CREDIT SCORING PADA DATA NASABAH

CIMB NIAGA

BY

SURABAYAMINING

MEMBERS

GRACE

JONATHAN

STEPHEN

MODEL DEVELOPMENT PROCESS

DATA EXPLORATION	→	DROP UNRELEVANT COLUMN	→	CREATE NEW COLUMN

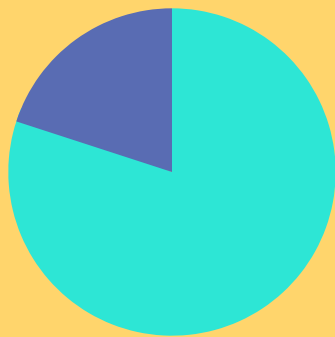| MODELING	←	MERGE DATASET	←	HANDLE MISSING VALUES
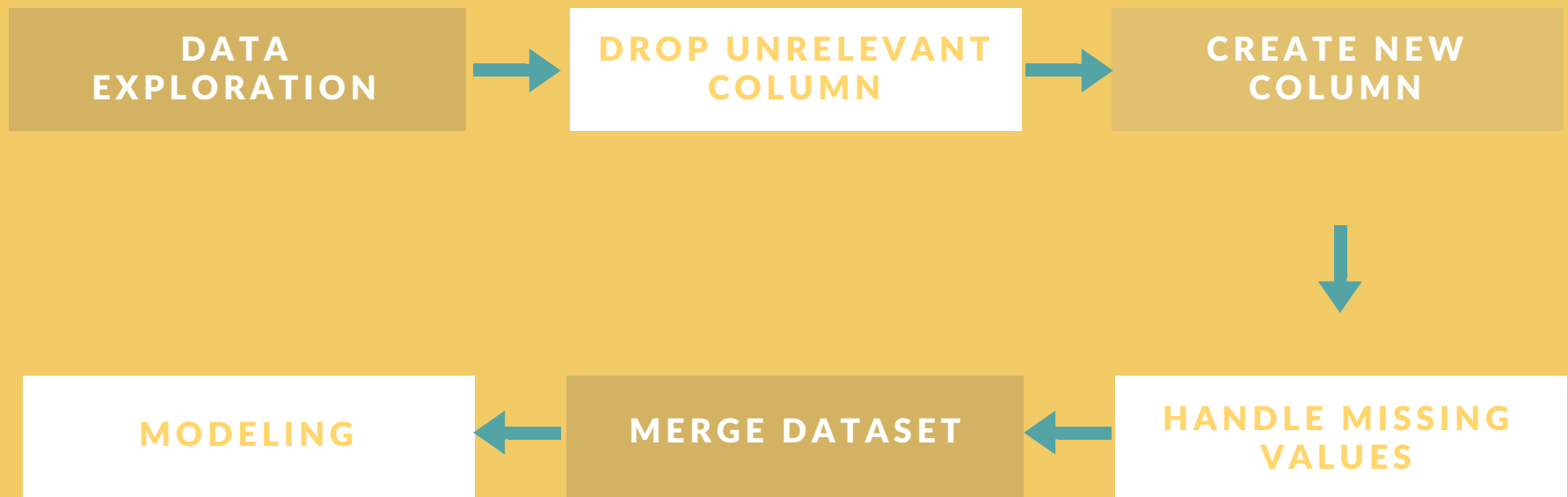
MODEL SELECTION PROCESS

SPLITTING DATASET

Validation Set
20%

Training Set
80%

We using MinMaxScaler for normalizing the data.

We use 3 different approach and model, which are:
1. Random Forest Classifier vs Extra Tree Classifier

2. SMOTE oversampling + (AdaBoost Classifier vs Random Forest Classifier vs Logistic Regression)

3. Model Stacking using Random Forest Classifier & Extra Tree Classifier

Evaluation metrics that are used for the models are Recall and Area Under Curve (AUC).

MODEL PERFORMANCE

	Recall	AUC
Approach 1 - Random Forest Classifier	0.77	0.72
Approach 1 - Extra Tree Classifier	0.80	0.70
Approach 2 - AdaBoost Classifier	0.20	0.69
Approach 2 - Random Forest Classifier	0.58	0.68
Approach 2 - Logistic Regression	0.18	0.66
Approach 3 - Random Forest Classifier & Extra Tree Classifier	0.82	0.71

Approach and Models Results

SURABAYA
MINING

FINAL MODEL

- Model Stacking using Random Forest Classifier & Extra Tree Classifier (Approach 3) has the most highest recall value and AUC value.

- Data is not being oversampled nor undersampled.

CONFUSION MATRIX

1286	1168
33	153

RECALL = 0.82

AUC = 0.71

BUSINESS PROBLEM & SOLUTION

Model accurately catch delinquency rate around 80%.

Reducing delinquency rate from 7% to 1.4%.

Using robust algorithm with a short time process.

DELINQUENCY RATE

Yes
7%

No
93%

THANK YOU!

"WITHOUT BIG DATA, YOU ARE BLIND AND DEAF AND IN THE MIDDLE OF A FREEWAY" – GEOFFREY MOORE