



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."



 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.



missing value

memory usage: 78.2+ KB

Exploratory Data Analysis

```
df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 10 columns):
                               Non-Null Count Dtype
     Column
    Daily Time Spent on Site 987 non-null
                                               float64
    Age
                               1000 non-null
                                               int64
                               987 non-null
                                               float64
     Area Income
    Daily Internet Usage
                               989 non-null
                                               float64
    Male
                               997 non-null
                                               object
    Timestamp
                               1000 non-null
                                               object
     Clicked on Ad
                               1000 non-null
                                               object
     citv
                               1000 non-null
                                               object
    province
                               1000 non-null
                                               object
    category
                              1000 non-null
                                               object
dtypes: float64(3), int64(1), object(6)
```

```
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)
```

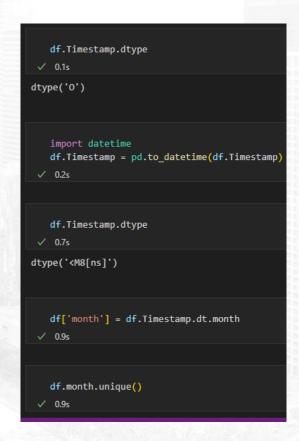
```
df1.isnull().sum()

Daily Time Spent on Site & Age & Area Income & Baily Internet Usage & Male & Baily Internet Usage & Baily Internet Us
```

Untuk selengkapnya, dapat melihat jupyter notebook disini



extract datetime data





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
```