

Predict Clicked Ads Customer Classification by Using Machine Learning



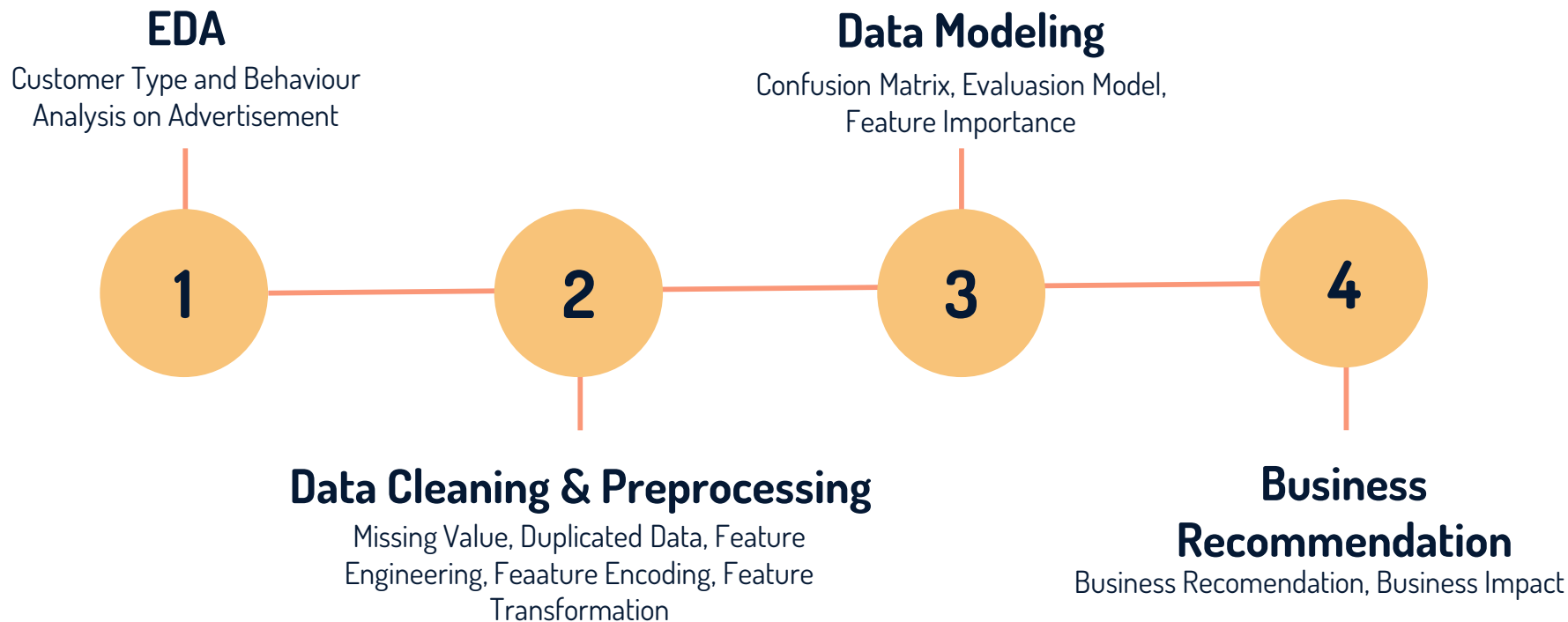
Overview



Sebuah perusahaan di Indonesia ingin mengetahui efektifitas sebuah iklan yang mereka tayangkan, hal ini penting bagi perusahaan agar dapat mengetahui seberapa besar ketercapainnya iklan yang dipasarkan sehingga dapat menarik customers untuk melihat iklan.

Dengan mengolah data historical advertisement serta menemukan insight serta pola yang terjadi, maka dapat membantu perusahaan dalam menentukan target marketing, fokus case ini adalah membuat model machine learning classification yang berfungsi menentukan target customers yang tepat. Dalam project ini kami menggunakan dataset yang disediakan oleh Rakamin Academy. Seluruh teknis pengerjaan menggunakan bahasa pemrograman python.

OUR PROCESS



01



Exploratory Data Analysis (EDA)

Customer Type and Behaviour
Analysis on Advertisement



Tahap awal yang kami lakukan pada dataset ini adalah melakukan Exploratory Data Analysis (EDA) dengan melihat statistik dari tiap-tiap variabel yang ada, dan menemukan insight dalam project ini. Disini kami melakukan analisis univariate, bivariate, dan juga multivariate.

Customer Type and Behaviour Analysis on Advertisement

Customer Type and Behaviour Analysis on Advertisement

1

```
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)
```

Terdapat 1000 baris data dengan 10 kolom. Beberapa kolom memiliki missing value yang dapat diisi dengan value tertentu. Tipe data pada kolom Timestamp masih berupa object, dapat diganti menjadi tipe data datetime.

2

	Daily_Time_Spent_on_Site	Age	Area_Income	Daily_Internet_Usage
count	987.000000	1000.000000	9.870000e+02	989.000000
mean	64.929524	36.009000	3.848647e+08	179.863620
std	15.844699	8.785562	9.407999e+07	43.870142
min	32.600000	19.000000	9.797550e+07	104.780000
25%	51.270000	29.000000	3.286330e+08	138.710000
50%	68.110000	35.000000	3.990683e+08	182.650000
75%	78.460000	42.000000	4.583554e+08	218.790000
max	91.430000	61.000000	5.563936e+08	267.010000

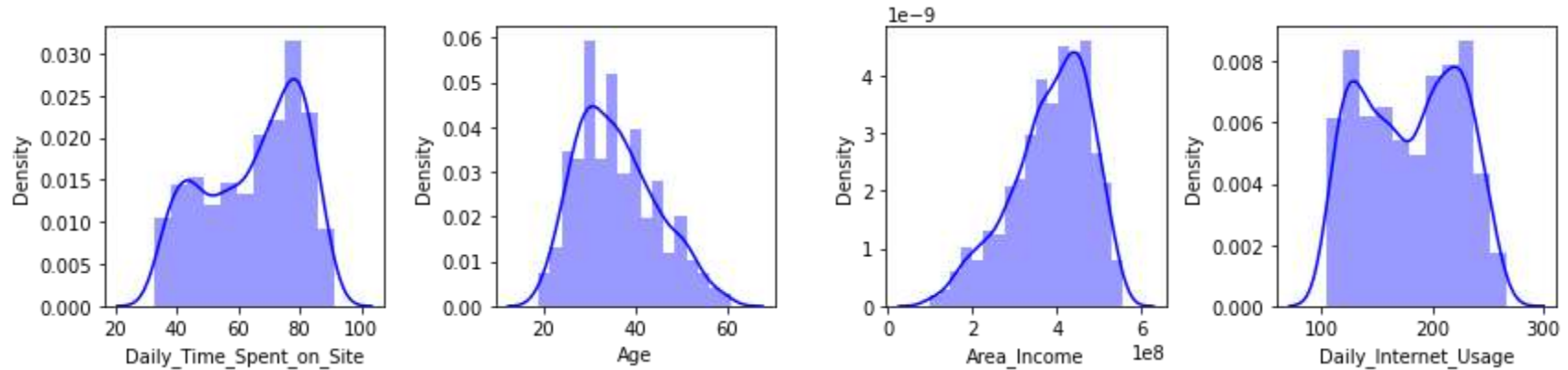
Tabel di samping menunjukkan statistik pada data numerik. Mulai dari rata-rata, nilai quartal, standard deviasi, dll. Terdapat visualisasi distribusi pada bagian selanjutnya.

3

	Gender	Timestamp	Clicked_on_Ad	city	province	category
count	997	1000	1000	1000	1000	1000
unique	2	997	2	30	16	10
top	Perempuan	5/26/2016 15:40	No	Surabaya	Daerah Khusus Ibukota Jakarta	Otomotif
freq	518	2	500	64	253	112

Tabel di samping menunjukkan statistik pada data numerik. Mulai dari rata-rata, nilai quartal, standard deviasi, dll. Terdapat visualisasi distribusi pada bagian selanjutnya.

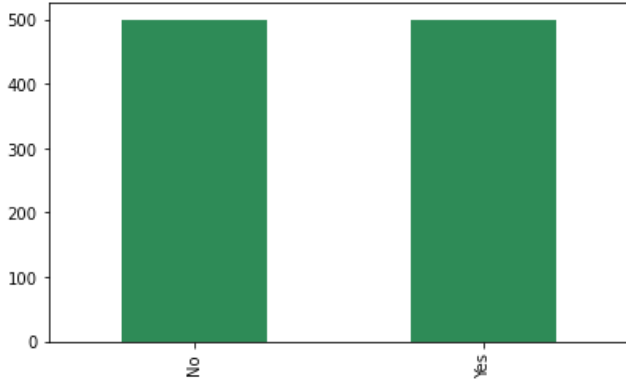
Customer Type and Behaviour Analysis on Advertisement



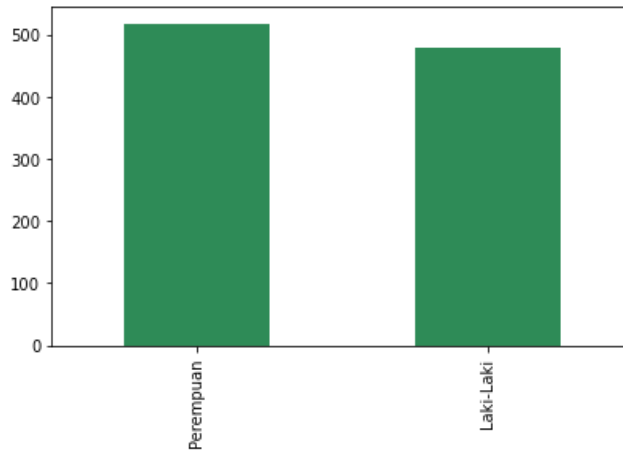
Pada grafik di atas terlihat bahwa feature-feature numerik pada data ini tidak terlalu normal atau masih skew. Namun apabila dilihat pada grafik dan analisis deskriptif sebelumnya perbedaan antara mean dan median pada feature-feature numerik ini tidak terlalu jauh, atau dengan kata lain tidak ada outlier yang terlalu besar. Pengguna internet pada kasus ini memiliki karakteristik yang cukup tersebar dari segi usia dan income,

Customer Type and Behaviour Analysis on Advertisement

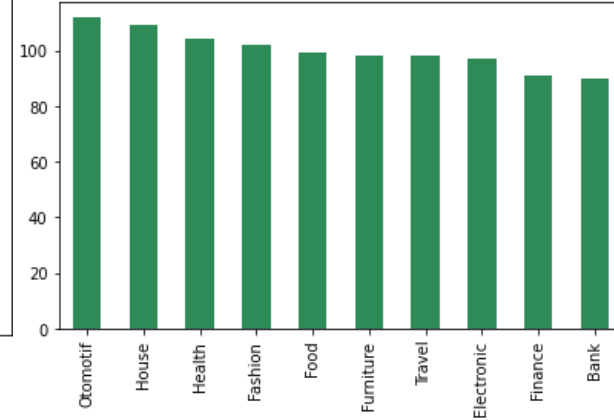
Clicked on Ads



Gender

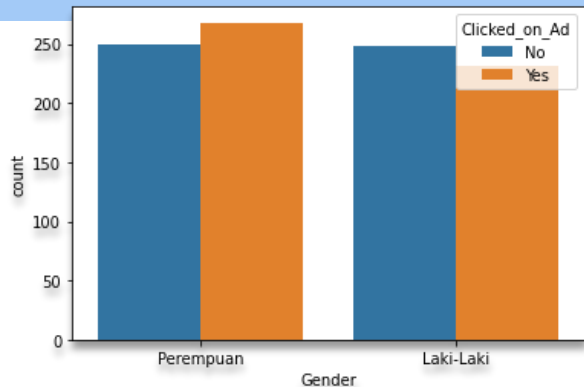


Category

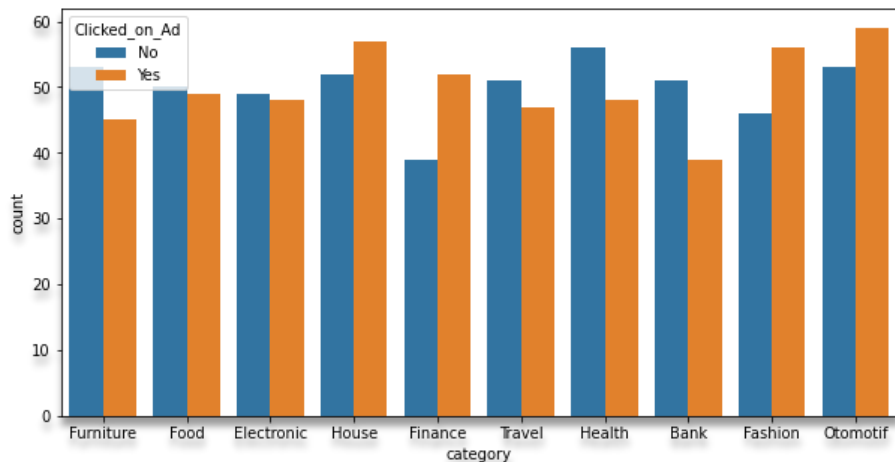


- 1 Pada countplot "Clicked on Ads" di atas, terlihat bahwa presentase pengguna yang mengklik iklan sama besarnya dengan yang tidak mengklik iklan. Presentase ini menunjukkan iklan yang ditampilkan cukup efektif untuk menarik perhatian pengguna internet.
- 2 Dalam kasus ini mayoritas pengguna internet adalah perempuan. Untuk meningkatkan efektifitas iklan, perusahaan dapat membuat iklan dengan tema atau hal-hal yang mungkin lebih diminati perempuan. Seperti fashion, health & beauty, atau hal-hal lainnya. Sehingga presentase pengguna yang mengklik iklan akan meningkat.
- 3 Jumlah kategori iklan tertinggi yang ada di perusahaan yaitu kategori otomotif. Tetapi jika kita lihat grafik tersebut, perbedaan jumlah kategori iklan tidak terlalu berbeda jauh.

Customer Type and Behaviour Analysis on Advertisement

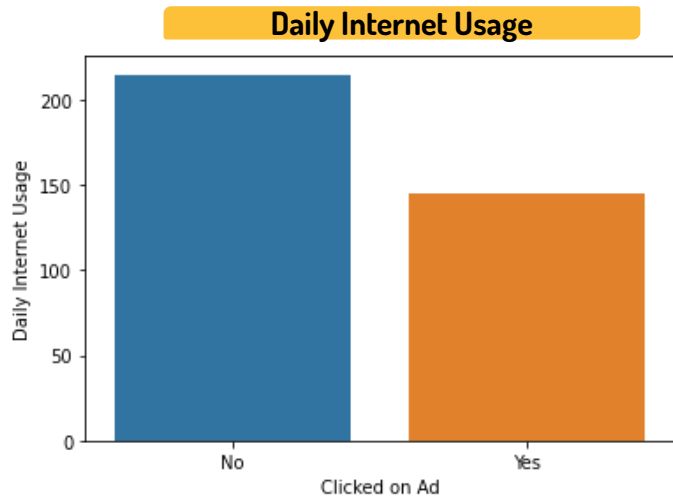
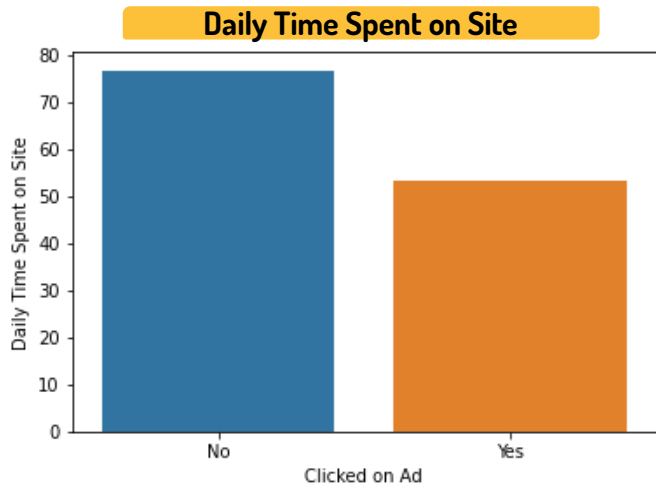


Grafik ini menunjukkan bahwa dibandingkan dengan perempuan, pengguna laki-laki lebih jarang mengklik iklan. Maka apabila perusahaan ingin meningkatkan efektifitas iklan, sejalan dengan analisis sebelumnya, perusahaan lebih baik menampilkan iklan dengan kategori atau tema yang diminati perempuan.



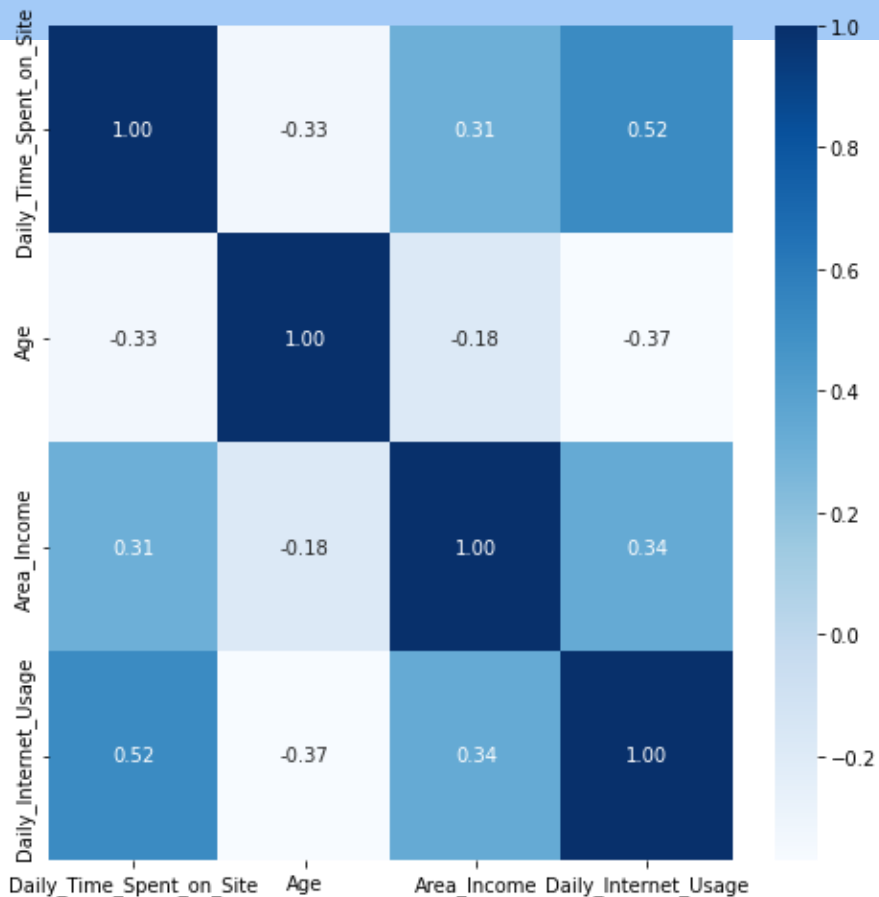
Terdapat beberapa kategori yang memiliki jumlah pengguna yang mengklik lebih sedikit daripada yang tidak mengklik. Untuk meningkatkan efektifitas iklan, kategori-kategori tersebut perlu membuat iklan yang lebih menarik.

Customer Type and Behaviour Analysis on Advertisement



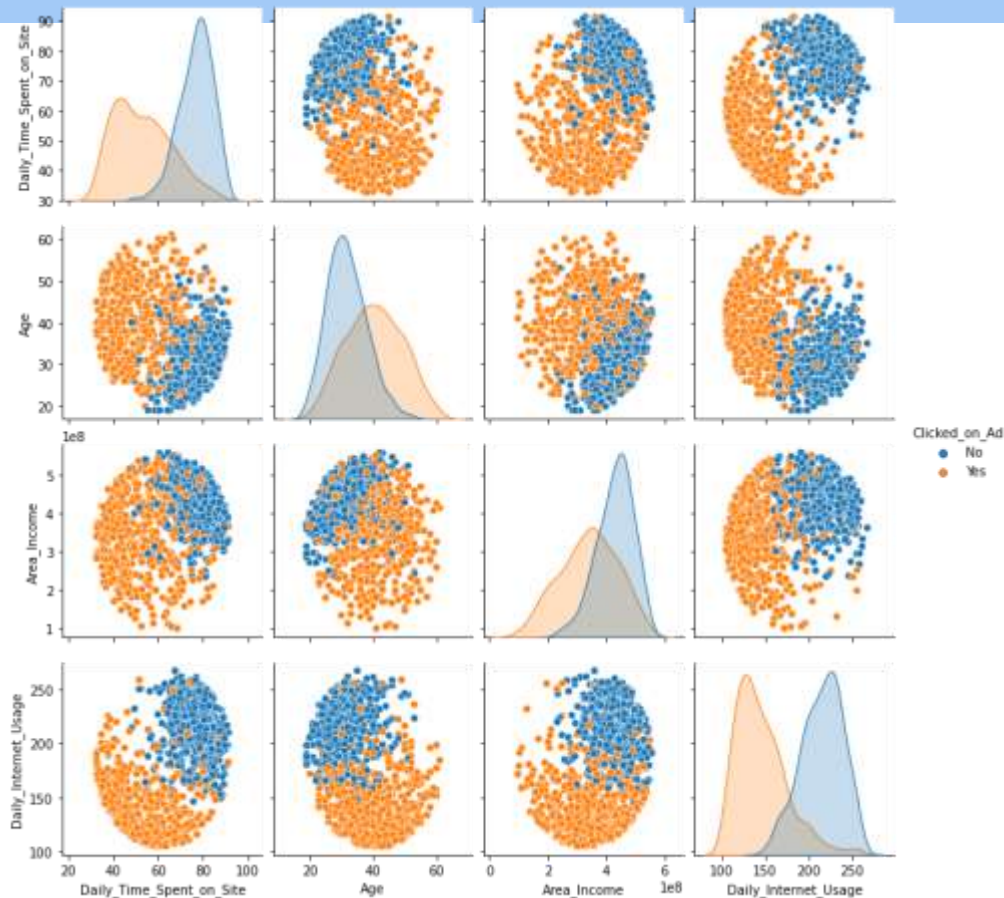
Dari dua barplot di atas dapat diketahui ternyata pengguna yang mengklik iklan justru pengguna yang penggunaan internet harian dan waktu kunjungan site hariannya lebih sedikit daripada pengguna yang tidak mengklik iklan. Perlu analisis lebih lanjut untuk mengetahui mengapa hal ini terjadi

Customer Type and Behaviour Analysis on Advertisement



- 1 Dari correlation heatmap tersebut terlihat bahwa semakin lama Daily Internet Usage maka semakin lama pula Daily Time Spent on Site nya.
- 2 Daily Internet Usage ternyata juga berkorelasi positif dengan Area Income. Dimana semakin besar income-nya maka semakin lama waktu penggunaan internet harian.
- 3 Fitur usia berkorelasi negatif dengan semua fitur lain. Semakin muda pengguna, semakin sedikit waktu penggunaan internet harian.

Customer Type and Behaviour Analysis on Advertisement



Disini kami menampilkan pairplot untuk memperlihatkan korelasi fitur satu dengan yang lain. Namun berbeda dengan correlation heatmap sebelumnya, disini kami juga membedakannya berdasarkan kolom Clicked_on_Ad untuk melihat perbedaan berdasarkan pengguna yang mengklik iklan dan yang tidak.

Predict Clicked Ads Customer Classification by using Machine Learning



Created by:

Dewi Ayu Rahmawati

ayudewi.ar@gmail.com

<https://linkedin.com/in/dewiayurahmawati>

I'm a Civil Engineering graduate from ITS. I'm a person who love to learn and start learning Data Science by joining Rakamin Bootcamp in early 2022. Now I want to shifting my career and looking for opportunities into the data field.