

Predict Clicked Ads Customer Classification by using Machine Learning

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



Created by:

Yanyan Gatot Mulyadi

yanyangatot1059@gmail.com

<https://www.linkedin.com/in/yanyan-gatot-348619177/>

About You

As a business graduate with a keen interest in data analysis, I possess a strong foundation in market research and SQL. I am adept at leveraging data to uncover valuable insights and develop impactful marketing campaigns. With a curious and innovative mindset, I constantly explore new ways to harness the power of data to drive business performance. My expertise in SQL allows me to efficiently manage and analyze large datasets, enabling data-driven decision-making. I am passionate about staying abreast of industry trends and leveraging data to achieve marketing success and deliver tangible results.

“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 ”

- Tulislah proses **Exploration Data Analysis** (EDA) yang mencakup **Statistical analysis** baik untuk data numerik maupun kategori, Selanjutnya buat visualisasi data untuk **Univariate** dan **Bivariate analysis**, serta **Multivariate analysis**
- Khusus untuk **Bivariate analysis**, tunjukkan hubungan antara kolom umur, daily internet usage, dan daily time spent on site.
- Tulislah juga **proses korelasi heatmap** untuk mengetahui tingkat korelasi antar kolom
- **Source code** yang sudah kamu buat, dapat ditampilkan dan berikan link untuk mengakses file tersebut. Contohnya seperti di pojok kanan bawah.

Terdapat 13 missing value pada kolom Daily Time Spent on Site, 13 missing value pada kolom Area Income, 11 missing value pada kolom Daily Internet Usage, dan 3 missing value pada kolom Male, dari sekitar 1000 baris yang ada. Karena jumlah missing value relative kecil (<5%) maka kami isi dengan mean, Male saya ganti menjadi gender.

```
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
dtype: int64
```

Data Cleaning & Preprocessing

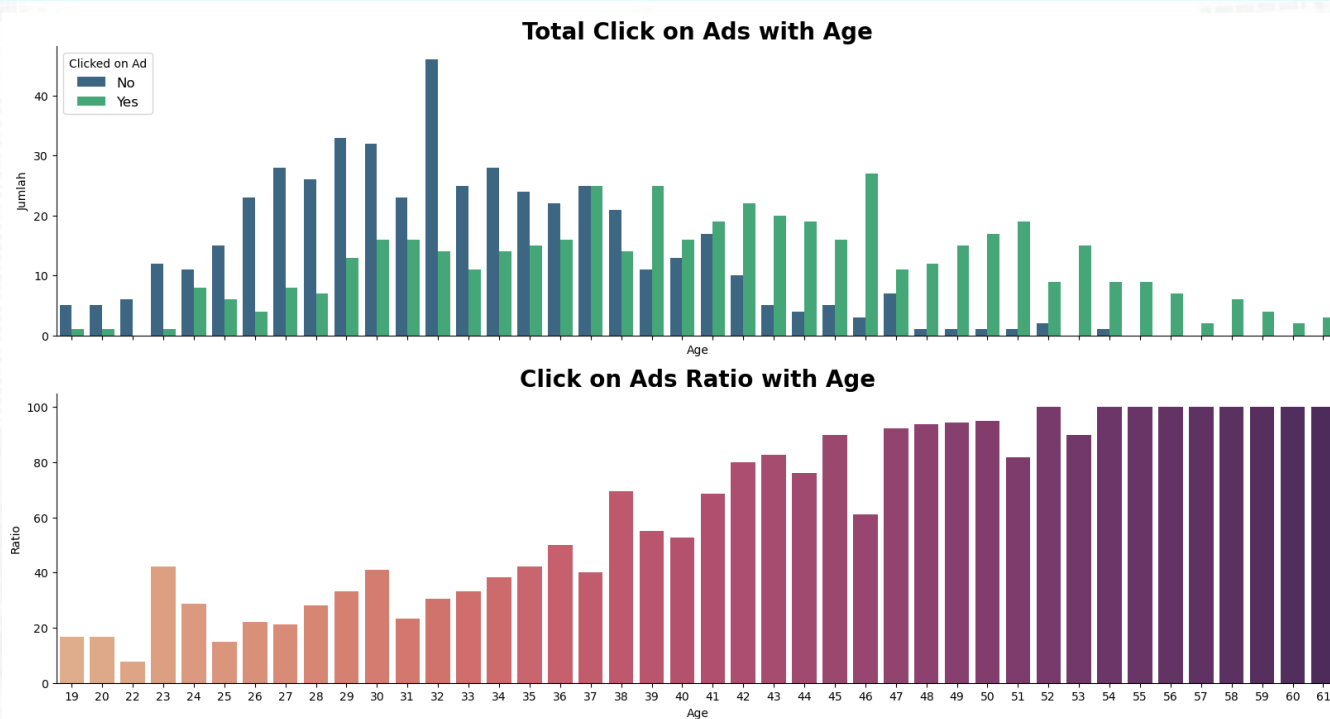
Kami juga melakukan feature engineering dengan mengekstrak data tahun bulan minggu dan tanggal pada kolom Timestamp untuk dapat digunakan sebagai featura pada model ML

	Daily Time Spent on Site	Age	Area Income	Daily Internet Usage	Gender	Timestamp	Clicked on Ad	city	province	category
0	68.95	35	432837300.0	256.09	1	2016-03-27 00:53:00	0	13	2	5
1	80.23	31	479092950.0	193.77	0	2016-04-04 01:39:00	0	8	0	4
2	69.47	26	418501580.0	236.50	1	2016-03-13 20:35:00	0	25	5	1
3	74.15	29	383643260.0	245.89	0	2016-01-10 02:31:00	0	4	9	7
4	68.37	35	517229930.0	225.58	1	2016-06-03 03:36:00	0	17	15	3



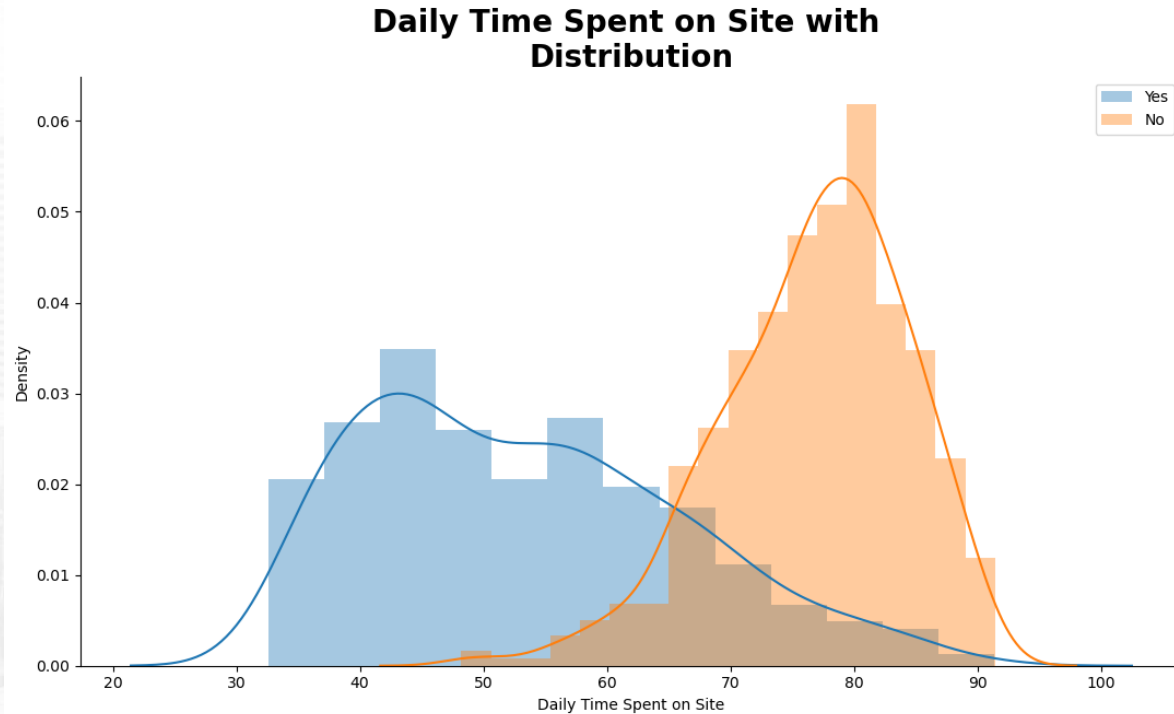
	Daily Time Spent on Site	Age	Area Income	Daily Internet Usage	Gender	Clicked on Ad	city	province	category	Tahun	Bulan	Tanggal	Pekan
0	68.95	35	432837300.0	256.09	1	0	13	2	5	2016	3	27	12
1	80.23	31	479092950.0	193.77	0	0	8	0	4	2016	4	4	14
2	69.47	26	418501580.0	236.50	1	0	25	5	1	2016	3	13	10
3	74.15	29	383643260.0	245.89	0	0	4	9	7	2016	1	10	1
4	68.37	35	517229930.0	225.58	1	0	17	15	3	2016	6	3	22

Customer Type and Behaviour Analysis on Advertisement



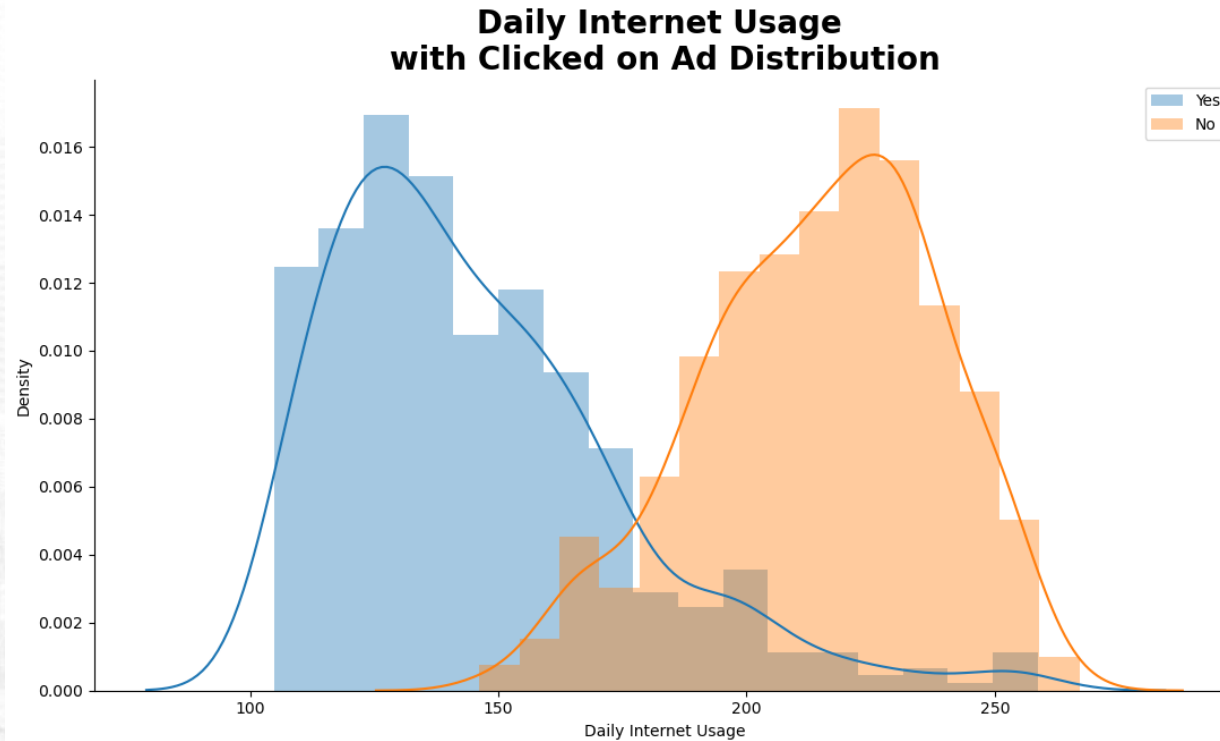
Meski mayoritas user memiliki usia 30 tahun-an,
akan tetapi user yang memiliki probabilitas
melakukan klik tertinggi berada di rentang usia 40
tahun-an.

Customer Type and Behaviour Analysis on Advertisement



Pengunjung situs lama cenderung tidak terlalu tertarik dengan iklan yang diberikan.

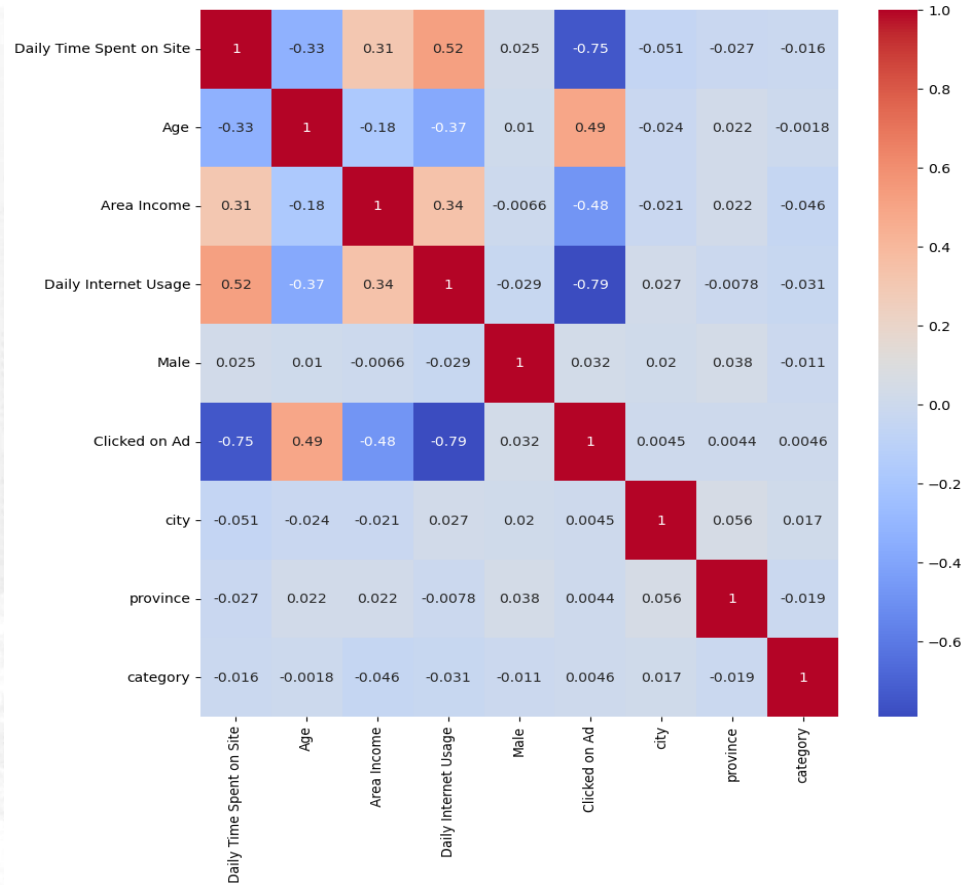
Customer Type and Behaviour Analysis on Advertisement



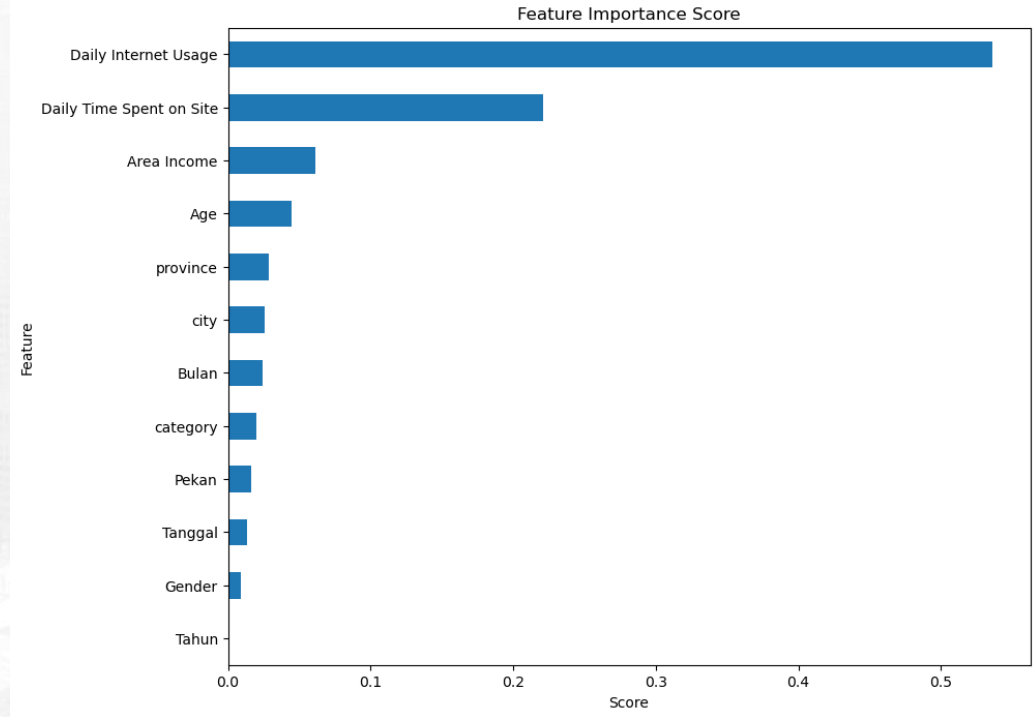
Pengguna internet lama cenderung tidak terlalu tertarik dengan iklan yang diberikan

Customer Type and Behaviour Analysis on Advertisement

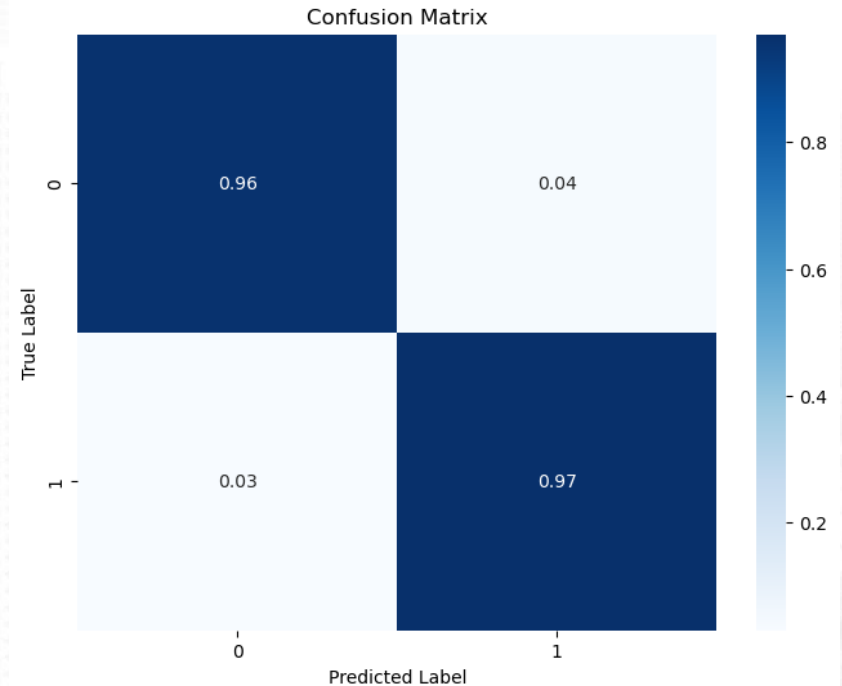
Kemungkinan klik iklan berbanding terbalik dengan Daily Internet Usage, Daily Time Spent on site, Area Income, dan berbanding lurus dengan Age



4 feature paling berpengaruh pada model machine learning ini adalah **daily internet usage**, **daily time spent on site**, **area income**, dan **age**.



Hasil confusion matrix model XGBClassifier di sampling menunjukkan performa yang baik, dengan 00.4% false positive dan 0.03% false negative dari model, Jumlah ini terbilang kecil jika dibandingkan dengan jumlah data test yang digunakan.



- 4 feature paling berpengaruh pada model machine learning ini adalah daily internet usage, daily time spent on site, area income, dan age. Maka dari itu kita dapat menfokuskan ads kepada kelompok dengan conversion rate click paling tinggi berdasarkan 4 feature tersebut
- Karena pengguna internet dan pengunjung situs senior cenderung tidak tertarik pada iklan, kita dapat menfokuskan iklan kepada pengguna internet dan pengunjung situs baru
- Karena pengguna internet dan pengunjung situs senior cenderung tidak tertarik pada iklan, kita dapat menfokuskan iklan kepada pengguna internet dan pengunjung situs baru