Feature Selection →

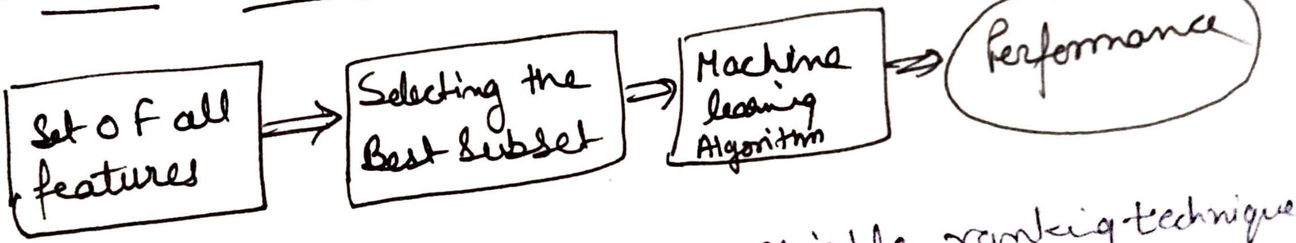
feature Selection is the Process of reducing the input variable to ~~g~~ the Model by using only relevant data.

The goal of feature Selection In Machine learing is to find the best Set of features that allows one to build Models

using feature Selection we Can Optimize our Models In Several ways.
1) Prevent learing from overfitting
2) Improved accuracy
3) Reduce traning time

Feature Selection techniques
(i) filter Methods
(ii) Wrapper Methods
(iii) Embedded Method.

(1) filter Method →



This Method uses the variable ranking technique for ordering and In order to Select the variables the Selection of features is independent of the classifier used.

Ranking Means How Much useful and Important each features is Expected to be for classification.

it basically Select the Subsets of variables as a Pre Processing step Independently of the chosen Predictor.

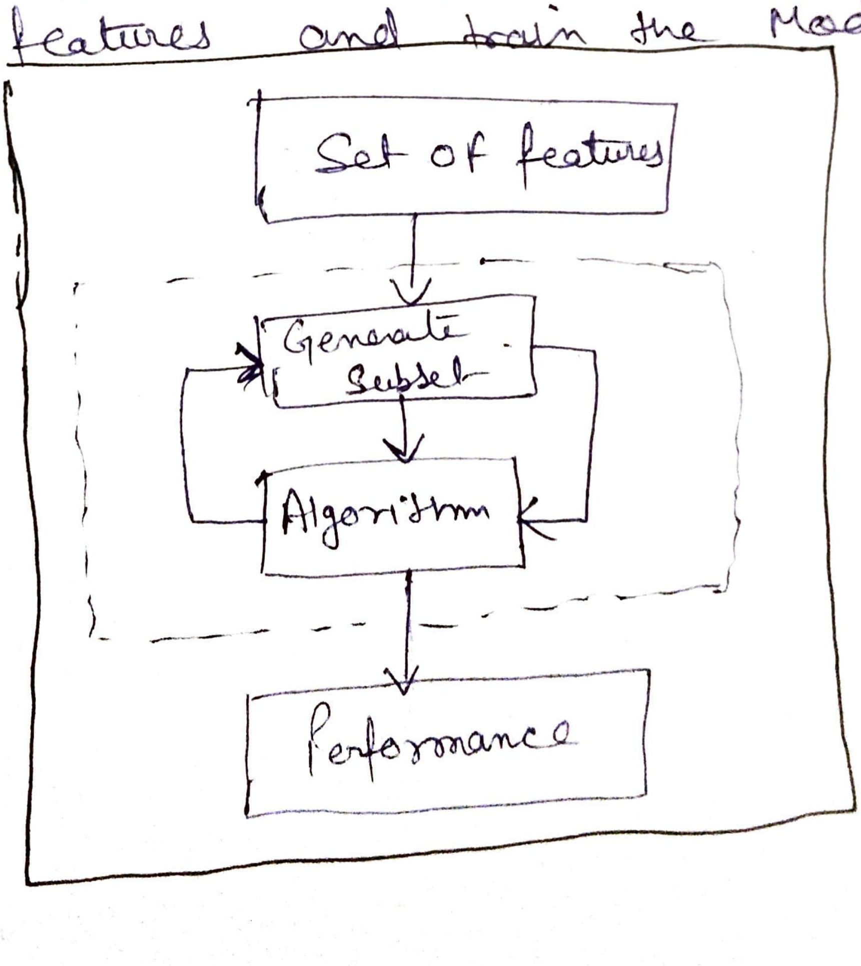
In this Method features are dropped based on their relation to the output or How they are correlating to the output.

Example

Name	No of times read	Condition of Book	Color

In Book classifier we dropped the color Column based on a Simple deduction.

2) Wrapper Method → we Split data in to Subsets and train Model using this. Based on the Output of the Model. we add and Subtract features and train the Model again.



for.

for Example

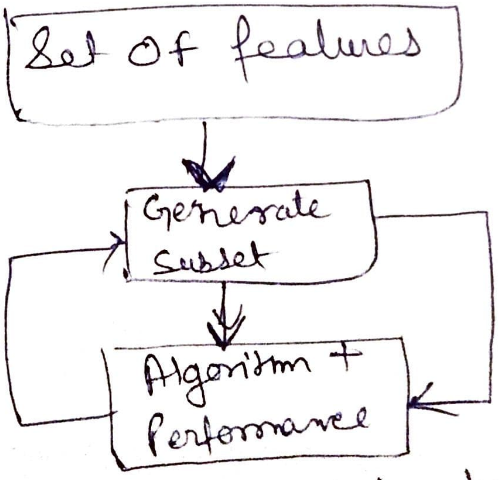
By using wrapper Method we would use a subset of different features to train the machine and adjust the subset according to output.

Name	No of time read	Condition of Book	Color

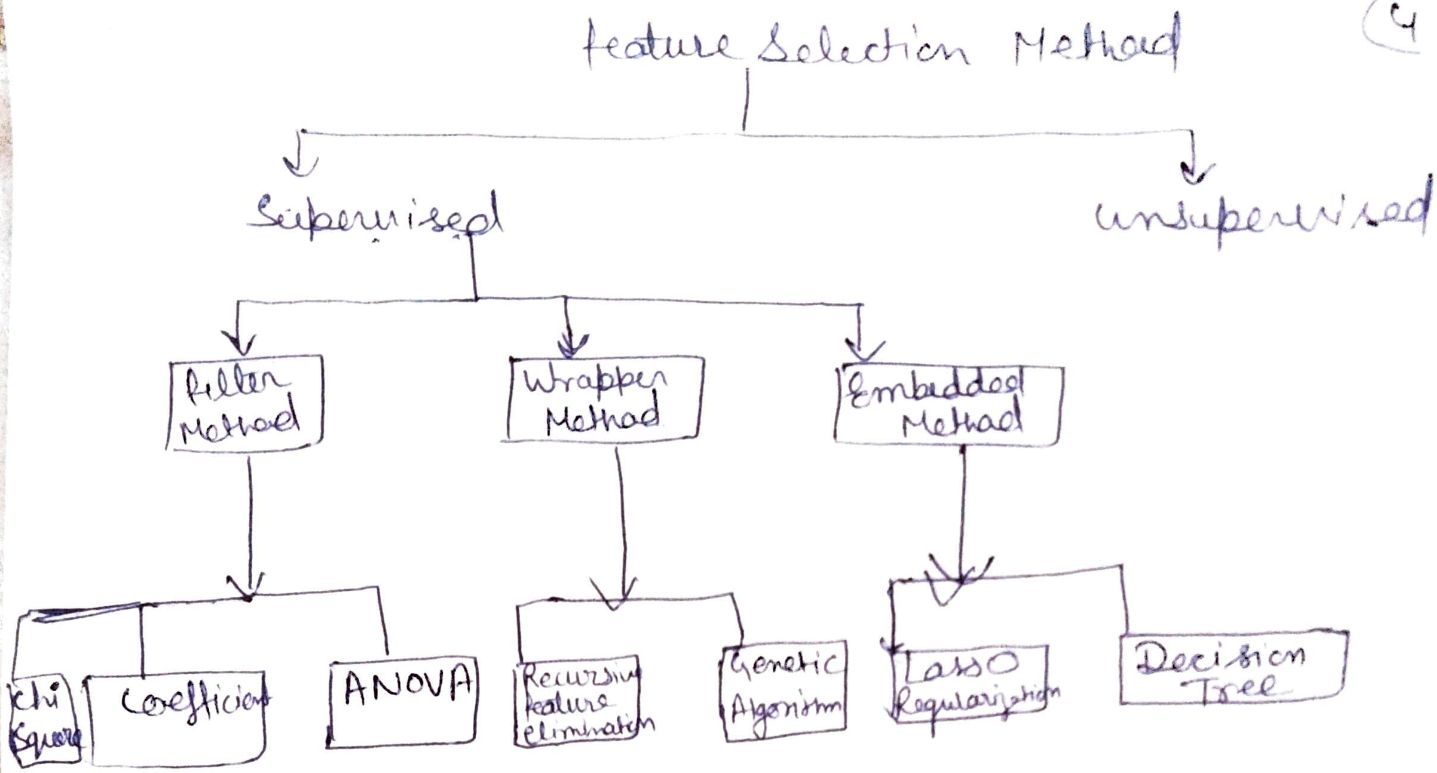
Name and No of times read.

Name, No of times read and Conditions after this Chock output.

(3) Embedded Method → This Method Combines the qualities of Both filter and wrapper Method to Create the best subset.



The Model will train and check the accuracy of different subsets and select the best among them.



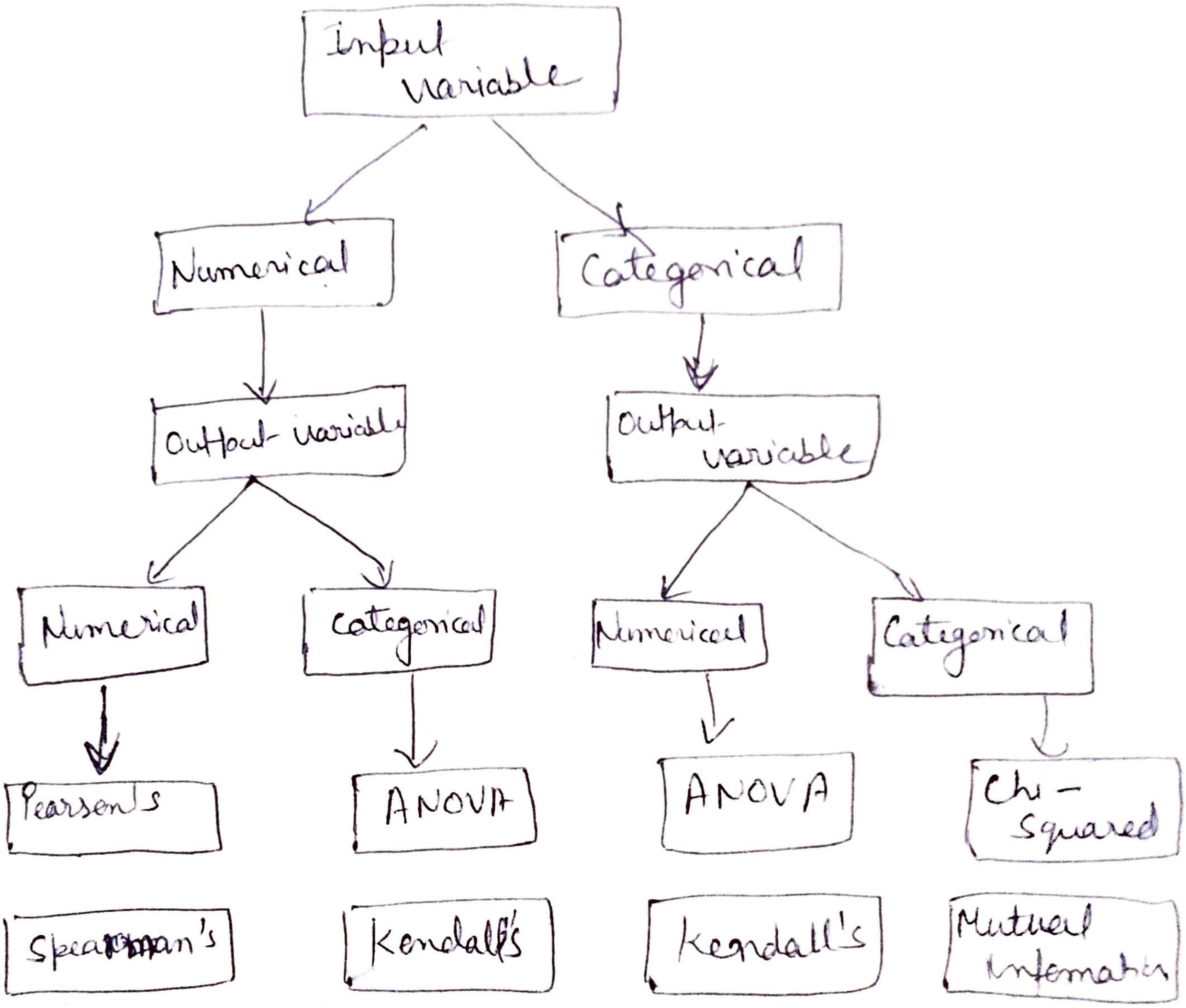
* filter Method
  (i) Chi Square
  (ii) Coefficient (~~Person~~ Correlation Coefficient)
  (iii) ANOVA

* Wrapper Method
  (i) Recursive feature elimination
  (ii) Genetic Algorithm

* Embedded Method
  (i) Lasso Regularization
  (ii) Decision Tree

Based on Input and output variables we can choose feature selection Model.

1. Numerical input & Numerical output
2. Numerical input & Categorical output
3. categorical input & Numerical output
4. categorical input & Categorical output



Chi Square Test → (x^2)

Chi Square Test is used to find the two variables are these related to each other or there is NO relationship.

For Ex

Table of observed value

Qualification Marital Status	Middle Class	High Class	Bachalor's	Masters	Ph.D	Total
Never Married	18	36	21	9	6	90
Married	12	36	45	36	21	150
Divorced	6	9	9	3	3	30
Widowed	3	9	9	6	3	30
Total	39	90	84	54	33	300

Table of Expected value

Qualification Marital Status	Middle class	High Class	Bachalor's	Masters	Ph.D	Total
Never Married	$\frac{90 \times 39}{300}$ $\Rightarrow 11.7$	$\frac{90 \times 90}{300}$ $\Rightarrow 27$	25.2	26.2	9.9	
Married	29.5	45	42	27	16.5	
Divorced	3.9	9	8.4	5.4	3.3	
Widowed	3.9	9	8.4	5.4	3.3	
Total						

$$\text{Chi Square } (x^2) = \frac{(\text{Observed value} - \text{Expected value})^2}{\text{Expected value}}$$

Observed value	Expected value	$(O-E)$	$(O-E)^2$	$\frac{(O-E)^2}{E}$
18	11.7	6.3	39.69	3.39
36	27	.9	81	3
21	25.2	−4.2	17.64	0.7
9	16.2	−7.2	51.84	3.2
6	9.9	−3.9	15.21	1.53
12	19.5	−7.5	56.25	2.88
36	45	−9	81	1.8
45	42	3	9	4.5
36	27	9	81	3
21	16.5	4.5	20.25	1.22
6	3.9	2.1	4.41	1.13
9	9	0	0	0
9	8.4	0.6	0.36	0.04
3	5.4	−2.4	5.76	1.06
3	3.3	−0.3	0.09	0.027
3	3.9	−0.9	0.81	0.207
9	9	0	0	0
9	8.4	0.6	0.36	0.04
6	5.4	−2.4	5.76	1.06
3	3.3	−0.3	0.09	0.02

$\boxed{x^2_{\text{Calculated}} = 23.57}$

$x^2 = \Sigma \frac{(O-E)^2}{E}$

$x^2 = 23.57$

Degree of freedom = (Columns-1)(Rows-1)

$\Rightarrow (5-1)(4-1)$

$\Rightarrow 12$

from

tabular value of Chi Square.

Significatnce level (α) = 0.05

$x^2_{tabular} = 21.03$

$x^2_{Calculated} = 23.57$

$x^2_{Calculated} > x^2_{tabular}$ (or $x^2_{critical}$)

then we reject Null hypothesis and accept alternate hypothesis.

Alternate hypothesis: there is significant relationship between Marital Status and Qualification.