# Predict Clicked Ads Customer Classification by using Machine Learning

Supported by:
Rakamin Academy
Career Acceleration School
www.rakamin.com

**Created by:**
**Haolia**
haoliaaul@gmail.com
https://www.linkedin.com/in/haolia/

“Geophysical engineering graduates who diligently want to fulfil a role where intellectual, integrity, and curiosity are highly valued. Motivated, able to research, design, implement new features and learn various software. Skill handling problems with unique ways to develop innovative solutions. Proficient using Python, SQL, Tableau and other statistical tools for data multi purposes. Looking for opportunities in data analyst, data science, data engineer and Business Intelligence. ”

# Data Cleaning & Preprocessing

- Pada tahap ***cleaning data***, tunjukan ***null*** atau ***missing value*** serta ***duplicated value*** pada dataset, serta cara penyelesaiannya.

- Tulislah pula proses **extract datetime data** sebelum dilakukan model machine learning.

- Tunjukan **Split Data** sebelum melakukan model machine learning

- Tulislah proses **feature encoding** pada tahap ini (gunakan get_dummy)

- **Source code** yang sudah kamu buat, dapat ditampilkan dan berikan link untuk mengakses file tersebut. Contohnya seperti di pojok kanan bawah.

Untuk selengkapnya, dapat melihat jupyter notebook disini

# Data Cleaning & Preprocessing

***missing value***

## Exploratory Data Analysis

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 10 columns):
 #   Column                    Non-Null Count  Dtype  
---  ------                    --------------  -----  
 0   Daily Time Spent on Site  987 non-null    float64
 1   Age                       1000 non-null   int64  
 2   Area Income               987 non-null    float64
 3   Daily Internet Usage      989 non-null    float64
 4   Male                      997 non-null    object 
 5   Timestamp                 1000 non-null   object 
 6   Clicked on Ad             1000 non-null   object 
 7   city                      1000 non-null   object 
 8   province                  1000 non-null   object 
 9   category                  1000 non-null   object 
dtypes: float64(3), int64(1), object(6)
memory usage: 78.2+ KB
```

```
df1.shape
```

```
(1000, 11)
```

## Missing Value :

```
df1.isnull().sum()
```

```
Daily Time Spent on Site    13
Age                          0
Area Income                 13
Daily Internet Usage        11
Male                         3
Timestamp                    0
Clicked on Ad                0
city                         0
province                     0
category                     0
month                        0
dtype: int64
```

```
# inputation median to null values for numerical feature
df1.fillna(df1.median(), inplace=True)

# inputation mode for categorical feature
df1['Male'].bfill(inplace=True)
```

```
df1.isnull().sum()
```

```
Daily Time Spent on Site    0
Age                         0
Area Income                 0
Daily Internet Usage        0
Male                        0
Timestamp                   0
Clicked on Ad               0
city                        0
province                    0
category                    0
month                       0
dtype: int64
```

Untuk selengkapnya, dapat melihat jupyter notebook disini

***extract datetime data***

```
df.Timestamp.dtype
```
✓ 0.1s
```
dtype('O')
```

```
import datetime
df.Timestamp = pd.to_datetime(df.Timestamp)
```
✓ 0.2s

```
df.Timestamp.dtype
```
✓ 0.7s
```
dtype('<M8[ns]')
```

```
df['month'] = df.Timestamp.dt.month
```
✓ 0.9s

```
df.month.unique()
```
✓ 0.9s

```
#ubah tipe data feature week
df1.week = df1.week.astype('int64')
```
✓ 0.9s

```
print(df1.year.dtype)
print(df1.month.dtype)
print(df1.week.dtype)
print(df1.day.dtype)
```
✓ 0.1s
```
int64
int64
int64
int64
```

Untuk selengkapnya, dapat melihat jupyter notebook disini

***feature encoding***

OHE pada feature `category`

```
cat_ohe = ['province', 'category']
```

✓ 0.1s

```
df_encode[cat_ohe].columns
```

✓ 0.1s

```
Index(['province', 'category'], dtype='object')
```

```
for cat in cat_ohe :
    onehots = pd.get_dummies(df_encode[cat], prefix=cat)
    df_encode = df_encode.join(onehots)
```

✓ 0.1s

Untuk selengkapnya, dapat melihat jupyter notebook disini