**Why do we need to do Feature Scaling:**

**A screenshot of a computer

Description automatically generated with medium confidence**

**Transform Missing Values using imputer in Sklearn:**

**A screenshot of a computer

Description automatically generated with medium confidence**

**imputer = SimpleImputer(strategy='constant', fill\_value=0)**

**sklearn.compose 🡪 ColumnTransformer**

**A screenshot of a computer

Description automatically generated with medium confidence**

**A screenshot of a computer

Description automatically generated with medium confidence**

**A screen shot of a computer screen

Description automatically generated with low confidence**

**Feature Scaling should be performed Before/After Test\_Split:**

Feature Scaling should always be performed after split because if we perform it before, there would be an information leakage from the test set. We always want to make sure test data appears like the real-world data to accurately predict the model performance.

**Random\_state in sklearn.model\_selection , train\_test\_split**

**A screenshot of a computer

Description automatically generated**

**Feature Scaling:**

Standardisation: x\_stand = (x -mean(x))/Standard deviation(x)

Normalisation: x\_norm = (x – min(x))/(max(x) – min(x))

Normalisation is recommended when data is normally distributed.

Standardisation will do the job all the time.

Standardisation (or) In general Feature scaling will not be applied on OneHotEncoded/Encoded Categorical variables.