***HACKATHON***

# *WAGE CLASSIFICATION*

**CHALERMCHON WONGSOPHA**

**WARINTORN NAWONG**

# OUR DATA

**32,561**

Number of Wage Information.

**6**

Number of Numerical Columns.

**7**

Number of Categorical Columns


We have single tools **"RANDOM FOREST CLASSIFIER"** to do Customer Segmentation by wage for finding the potential customers in our company.

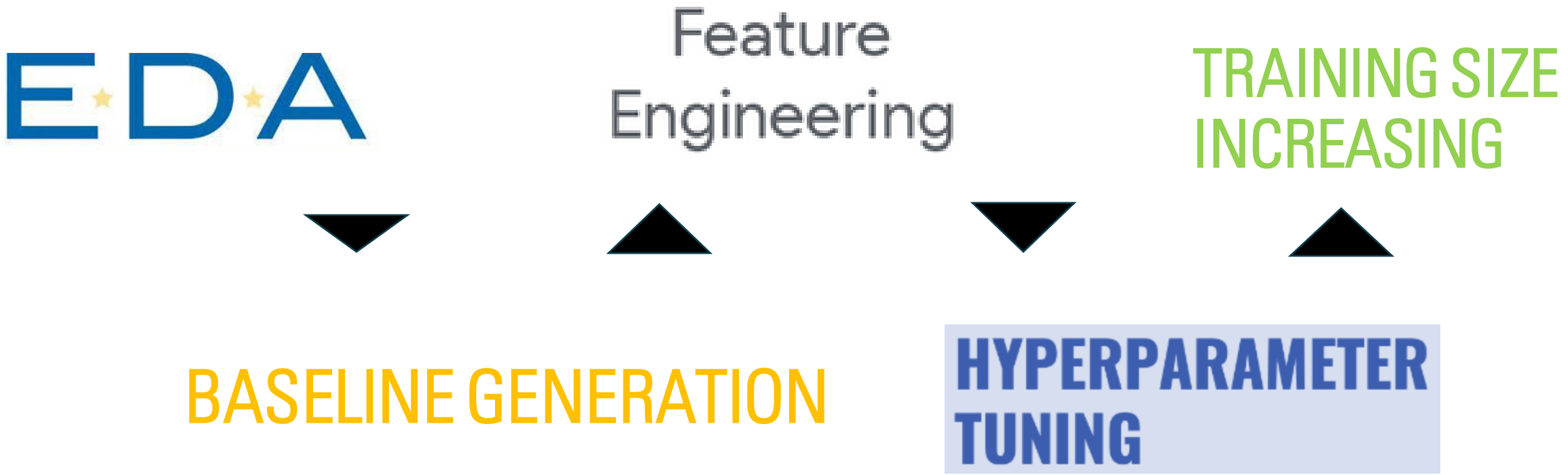
# OUR CONSTRAINTS

Our Data Scientist has studied from the **" Low Grade Online platform"** which they taught only **"Random Forest Classifier"** and we have no time to teach them more algorithms.


- Loss of Explainability.
- Inaccessible to more complex,
  high predictive power algorithms.

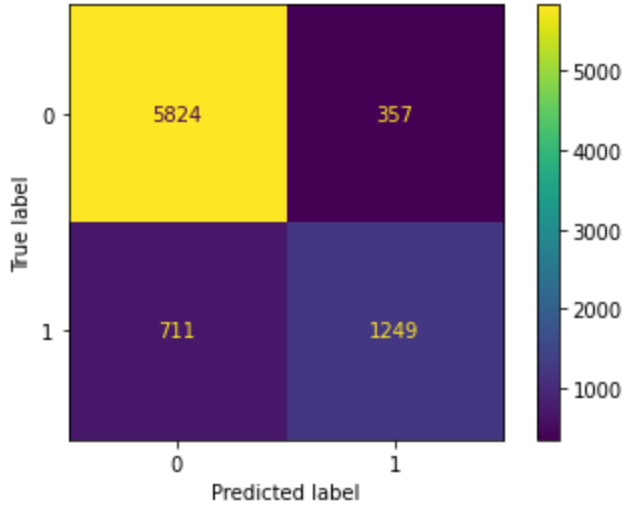

- More time to concentrate on others issue i.e. Data Quality , Feature engineering.

# PROCESS & DECISION ROADMAP



# FINAL MODEL SUMMARY

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.89 | 0.94 | 0.92 | 6181 |
| 1 | 0.78 | 0.64 | 0.70 | 1960 |
| accuracy | | | 0.87 | 8141 |
| macro avg | 0.83 | 0.79 | 0.81 | 8141 |
| weighted avg | 0.86 | 0.87 | 0.86 | 8141 |




## HYPERPARAMETER.

MAX DEPTH = 40

MIN SAMPLE LEAF = 3

MIN SAMPLE SPLIT = 12

N_ESTIMATORS = 100

## SUBMISSION SCORES (MACRO AVG)

1. ***FEATURE ENGINEERING WITHOUT GRIDSEARCH : 85.26 %***
2. ***FEATURE ENGINEERING WITH GRIDSEARCH : 89.87 %***
3. ***COMBINE ALL TRAINING DATA : 89.94 %.***

# *THANK YOU*