Feature Selection Techniques:

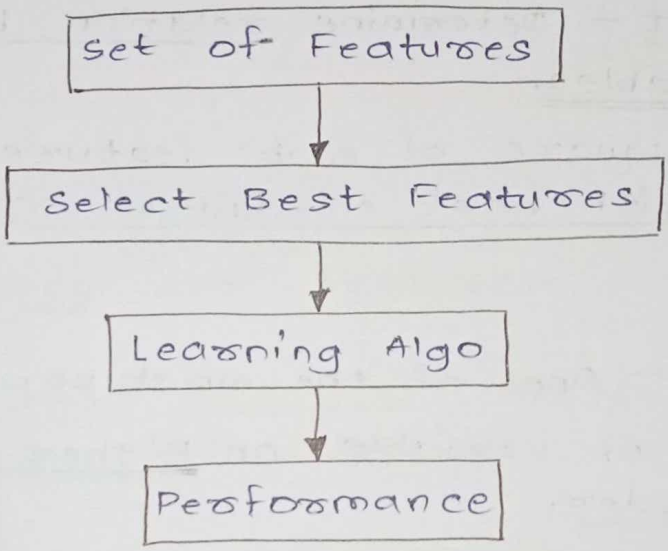
1) Filter Methods:-

↳ Select features on basis of statistics measures.

↳ Not depend on learning algo & choose feature as pre-processing step.

↳ Filter out irrelevant features & redundant column.

↳ Low computational time & not overfit data.



- Steps in Filter Methods:

(1) Calculate statistical measures for each features w.r.t. target variable.

(2) Rank / score features based on these measures.

(3) Select top-ranked features according to threshold / fixed num. of features.

(4) Use selected features subset for model training.

- Some Common Techniques:

① Correlation - Rank features based on correlation with target variable in classification / Regression.

② Information Gain - Determine reduction in entropy by transforming dataset.
↳ Calculate Info Gain of each variable w.r.t. target variable.

③ Chi-Square Test - Determine relation between categorical variables.
↳ Calculate chi-square of each feature & target, select feature with best chi-square score.

④ Fisher's Score - One of the most popular.
↳ Return rank of variable on Fisher criteria in descending order.
↳ Then select large Fisher score variable.

⑤ Missing Value Ratio -
↳ Evaluate feature set against threshold value.
↳ Drop variable with missing values more than thresh value.

$$MVR = \frac{\text{Num. of missing values} * 100}{\text{Total num. of observations}}$$

- **Advantages of Filter Methods:**

① Computational Efficiency — Efficient, because do not require training on entire dataset.

② Independence of Model — can use along with any machine learning models.

③ Interpretability — Easy understand importance of each feature relation with target variable.

- **Limitations of Filter Methods:**

① Independence Assumptions — consider each feature in isolation & not consider interaction/ dependency betⁿ features.

② Threshold Selection — Require domain knowledge for threshold value selection.

③ Limited to Univariate Relationships — May not capture complex relationships betⁿ imp features.