

## Feature Engineering:

- > Most important technique used in creating machine learning models.
- > A basic term used to cover many operations that are performed on variables(features) to fit them into the algorithm.
- > It helps in increasing the accuracy of the model thereby enhances the results of the predictions.
- > Perform better on data than basic machine learning models.

## List of Techniques of Feature engineering :

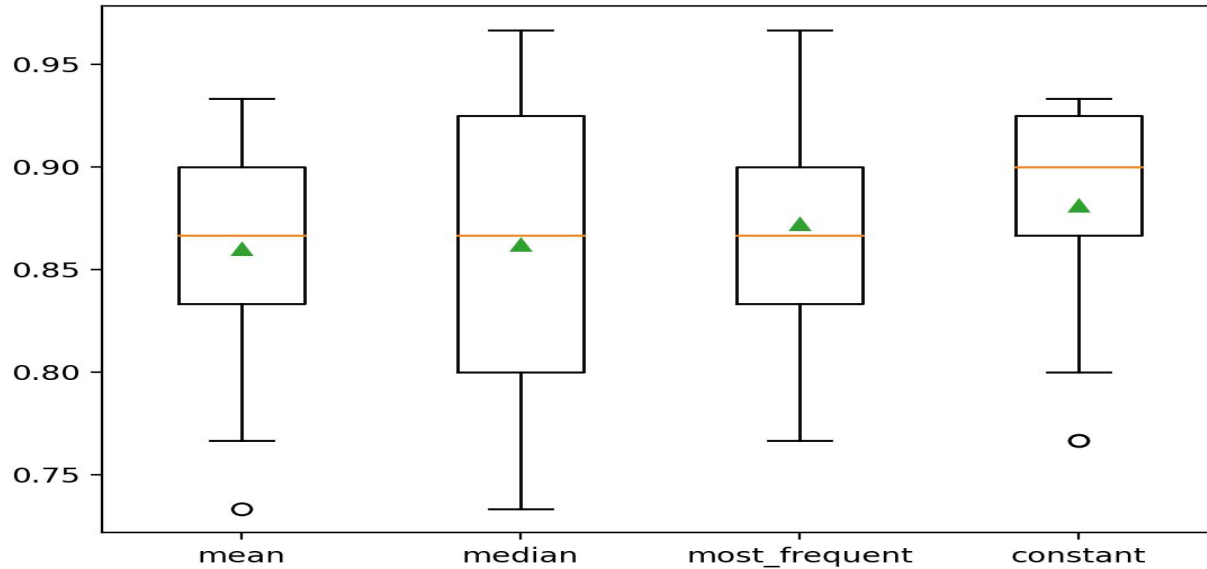
1) Imputation.

2) Encoding.

## 1) Imputation :

-> Imputation is the technique in ML that is used to handle missing values in our dataset.

-> The statistical approach of handling missing values uses the mean, median, or mode imputations.



# Imputing Missing Data: Machine Learning

- KNN: Find K “nearest” (most similar) rows and average their values
  - Assumes numerical data, not categorical
  - There are ways to handle categorical data (Hamming distance), but categorical data is probably better served by...
- Deep Learning
  - Build a machine learning model to impute data for your machine learning model!
  - Works well for categorical data. Really well. But it’s complicated.
- Regression
  - Find linear or non-linear relationships between the missing feature and other features
  - Most advanced technique: MICE (Multiple Imputation by Chained Equations)



Input Data  
Validation

Predicted Value  
Imputation

Distribution-based  
Imputation

Reduced Feature  
Models

Unique Value  
Imputation

**Missing Data Imputation Techniques (During Model Prediction)**

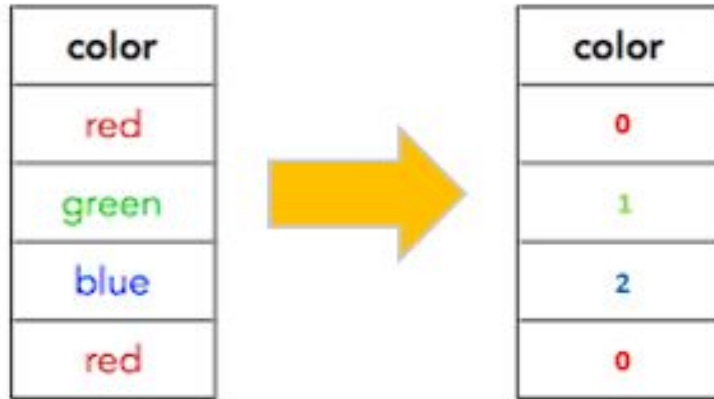
## 2)Encoding.

-> Feature Encoding is the conversion of Categorical features to numeric values as Machine Learning models cannot handle the text data directly.

Feature Engineering are basically of 2 types:

### 1. Label Encoding:

-> Label Encoding refers to converting the labels into numeric form so as to convert it into the machine-readable form.



## 2. One Hot Encoding:

-> It uses to overcome the disadvantage label encoding.

-> One-hot encoding is processed in 2 steps:

a) Splitting of categories to different columns.

b) Put '0' for others and '1' as an indicator for the appropriate column.

Example 1:

Label Encoding

Food Name	Categorical #	Calories
Apple	1	95
Chicken	2	231
Broccoli	3	50



One Hot Encoding

Apple	Chicken	Broccoli	Calories
1	0	0	95
0	1	0	231
0	0	1	50

Example 2:

Type	Onehot encoding →	Type	AA_Onehot	AB_Onehot	CD_Onehot
AA		AA	1	0	0
AB		AB	0	1	0
CD		CD	0	0	1
AA		AA	0	0	0