

# Feature Selection 101: The Manual For Beginners

[BEGINNER](#)[VIDEOS](#)

This article was published as a part of the [Data Science Blogathon](#).

## Introduction

With the overwhelming hype of feature selection in machine learning and data science today, you might wonder why you should care about feature selection. The answer is that most machine-learning models require a large amount of training data. If you don't have enough data, you will have difficulty training the model. In addition, having too many features means you're likely to get overfit. Overfitting occurs when a model learns from noise instead of the true data. Hence, it is essential to choose some or a limited number of the most significant data features to train our models. Hence the concept of 'Feature Selection' comes into the picture.

Let us start by answering the basic question, 'What is Feature Selection?'

## What is Feature Selection?

As the name suggests, feature selection is the process of choosing an optimal subset of attributes according to a certain criterion and is essentially the task of removing irrelevant features from the dataset.

The criterion for choosing the features depends on the purpose of performing feature selection. Given the data and the number of features, we need to find the set of features that best satisfies the criteria. Ideally, the best subset would be the one that gives the best performance.

## Why do we Need to Perform Feature Selection?

In real-time, the data that we use for our machine learning and data science applications has many drawbacks to it. The three main problems are:

1. Having too much data can make the learning system (machine learning model) incapable of handling the data, and consequently, it cannot learn anything.
2. Too little data can make the model learning nothing meaningful and leads to many unnecessary assumptions being made.
3. Noisy data can cause unwanted distractions during the learning process.

Therefore, choosing and feeding the machine learning model with only optimal features that best influence the target variable is crucial.

- The number of features/variables/attributes plays a very vital role in the size of the hypothesis space. (Hypothesis is a learning function that predicts the results based on the data provided). As the number of features increases linearly, the hypothesis space grows exponentially. And smaller the functional space, the easier it would be for the model to predict the results. This feature selection helps remove

unnecessary variables from the dataset, thereby minimizing the hypothesis space. This makes the learning process a way lot easy and simple.

- It improves the data quality.
- Feature selection makes the algorithms learn and work faster on large datasets.
- It enhances the comprehensibility of the outcome.
- Feature selection is a booster for ML models even before they are built.

Having understood why it is important to include the feature selection process while building machine learning models, let us see what are the problems faced during the process.

## Problems faced during Feature Selection

When performing feature selection, people frequently struggle with the following questions:

1. How do you decide which features are the “best”?

The first issue that anyone faces during feature selection is determining how to identify and select the most optimal features in the data set. In this context, “best” features have a greater influence on the target variable and the outcome than others.

2. What evaluation criteria should you use when selecting features?

Once you’ve determined how to select the appropriate attributes for your solution, the next question is how you’ll evaluate the features and determine which ones are the best. Various techniques can be used to evaluate features.

3. How should new features be selected, added, or removed?

In some cases, you may need to add new features or remove existing ones. Again, another big question is how you will perform this addition or elimination.

4. How does the application determine the feature selection process?

Feature selection has no fixed steps that can be implemented. It is a very flexible task, and the process and steps involved are highly dependent on the business problem and the end application. Figuring out the right process that benefits your application is, of course, a tedious task.

Understanding the various types of feature selection techniques can help solve the above problems to some extent. Let us investigate the various types of techniques in feature selection.

## Types of Feature Selection methods

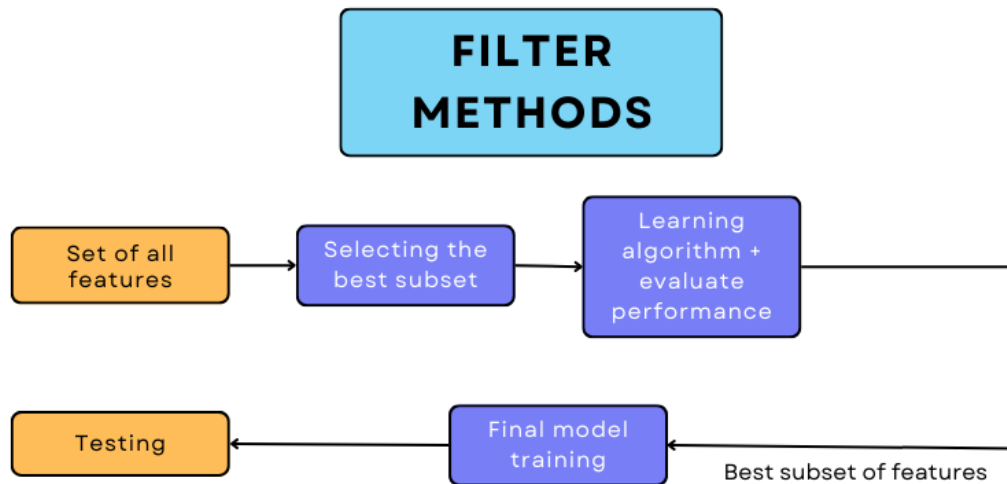
Feature selection can be made using numerous methods. The three main types of feature selection techniques are:

1. Filter methods
2. Wrapper methods
3. Embedded methods

Let us look into each of these methods in detail. There are generally two phases in filter and wrapper methods – the feature selection phase ( Phase 1) and the feature evaluation phase (Phase 2).

## Filter methods

Feature selection using filter methods is made by using some information, distance, or correlation measures. Here, the features' sub-setting is generally done using one of the statistical measures like the Chi-square test, ANOVA test, or correlation coefficient. These help in selecting the attributes that are highly correlated with the target variable. Here, we work on the same model by changing the features.



Source: Author

Why should you be choosing the filter method?

- It does not rely on the model's bias and instead depends only on the characteristics of the data. Hence, the same feature subset can be used to train different algorithms.
- The time taken by information or distance-related measures is very; hence, a filter method can produce subsets faster.
- They can handle large amounts of data.

## Wrapper methods

Source: Author

In wrapper methods, we generate a new model for each feature subset that is generated. The performance of each of these is recorded and the features which produce the best performance model are used for training and testing the final algorithm. Unlike filter methods that use distance or information-based measures for feature selection, wrapper methods use many simple techniques for choosing the most significant attributes. They are:

### **(1) Forward Selection**

It is an iterative greedy process where you start with absolutely no features and in each iteration, you keep adding one most significant feature. Here, the variables are added in the decreasing order of their correlation with the target variable.

Source: Author

This addition of a new attribute is done until the model's performance does not increase on further adding other features that are when you reach the point where you get the best possible performance.

## **(2) Backward Elimination**

As the name suggests, here we start with all the features present in the dataset, and with each iteration, we remove one least significant variable.

We remove the attributes until there is no improvement in the model's performance on eliminating features. The least correlated feature with the target variable is chosen based on certain statistical measures. In contrast to the filter methods, the features are removed in the increasing order of correlation with the target variable.

Source: Author

It is also possible to combine both these methods. This is often called Bidirectional Elimination. This is similar to forward selection but the only difference is that if it finds any already added feature to be insignificant at a later stage when a new feature is added, it removes the former through backward elimination.

It is worth noting that wrapper methods may work very effectively for certain learning algorithms. However, the computational costs are very high when these wrapper methods as compared to filter methods.

## **Embedded methods**

Source: Author

In embedded methods, all the combinations of the features are generated. Then each of these combinations of attributes is used to train the model, and as usual, its performance is observed. The combination which gives the best performance is chosen for the final training.

**The choice of technique used for feature selection depends on the application and the dataset's size and requires an in-depth understanding of the dataset. As mentioned before,**

## Conclusion

With this, we conclude our discussion of feature selection. To summarize, we began by defining feature selection and comprehending its significance. Later on, we looked at the problems encountered during it and how knowing different attribute selection methods can help us overcome those problems.

The main takeaways from this article are:

- Feature selection is a simple yet critical step that involves removing unwanted attributes and selecting the most powerful features from the dataset.
- It helps us improve prediction accuracy and enhance the data's quality and understandability.
- In filter methods, we use a single model and keep retraining it with different attribute subsets, while in wrapper methods, we make a new model for each subset.
- Forward selection, [backward elimination](#), and bidirectional elimination are the three techniques used for selecting feature subsets in wrapper methods.
- The dataset size, understanding of the data, and the purpose of feature selection determine which technique should be used for selecting the most optimal features.

I hope you liked my article. If you have any opinions or questions, then comment below.

Connect with me on [LinkedIn](#) for further discussion.

**The media shown in this article is not owned by Analytics Vidhya and is used at the author's discretion.**

---



**Yamini Ane**