Different standardization EDA techniques for different algorithms?

Why do we apply “transform” to the test data?

Ans:- concepts like mean, mu, or standard deviation are to be calculated only for train data and not for test data, applying “transform” to the test data will prevent the test to calculate the above mathematical concept.  
Hence to avoid data leakage, We apply “transform” to test data.

regression.fit() here “fit()” means only applyng the formula. fit\_transform means applying some formula and

changing the data.

**Type of Encoding:**

**Nominal Encoding:**

Here we are not concerned about the order of the categorical variable. (State, Countries, Male or Female)

1. One hot Encding
2. One hot encoding with many categorical variables.
3. Mean encoding

**Ordinal Encoding:**

Here we need to consider the rank or the order of the categorical variable. (Grandes, education degree)

1. Label Encoding.
2. Target Guided Encoding.

**Feature selection:**

**Filter method:**

1. Annova test:
2. Chi-square test:

**Wrapper method:**

1. Forward selection:
2. Backward Elimination:
3. Recursive Feature Elimination:

**Embedded method:**

**Methods to be used in models:**

1. Univariate Selection
2. Feature Importance
3. Correlation Matrix with Heatmap