Feature Selection in Machine Learning with Python

DataTalks.Club

August 2022

Soledad Galli, PhD

About me

- Data science instructor: www.trainindata.com
- Open-source developer: Feature-engine https://feature-engine.readthedocs.io/en/latest
- **Book**: Feature selection in machine learning: https://leanpub.com/feature-selection-in-machine-learning/

🐦 @Soledad_Galli

in in/soledad-galli/

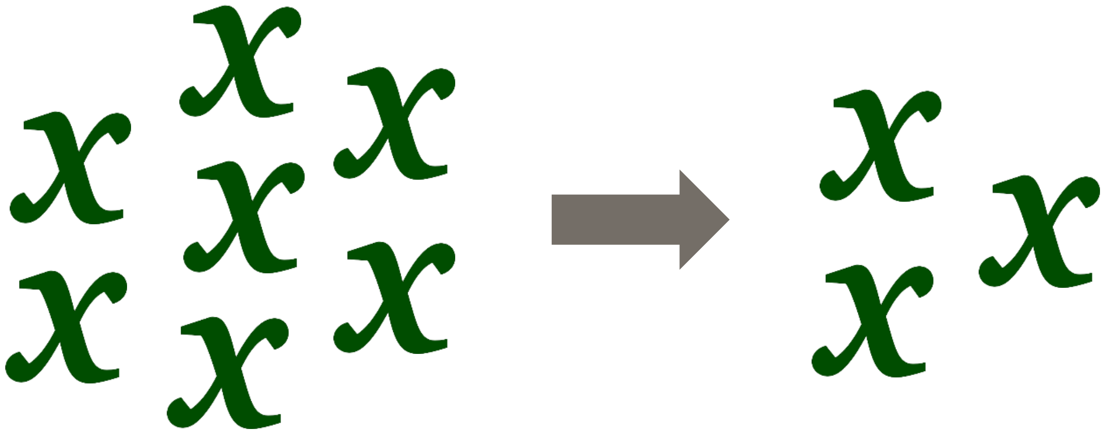
About this talk

Slides and code:

https://github.com/solegalli/DataTalks.Club2022

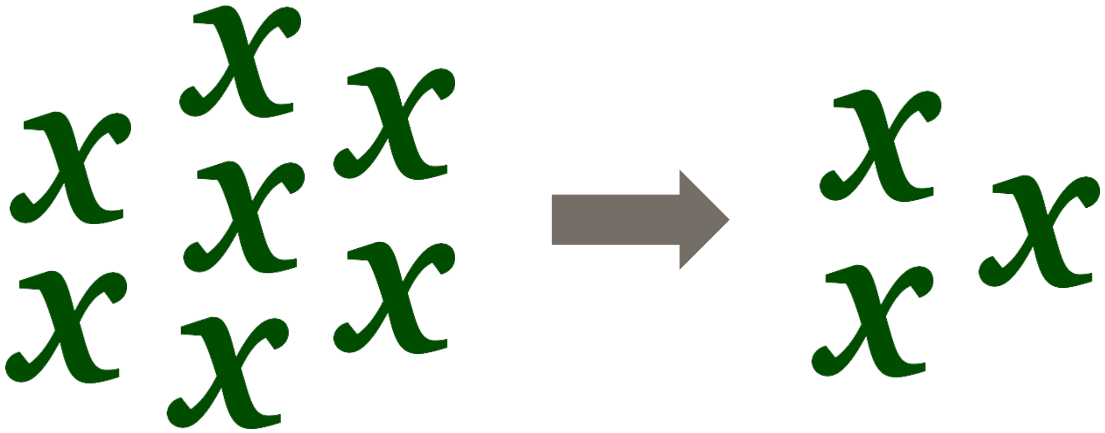
Feature selection

Feature selection is the process of selecting a subset of features to train machine learning models.



Feature selection vs dimension reduction

- Feature selection is not the same as dimensionality reduction.
- In **feature selection** the nature of the features is not changed.



Why do we select features?

Simpler models are:

- ✓ Easier to understand.
- ✓ Faster.
- ✓ Less storage.
- ✓ Easier to maintain.

Uses of machine learning models

Insurance Claim

An insurance claim is

A formal request to an ins

payment based on the te

Insurance claims are re

ut to the insured

Insurance Claims

Fraud

Credit Risk

Marketing

Premium

Customer Churn

How do we select features?

Many feature selection algorithms

```
Many feature
selection
algorithms
```

Many feature “optimal” subsets

Many “optimal” feature subsets.

Python open-source - feature selection

Train In Data

First step

First: variable redundancy

Constant variables
Only 1 value per variable

Quasi – constant Variables
> 99% of observations show same value

Duplication
Same variable multiple times in the dataset

First: variable redundancy

Constant variables
Only 1 value per variable

Quasi – constant Variables
> 99% of observations show same value

Duplication
Same variable multiple times in the dataset

scikit learn

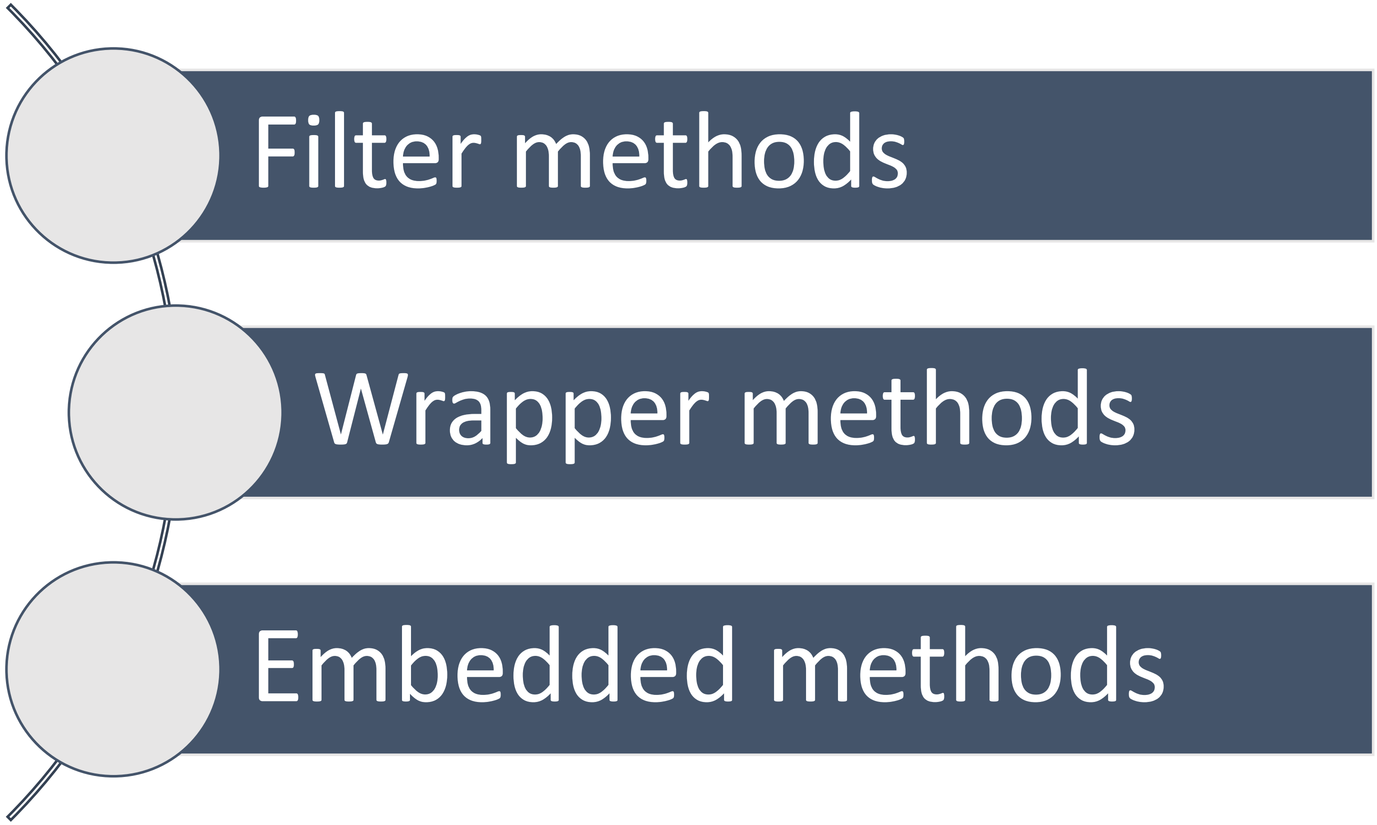
Feature-engine

Train In Data

Feature selection methods

Feature selection methods

Based on algorithms characteristics.



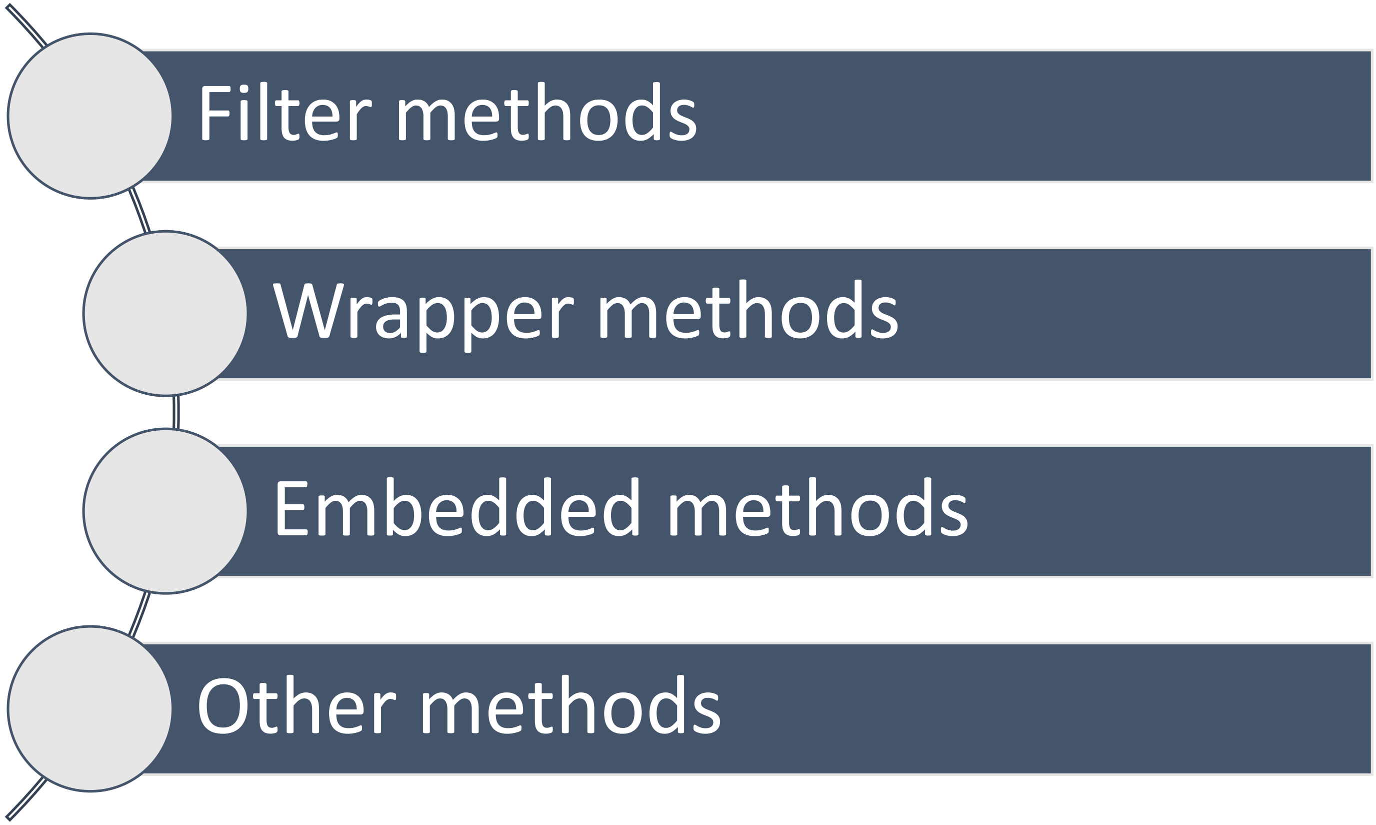
Filter methods

Wrapper methods

Embedded methods

Feature selection methods

Based on algorithms characteristics.



Filter methods

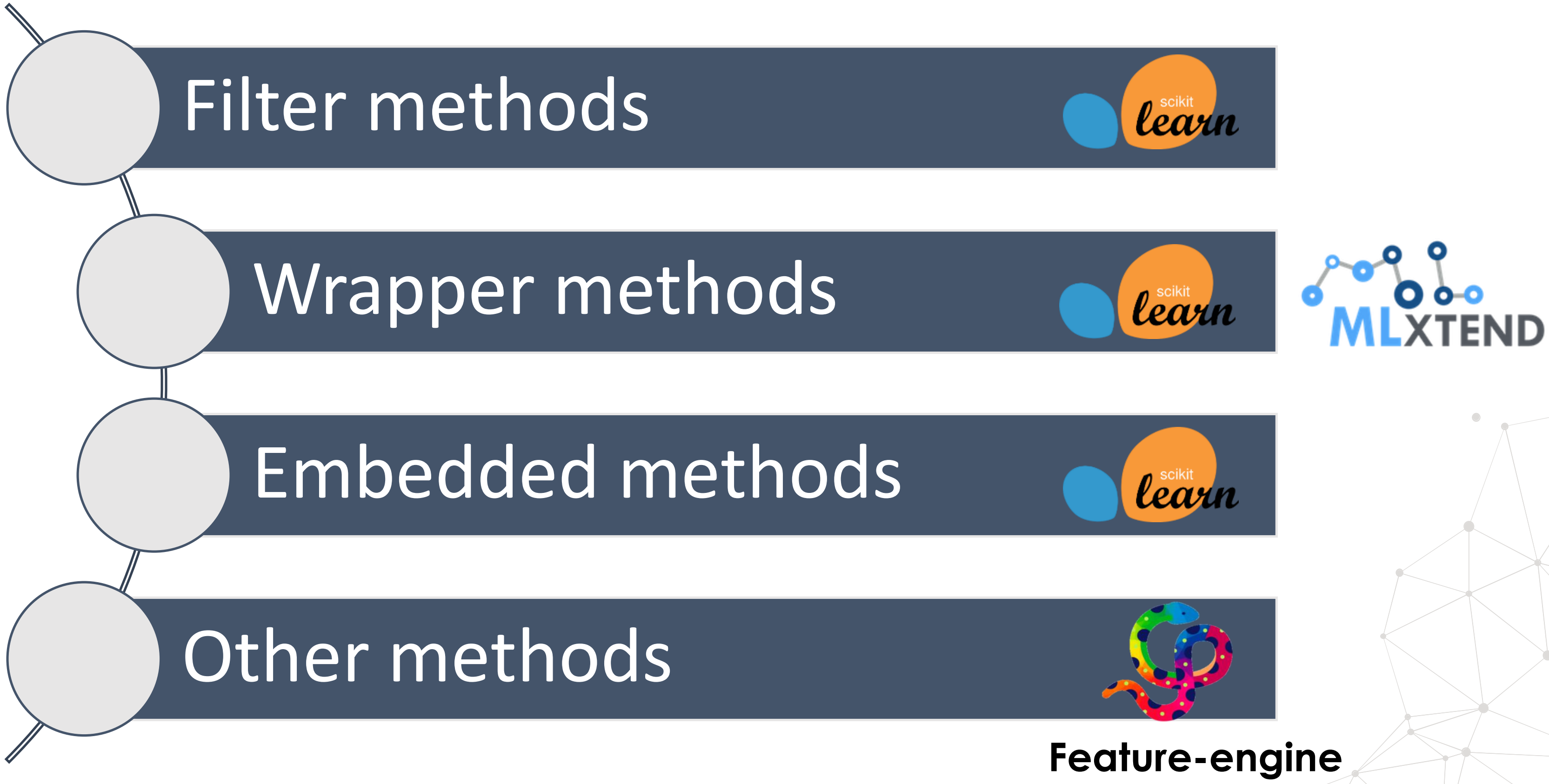
Wrapper methods

Embedded methods

Other methods

Feature selection methods

Based on algorithms characteristics.



Filter methods

Wrapper methods

Embedded methods

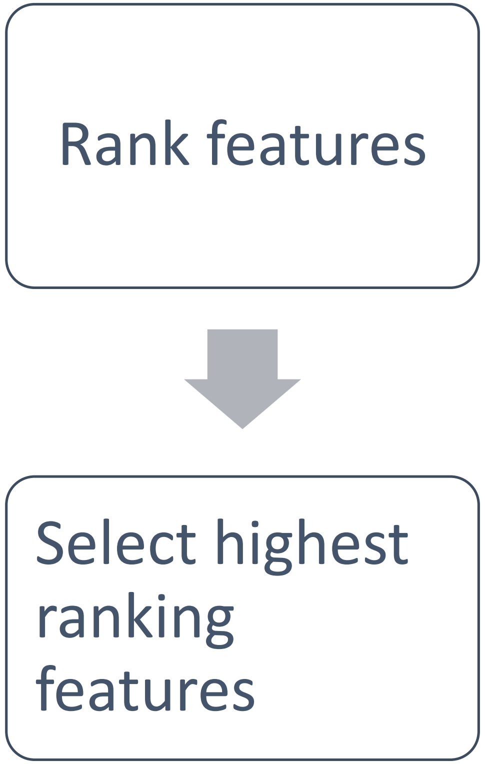
Other methods

Feature-engine

Filter methods

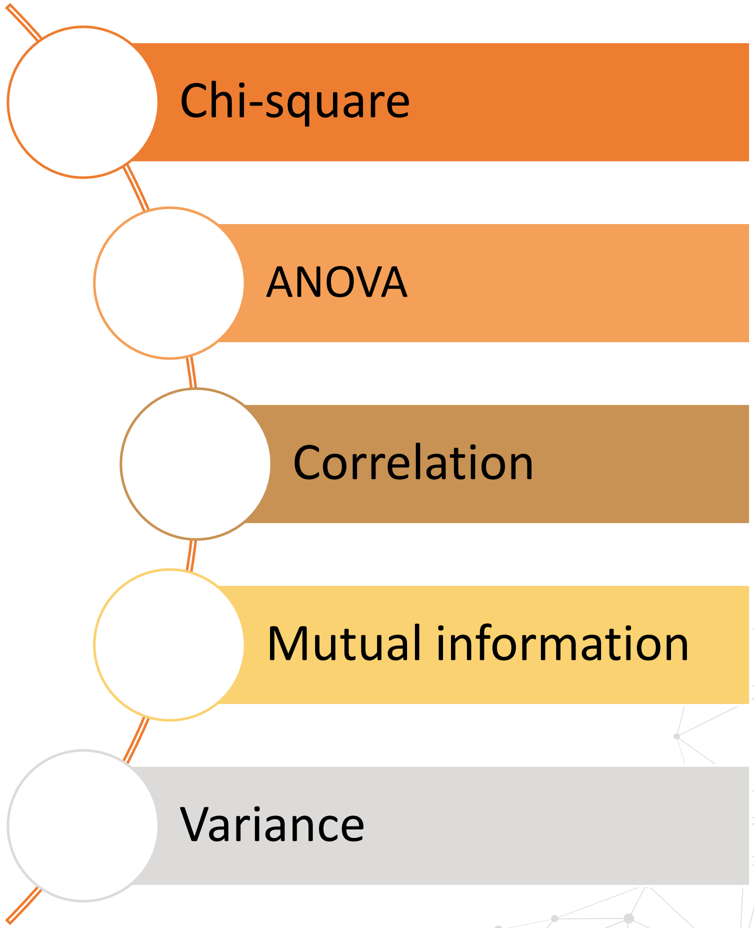
AKA: Ranking methods

Filter methods



Rank features

Select highest ranking features



Chi-square

ANOVA

Correlation

Mutual information

Variance

Statistical tests

Chi-square

- ✓ Categorical variables
- ✓ Categorical target

ANOVA

- ✓ Continuous variables
- ✓ Categorical target

Correlation

- ✓ Continuous variables
- ✓ Continous target

Null hypothesis: the populations are the same / no correlation.

Ranking criteria: p-value.

These tests make assumptions on the data.

Chi-square

$\chi 2$ = sum (Observed – expected)^2 / expected

Observed

	Female	Male
Died	120	60
Surived	92	30

Expected

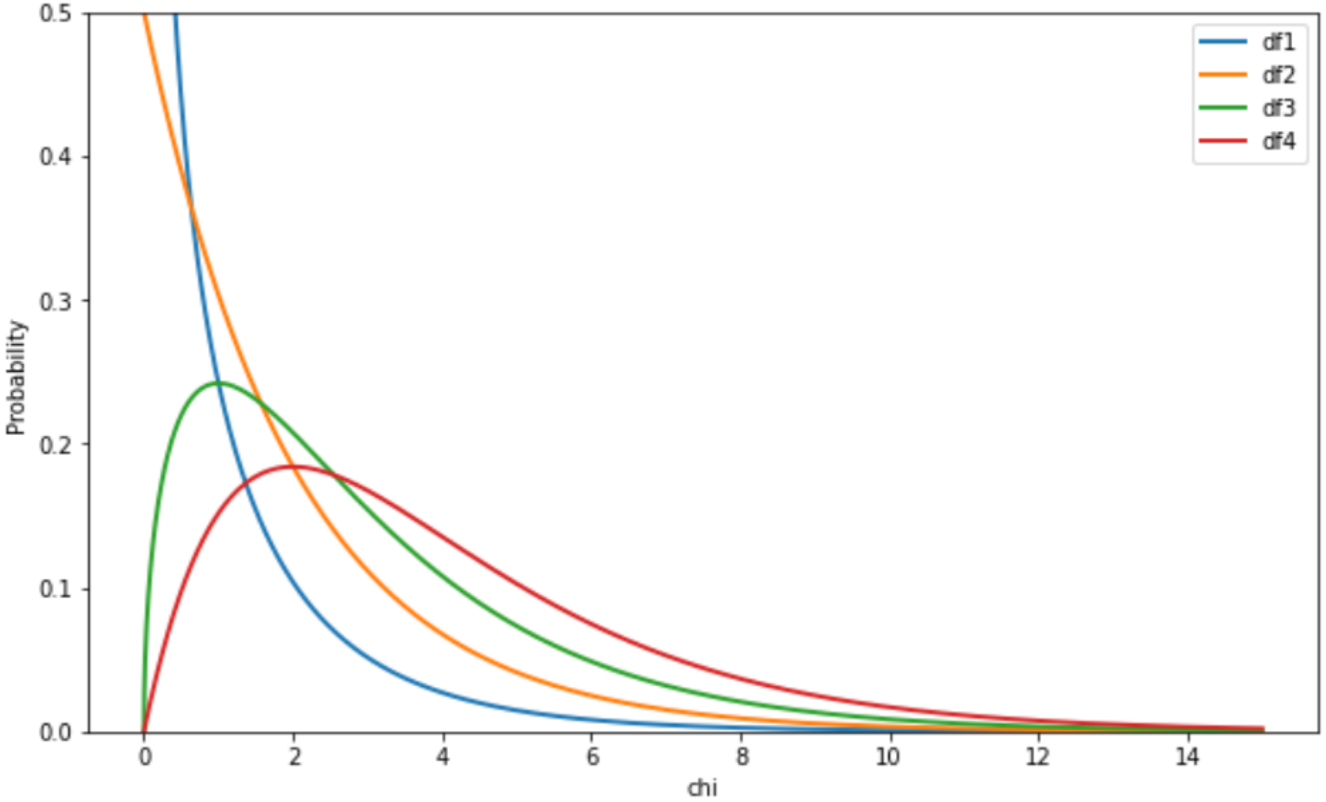
	Female	Male
Died	120	53
Surived	85	36

E = (Row x Column) / Total

Data consists of 200 women and 100 man

Chi-square

$\chi 2$ = sum (Observed – expected)^2 / expected

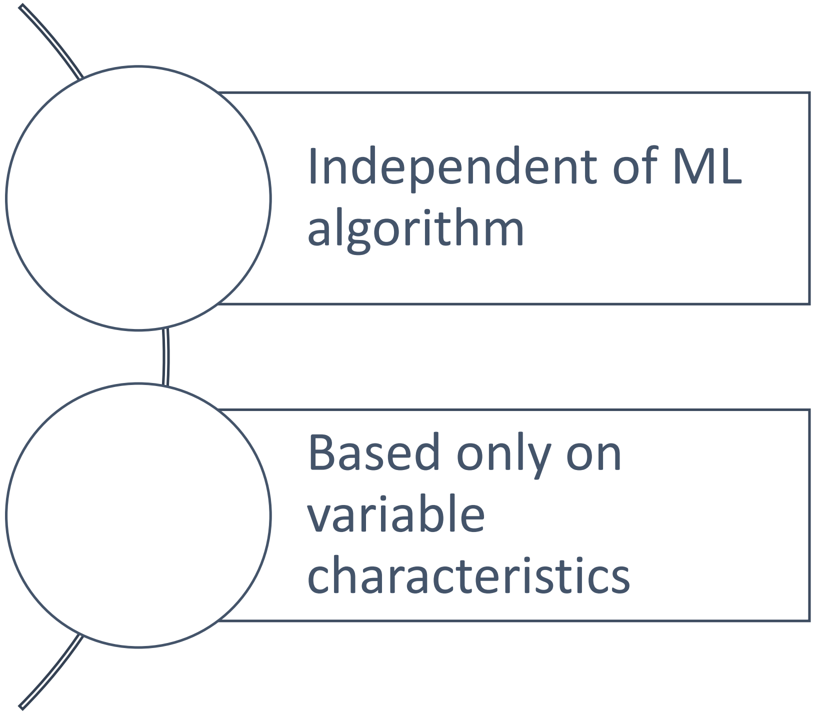


Chi-square: use scipy

Scikit-learn’s chi2 implementation is not suitable for categorical variables:

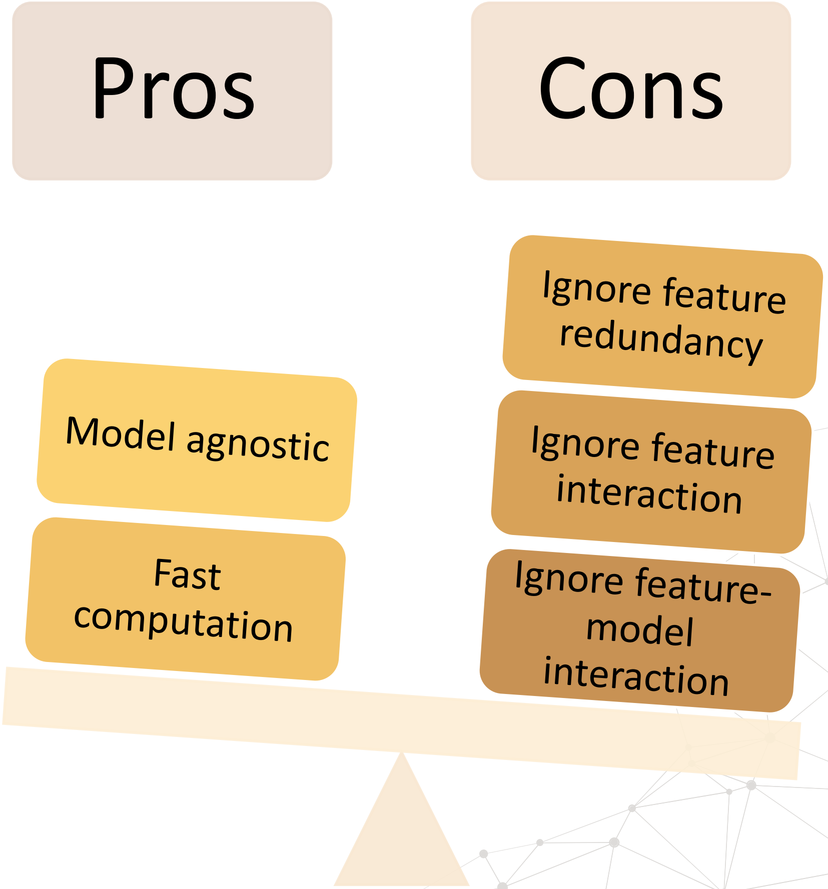
https://github.com/scikit-learn/scikit-learn/issues/21455

Filter methods - characteristics



Independent of ML algorithm

Based only on variable characteristics



Pros

Model agnostic

Fast computation

Cons

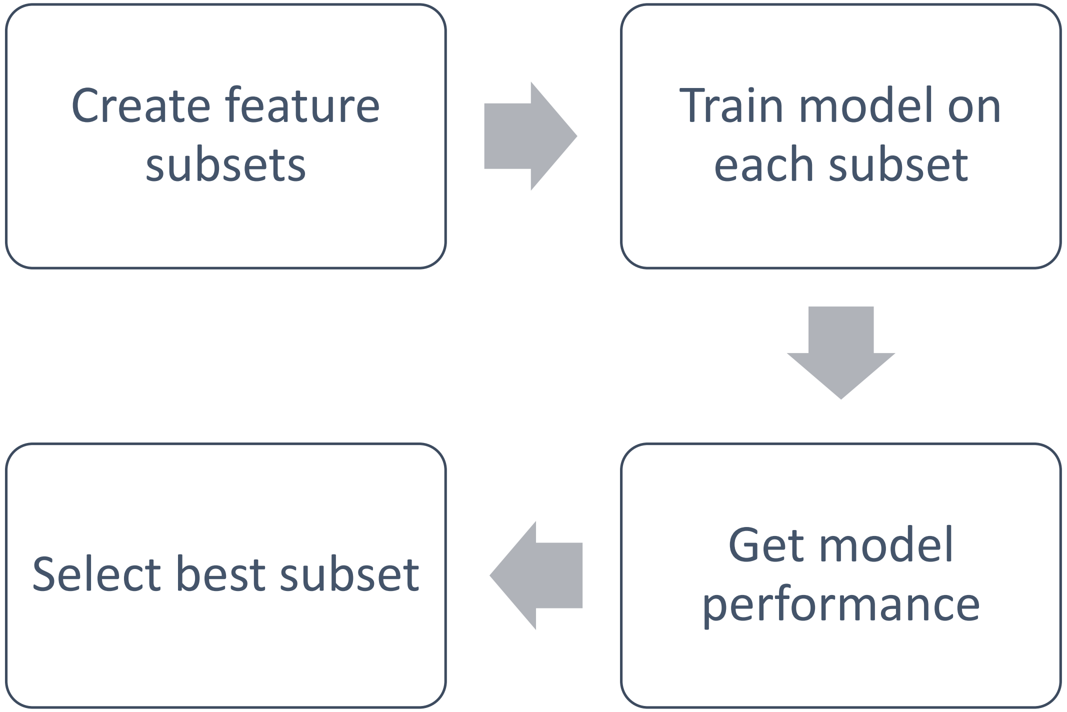
Ignore feature redundancy

Ignore feature interaction

Ignore feature-model interaction

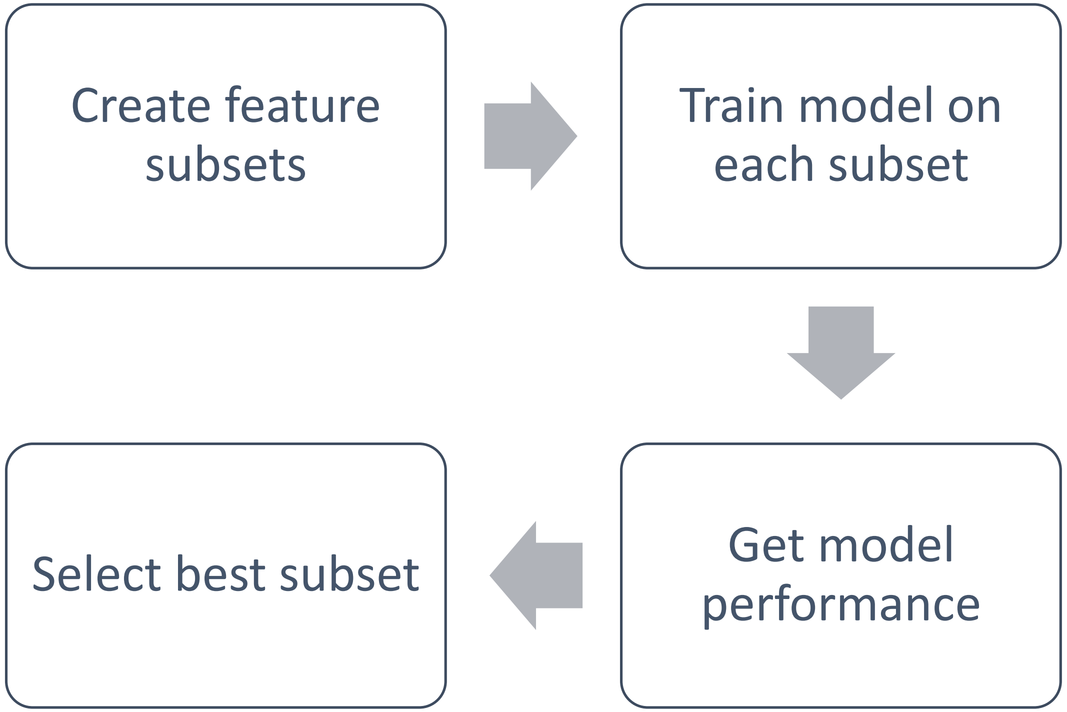
Wrapper methods

Wrapper methods



- Exhaustive search
- Forward search
- Backward search

Wrapper methods

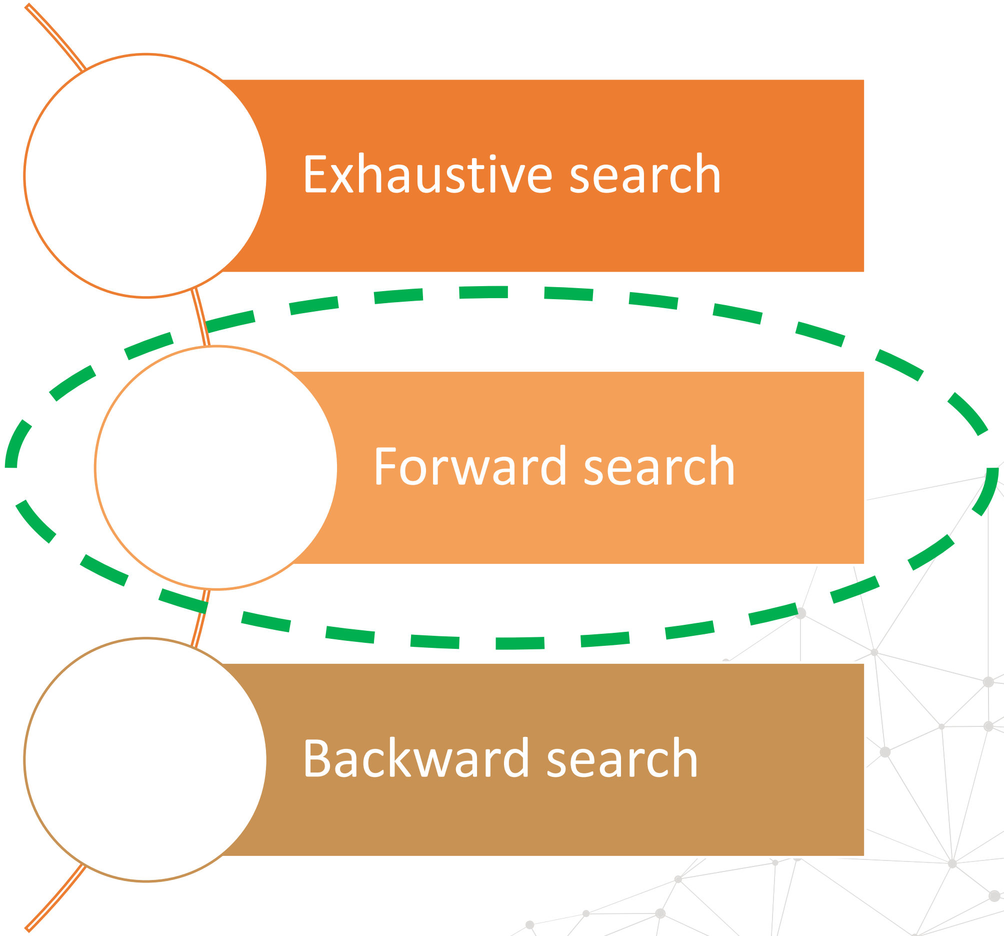


Create feature subsets

Train model on each subset

Get model performance

Select best subset

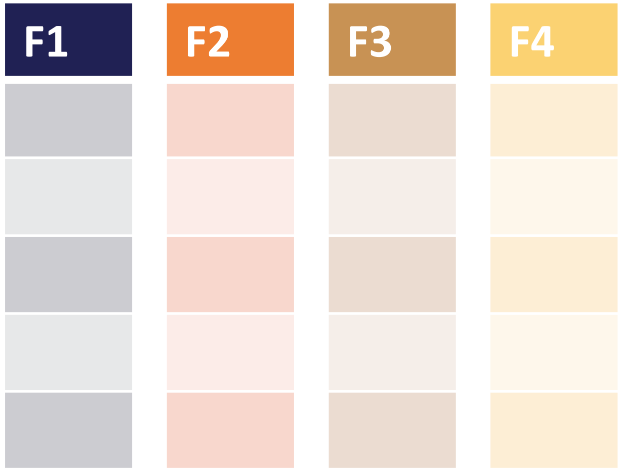


Exhaustive search

Forward search

Backward search

Forward feature selection



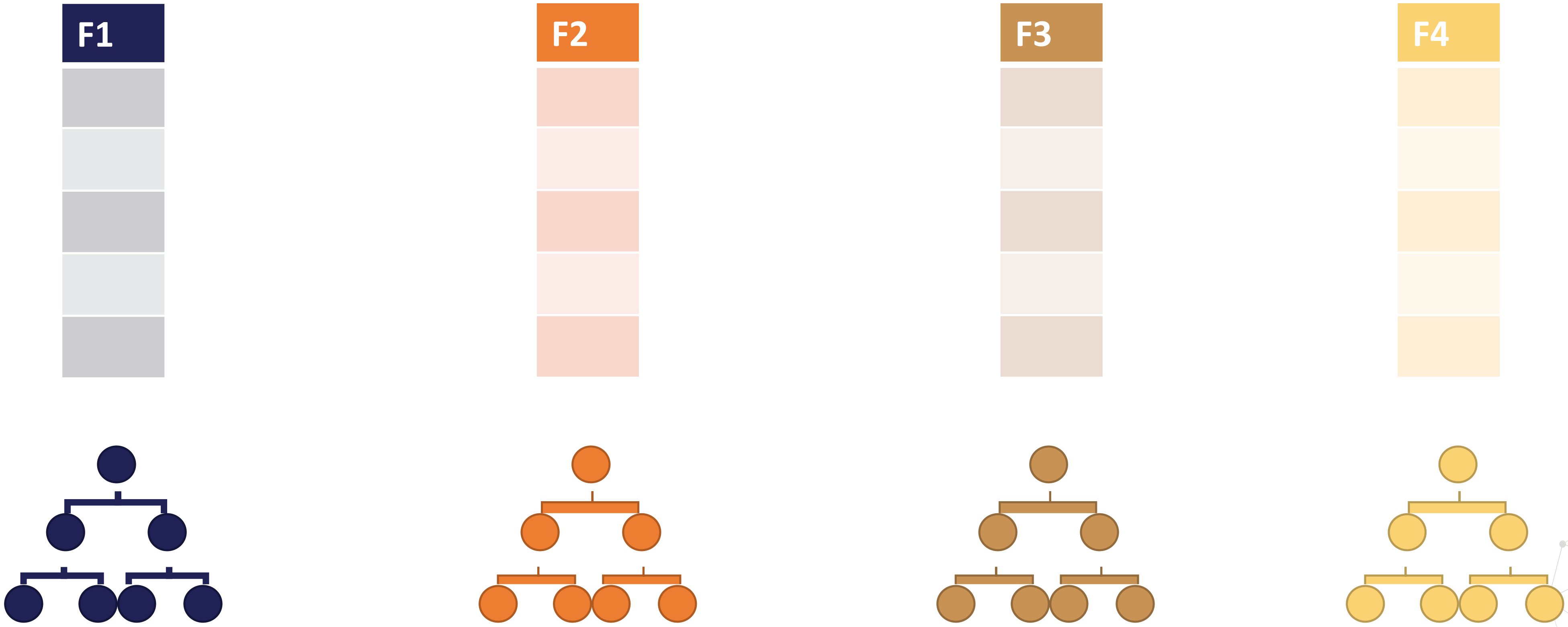
F1

F2

F3

F4

Forward feature selection



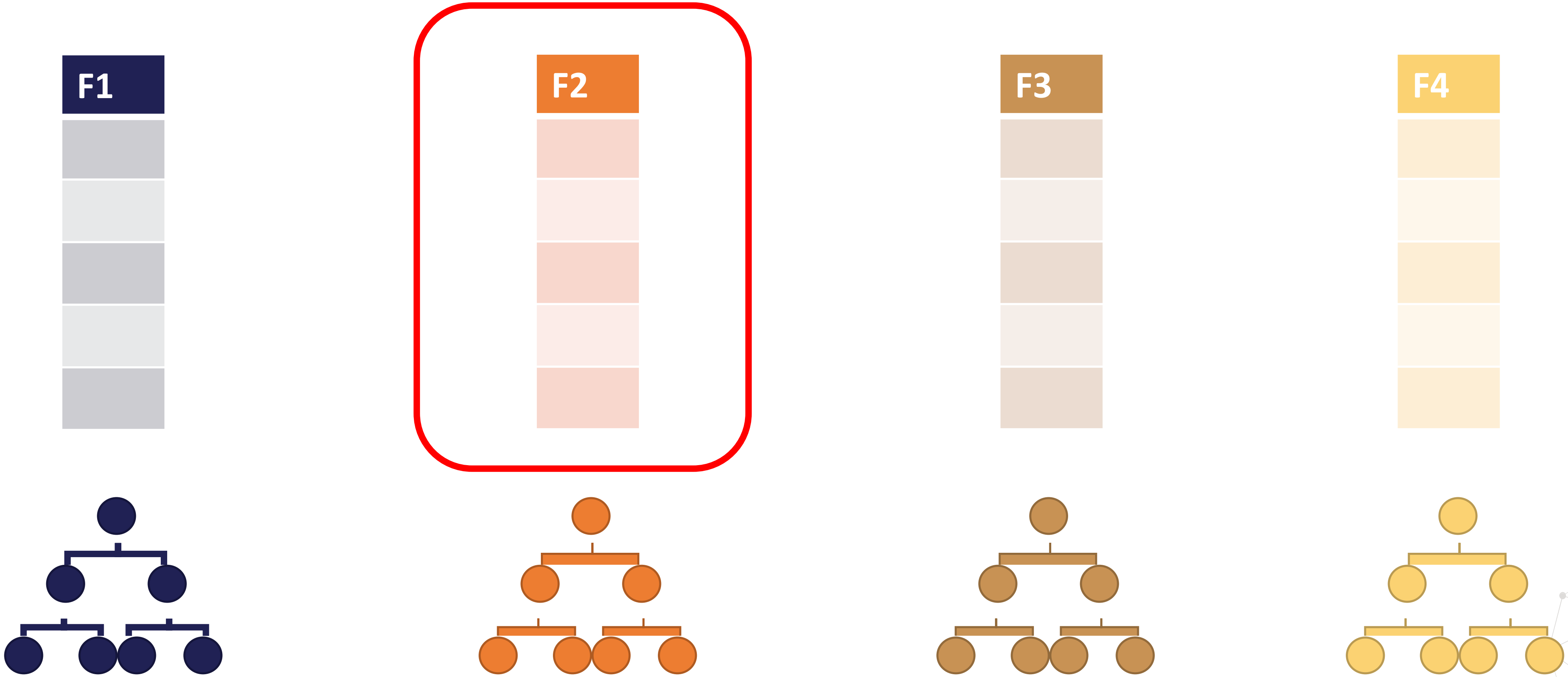
Roc-auc
0.62

Roc-auc
0.72

Roc-auc
0.65

Roc-auc
0.59

Forward feature selection



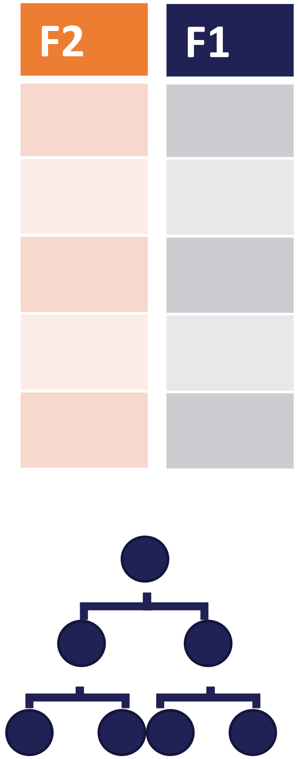
Roc-auc 0.62

Roc-auc 0.72

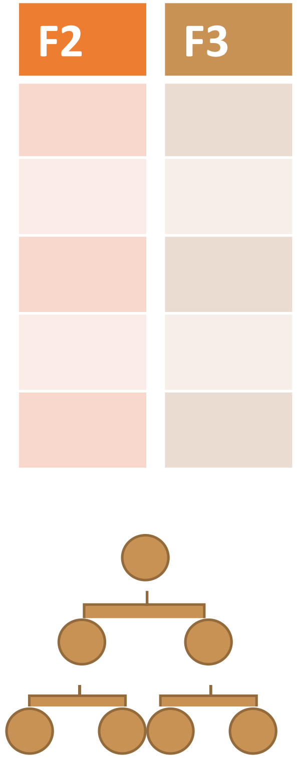
Roc-auc 0.65

Roc-auc 0.59

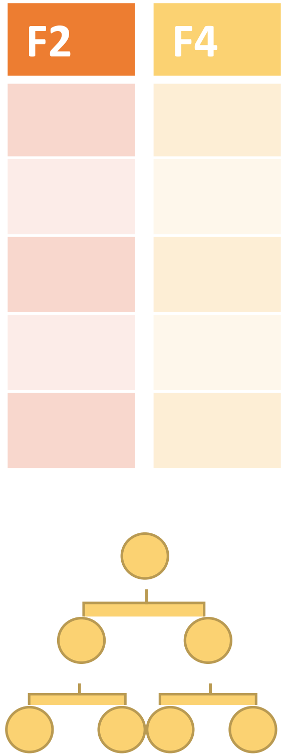
Forward feature selection



Roc-auc
0.74

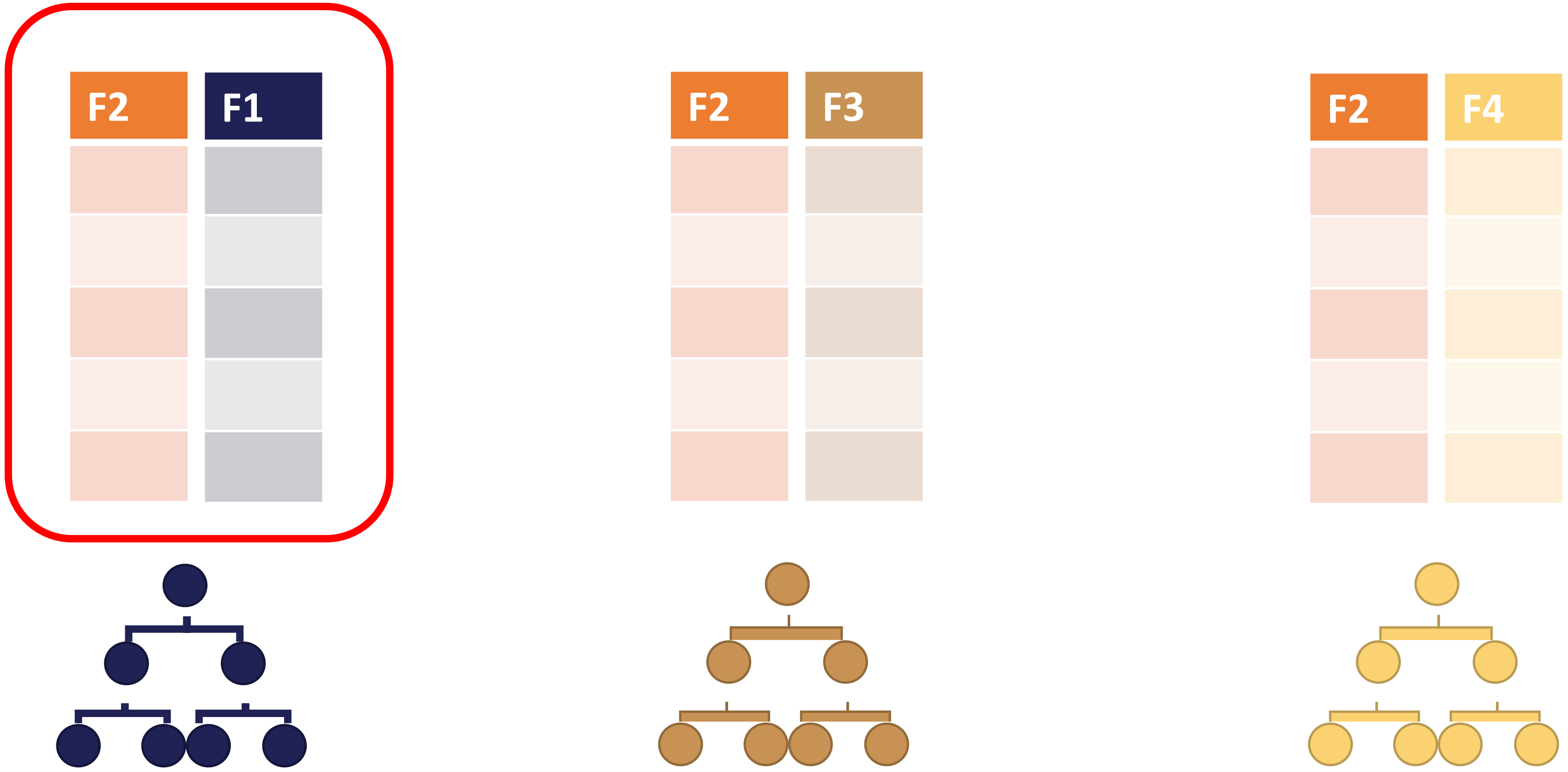


Roc-auc
0.72



Roc-auc
0.72

Forward feature selection

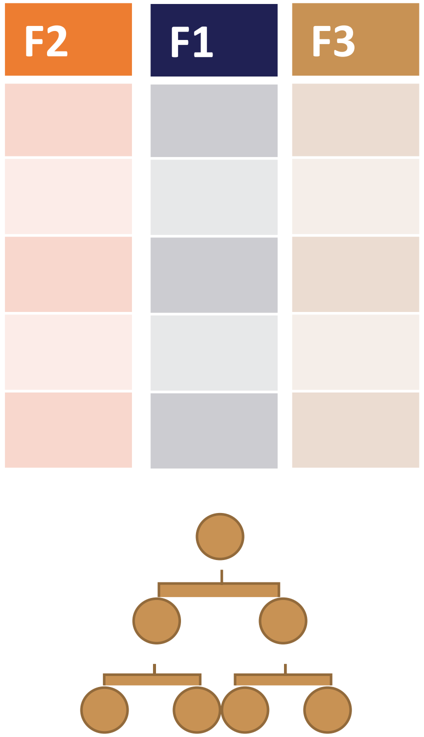


Roc-auc
0.74

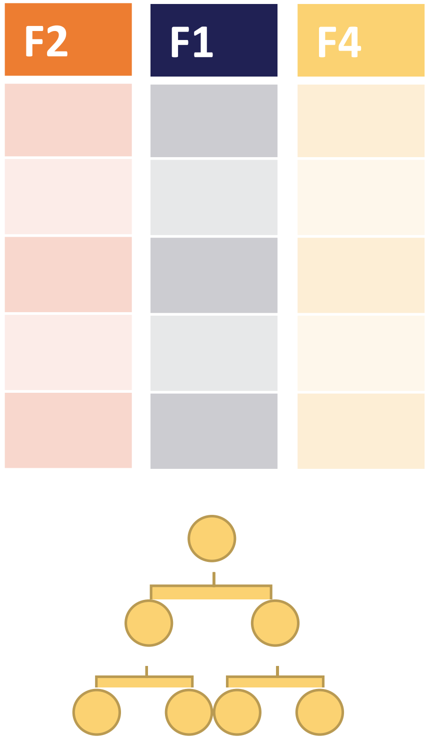
Roc-auc
0.72

Roc-auc
0.72

Forward feature selection

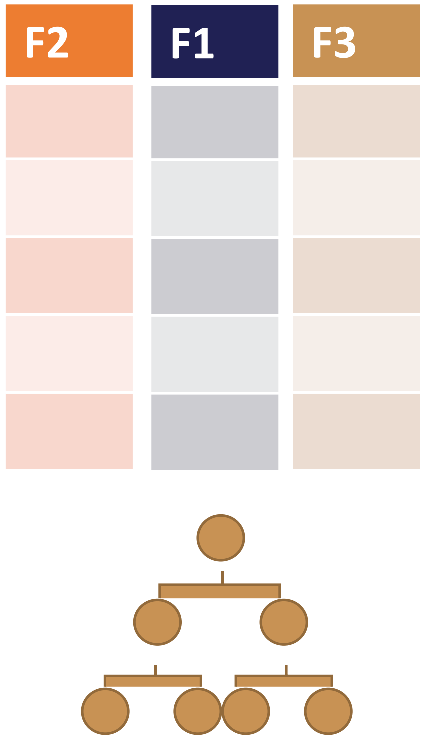


Roc-auc
0.75

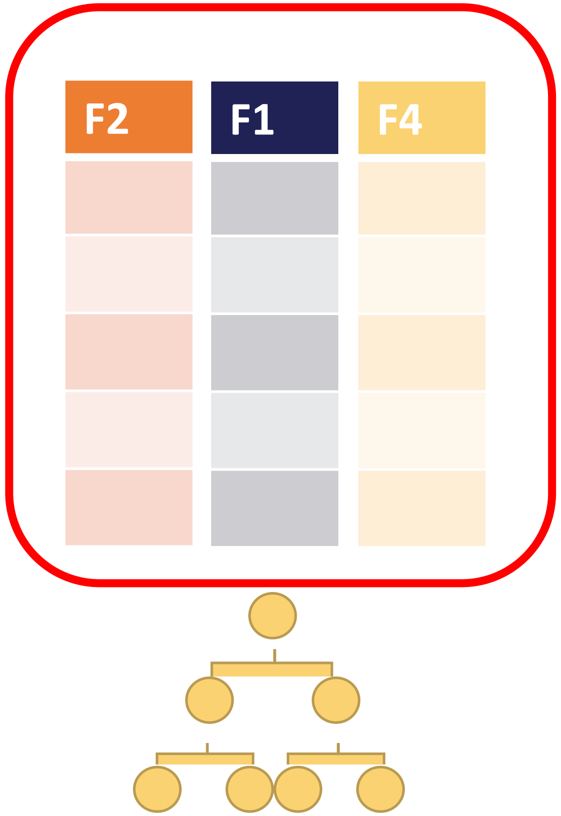


Roc-auc
0.76

Forward feature selection

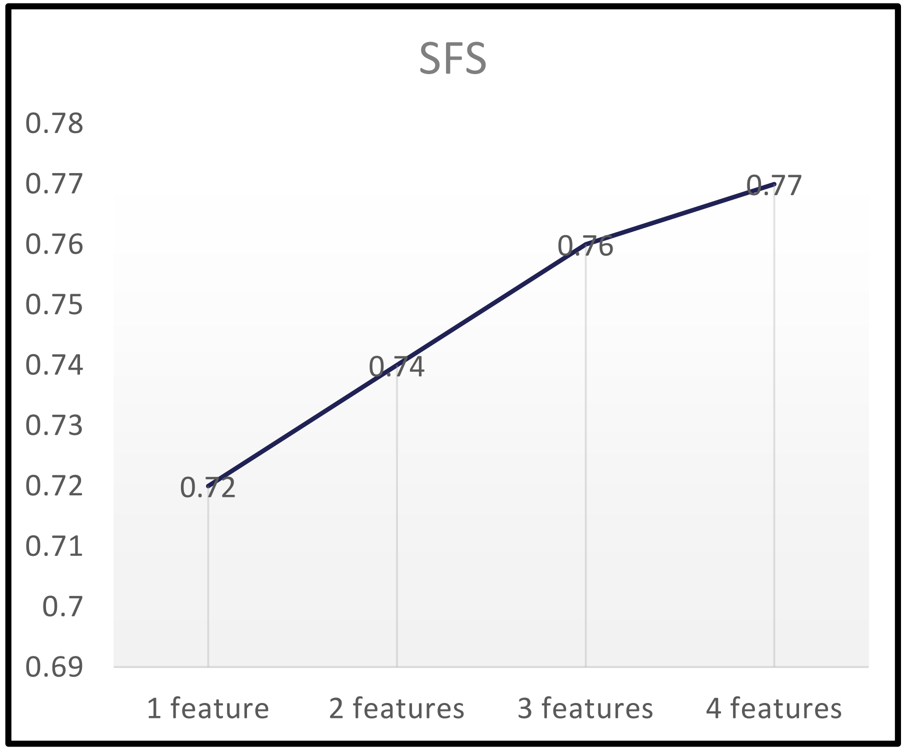


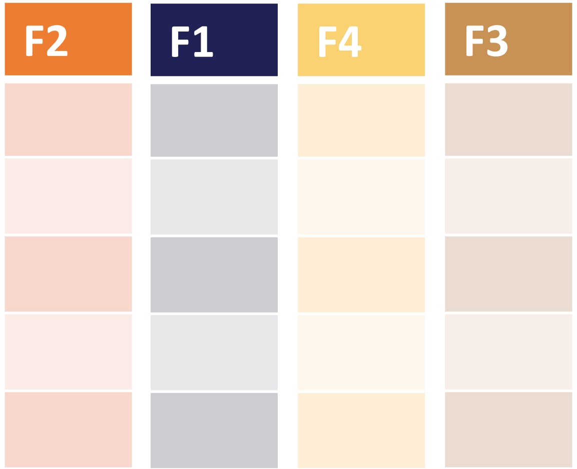
Roc-auc
0.75

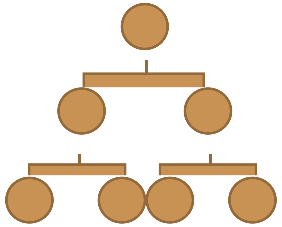


Roc-auc
0.76

Forward feature selection







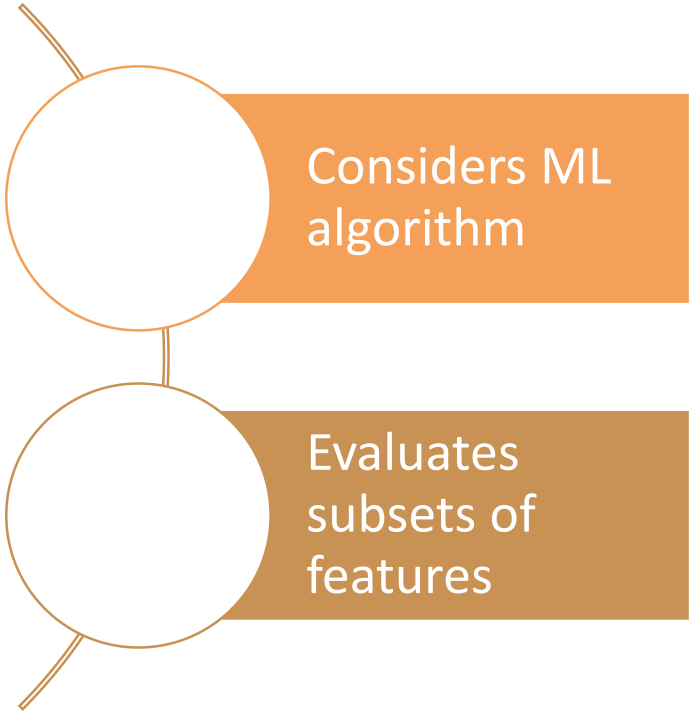
Roc-auc
0.77

When to stop the search

- **Ideal**: When performance does not increase beyond a threshold
  - ✓ Threshold to be defined by the user

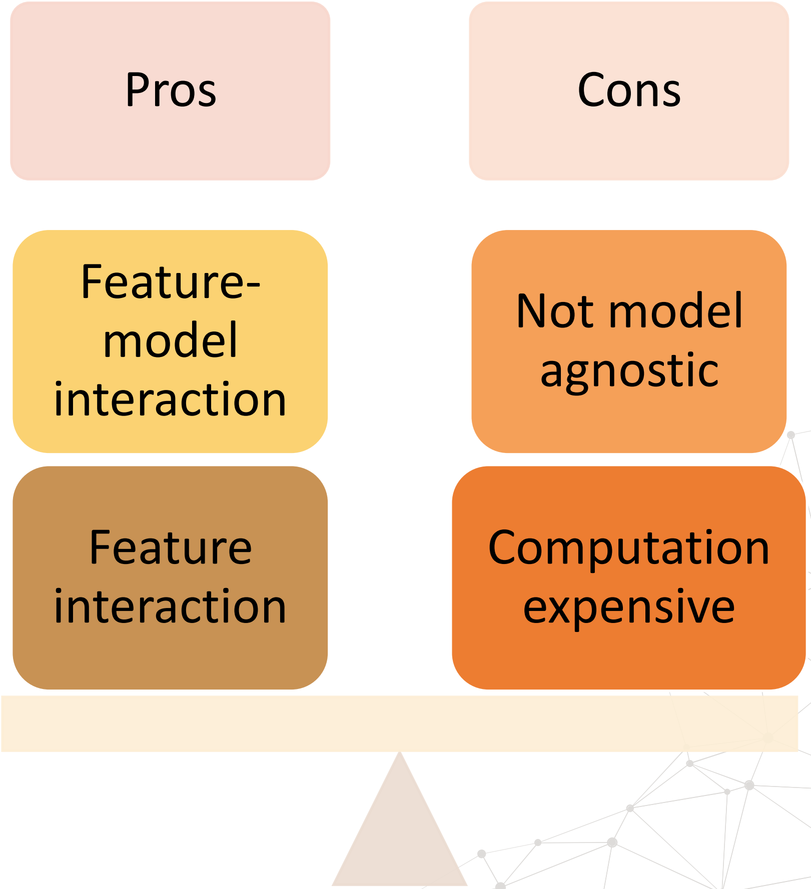
- **MLXtend implementation**: when certain number of features is reached
  - ✓ Number of features defined by the user

Wrapper methods - characteristics



Considers ML algorithm

Evaluates subsets of features



Pros

Cons

Feature-model interaction

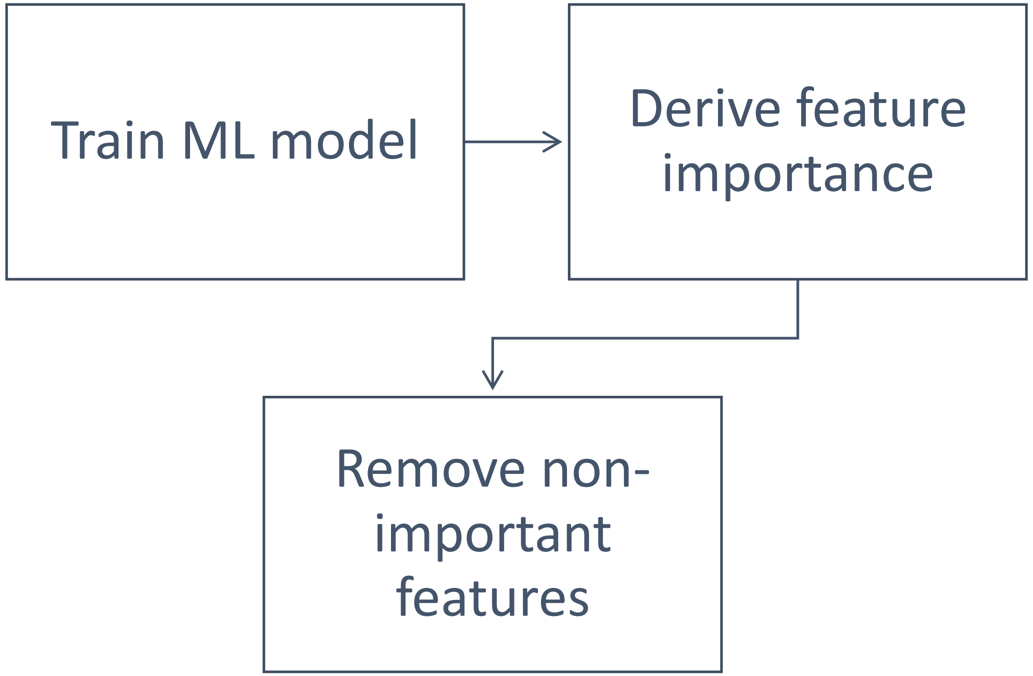
Not model agnostic

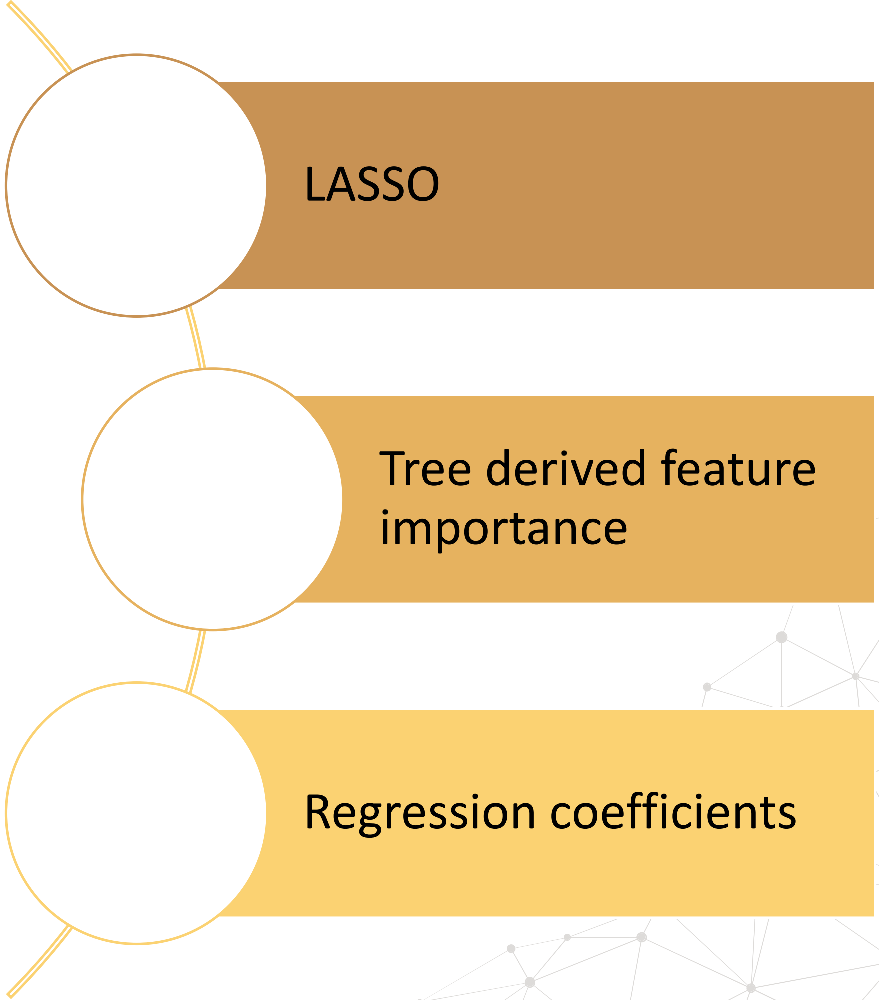
Feature interaction

Computation expensive

Embedded methods

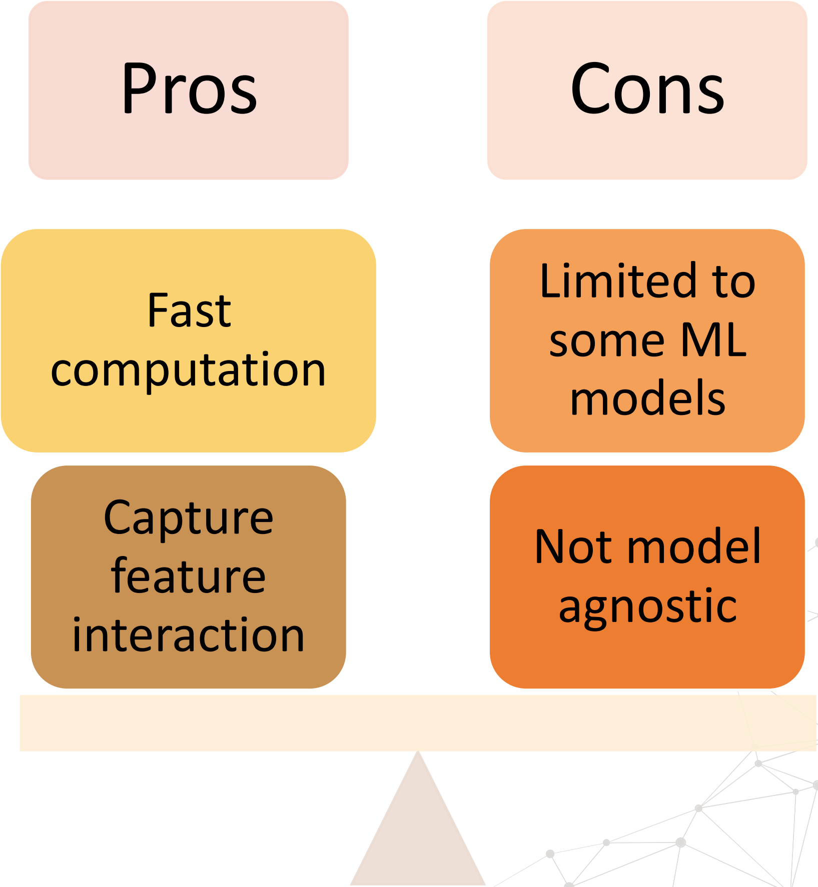
Embedded methods





Embedded methods - characteristics

Feature selection during model training



Pros

Cons

Fast computation

Capture feature interaction

Limited to some ML models

Not model agnostic

Other methods

Other feature selection methods

Other methods

- □ Feature permutation / shuffling
- □ Probe features
- □ MRMR
- □ RFE / RFA
- □ CBFS

Statistics

- □ Population Stability Index
- □ Information value

Other feature selection methods

Other methods

☑ Feature permutation / shuffling

□ Probe features

□ MRMR

☑ RFE / RFA

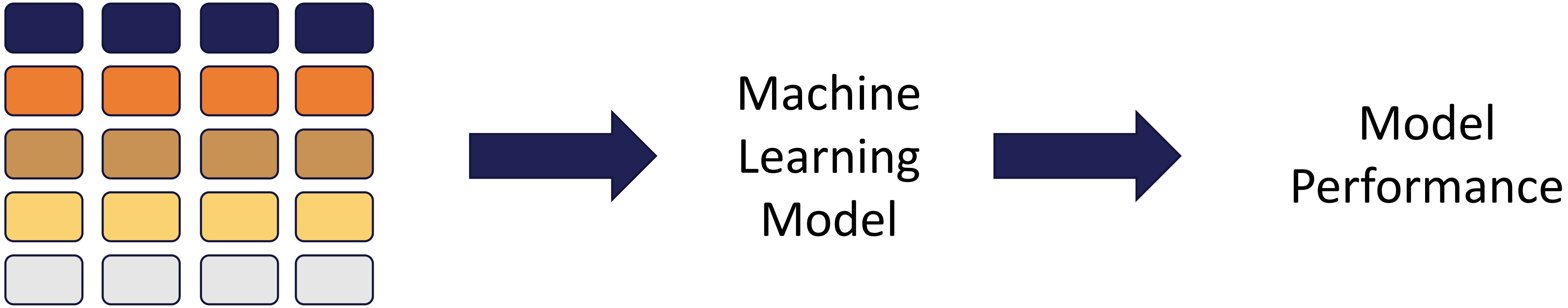
□ CBFS

Statistics

☑ Population Stability Index

□ Information value

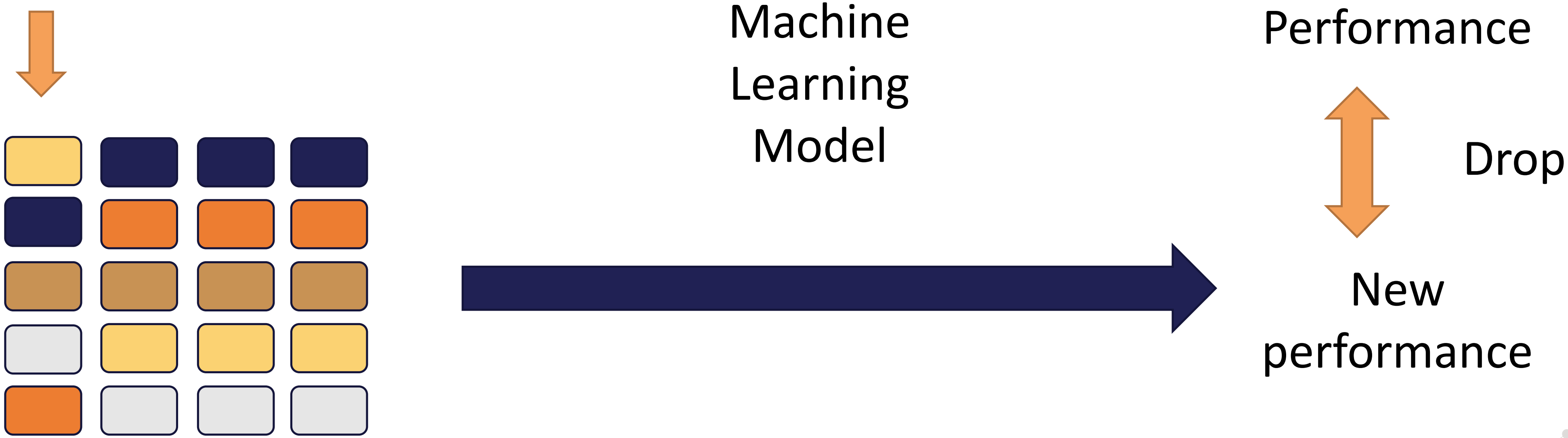
Feature Shuffling



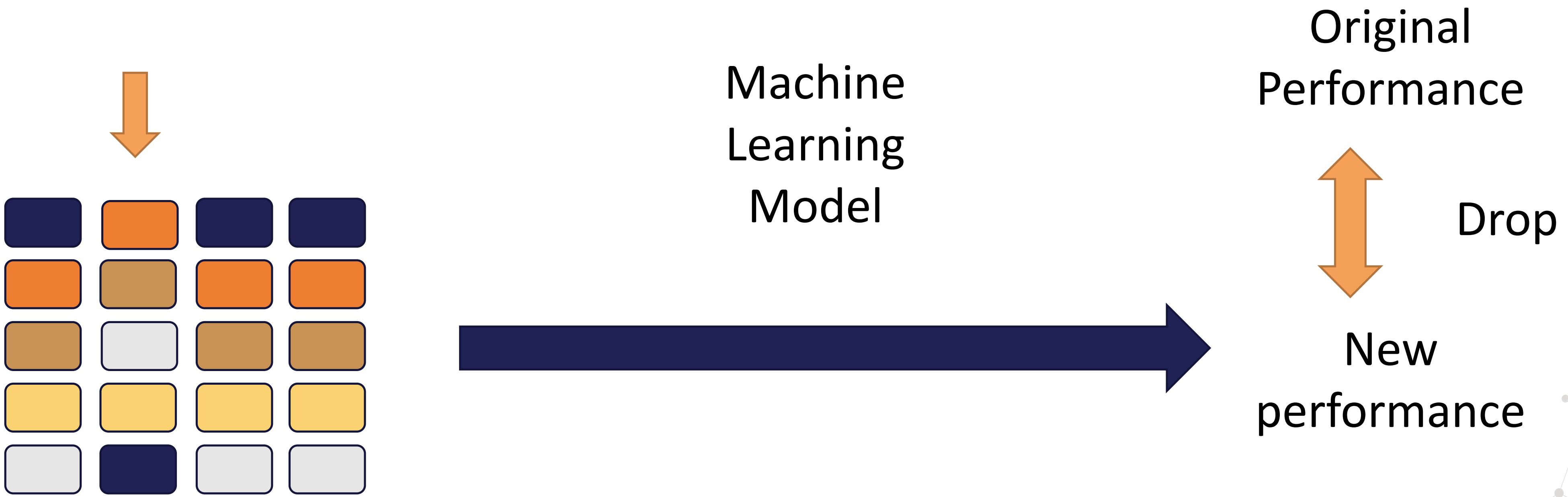
Machine Learning Model

Model Performance

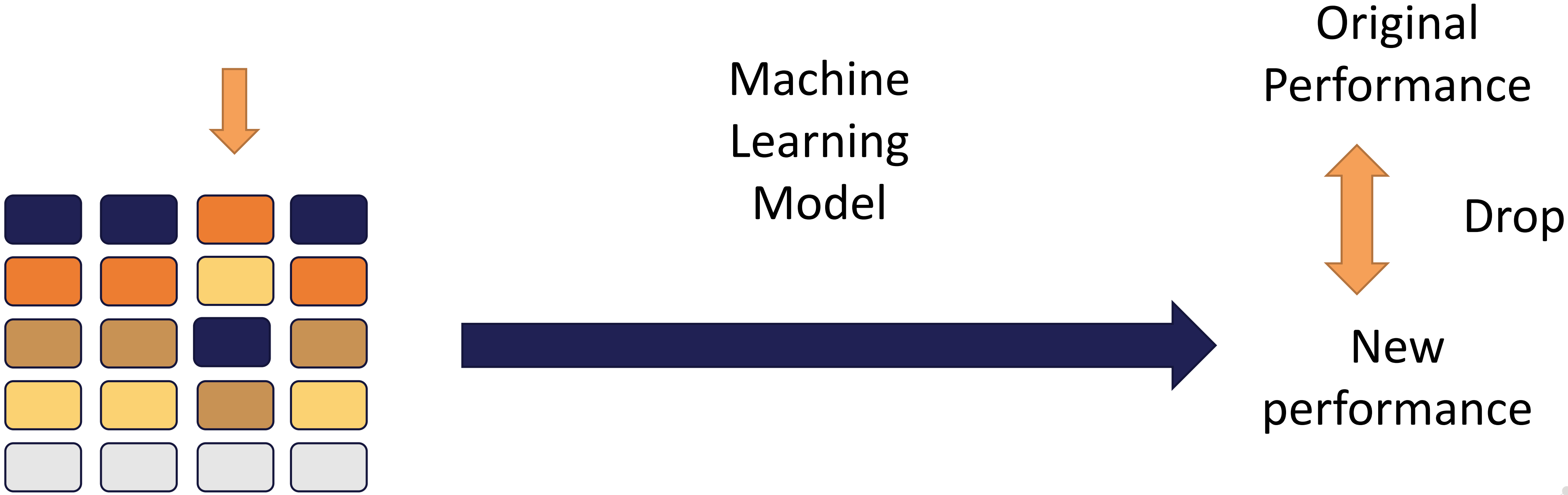
Feature Shuffling



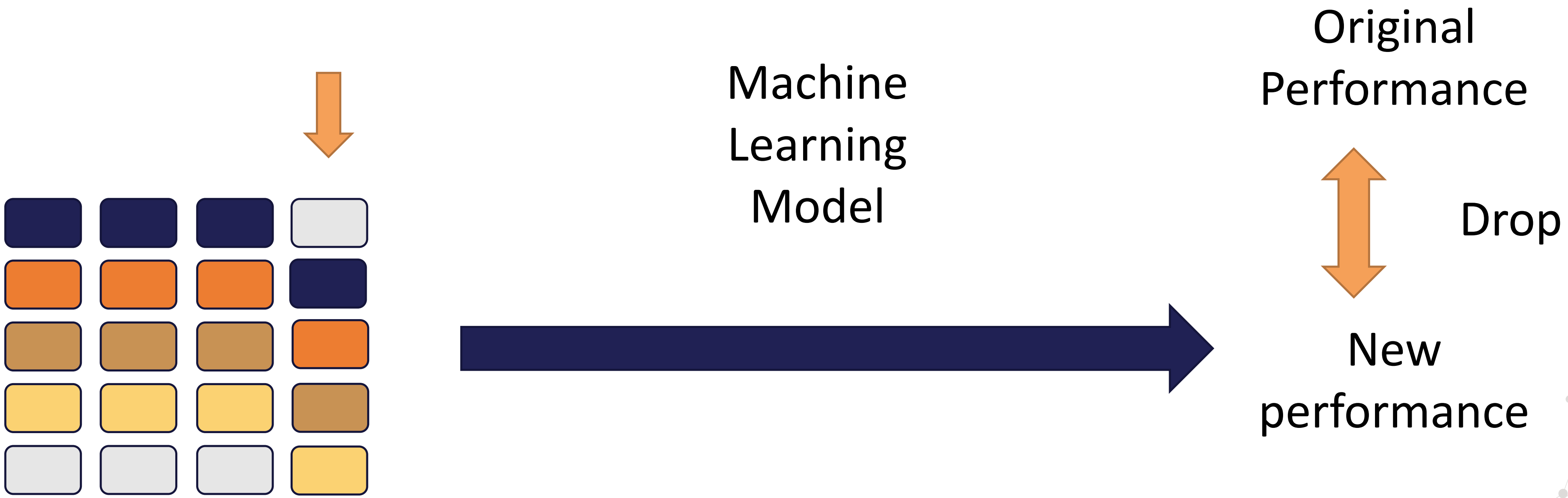
Feature Shuffling



Feature Shuffling



Feature Shuffling



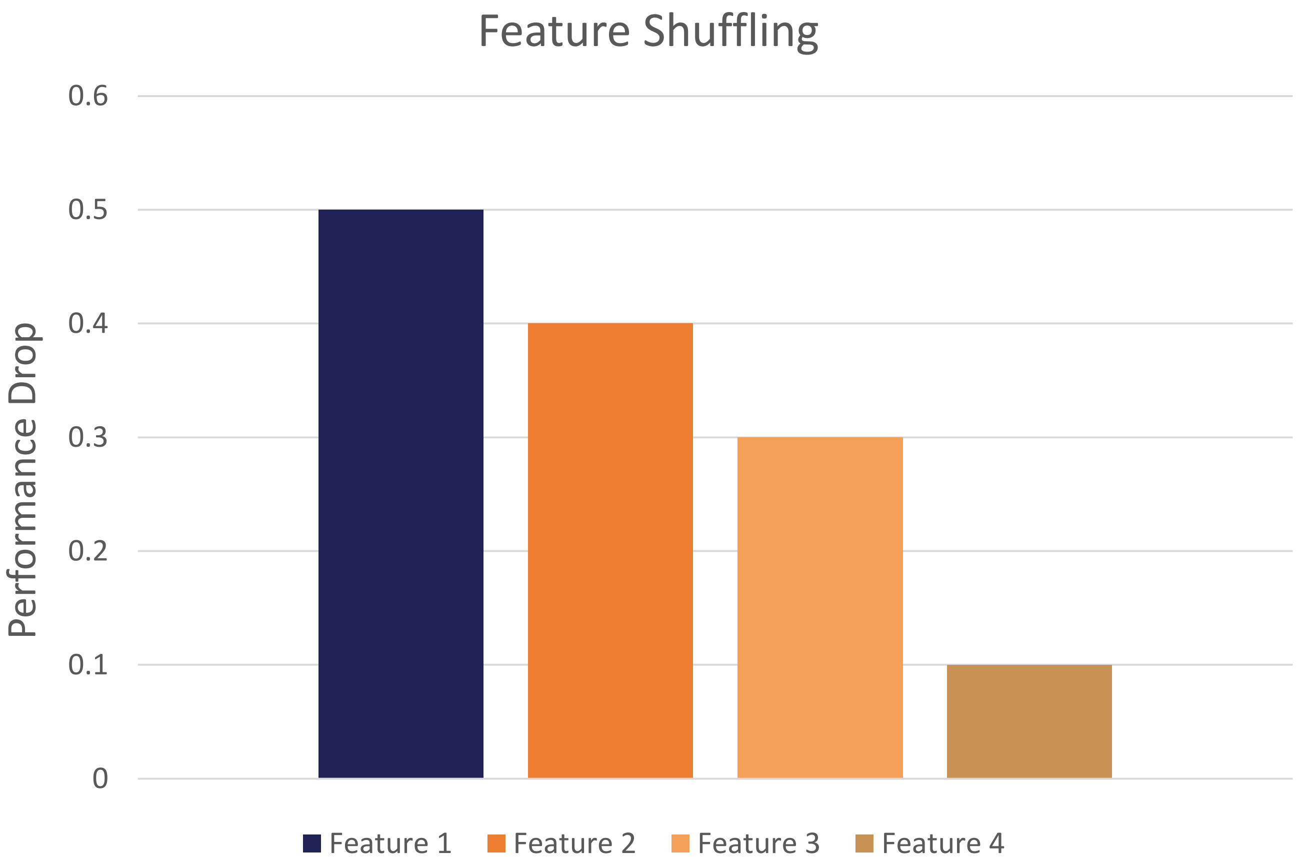
Machine Learning Model

Original Performance

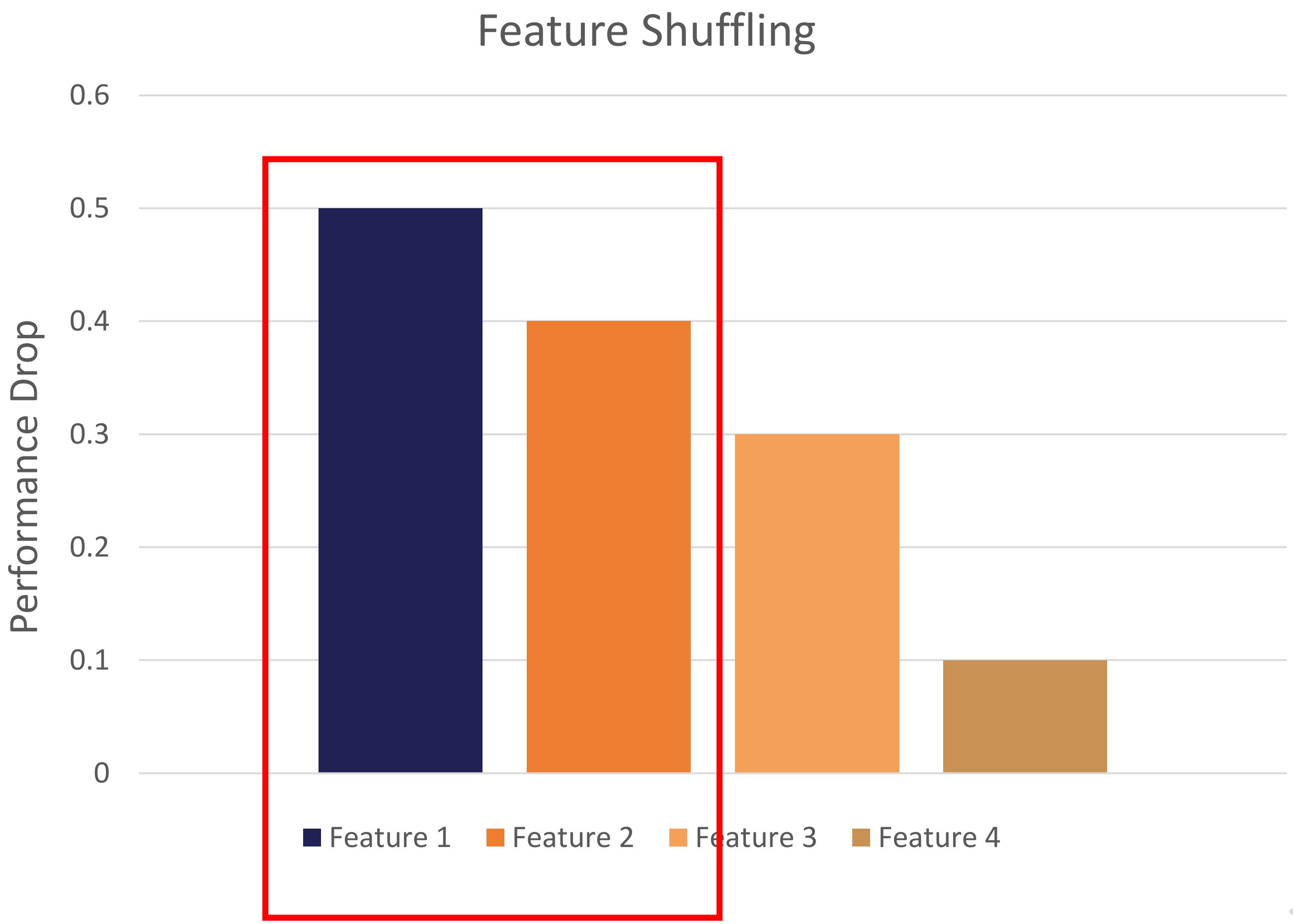
Drop

New performance

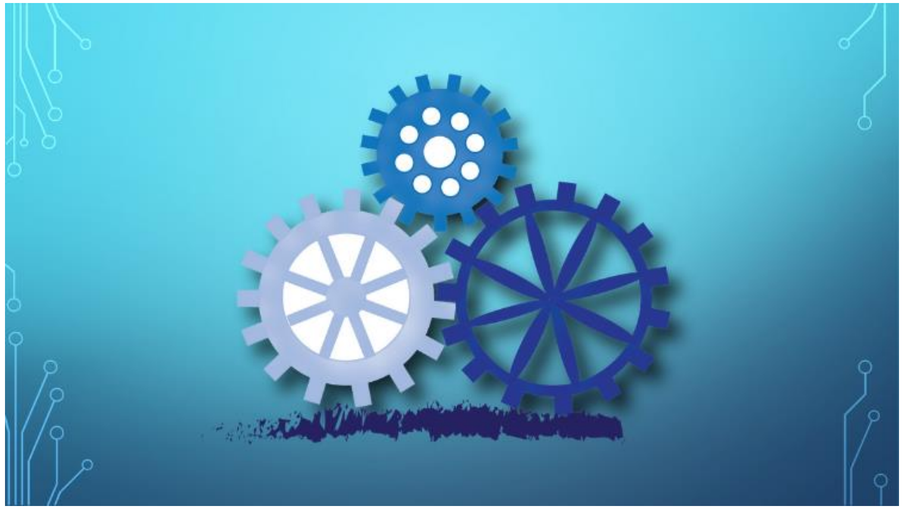
Feature Shuffling



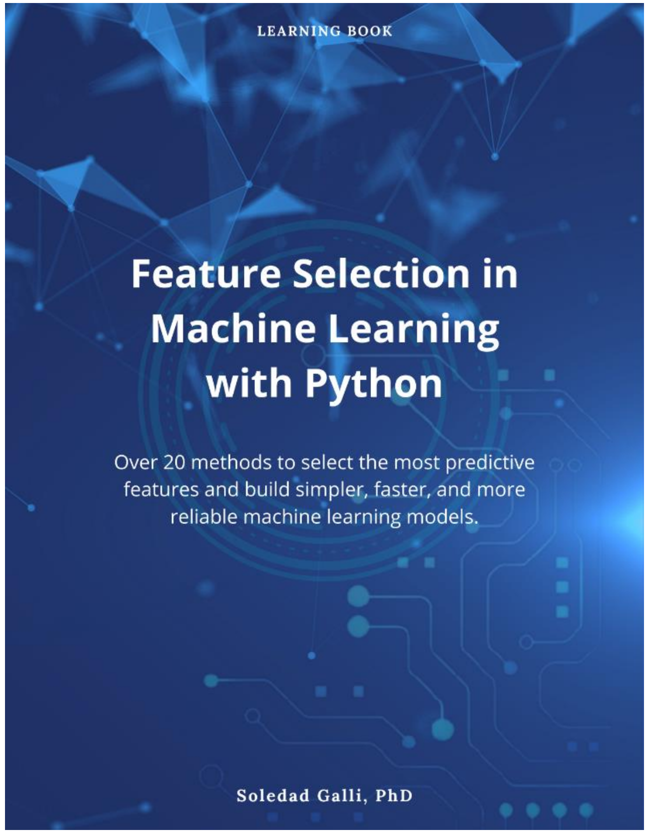
Feature Shuffling



Thank you



https://www.trainindata.com/p/feature-selection-for-machine-learning



LEARNING BOOK

Feature Selection in Machine Learning with Python

Over 20 methods to select the most predictive features and build simpler, faster, and more reliable machine learning models.

Soledad Galli, PhD

https://leanpub.com/feature-selection-in-machine-learning/