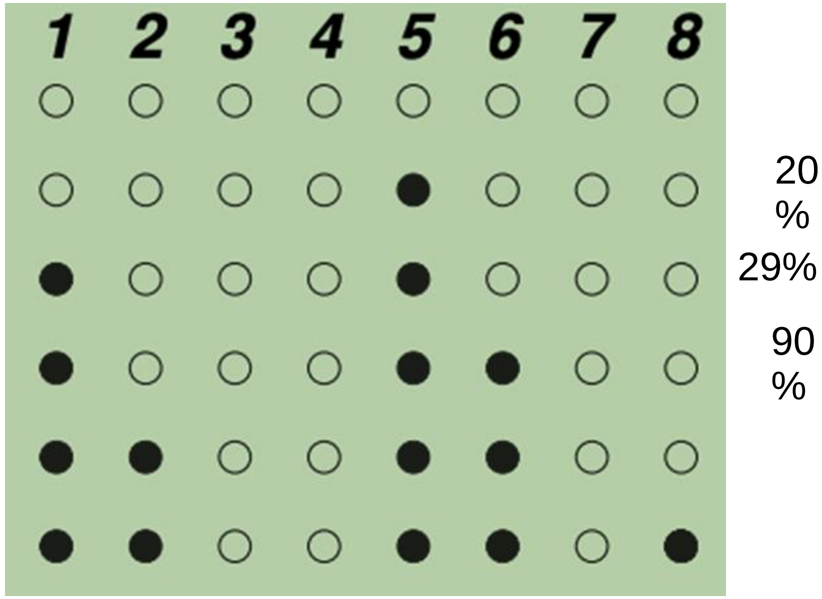


- The Wrapper methodology considers the ***selection of feature*** sets as a ***search problem***, where ***different combinations*** are ***prepared***, ***evaluated***, and ***compared*** to other combinations.
- A ***predictive model*** is used to ***evaluate*** a ***combination*** of ***features*** and ***assign*** model ***performance scores***.
- The ***performance*** of the Wrapper method ***depends*** on the ***classifier***.
- The ***best subset*** of features is selected based on the ***results*** of the ***classifier***.



+ 2. Wrapper Supervised Feature Selection

- ***Forward selection***
  - Forward selection is an ***iterative process***, which begins with an ***empty*** set of features.
  - After each ***iteration***, it keeps ***adding on*** a feature and ***evaluates*** the ***performance*** to check whether it is ***improving*** the performance or ***not***.
  - The process continues until the ***addition*** of a new variable/feature ***does not improve*** the performance of the model.



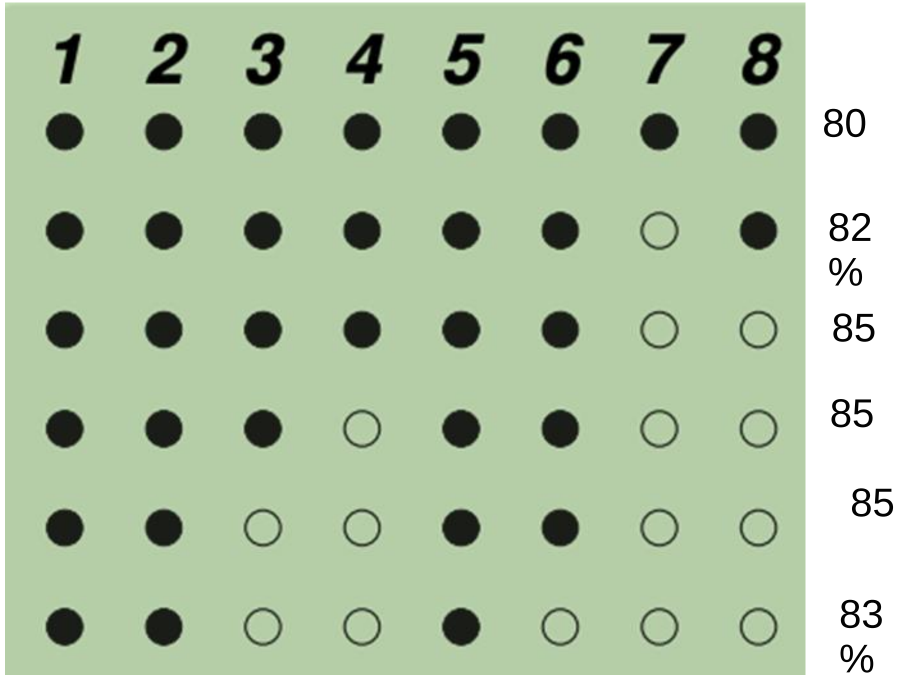
+ 2. Wrapper Supervised Feature Selection

- ***Backward elimination***

- ***Backward*** elimination is also an ***iterative*** approach, but it is the ***opposite*** of forward selection.

- This technique begins the process by ***considering all*** the ***features*** and ***removes*** the ***least significant feature***.

- This elimination process ***continues*** until ***removing*** the features ***does not*** improve the performance of the model.



+ 2. Wrapper Supervised Feature Selection

- ***Exhaustive Feature Selection***
  - Exhaustive feature selection is one of the ***best feature selection methods***, which evaluates each ***feature set as brute-force***.
  - It means this method ***tries & make*** each ***possible combination*** of features and ***return*** the best performing feature set.
- ***Recursive Feature Elimination***
  - Recursive feature elimination is a recursive ***greedy optimization*** approach, where features are selected by ***recursively*** taking a ***smaller and smaller*** subset of features.
  - Now, an ***estimator*** is trained with ***each set of features***, and the ***importance*** of each feature is determined using ***coef_attribute*** or through a ***feature_importances_attribute***.
  - This process repeats recursively on ***pruned set until*** the desired num of ***features reached***

+ 2. Wrapper Method

- ***Computationally expensive*** than filter method
- Perform ***better*** than filter method
- Not recommended on ***high number of features.***
- Before applying the wrapper feature selection, we must specify:
- ***What model type and which learning algorithm shall be used?***
- ***How to evaluate the model accuracy?***
  - Based on testing data, or using k-fold cross-validation?

+ Filter Vs. Wrapper

- Wrapper methods are ***computationally more expensive*** than filter methods, ***due*** to the ***repeated learning*** steps and ***cross-validation***.
- However, these methods are more ***accurate*** than the ***filter*** method

+ Embedded Feature Selection

- In the Embedded method, there are ***ensemble learning*** and ***hybrid learning*** methods for feature selection.

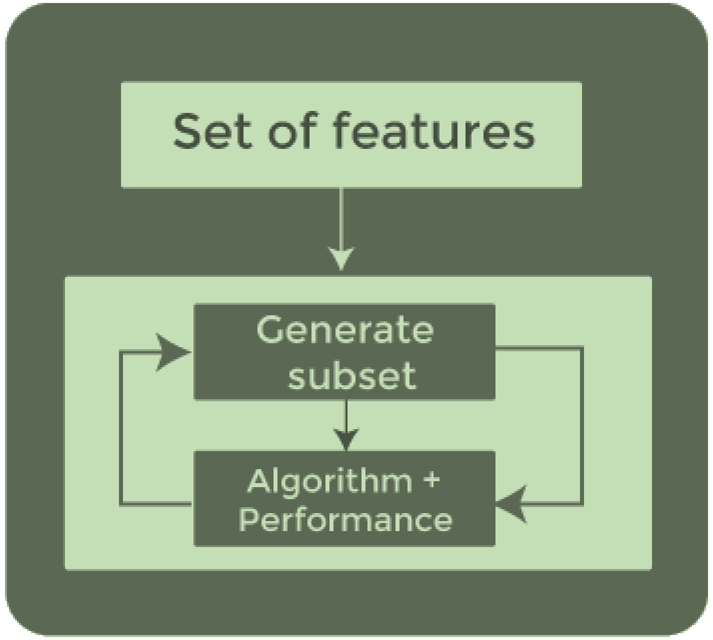
- Since it has a ***collective decision***, its performance is ***better*** than the other ***two models***.

اخف

- It is ***computationally less intensive*** than ***wrapper*** methods. However, this method has a ***drawback specific to a learning model***.

- ***Random forest is one such example of wrapper and embedded feature selection.***

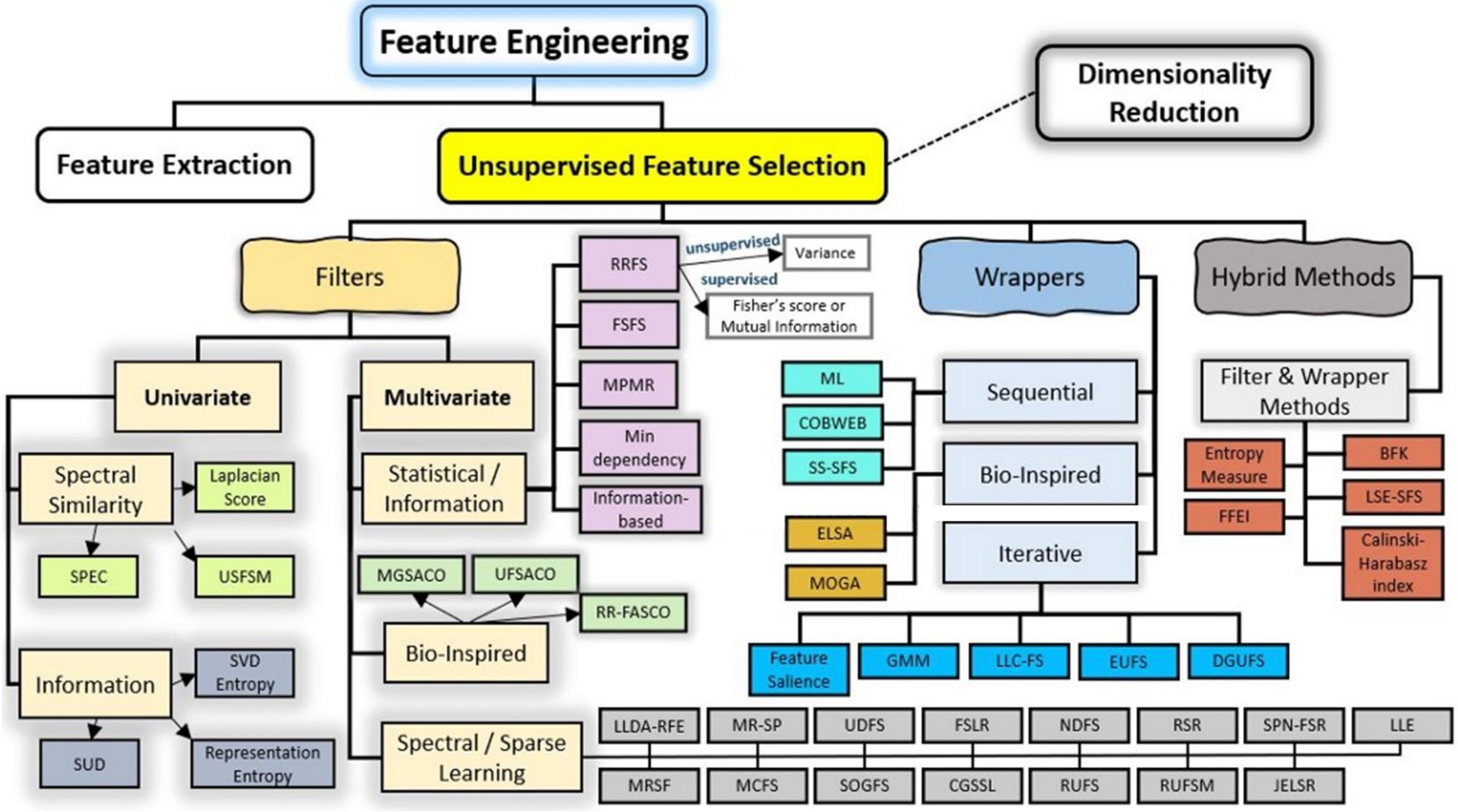
- ***Example***: : LASSO Regularization L1, Random Forest.



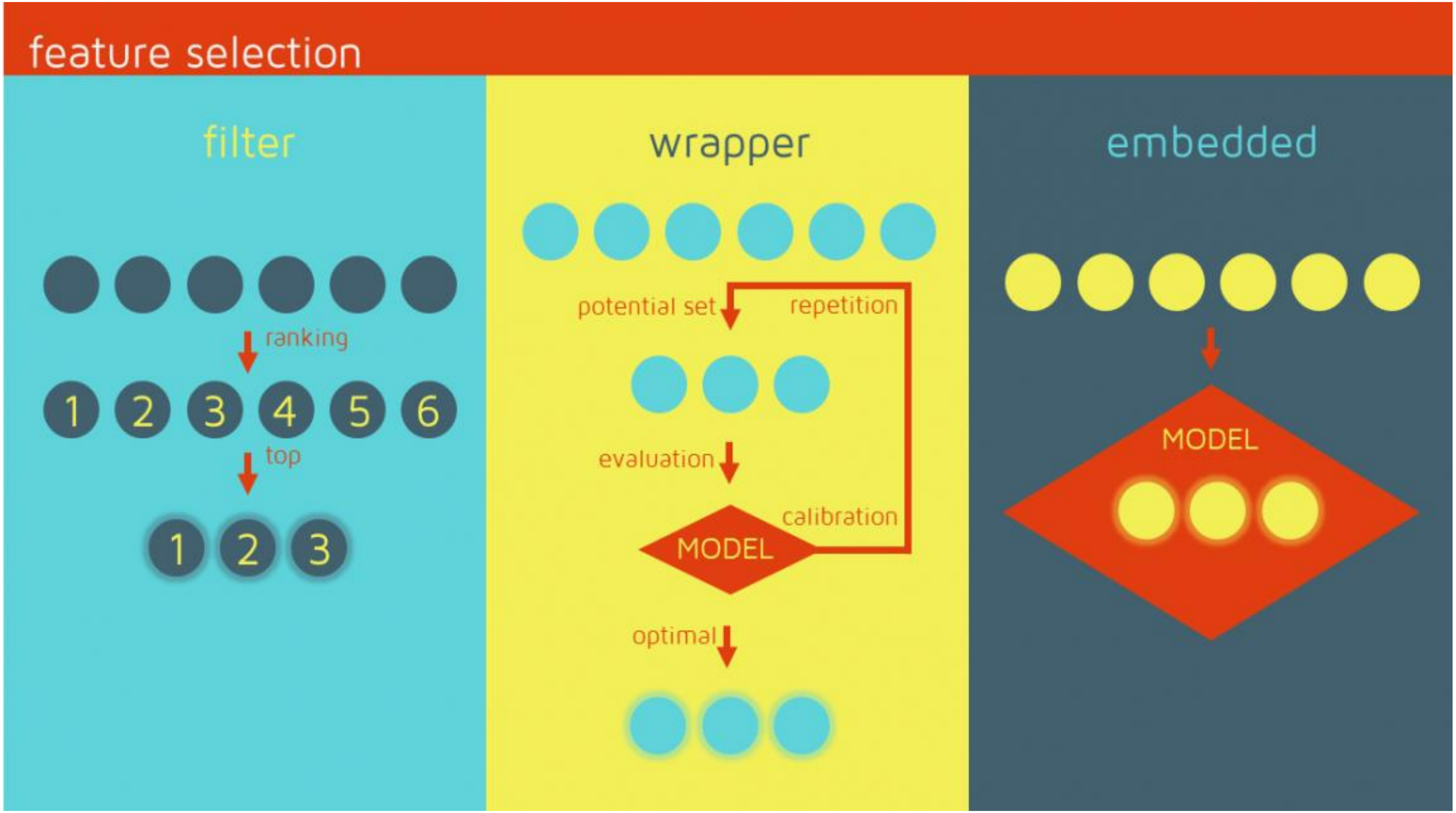
+ Hybrid Feature Selection

- The process of creating ***hybrid feature selection methods*** depends on ***what you choose to combine***.
- The main priority is to ***select the methods*** you're going to use, then ***follow*** their ***processes***.
- The idea here is to use these ***ranking methods*** to ***generate*** a ***feature ranking*** list in the ***first step***, then use the ***top k features*** from this list to perform ***wrapper methods***.
- With that, we can ***reduce the feature space*** of our dataset using these ***filter–based*** rangers to improve the ***time complexity*** of the wrapper methods.

Unsupervised Feature Selection



Review



https://quantdare.com/what-is-the-difference-between-feature-extraction-and-feature-selection/

End of Lecture – 03