Rakamin
Academy

Predict Clicked Ads Customer Classification by using Machine Learning

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

Created by:
Muhammad Cikal Merdeka
Email : mcikalmerdeka@gmail.com
LinkedIn : linkedin.com/in/mcikalmerdeka
Github : github.com/mcikalmerdeka

Dedicated entry-level data scientist with analytical and experimental background of Physics. My graduation 2023, a pivotal year marked by significant advancements in artificial intelligence with the introduction of GPT-4 and other generative AI models, has fueled my curiosity and excitement to delve into the field of data. I have comprehensive grasp of data science methodology from business understanding to modelling process with proficiency in **Python, SQL, Tableau, Power BI, Looker Studio and other tools** related to data analytics workflow from several coursework and bootcamps.

Project Overview

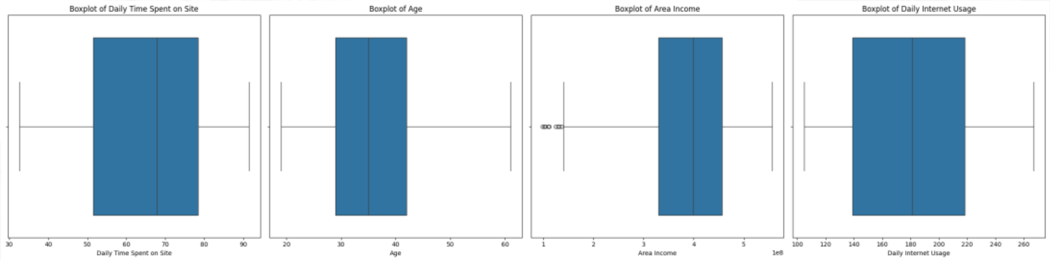
A company engaged in digital marketing in Indonesia wants to determine the effectiveness of an advertisement they have aired. This is crucial for the company to gauge the reach of their marketed ads and attract customers to view them.

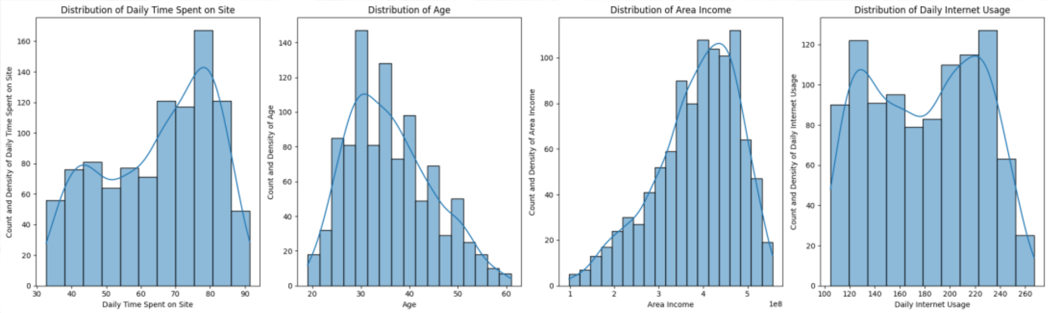
By analyzing historical advertisement data and identifying insights and patterns, it can assist the company in determining target marketing. The focus of this project is to develop a machine learning classification model that can accurately determine the target customers.

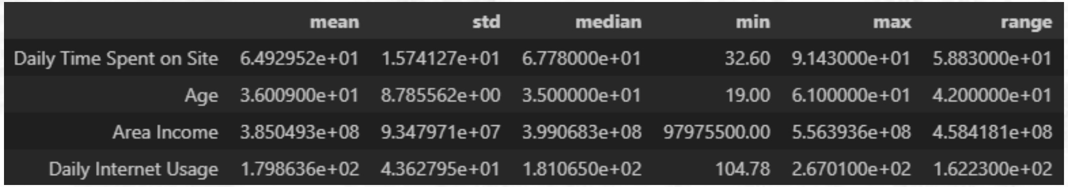
For more details, you can view the Jupyter notebook here.

Exploratory Data Analysis (EDA)

Univariate Analysis and Statistical Summary – Numerical Features







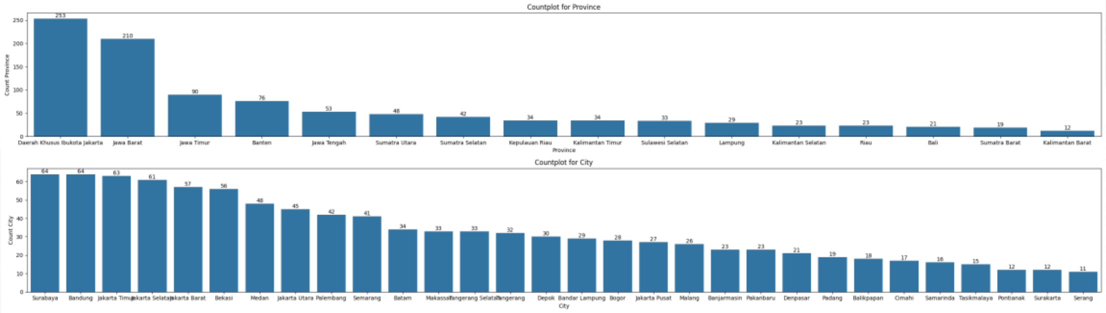
	mean	std	median	min	max	range
Daily Time Spent on Site	6.492952e+01	1.574127e+01	6.778000e+01	32.60	9.143000e+01	5.883000e+01
Age	3.600900e+01	8.785562e+00	3.500000e+01	19.00	6.100000e+01	4.200000e+01
Area Income	3.850493e+08	9.347971e+07	3.990683e+08	97975500.00	5.563936e+08	4.584181e+08
Daily Internet Usage	1.798636e+02	4.362795e+01	1.810650e+02	104.78	2.670100e+02	1.622300e+02

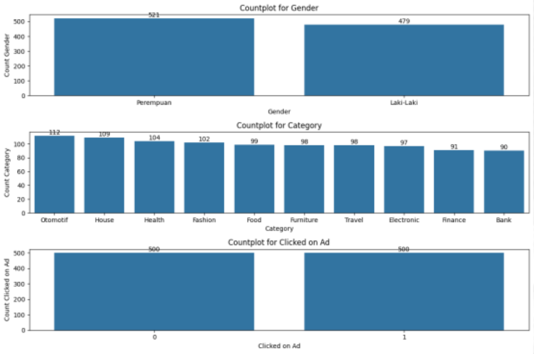
- The distribution of our customer's Daily Time Spent on Site, Age, and Area Income are moderately normal distribution while Daily Internet Usage is uniform distribution.

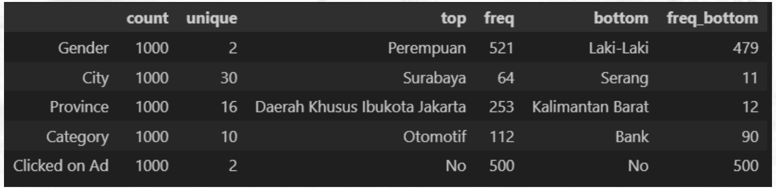
- Our customers usually spent 64 minutes per day on our site, have an average of age of 36, the area income of our customers is Rp.38,000,000, and average their daily internet usage is 3 hours.

Exploratory Data Analysis (EDA)

Univariate Analysis and Statistical Summary – Categorical Features







	count	unique	top	freq	bottom	freq_bottom
Gender	1000	2	Perempuan	521	Laki-Laki	479
City	1000	30	Surabaya	64	Serang	11
Province	1000	16	Daerah Khusus Ibukota Jakarta	253	Kalimantan Barat	12
Category	1000	10	Otomotif	112	Bank	90
Clicked on Ad	1000	2	No	500	No	500

- The majority of our customers are from DKI Jakarta and Jawa Barat province, while as for the city is quite uniformly distributed.

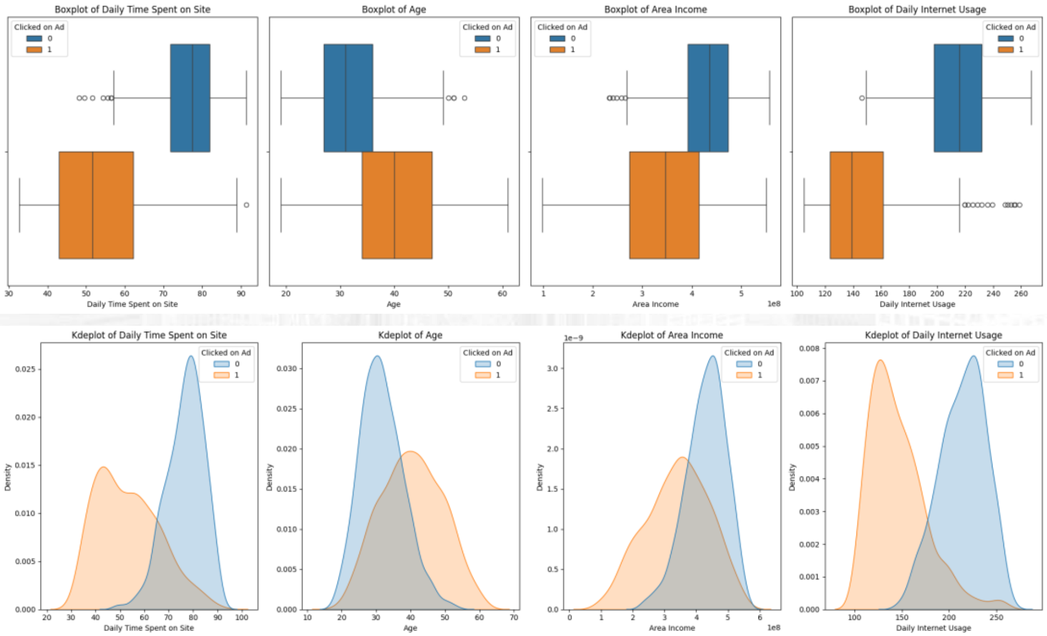
- The distribution of male and female Gender of our customers is almost equal.

- Clicked on Ad target variable has an equal distribution of No and Yes, we want to increase the click through rate (CTR).

- Ad category that is viewed is almost equally distributed among our customers.

Exploratory Data Analysis (EDA)

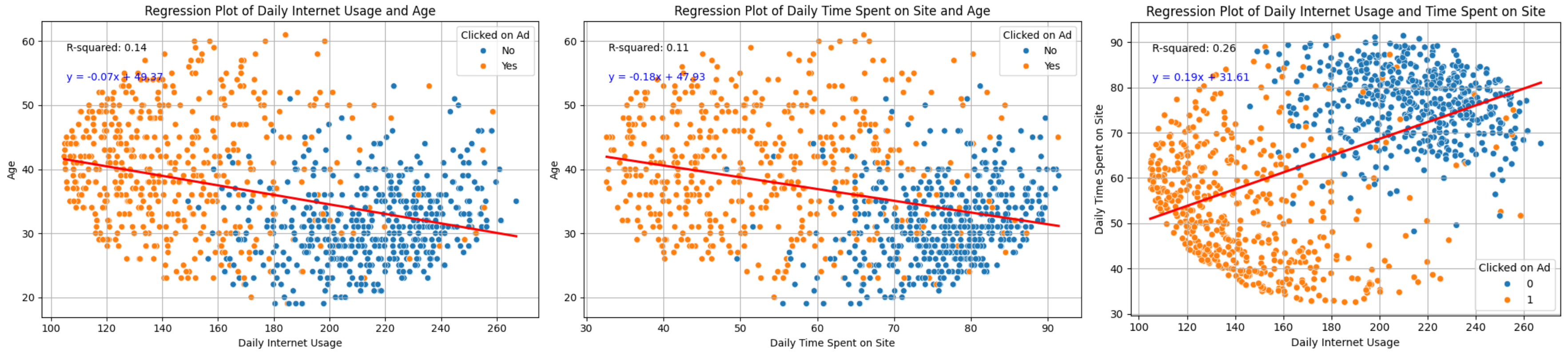
Bivariate Analysis – Numerical Features



- The more time customers spent time on the site, the less likely they will click on an ad.
- The average age of customers that clicked on an ad is 40, while the average for those that didn't is 31.
- The average income of customers that clicked on an ad is considerably lower than those that didn't.
- The more time customers use the internet daily, the less likely they will click on an ad.

Exploratory Data Analysis (EDA)

Bivariate Analysis – Regression Plot



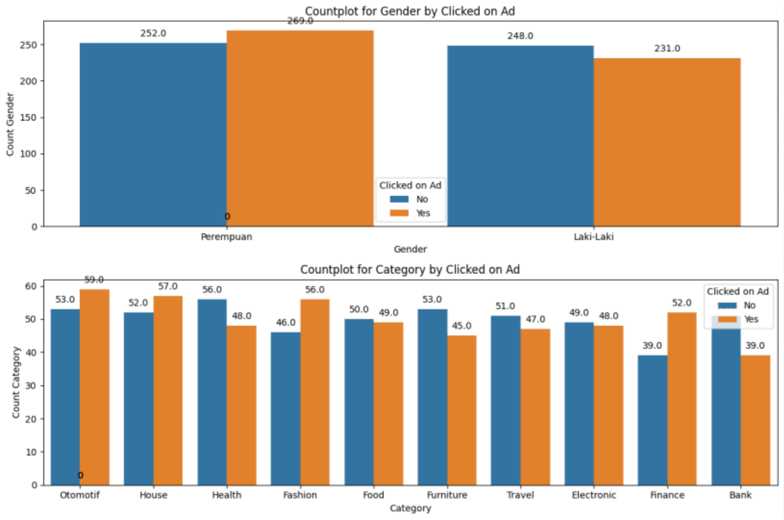
- Age is slightly negatively correlated with Daily Internet Usage. Older customers spend less time on the internet on average compared to younger customers.

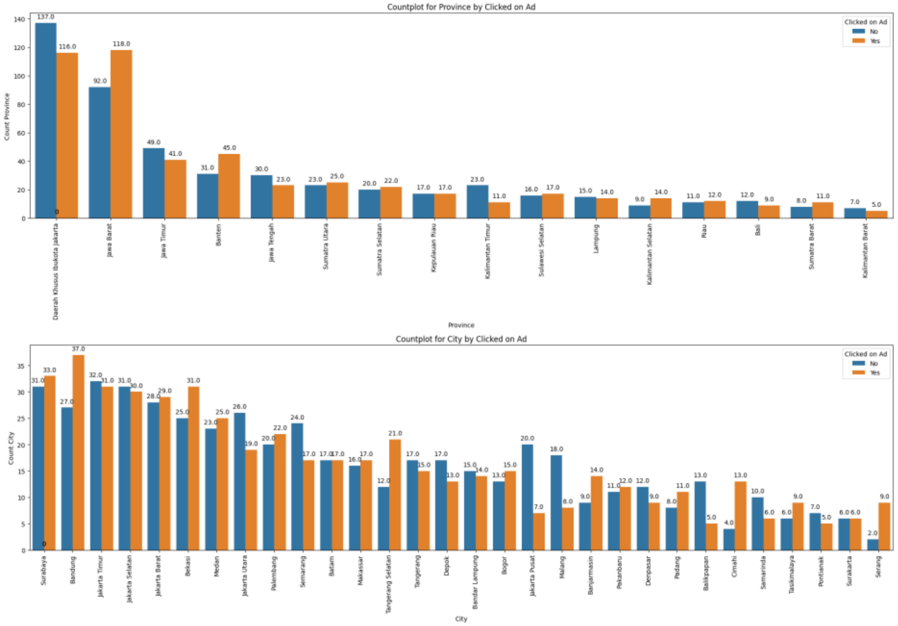
- Age is also slightly negatively correlated with Daily Time Spent on Site.

- Internet usage is slightly positively correlated with time spent on site. But there is a quite clear separation between two clusters of data. One cluster is less active and the other more so. Less active customers have a higher tendency to click on an ad compared to more active customers.

Exploratory Data Analysis (EDA)

Bivariate Analysis – Categorical Features

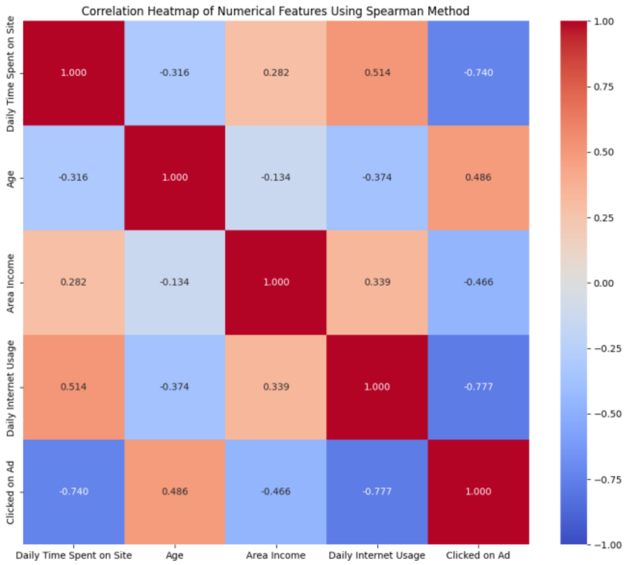


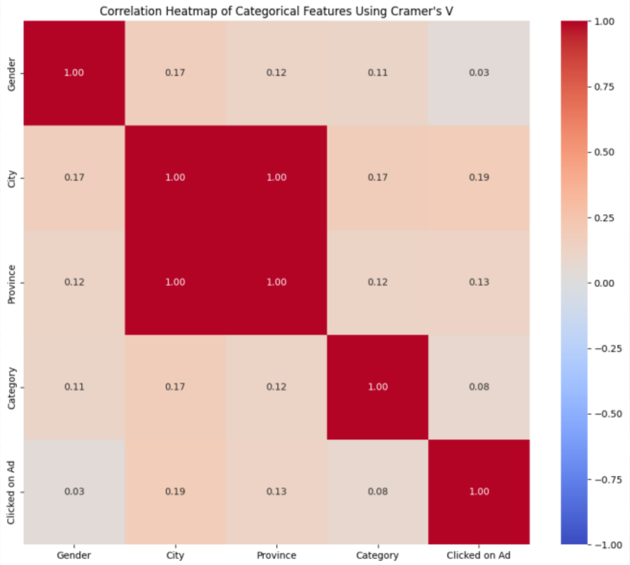


- Females clicked on an ad slightly more than males overall even though the difference is not significant.
- Ad categories click rates are pretty equal with none below 40% and none above 60%.
- Province with the highest amount of click on ad is Jawa Barat, while the lowest is Kalimantan Barat.
- City with the highest amount of click on ad is Bandung, while the lowest are Balikpapan and Pontianak.

Exploratory Data Analysis (EDA)

Multivariate Analysis – Numerical & Categorical Features





- Numerical features have high to moderate correlation to target, while all categorical features have really low correlation.
- There is no redundant case (high coefficient to each other) for numerical features, while for categorical features City and Province have perfect correlation.
- A value of 1 indicates a perfect positive correlation (as one variable increases, the other variable increases) and -1 indicates a perfect negative correlation (as one variable increases, the other variable decreases).