One of the most important steps in data preprocessing is feature scaling. That step involves transforming features to a similar scaling. This step helps improving model performance, reducing the impact of outliers, and ensuring that the data is on the same scale.

Feature scaling techniques:

1-Standardization:

Formula: z = (x – mean) / standard deviation

This technique is appropriate when the data approximately follows a Gaussian distribution. It is also less sensitive to outliers and preserves the relationship between the data points even though it changes the shape of original distribution. Standardization is not suitable for data that have outliers. It adjusts the mean to 0.

2-Normaliztion:

Formula: z = (x - min_value) / (max_value – min_value)

This technique is useful when the distribution of data is unknown or not Gaussian distribution. It also retains the shape of original distribution. However, it is sensitive to the outliers and it may not preserve the relationship between the data points. Normalization is not suitable for data that have outliers. It adjusts the data between 0 and 1.

3-Robust Scalar:

Formula: z = (x – median) / IQR

This technique removes the median and scales the data according to the quartile range. The IQR is the range between the 1st quartile and the 3rd quartile. It absorbs the effects of outliers while scaling. It is the best choice if there are many outliers that were not removed. Robust scalar is more suitable for data that have outliers.

4-Max-absolute Scalar:

Formula: z = x / (max (| x |))

Max-absolute scalar is not suitable for data that have outliers. It adjusts the data between -1 and 1.