Targeted Campaign to Optimize operational cost

Aqilah, Feby Genta, M Fadly

Table of Content

Background (I)

- A company just finished the new product campaign to all eligible customers
- Historically, 31.5K eligible customers are being contacted via Telemarketing every month
- Only handful of customers that took the offers (13%)
- Cost breakdown of Campaign:

Type of Campaign	#Tele Agent	Cost Yearly / Tele Agent	Yearly Cost
Telemarketing	11*	Rp. 52.800.000*	Rp. 580.800.000

based on common capacity 130 calls / day / agent with minimum wage around Rp 4.400.000 for 31.500 customer per month
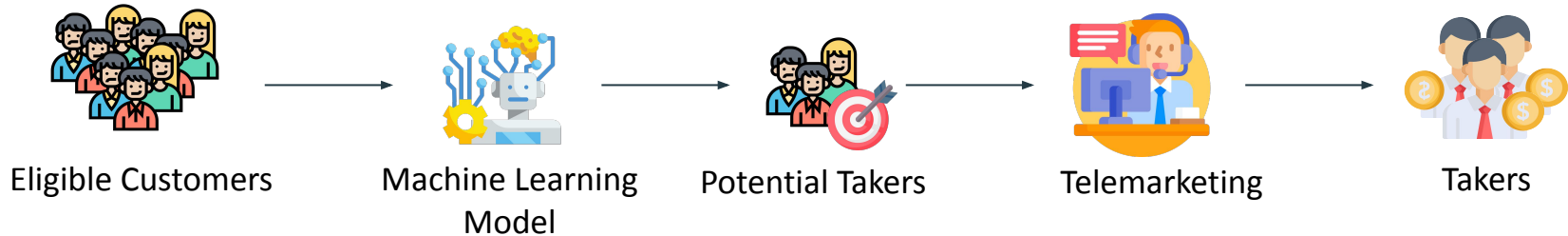
Background (II)

- Company need to find a way to optimize the next campaign based on the current result
- One of the strategy is to do more targeted campaign using machine learning model
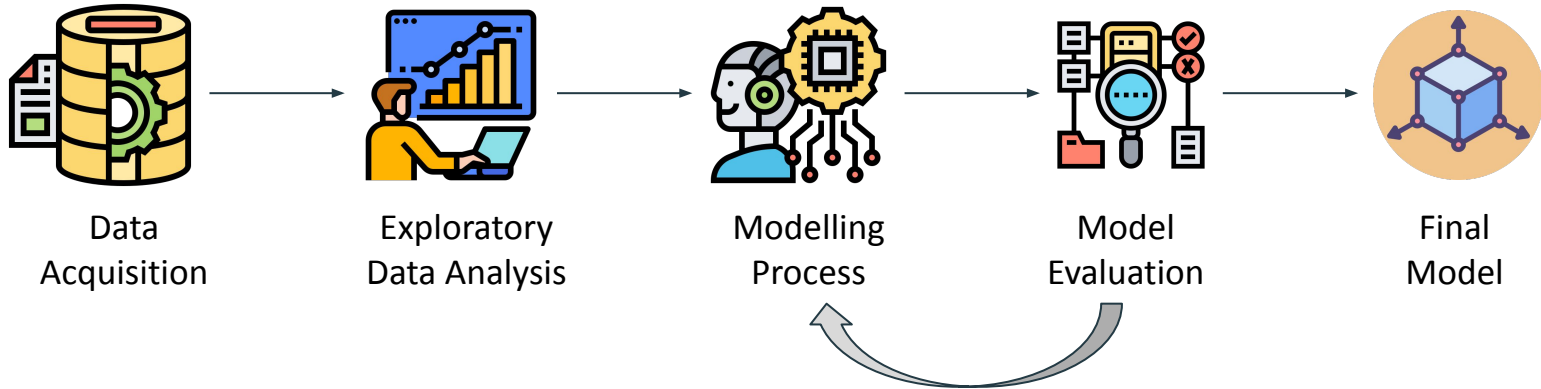
Current Campaign Flow

Eligible Customers → Telemarketing → Takers

Proposed New Campaign Flow

Eligible Customers → Machine Learning Model → Potential Takers → Telemarketing → Takers

Objective

- Generate machine learning model to design more targeted campaign

Workflow

Data Acquisition → Exploratory Data Analysis → Modelling Process → Model Evaluation → Final Model

Data Acquisition

- Source: Marketing campaign result database (obtained from internet)
- Target is imbalance: Only 13% customers took the campaign
- Consist of : Demography, financial condition, campaign history

```
RangeIndex: 31480 entries, 0 to 31479
Data columns (total 20 columns):
 #   Column                        Non-Null Count   Dtype
---  ------                        --------------   -----
 0   id                            31480 non-null   int64
 1   target                        31480 non-null   object
 2   day                           31480 non-null   int64
 3   month                         31480 non-null   object
 4   duration                      31480 non-null   int64
 5   contactId                     31480 non-null   int64
 6   age                           31480 non-null   int64
 7   gender                        31480 non-null   object
 8   job                           31480 non-null   object
 9   maritalStatus                 31480 non-null   object
 10  education                     31480 non-null   object
 11  creditFailure                 31480 non-null   object
 12  accountBalance                31480 non-null   int64
 13  house                         31480 non-null   object
 14  credit                        31480 non-null   object
 15  contactType                   31480 non-null   object
 16  numberOfContacts              31480 non-null   int64
 17  daySinceLastCampaign          5738 non-null    float64
 18  numberOfContactsLastCampaign  31480 non-null   int64
 19  lastCampaignResult            31480 non-null   object
```

Data Preprocessing

- Ordinal variable converted and encoded into Categorical column :

Education	Credit Failure	House	Month

- Nominal variable is encoded into dummy columns :

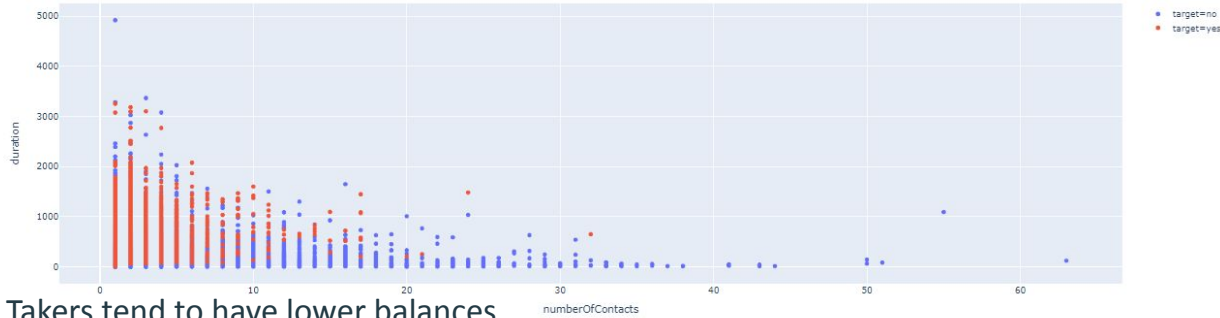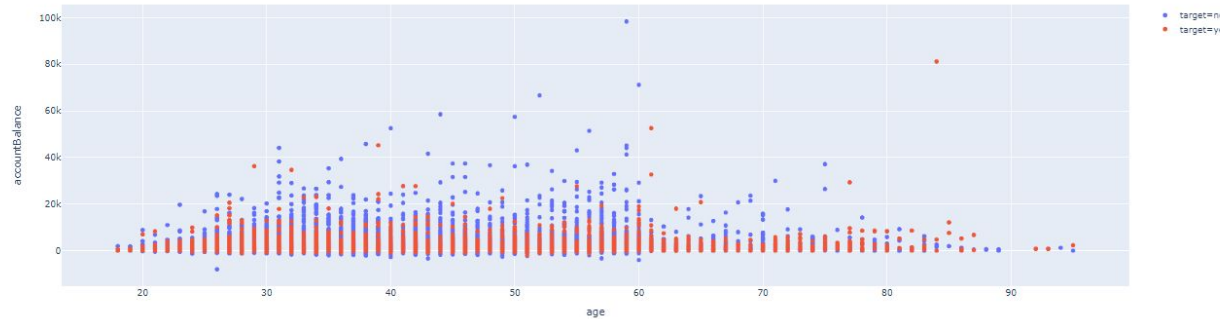|				
Gender	Marital Status	Contact Type	Job	Last Campaign Result

Exploratory Data Analysis (EDA)
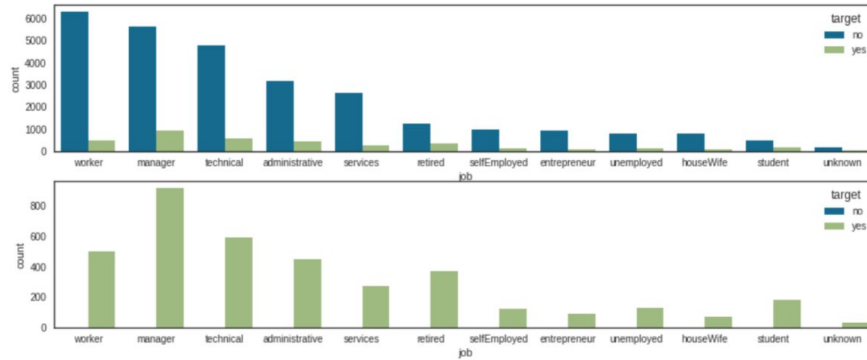
- Takers tend to spend more time to talk with Tele Agent
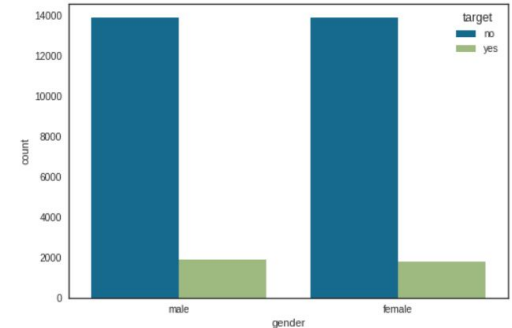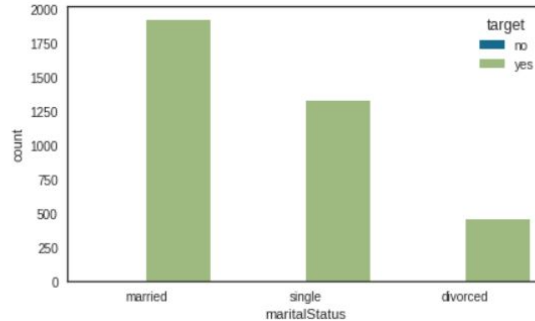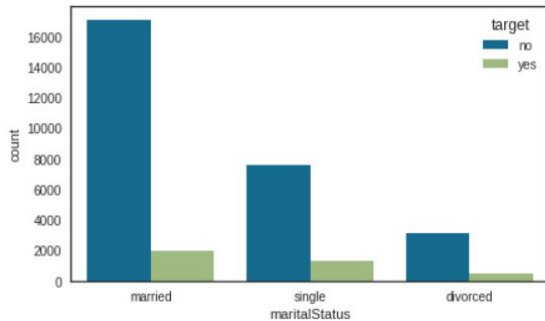

- Takers tend to have lower balances

EDA (II)

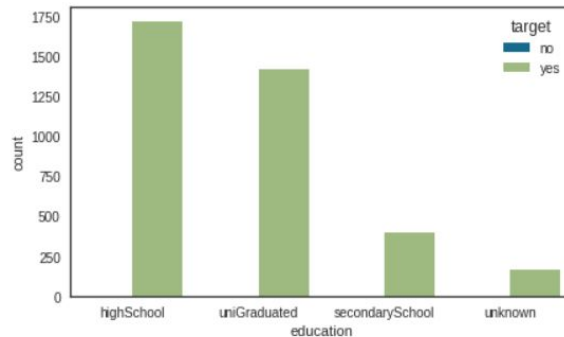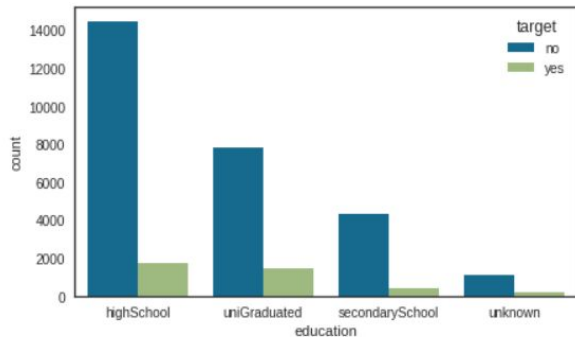● Workers got the campaign the most while manager and technician took the campaign more

● No visible relationship between marital status or gender with campaign result
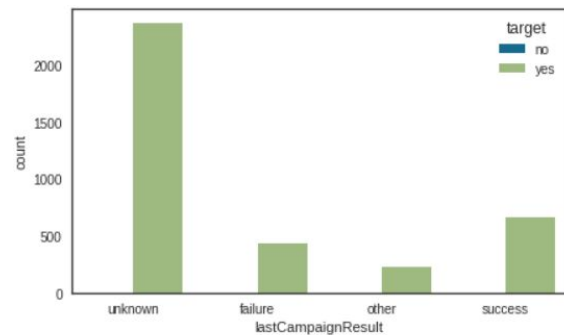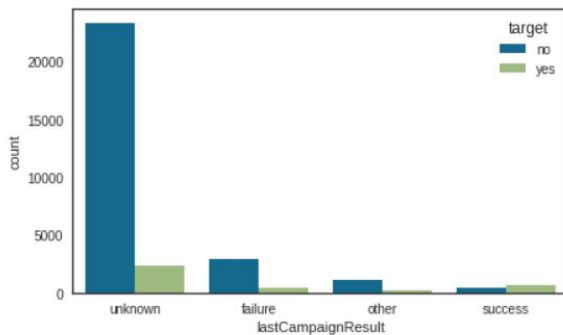
EDA (III)

- No visible relationship between education and campaign result

- No visible relationship between last campaign result and current campaign result

EDA (IV)

- There's no multicollinearity between variables
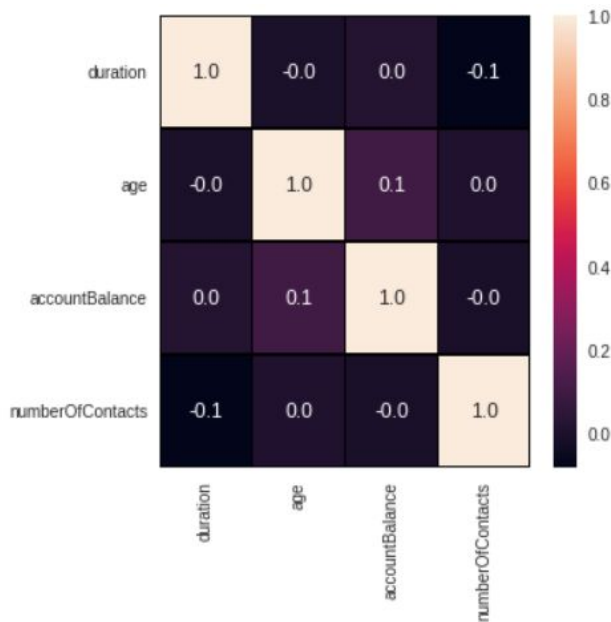
- Top 10 variable with highest correlation with target variable

	index	target_yes
0	target_yes	1.000000
1	duration	0.394059
2	lastCampaignResult_success	0.299590
4	house_no	0.136624
5	contactType_cellPhone	0.135108
6	numberOfContactsLastCampaign	0.084866
7	job_retired	0.080035
9	education_uniGraduated	0.072252
10	job_student	0.071846

Data Modelling

- Data is split into train set (70%) and test set (30%)
- SMOTE is applied into train-set to optimize the training
- Training is done through 13 different statistical model:

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
lightgbm	Light Gradient Boosting Machine	0.9126	0.9330	0.4830	0.6336	0.5472	0.4998	0.5060	1.065
gbc	Gradient Boosting Classifier	0.9090	0.9227	0.4860	0.6069	0.5387	0.4890	0.4933	9.596
rf	Random Forest Classifier	0.9075	0.9241	0.3527	0.6427	0.4548	0.4089	0.4313	4.239
ada	Ada Boost Classifier	0.9034	0.9003	0.4550	0.5744	0.5073	0.4545	0.4586	2.217
et	Extra Trees Classifier	0.9031	0.9103	0.3510	0.5974	0.4411	0.3921	0.4093	3.620
dummy	Dummy Classifier	0.8907	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000	0.100
dt	Decision Tree Classifier	0.8781	0.6973	0.4659	0.4452	0.4545	0.3861	0.3867	0.436
ridge	Ridge Classifier	0.8587	0.0000	0.7815	0.4219	0.5477	0.4729	0.5046	0.110
lda	Linear Discriminant Analysis	0.8587	0.9048	0.7815	0.4219	0.5477	0.4729	0.5046	0.360
lr	Logistic Regression	0.8361	0.9029	0.8208	0.3840	0.5230	0.4395	0.4859	4.969
nb	Naive Bayes	0.8095	0.8077	0.6403	0.3170	0.4238	0.3251	0.3541	0.108
knn	K Neighbors Classifier	0.7741	0.7478	0.6023	0.2653	0.3683	0.2553	0.2872	0.599
svm	SVM - Linear Kernel	0.7552	0.0000	0.7313	0.2866	0.4024	0.2908	0.3451	0.695
qda	Quadratic Discriminant Analysis	0.1093	0.5000	1.0000	0.1093	0.1971	0.0000	0.0000	0.183

Model Evaluation

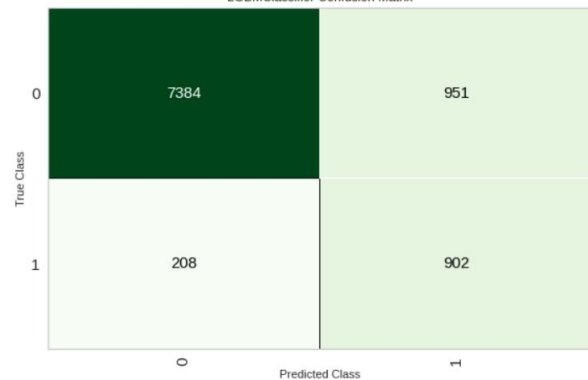- Hyperparameter Tuning is done on LGBM to optimize the score

Classification Report
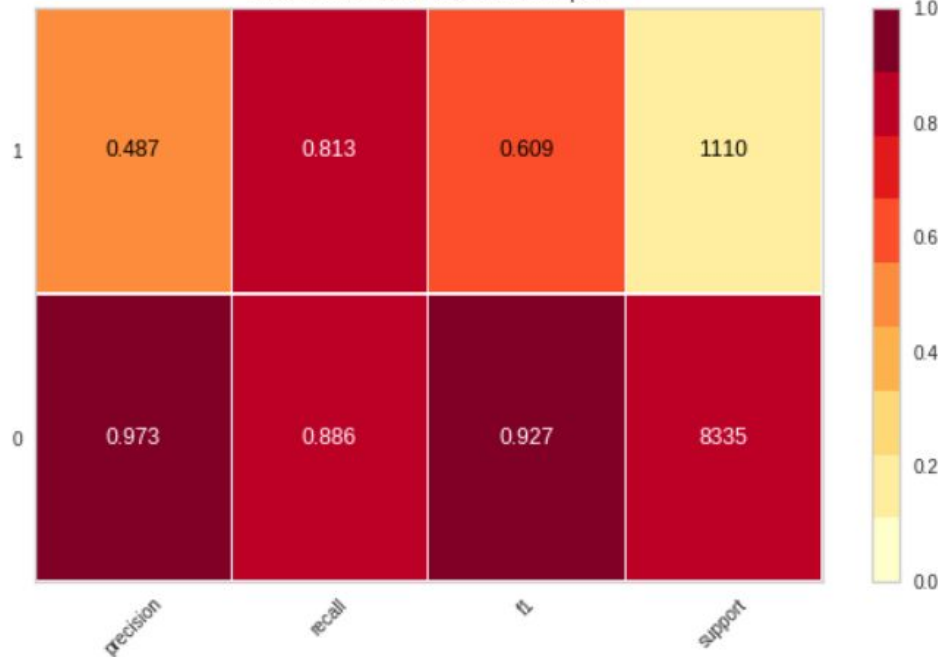
Confusion Matrix

Top 5 Features

Duration   Account Balance   Day   Age   Day Since Last Campaign

Conclusion

- The most optimal algorithm is LGBM (Light Gradient Boosting Machine)
- The best performance is 81% (Recall score)
- Only 19% of population that is need to be targeted on the next campaign
- Total cost saving up to Rp 475.200.000 (81%)

Parameter	Current	Proposed
Population	31.500	5.985
Need in Tele Agent	11	2
Cost Yearly / Tele Agent	Rp 52.800.000	Rp 52.800.000
Total Cost	Rp 580.800.000	Rp 105.600.000
Yearly saving	Rp 475.200.000 (81%)	

Appendix

Tele Agent Cost Calculation - Current

Leads:	31,500 / month
Leads / day	1,400
Capacity / Agent / day:	130
Minimum Agent to achieve 31.500 lead in a month:	1,400 : 130 = 11 Person
Minimum Wage:	4,400,000
Yearly cost of 11 Person in a year	580,800,000

Tele Agent Cost Calculation - Proposed

Leads:	5,985 / month
Leads / day	272
Capacity / Agent / day:	130
Minimum Agent to achieve 31.500 lead in a month:	272 : 130 = 2 Person
Minimum Wage:	4,400,000
Yearly cost of 2 Person in a year	105,600,000