Okay so you are having some doubts in Feature selection,

It is kind of last step of data preprocessing step ok.

When you are all done dealing with missing values, outliers and data transformation. This is also called Feature Engineering, Don't get confused by this

Now you decide which feature among the independent features is less relevant to our model and remove it so that we can get much more accurate results

This is called Feature Selection

Definition:

The process of selecting a subset of relevant features from a larger set of available features

Are you now clear with the definition??

This is done so that, By removing irrelevant or redundant features, you can simplify the model, reduce overfitting, improve interpretability, and potentially enhance prediction accuracy

Now These are some of the common techniques

- Univariate Selection: This technique evaluates each feature independently and selects the most relevant ones based on statistical tests. It assesses the correlation between each feature and the target variable and selects the features with the highest scores.
- Feature Importance Ranking: This method uses algorithms such as decision trees, random forests, or gradient boosting to assign importance scores to each feature. Features with higher scores are considered more relevant and are retained, while less important ones can be discarded.
- Correlation Analysis: Correlation analysis measures the relationship between pairs of features and the target variable. Features that are highly correlated with the target variable are likely to be valuable. However, if two

features are strongly correlated with each other, it may be redundant to keep both, and one of them can be removed.

- Recursive Feature Elimination: This technique starts with all features and iteratively eliminates the least important features based on a model's performance. It repeatedly trains the model on a subset of features, evaluates its performance, and removes the least significant feature until a desired number of features is reached.
- Regularization Methods: Regularization methods, such as Lasso or Ridge regression, can be employed to introduce a penalty for the number of features used. These methods encourage the model to select only the most relevant features by assigning them higher coefficients and shrinking the coefficients of less important features towards zero.
- Embedded Methods: Some machine learning algorithms have built-in feature selection mechanisms. For example, decision trees and random forests inherently measure feature importance during their training process. Similarly, support vector machines (SVM) have feature weights associated with support vectors, which can be used for feature selection.

IF ITs STILL NOT CLEAR LET ME GIVE YOU AN REAL-LIFE EXAMPLE

Imagine you are cooking

now you have washed the vegetables, removed the bad ones, so now you have processed the raw material - this raw material are you features (Independent) and this process is called Feature Engineering

Now you select which ingredients you need/required for you to prepare a dish for your lunch.

This process of selection of the processed raw material is what you call Feature Selection

Hope you have understood it, go through the above-mentioned feature selection techniques and tell me if you have any doubts.