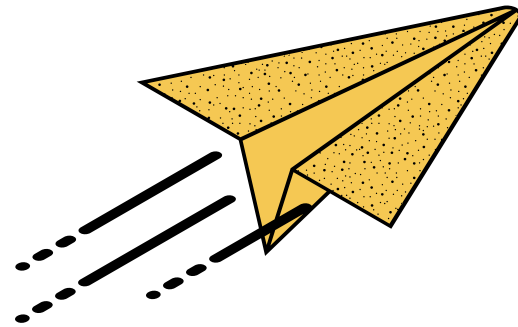
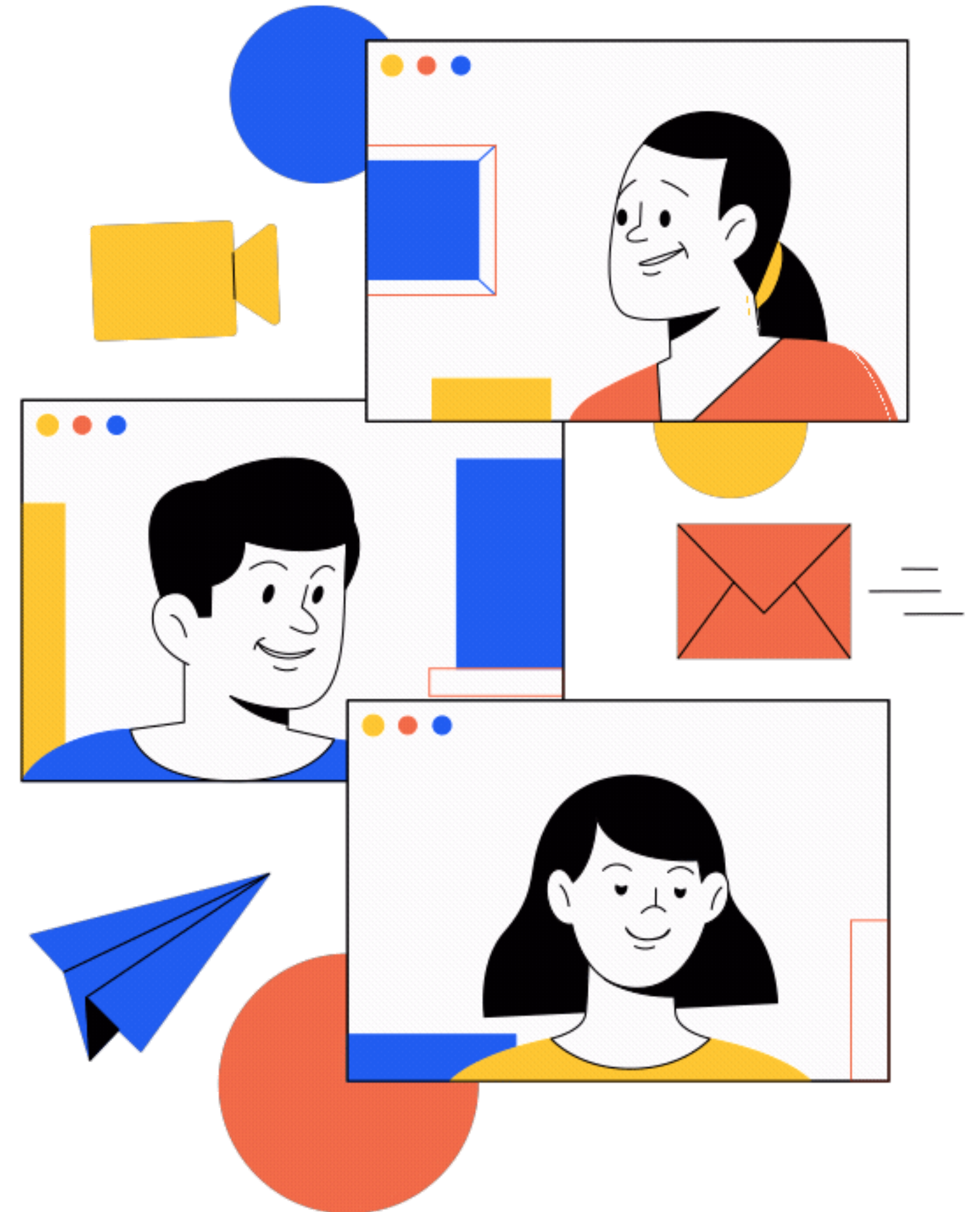


SAINS DATA



# MEMPREDIKSI PENINGKATAN PEMASARAN BANK

Kecerdasan Artificial - Kelompok 5



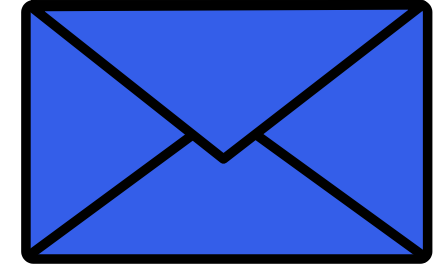


# **KELOMPOK 5**

**Ahmad Rafi Syaifudin 22031554030**

**Riva Dian Ardiansyah 22031554043**

**Analicia 2201444007**



# KONTRIBUSI KELOMPOK 5



## ANALICIA

- *LAPORAN*
- *MODELLING*
- *SPLITING DATA*
- *PPT*
- *DATA PREPROCESSING  
(ENCODING)*
- *HYPERPARAMETER  
TUNING*

## RIVA

- *STREAMLIT*
- *DATA COLLECTING*
- *POWER POINT*
- *CONFUSION MATRIX*
- *FEATURE  
IMPORTANCE*

## RAFI

- *STREAMLIT*
- *CLEANING*
- *POWER POINT*
- *EDA*
- *UNDERSAMPLING*
- *HYPERPARAMETER  
TUNING*

# AGENDA

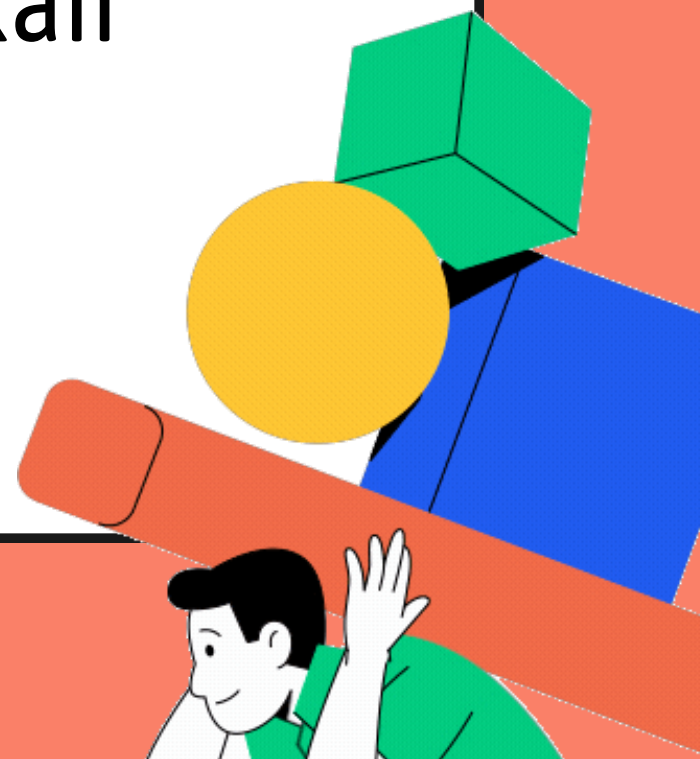
1. DATA CLEANING
2. EVALUASI MODEL
3. KOMPARASI MODEL
4. VALIDASI MODEL
5. DEPLOYMENT STREAMLIT



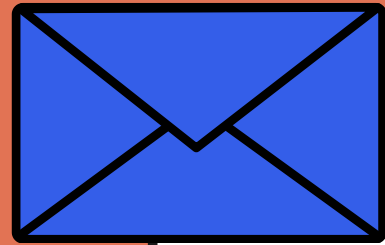
# LATAR BELAKANG

Project ini membahas berbagai jenis model algoritma untuk pengklasifikasian dengan menggunakan beberapa model, yaitu Random Forest, Decission Tree, dan Logistic Regression. Dengan tujuan untuk mengevaluasi dan menerapkan metode tersebut.

Data yang digunakan berisi pemasaran suatu institusi bank. Project ini memprediksi terhadap suatu klien bank yang akan berlangganan deposito berjangka panjang (variabel  $y$ ) berdasarkan atribut pemasaran oleh institusi bank.







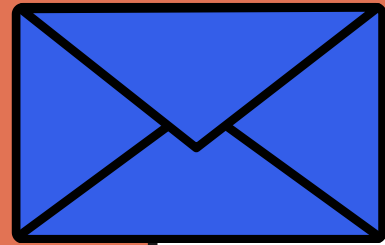
- Dimensi Data

# CLEANING DATA

Dataset nya terdiri dari 17 kolom 41rb baris

```
[7] # Ringkasan dataset (tipe masing-masing kolom pada data)  
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 41188 entries, 0 to 41187  
Data columns (total 21 columns):  
#   Column                Non-Null Count  Dtype  
---  ---  
0   age                    41188 non-null  int64  
1   job                    41188 non-null  object  
2   marital                41188 non-null  object  
3   education              41188 non-null  object  
4   default                41188 non-null  object  
5   housing                41188 non-null  object  
6   loan                   41188 non-null  object  
7   contact                41188 non-null  object  
8   month                  41188 non-null  object  
9   day_of_week            41188 non-null  object  
10  duration                41188 non-null  int64  
11  campaign                41188 non-null  int64  
12  pdays                  41188 non-null  int64  
13  previous                41188 non-null  int64  
14  poutcome               41188 non-null  object  
15  emp.var.rate            41188 non-null  float64  
16  cons.price.idx          41188 non-null  float64  
17  cons.conf.idx           41188 non-null  float64  
18  euribor3m               41188 non-null  float64  
19  nr.employed             41188 non-null  float64  
20  y                       41188 non-null  object  
dtypes: float64(5), int64(5), object(11)  
memory usage: 6.6+ MB
```



- **Dimensi Data**

# DATA EXPLORATION

Dataset nya terdiri dari 17 kolom 41rb baris

```
[7] # Ringkasan dataset (tipe masing-masing kolom pada data)  
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 41188 entries, 0 to 41187  
Data columns (total 21 columns):  
#   Column                Non-Null Count  Dtype    
---  ---                  
0    age                   41188 non-null  int64    
1    job                   41188 non-null  object   
2    marital               41188 non-null  object   
3    education             41188 non-null  object   
4    default               41188 non-null  object   
5    housing               41188 non-null  object   
6    loan                  41188 non-null  object   
7    contact               41188 non-null  object   
8    month                 41188 non-null  object   
9    day_of_week           41188 non-null  object   
10   duration               41188 non-null  int64    
11   campaign               41188 non-null  int64    
12   pdays                 41188 non-null  int64    
13   previous               41188 non-null  int64    
14   poutcome               41188 non-null  object   
15   emp.var.rate           41188 non-null  float64   
16   cons.price.idx          41188 non-null  float64   
17   cons.conf.idx           41188 non-null  float64   
18   euribor3m              41188 non-null  float64   
19   nr.employed             41188 non-null  float64   
20   y                       41188 non-null  object   
dtypes: float64(5), int64(5), object(11)  
memory usage: 6.6+ MB
```

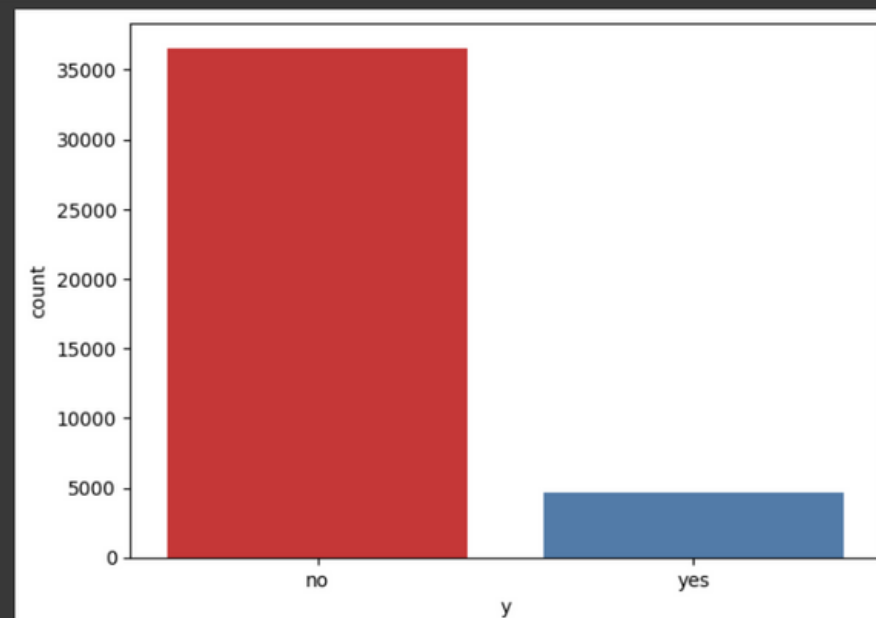
# DATA VISUALIZATION

```
[13] df['y'].value_counts()
```

```
no    36537  
yes    4639  
Name: y, dtype: int64
```

```
[14] # Membuat visualisasi data
```

```
f, ax = plt.subplots(figsize=(7, 5))  
ax = sns.countplot(x="y", data=df, palette="Set1")  
plt.show()
```



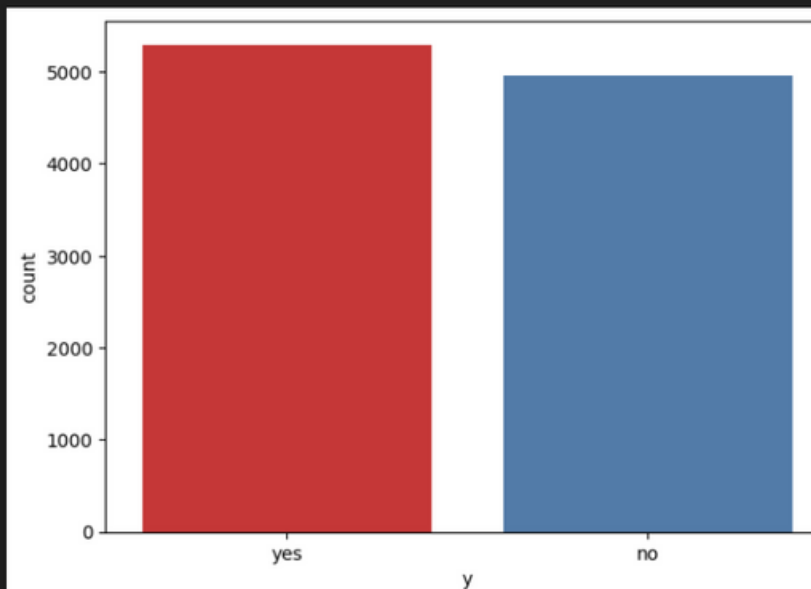
Data sebelum di resample diketahui "no" terdapat 36537 sedangkan "yes" 4639

```
df['y'].value_counts()
```

```
... yes    5289  
    no    4952  
    Name: y, dtype: int64
```

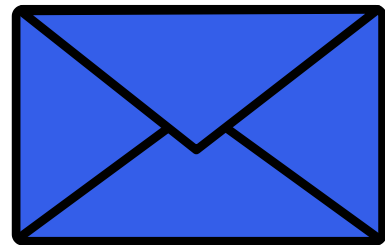
```
# Membuat visualisasi data
```

```
f, ax = plt.subplots(figsize=(7, 5))  
ax = sns.countplot(x="y", data=df, palette="Set1")  
plt.show()
```



Data sesudah di resample diketahui "no" terdapat 4952 sedangkan "yes" 5289





# EVALUASI MODEL



## • Random Forest

Classification Report:

	precision	recall	f1-score	support
0	0.83	0.80	0.82	1402
1	0.83	0.85	0.84	1520
accuracy			0.83	2922
macro avg	0.83	0.83	0.83	2922
weighted avg	0.83	0.83	0.83	2922

Accuracy Score:- 0.8299110198494182  
Precision Score:- 0.8255887969446213  
Recall Score:- 0.8532894736842105

Pada random forest dapat digunakan untuk menganalisis faktor dan strategi pemasaran bank terhadap pelanggan. Selain itu, dapat membantu dalam memprediksi risiko pelanggan terhadap deposito sehingga dari pihak bank dapat membuat keputusan yang lebih baik. Maka random forest ini dapat digunakan untuk memprediksi dan analisis terhadap peningkatan strategi pemasaran.

## • Logistic Reggresion

Classification Report:

	precision	recall	f1-score	support
0	0.79	0.84	0.82	1402
1	0.85	0.80	0.82	1520
accuracy			0.82	2922
macro avg	0.82	0.82	0.82	2922
weighted avg	0.82	0.82	0.82	2922

Accuracy Score:- 0.8193018480492813  
Precision Score:- 0.8463687150837989  
Recall Score:- 0.7973684210526316

Logistic regression adalah suatu model statistik umum tentunya digunakan untuk menganalisis. Logistic regression digunakan untuk memodelkan suatu hubungan antara variable dependen yang memiliki dua nilai yang mungkin, seperti ya atau tidak, 0 atau 1, dan lain-lain. Logistic regression juga digunakan untuk memahami faktor-faktor yang signifikan terhadap strategi dalam pemasaran.

## • Decision Tree

Classification Report:

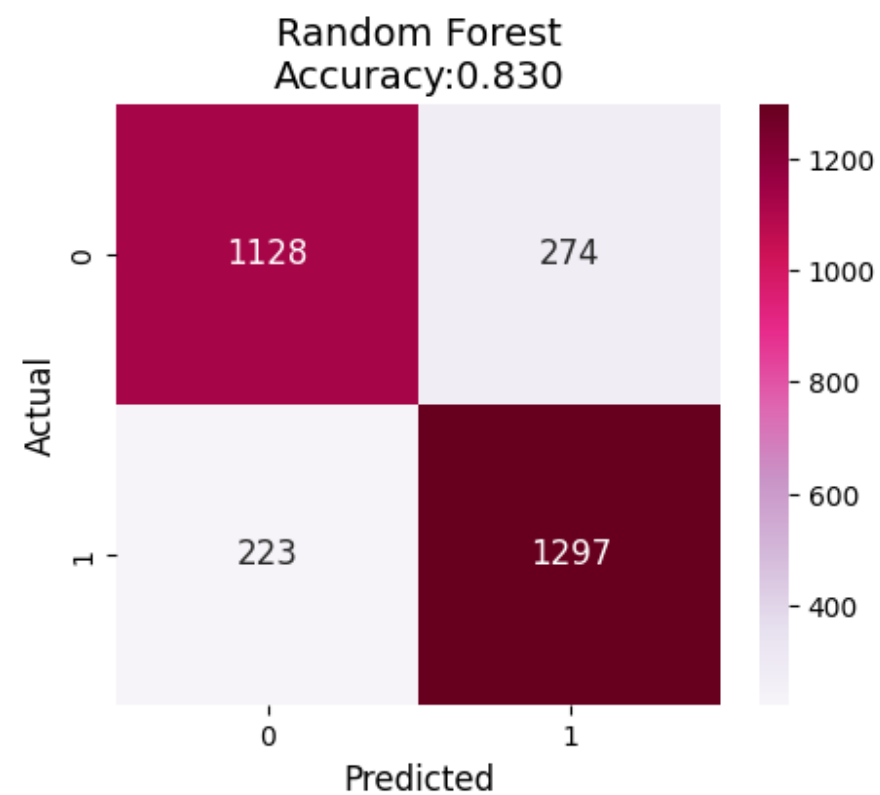
	precision	recall	f1-score	support
0	0.74	0.74	0.74	1402
1	0.76	0.76	0.76	1520
accuracy			0.75	2922
macro avg	0.75	0.75	0.75	2922
weighted avg	0.75	0.75	0.75	2922

Accuracy Score:- 0.7494866529774127  
Precision Score:- 0.7612732095490716  
Recall Score:- 0.7552631578947369

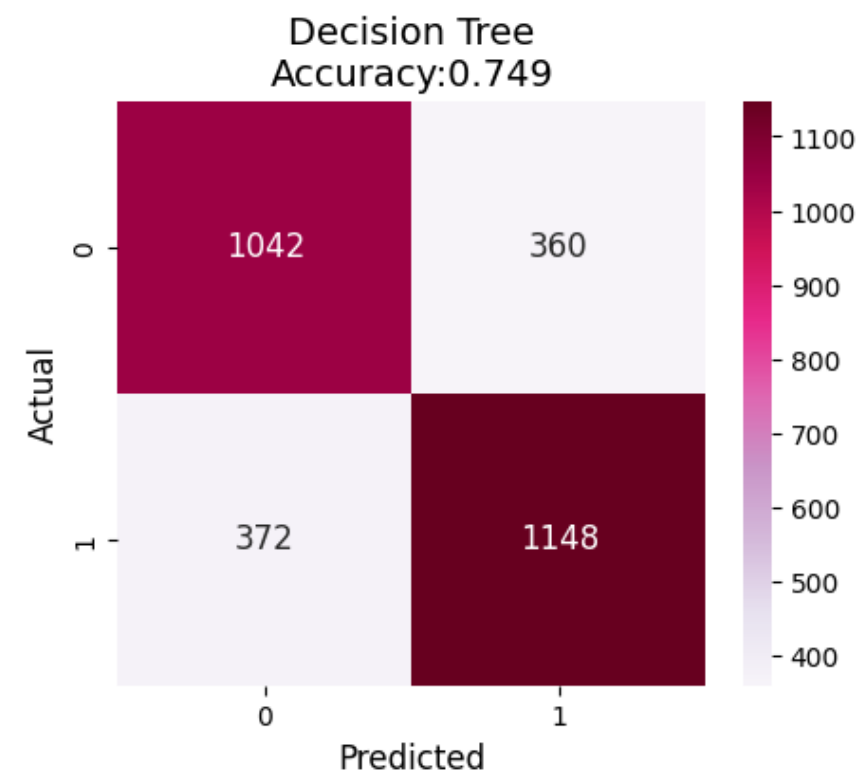
Decision tree merupakan machine learning yang juga digunakan untuk memprediksi atau mengklasifikasi data dengan mengambil keputusan dengan cara yang mirip dengan pohon keputusan atau random forest. Decision tree mudah melakukan interpretasi dan visualisasi yang memudahkan dalam pengambilan keputusan, juga tidak memiliki tahap preprocessing yang rumit dan memiliki fitur otomatis dengan cara mengidentifikasi fitur yang penting dalam pembuatan keputusan.

# KORELASI MATRIX

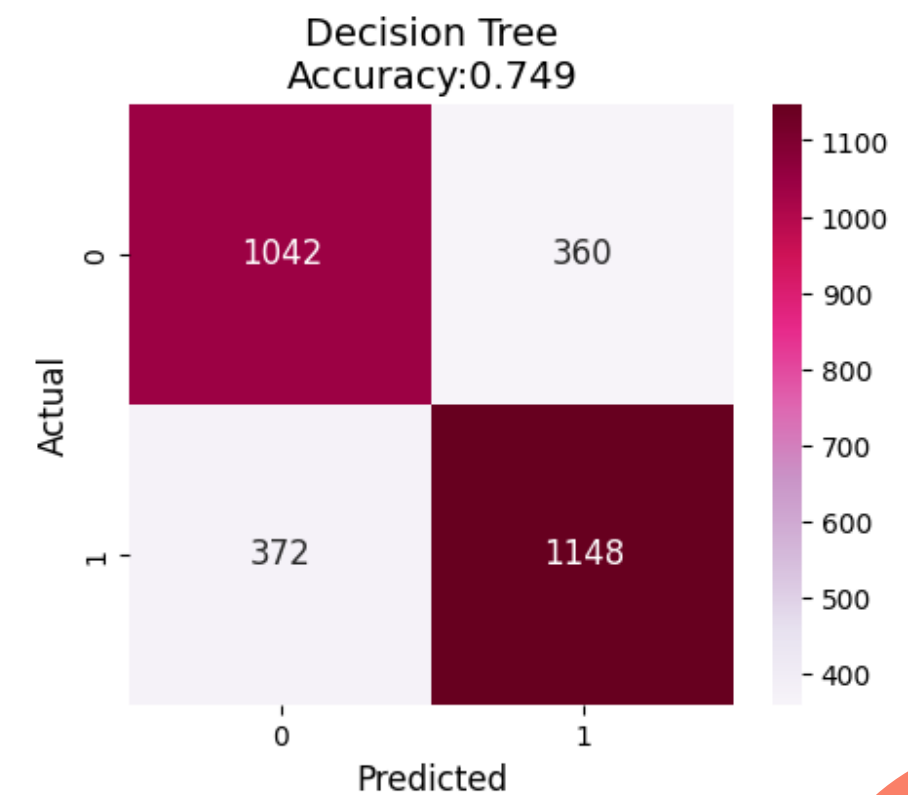
## • RANDOM FOREST



## • LOGISTIC REGRESSION



## • DECISION TREE



# HASIL HYPERPARAMETER TUNING

## • RANDOM FOREST

[[1127 275]  
[ 217 1303]]

	precision	recall	f1-score	support
0	0.84	0.80	0.82	1402
1	0.83	0.86	0.84	1520
accuracy			0.83	2922
macro avg	0.83	0.83	0.83	2922
weighted avg	0.83	0.83	0.83	2922

Accuracy: 0.8316221765913757

## • DECISION TREE

[[ 959 443]  
[ 236 1284]]

	precision	recall	f1-score	support
0	0.80	0.68	0.74	1402
1	0.74	0.84	0.79	1520
accuracy			0.77	2922
macro avg	0.77	0.76	0.76	2922
weighted avg	0.77	0.77	0.77	2922

Accuracy: 0.7676249144421629

## • LOGISTIC REGRESSION

Akurasi: 0.8186173853524983  
Presisi: 0.849083215796897  
Recall: 0.7921052631578948  
F1 Score: 0.8196051735874744

Classification Report:

	precision	recall	f1-score	support
0	0.79	0.85	0.82	1402
1	0.85	0.79	0.82	1520
accuracy			0.82	2922
macro avg	0.82	0.82	0.82	2922
weighted avg	0.82	0.82	0.82	2922

# STREAMLIT



### Settings

Select Algorithm

Random Forest

### Random Forest Hyperparameters

Number of trees:

100

1 300

Maximum depth of trees:

10

1 20

### Make a Prediction

age

-0.21 - +

duration

-0.34 - +

## Hallo Selamat Datang di Dashboard saya

Perkenalkan kami kelompok 5 Ahmad Rafi Syaifudin (22031554030), Riva Dian Ardiansyah (22031554043), Analicia (22031554007)

Kita ingin membuat prediksi atas Pemasaran bank untuk mengetahui deposit pengguna

## Deposit Prediction App

X\_data shape:

(2923, 39)

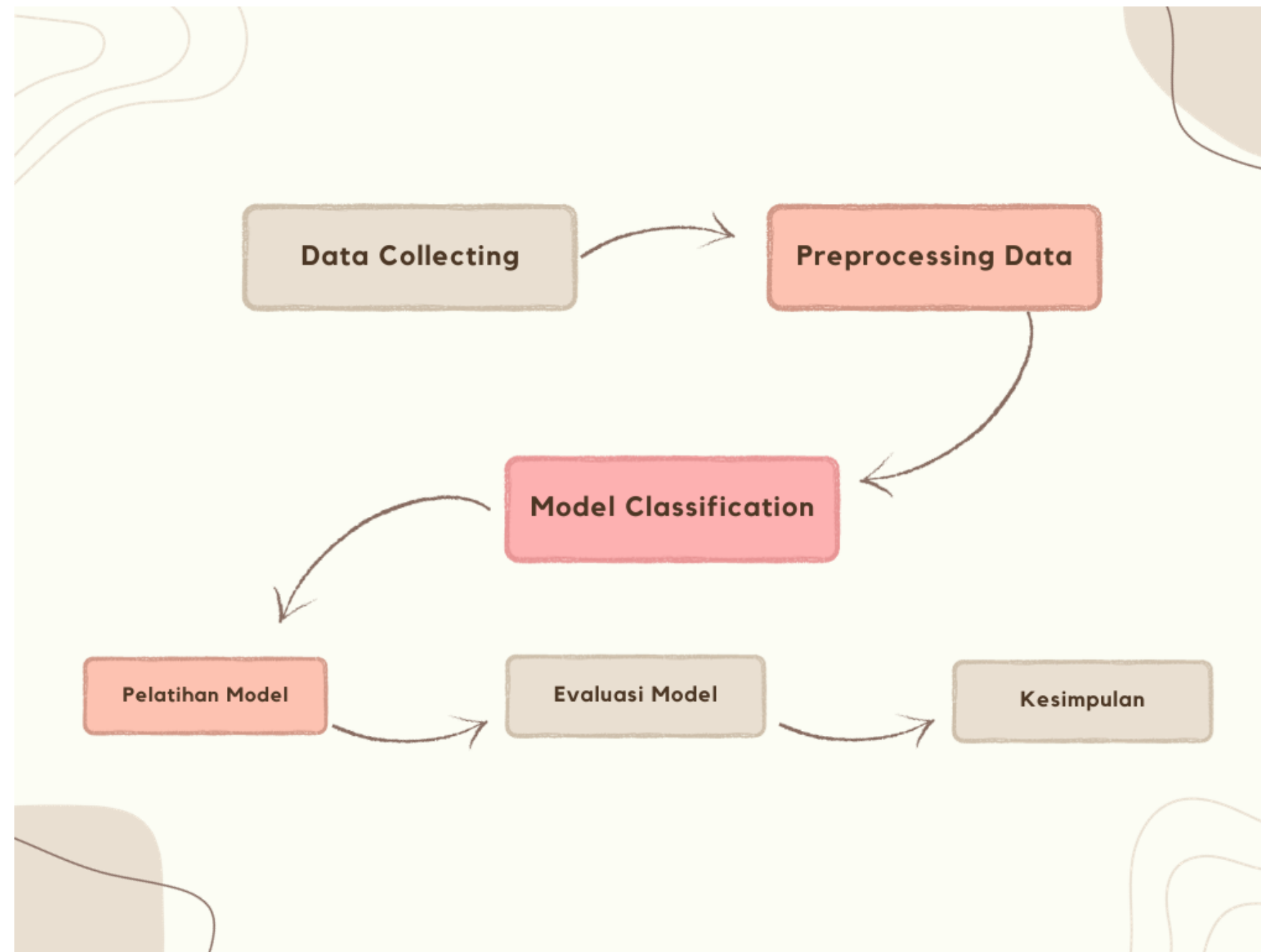
y\_data shape:

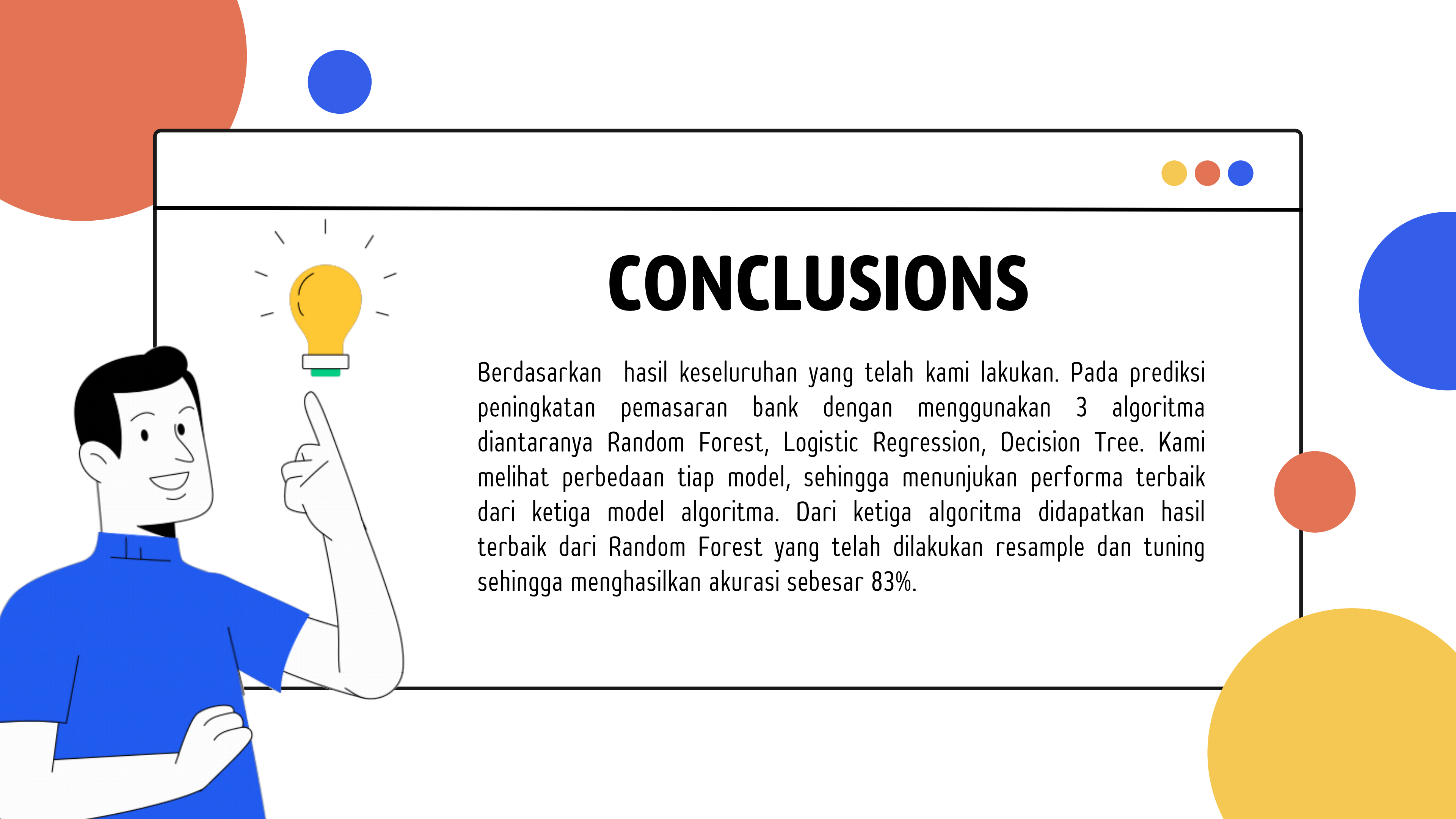
(2923, 1)

Model train yang digunakan Random Forest.



# DIAGRAM ALIR



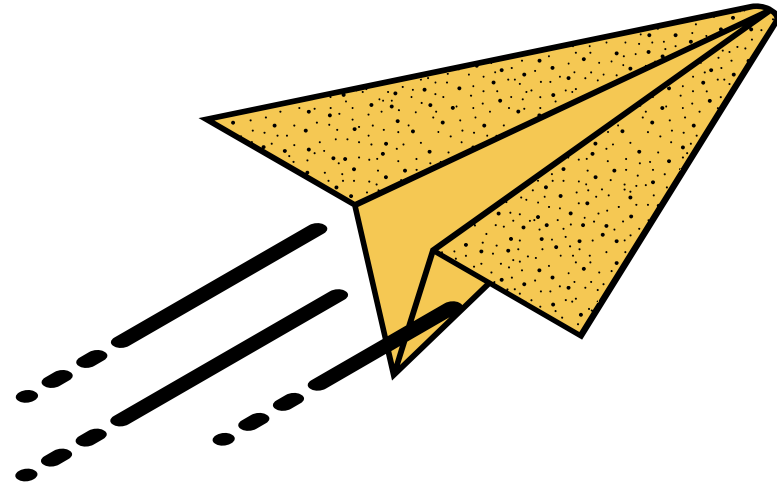


# CONCLUSIONS

Berdasarkan hasil keseluruhan yang telah kami lakukan. Pada prediksi peningkatan pemasaran bank dengan menggunakan 3 algoritma diantaranya Random Forest, Logistic Regression, Decision Tree. Kami melihat perbedaan tiap model, sehingga menunjukan performa terbaik dari ketiga model algoritma. Dari ketiga algoritma didapatkan hasil terbaik dari Random Forest yang telah dilakukan resample dan tuning sehingga menghasilkan akurasi sebesar 83%.



[www.reallygreatsite.com](http://www.reallygreatsite.com)



# THANK YOU!

