Methods to Handle Multicollinearity in Machine Learning (Other Than VIF)

1) Removing Highly Correlated Features

Calculate the correlation matrix of the dataset.

Identify feature pairs with high correlation (e.g., correlation coefficient > 0.8).

Remove one feature from each highly correlated pair to reduce redundancy and simplify the model.

2) Principal Component Analysis (PCA)

PCA is a dimensionality reduction technique that transforms correlated features into a set of uncorrelated principal components.

It retains the most important variance in the data while eliminating multicollinearity.

Particularly useful when working with a large number of features.

3) Feature Selection Methods

Utilize automated feature selection techniques to remove redundant features:

Recursive Feature Elimination (RFE).

SelectKBest.

Tree-based feature importance (e.g., from Random Forest or Gradient Boosting).

These methods help identify and retain only the most relevant variables.

4) Collecting More Data

Increasing dataset size may reduce multicollinearity by providing more variation between variables.

Though not always practical, it is an option when feasible.