Imagine you have a big bag of Legos. You want to build a house, but you don't know which Legos to use. You could just start building, but you might end up using the wrong Legos and your house might not turn out right.

Feature selection is like choosing the right Legos to build a house. You have a big dataset of information, and you want to choose the features that are most important for your machine learning model. If you choose the wrong features, your model might not work very well.

There are different ways to choose features for a machine learning model. One way is to use filter methods. Filter methods look at each feature and score it based on how important it is. The features with the highest scores are the most important.

Another way to choose features is to use wrapper methods. Wrapper methods build a machine learning model and then see how well the model works with different subsets of features. The subset of features that makes the model work the best is the best set of features.

Feature selection is an important part of machine learning. It can help to improve the performance of machine learning models.

Concept with Real Life example

Imagine you are trying to build a machine learning model to predict whether a patient will have a heart attack. You have a dataset of information about patients, including their age, weight, blood pressure, and cholesterol levels.

You could just use all of this information to train your model, but that would be a waste of time and resources. Some of this information is not very important for predicting heart attacks, and including it in your model would only make it more complex and harder to train.

This is where feature selection comes in. Feature selection is the process of choosing the most important features from a dataset. This can help to improve the performance of your machine learning model by making it more accurate and easier to train.

As said earlier there are two main types of feature selection techniques: filter methods and wrapper methods.

Filter methods are a type of feature selection technique that uses statistical measures to score each feature and select the features with the highest scores. One common filter method is **correlation-based feature selection**, which scores each feature based on its correlation with the target variable just like heatmap values we use for multivariate analysis. The features with the highest scores are the most relevant to predicting heart attacks.

Wrapper methods are a type of feature selection technique that uses a machine learning model to evaluate the performance of different feature subsets. One common wrapper method is **recursive feature elimination**, which starts with all of the features in the dataset and then iteratively removes the feature that has the least impact on the performance of the machine learning model. The process is repeated until the desired number of features is left.

In the case of the heart attack prediction model, you could use correlation-based feature selection to score each feature based on its correlation with the target variable, which is whether or not the patient had a heart attack. The features with the highest scores would be the most relevant to predicting heart attacks.

You could also use recursive feature elimination to start with all 10 features in the dataset and then iteratively remove the feature that has the least impact on the performance of a decision tree classifier. The process would be repeated until you are left with a subset of 5 features.

The best feature selection technique to use depends on the specific problem we are trying to solve. In general, filter methods are a good choice if you are looking for a quick and easy way to reduce the number of features in your dataset. Wrapper methods are a good choice if you are looking for a more accurate way to select features and you are willing to spend the extra time and computational resources.