

NOAMUNDI TASK

# SCAN DATA

Ponder Problem statement

Observe features

Look over data

# HYPOTHESIS

Form possible hypothesis based on intuition  
Get data ready to verify hypothesis.

# PREPROCESSING

Observe columns with null values and their occurrences.

Check for uniformity of Datatype

Remove redundant columns (ex: skip\_logic\_failed)

# VALIDATIONS

Age having negative values (Set range  $>4$   $<114$ )

Age  $>$  Dropout age + 3

Jandhan yojna yes, bank acc no

# FINDING DEPENDENCIES

Univariate Feature selection (Correlation matrix)

PCA-principal components

DecisionTree.tree\_.decision\_path

RandomForest.Feature\_importances

# INSIGHTS

1. Scheme name: 36.3% and 29.4% Scholarship in Dropped out.
2. 97.24% Taken Scholarship didn't drop out.
3. Age: 88.6 % Males dropout-57% after the age of 18.
4. 88.8% Female dropout-59.3% after the age of 18.
5. Lifeins (Life insurance): 0.12% took insurance and then dropped out.
6. 99% taking insurance dropped out too.
7. Groupname\_14: Tribal\_origination maximum dropouts %.
8. Malaria: 50% dropouts due to Malaria.
9. 89% Malaria caused dropouts.
10. Hearing problem, suddenillness maximum disease type caused dropouts,  
90.45% Hearing problems ,dropped out
11. OccChild: (Max correlation) Maximum family business.
12. Occupation of the members: Maximum dropout from earnyn - 96.39% of government jobs dropouts.