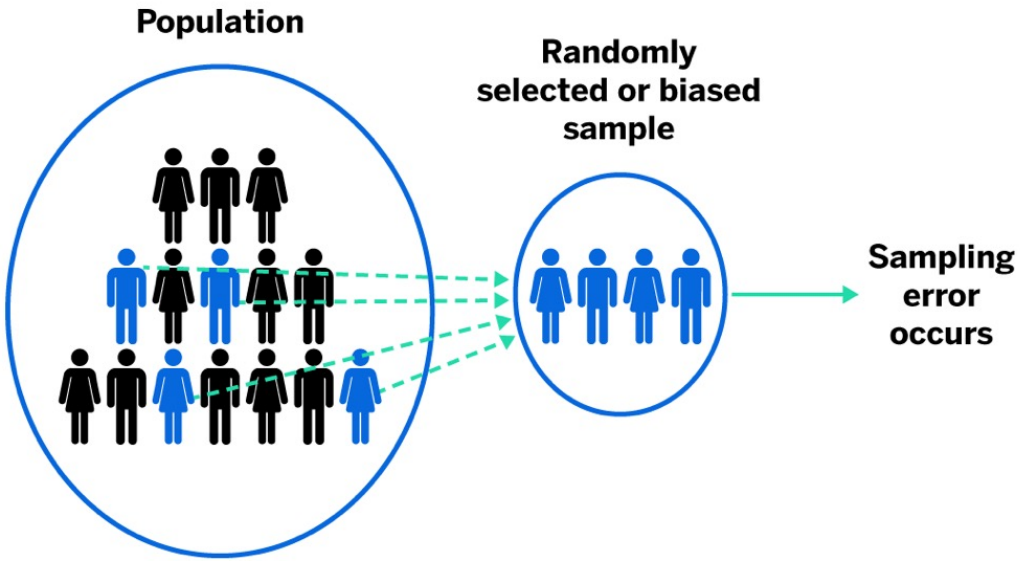
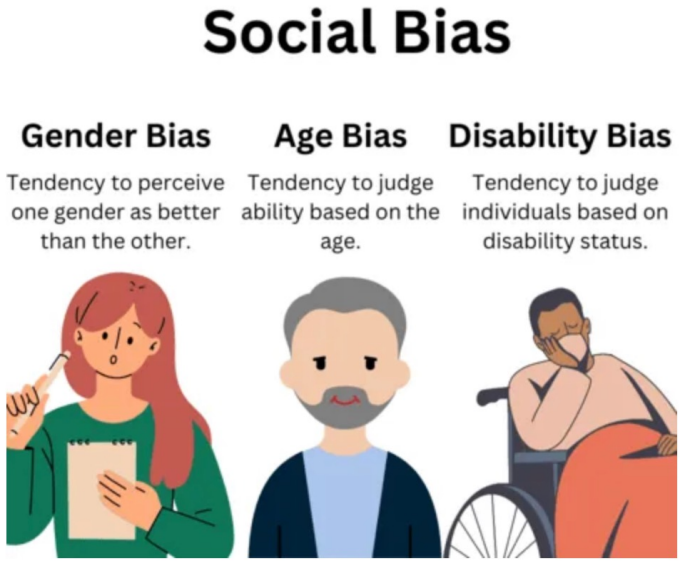
Data Bias

Sampling Bias: Overrepresentation or underrepresentation in datasets.

Label Bias: Mislabeling due to subjective interpretations.

Historical Bias: Embedded societal inequalities.





Sampling Bias

Lable Bias

Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%	100%	98.3%	20.8%
FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
IBM	88.0%	65.3%	99.7%	92.9%	34.4%

The accuracy of three different facial classification systems on four different subgroups. Table sourced from the Gender Shades overview website.

Studies have shown that facial recognition technologies from companies like Microsoft and IBM have higher error rates for darker-skinned individuals, particularly women.

Transparency???

AI in the News

Millions of black people affected by racial bias in health-care algorithms

Artificial Intelligence has a gender bias problem – just ask Siri

GM's Cruise Recalls Self-Driving Software Involved in June Crash

Microsoft Plans to Eliminate Face Analysis Tools in Push for 'Responsible A.I.'

The New Chatbots Could Change the World. Can You Trust Them?

Tesla 'full self-driving' triggered an eight-car crash, a driver tells police

Minority Patients Often Left Behind By Health AI

Many Facial-Recognition Systems Are Biased, Says U.S. Study

Risks Rise As Robotic Surgery Goes Mainstream

Gemini image generation got it wrong. We'll do better.

Feb 23, 2024
2 min read

We recently made the decision to pause Gemini's image generation of people while we work on improving the accuracy of its responses. Here is more about how this happened and what we're doing to fix it.

Prabhakar Raghavan
Senior Vice President

Share

Google's AI image generator, part of its Gemini suite, faced criticism for producing racially diverse depictions of historical figures, such as Vikings and Nazi soldiers, which some users found historically inaccurate.

Ongoing issue…

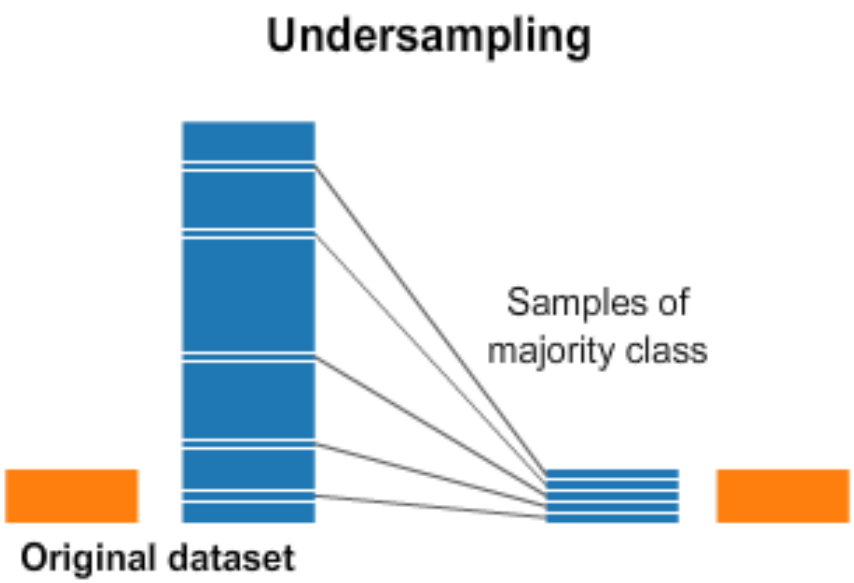
Impacts of Bias

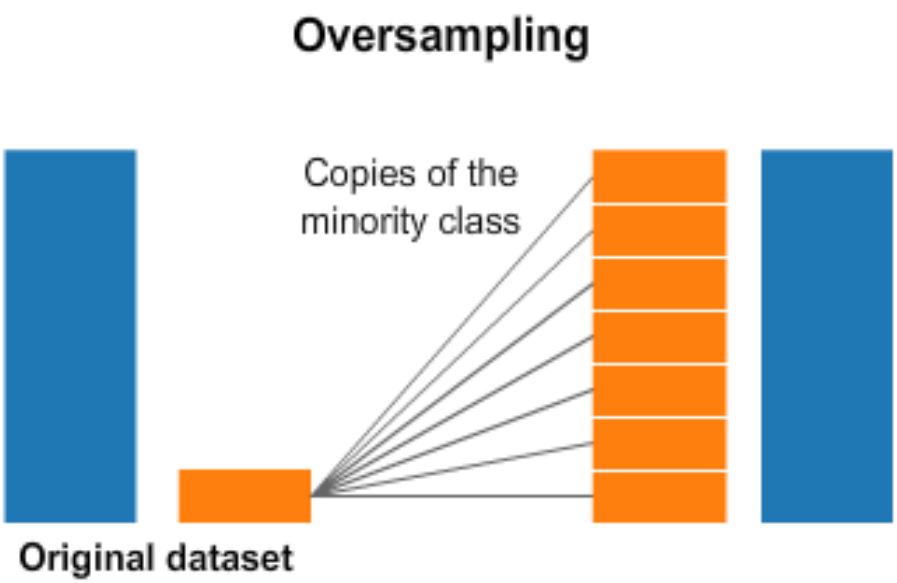
Ethical Issues: Discrimination in decisions (e.g., hiring, loans).

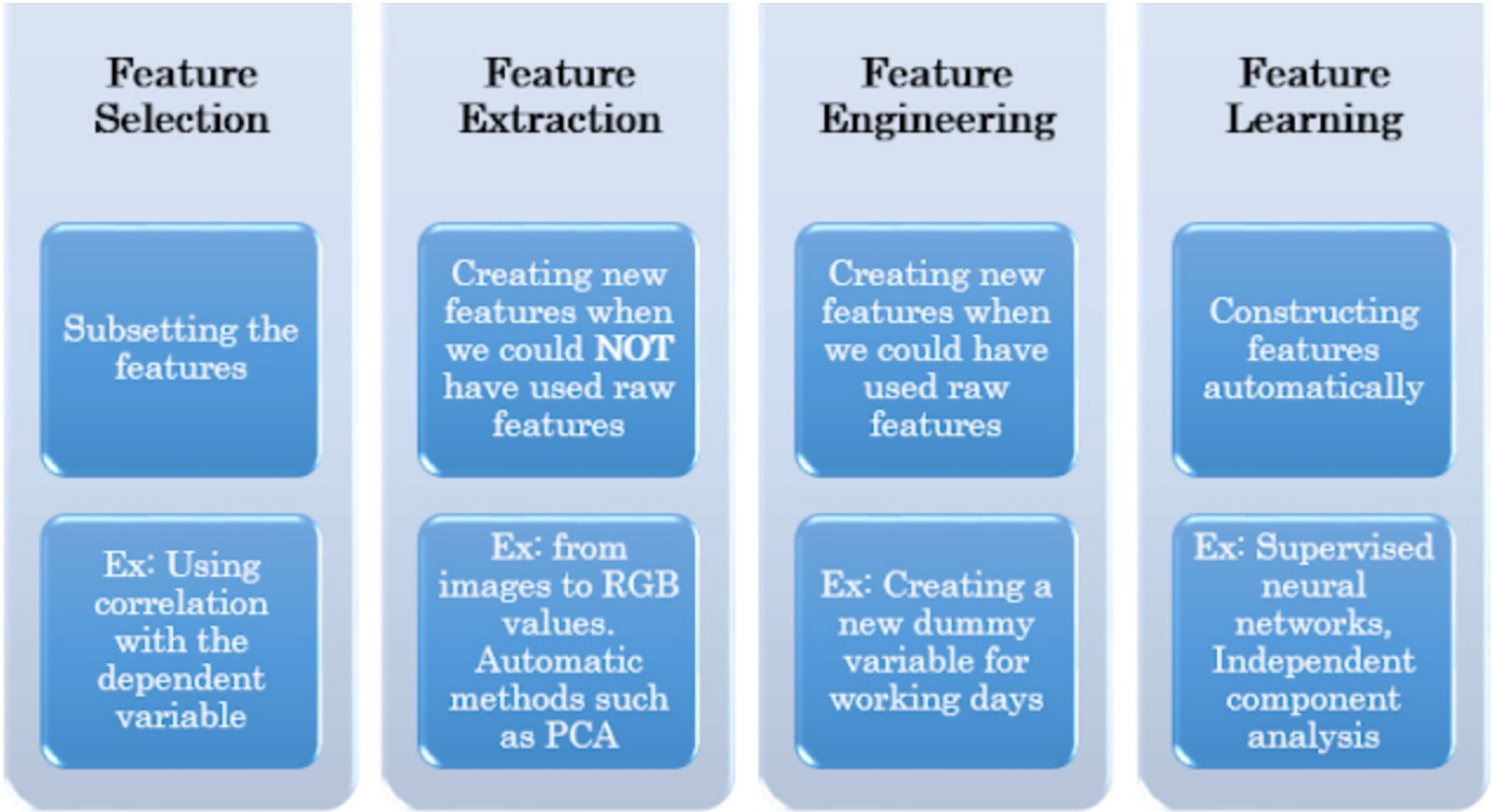
Reputational Damage: Loss of trust in AI systems.

Legal Consequences: Non-compliance with anti-discrimination laws.

Economic Losses: Inefficient or incorrect predictions.


Undersampling
Samples of
majority class
Original dataset


Oversampling
Copies of the
minority class
Original dataset


Feature Selection
Subsetting the features
Ex: Using correlation with the dependent variable
Feature Extraction
Creating new features when we could NOT have used raw features
Ex: from images to RGB values. Automatic methods such as PCA
Feature Engineering
Creating new features when we could have used raw features
Ex: Creating a new dummy variable for working days
Feature Learning
Constructing features automatically
Ex: Supervised neural networks, Independent component analysis

Feature Selection

Full Feature Set

Identify Useful Features

Selected Feature Set

Feature Selection Methods

- **Filter Methods** : filter methods are generally used as a preprocessing step. The selection of features is independent of any machine learning algorithm. Instead the features are selected on the basis of their scores in various statistical tests for their correlation with the outcome variable. Some common filter methods are Correlation metrics (Pearson, Spearman, Distance), Chi-Squared test, Anova, Fisher's Score etc.
- **Wrapper Methods** : in wrapper methods, you try to use a subset of features and train a model using them. Based on the inferences that you draw from the previous model, you decide to add or remove features from the subset. Forward Selection, Backward elimination are some of the examples for wrapper methods.
- **Embedded Methods** : these are the algorithms that have their own built-in feature selection methods. LASSO regression is one such example.

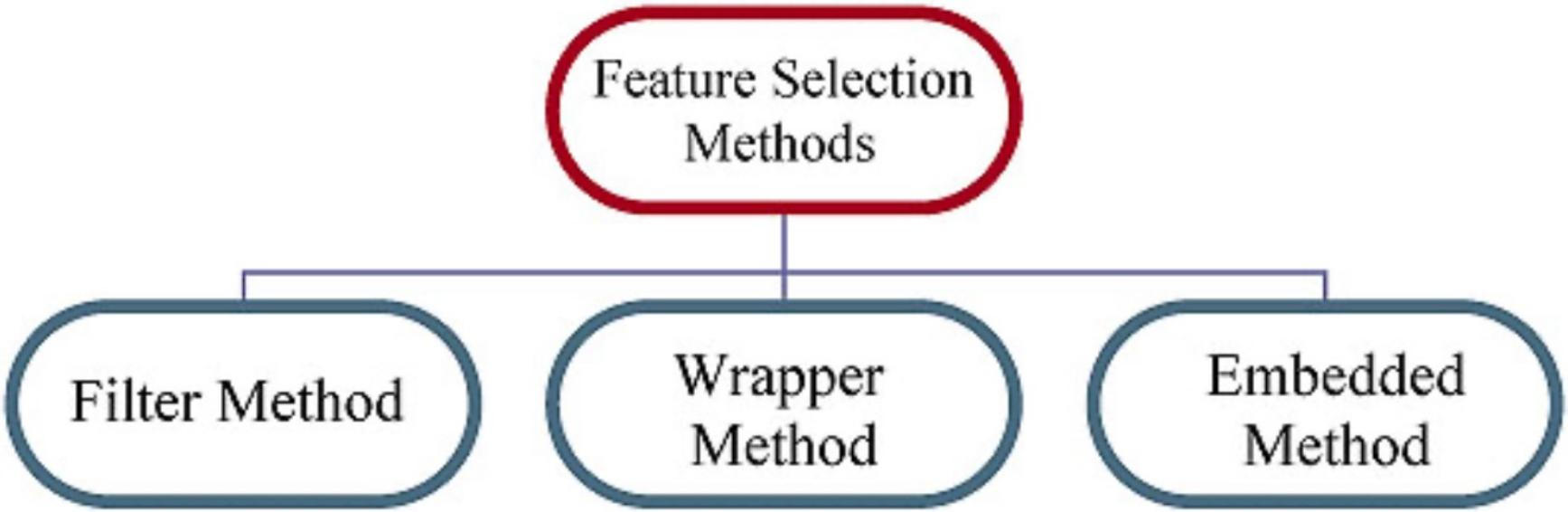
Exhaustive – worst time complexity

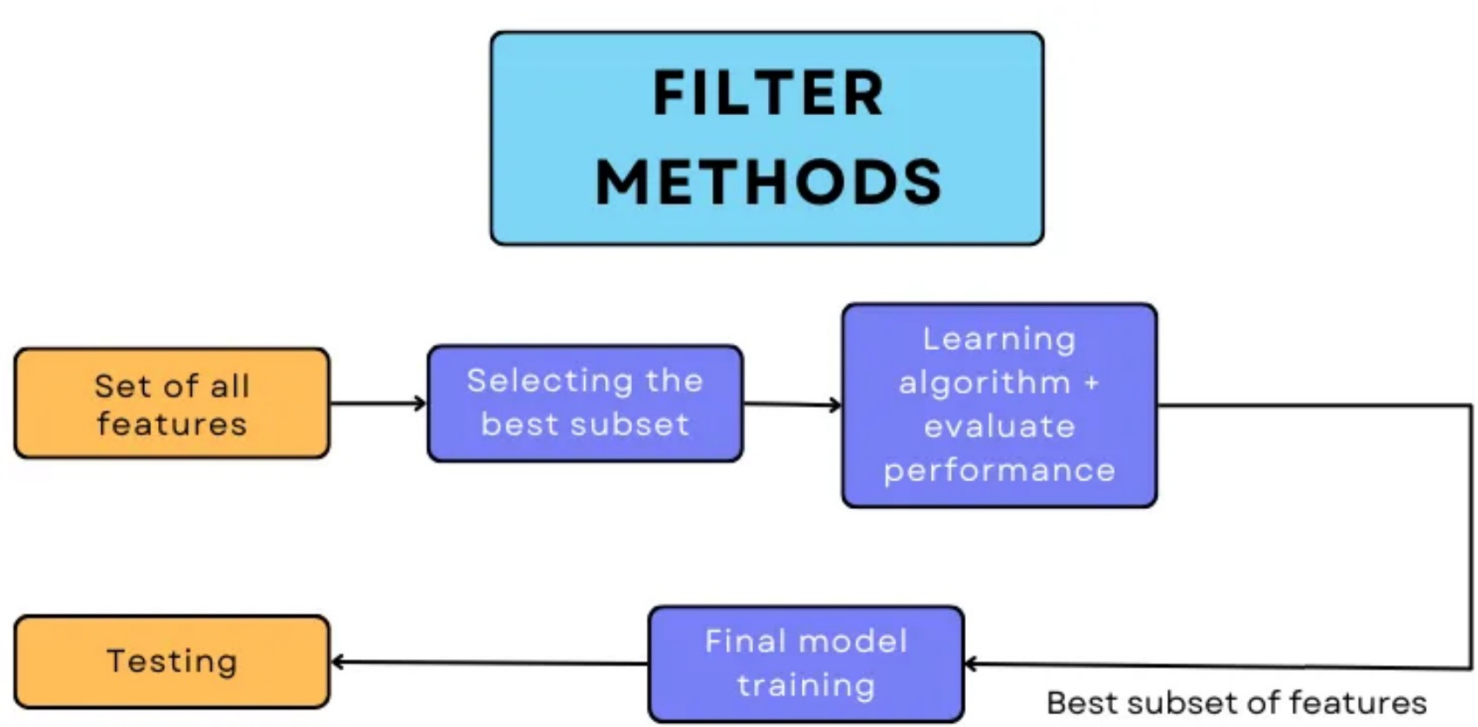
Let's consider an example to understand this better. Suppose we have three features: f1, f2, and f3, and we want to predict a target variable y. The exhaustive feature selection process would involve evaluating all possible combinations of these three features.

2^3-1=7 subsets we have to make

The possible combinations (subsets) of features are:

1. f1
2. f2
3. f3
4. f1, f2
5. f1, f3
6. f2, f3
7. f1, f2, f3


Feature Selection Methods
Filter Method
Wrapper Method
Embedded Method

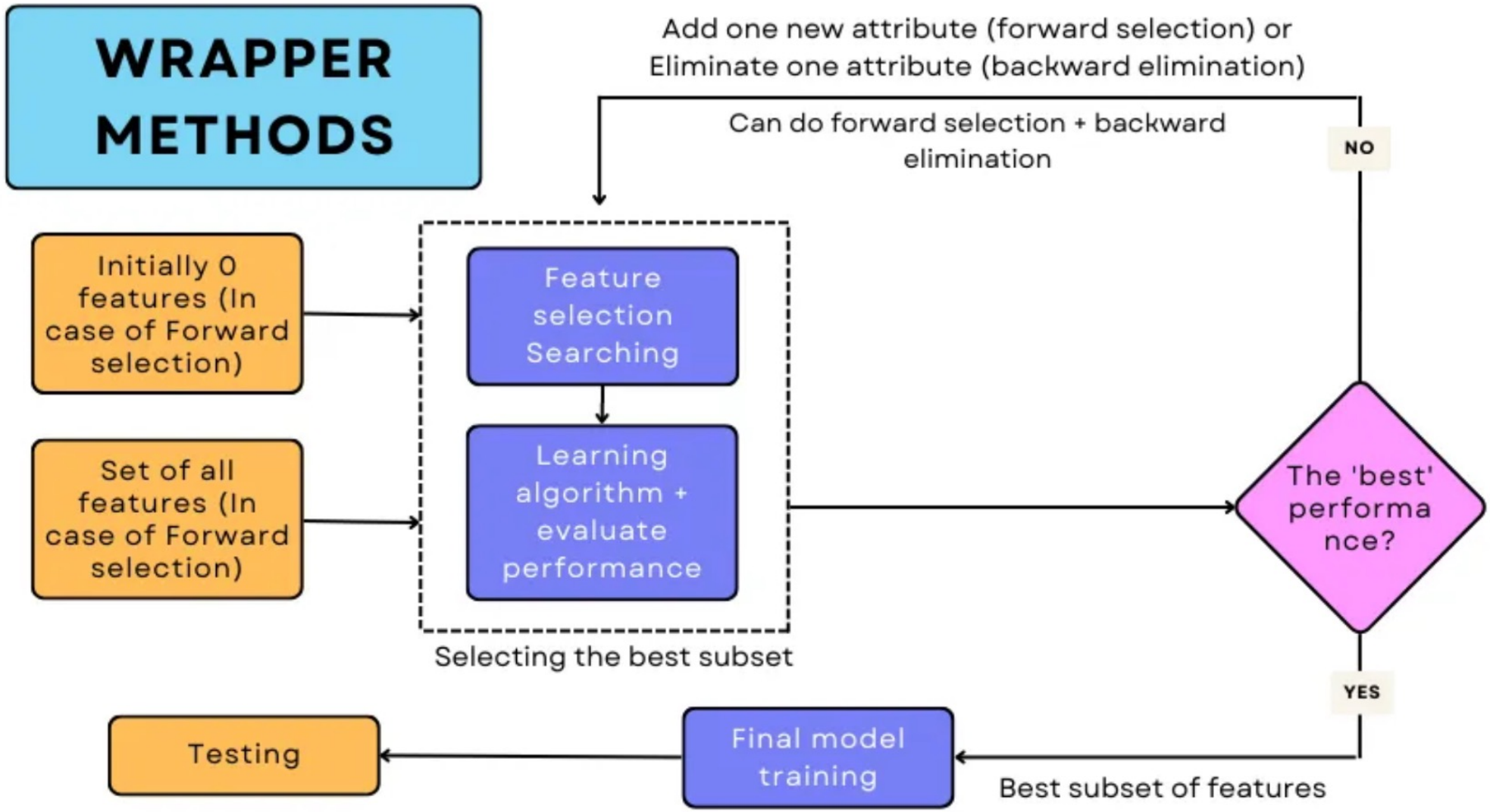


Independently analyze each feature based on a pre-defined metric, such as correlation with the target variable or information gain.

In filter methods, the variables that have the most impact on the output or target variable are considered important. The variables are ranked based on their significance towards the output. The top-ranking features are selected and the irrelevant features are removed.

There are multiple statistical tests that we can use to find associations between features and the target variable. These tests rank features based on their “importance”. The most commonly used statistical tests in data science projects are:

- Chi-squared
- Anova
- Correlation


WRAPPER METHODS
Add one new attribute (forward selection) or
Eliminate one attribute (backward elimination)
Can do forward selection + backward elimination
NO
Initially 0 features (In case of Forward selection)
Set of all features (In case of Forward selection)
Feature selection Searching
Learning algorithm + evaluate performance
Selecting the best subset
The 'best' performance?
YES
Testing
Final model training
Best subset of features

backward elimination

1. Start with all features included in the model.
2. Train a model using all the features and evaluate its performance using a performance metric such as accuracy, precision, recall, or F1 score.
3. Remove one feature at a time and train the model again using the remaining features.
4. Evaluate the performance of the model with the removed feature and compare it to the performance of the previous model that used all the features.
5. If the model's performance improves with the removal of the feature, keep the feature removed. If not, keep the feature included in the model.
6. Repeat steps 3–5 for all remaining features until no further improvement in the model's performance is observed.

Let's consider an example to illustrate the backward elimination method. Suppose we have a dataset of customer information for a company that sells products online. The dataset contains the following features:

- Age
- Gender
- Income
- Education level
- Time spent on the website
- Number of products purchased
- Customer satisfaction rating

The goal is to predict whether a customer will purchase a product or not based on these features.

To use the backward elimination method, we first start with all features included in the model. We train a machine learning model, such as logistic regression or decision tree, on the entire dataset and evaluate its performance using a performance metric such as accuracy.

Suppose the accuracy of the model is 85%. We then remove one feature at a time and train the model again using the remaining features. We evaluate the performance of the model with the removed feature and compare it to the performance of the previous model that used all the features.

Suppose we remove the “Time spent on the website” feature and train the model again. We evaluate the performance of the model and find that its accuracy drops to 80%. Since the model’s performance has worsened with the removal of this feature, we keep the feature included in the model.

Next, we remove the “Education level” feature and train the model again. We evaluate the performance of the model and find that its accuracy remains at 85%. Since the model’s performance does not improve with the removal of this feature, we remove it from the model.

We repeat these steps for each remaining feature until no further improvement in the model’s performance is observed.

Sequential Forward Selection

1. Initialization: Start with an empty feature set.
2. Evaluation of Single Features: Evaluate the performance of each individual feature when added to the empty set. In this case, evaluate the performance of f1, f2, f3, and f4 separately.

- Add f1 to the empty set and evaluate the performance of the model.
- Add f2 to the empty set and evaluate the performance of the model.
- Add f3 to the empty set and evaluate the performance of the model.
- Add f4 to the empty set and evaluate the performance of the model.

3. Based on the predefined criterion, select the feature that performs the best when added individually. Let's assume that f2 performs the best.

Iteration 1: Add the best performing feature from step 2 (f2) to the feature set.

- Add f2 to the feature set.
- Evaluate the performance of the model with f2.
- Calculate the performance metric (e.g., accuracy, mean squared error, etc.).

4. Evaluation of Feature Pairs: Evaluate the performance of all possible pairs of the current feature set and the remaining features (f1, f3, and f4). In this case, evaluate the performance of (f2, f1), (f2, f3), and (f2, f4).

- Add (f2, f1) to the feature set and evaluate the performance of the model.
- Add (f2, f3) to the feature set and evaluate the performance of the model.
- Add (f2, f4) to the feature set and evaluate the performance of the model.

5. Select the pair that gives the best performance according to the predefined criterion. Let's assume that (f2, f4) performs the best.

Iteration 2: Add the best performing feature pair from step 4 ((f2, f4)) to the feature set.

- Add (f2, f4) to the feature set.
- Evaluate the performance of the model with (f2, f4).
- Calculate the performance metric.

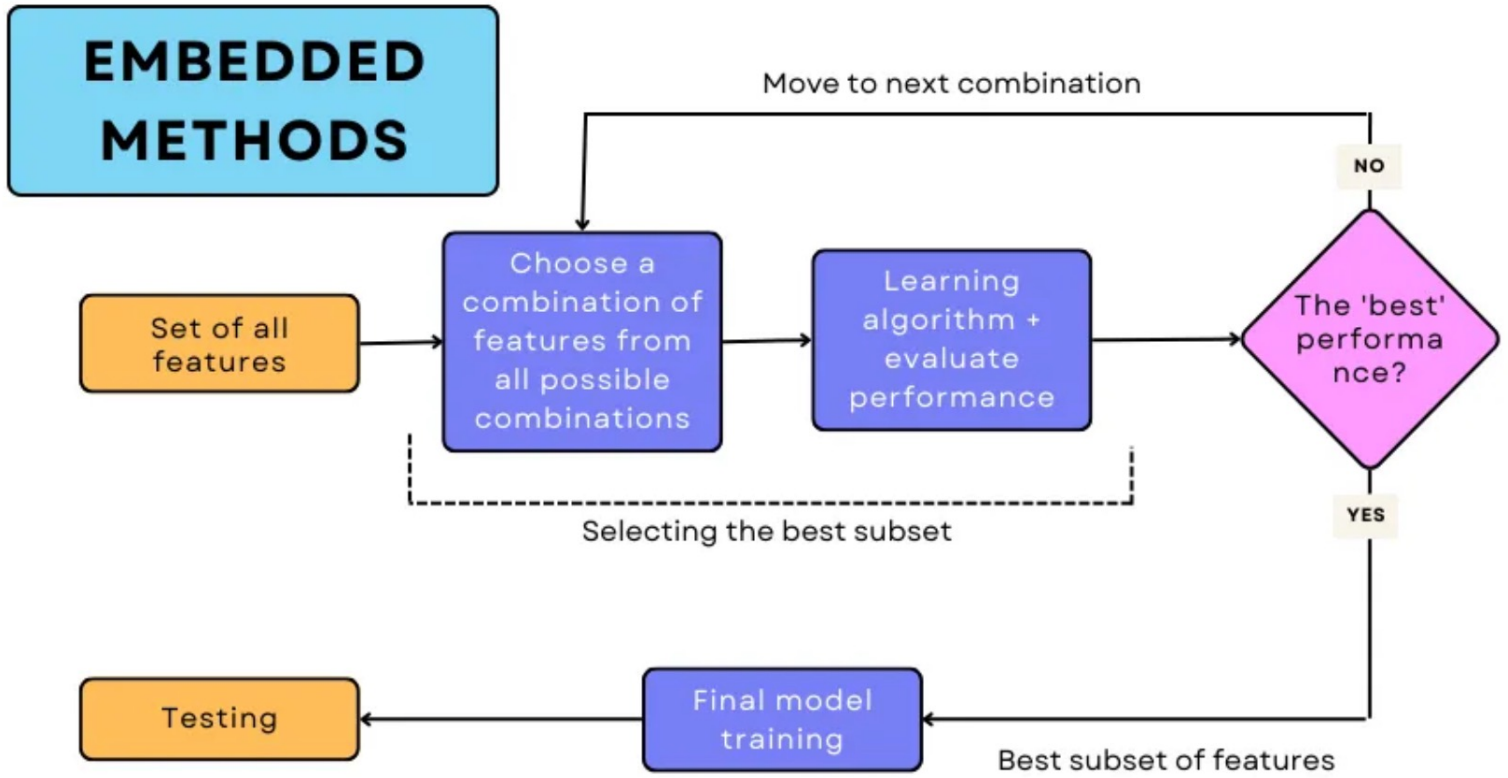
Repeat Steps 4 and 5: Continue the process by evaluating all possible combinations of the current feature set and the remaining features, selecting the best performing feature combination, and adding it to the feature set.

- Evaluate the performance of (f2, f4, f1) and (f2, f4, f3).
- Select the best performing feature combination.
- Add it to the feature set.

Termination: The process continues until a predefined stopping criterion is met. This criterion can be the desired number of features or a threshold on the performance metric.

- Evaluate the performance of (f2, f4, f1, f3).
- Determine if the stopping criterion is met.
- If not, add it to the feature set and continue the process.

The final result of the SFS algorithm will be the selected subset of features that provides the best performance according to the predefined criterion. In this example, it could be (f2, f4, f1, f3), indicating that these four features together provide the most predictive power for the model.



approach. They **incorporate feature selection directly into the model training process**. This allows the model to learn not only the relationship between features and the target variable, but also which features are most relevant.

Filter methods	Wrapper methods	Embedded methods
Generic set of methods which do not incorporate a **specific machine learning algorithm.**	Evaluates on a **specific machine learning algorithm** to find optimal features.	Embeds (fix) features during **model building process.** Feature selection is done by observing each iteration of model training phase.
Much **faster** compared to Wrapper methods in terms of time complexity	**High computation time** for a dataset with many features	Sits **between Filter methods and Wrapper methods** in terms of time complexity
Less prone to **over-fitting**	High chances of **over-fitting** because it involves training of machine learning models with different combination of features	Generally used to reduce **over-fitting** by **penalizing** the coefficients of a model being too large.
Examples – **Correlation, Chi-Square test, ANOVA, Information gain** etc.	Examples - **Forward Selection, Backward elimination, Stepwise selection** etc.	Examples - **LASSO, Elastic Net, Ridge Regression** etc.