**Predict Clicked Ads Customer Classification by using Machine Learning: Overview**

Some company in Indonesia want to know the effective about some ad that their already publish, this is an important thing for company to know how much this ad can interactive the customer when their look at the ad.

With processing the historical data advertisement, found an insight, and also path that happened, so can help the company to found the target marketing , focus in this case is made a classification machine learning model that can help marketing found the best and accurate customer for the company purpose.

**About Dataset**

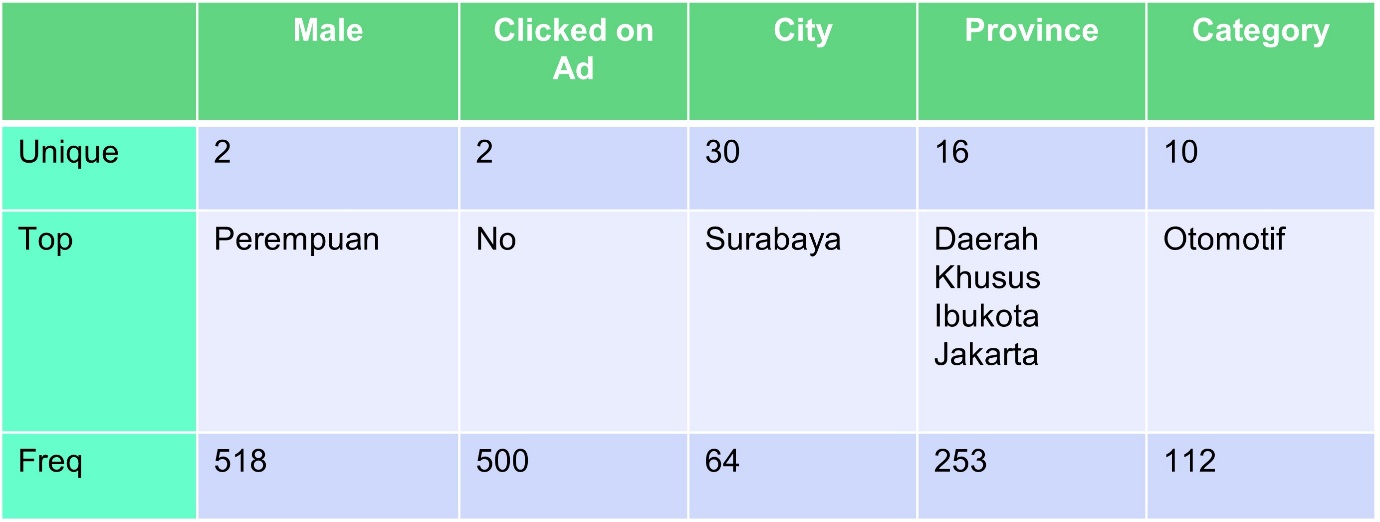
* This data have 1000 data, such as 7 numerics and 5 categoricals.
* There are some uncompleted data in some columns.

**EDA**

1. ****Numeric data

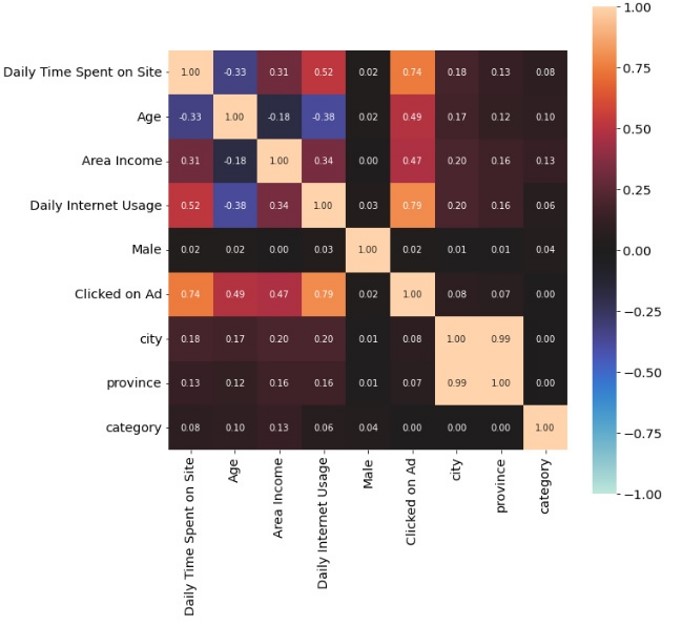
* Daily Time Spent on Site have mean 64.929 and Age 36 years old.
* Minimum people usage internet daily is 104.7.
* Maximum income people in 5.563.

1. Categoric data

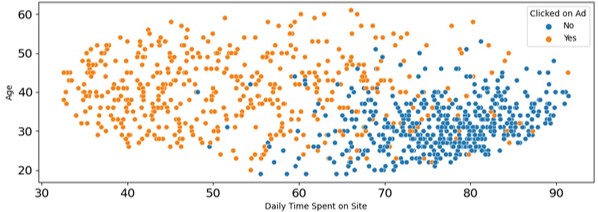
****

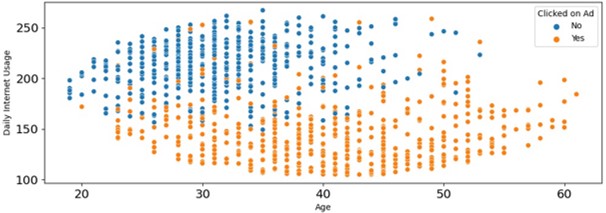
* Female have many time to surfing in internet.
* Surabaya is the most city that get into the internet.
* Many people go to automotive section in internet.

1. Correlation

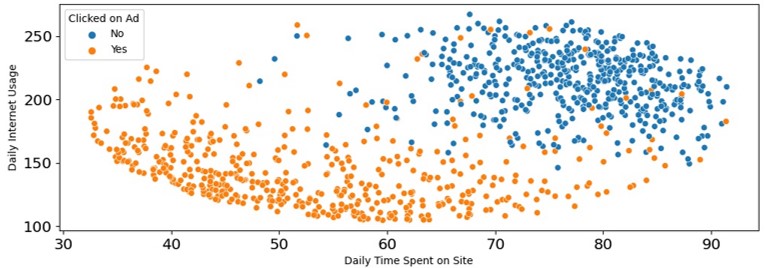


The correlation between Clicked on Ad and with another columns

1. Daily Time Spent on Site 0.74
2. Age 0.49
3. Area indome 0.47
4. Daily Internet Usage 0.79
5. Male 0.02
6. City 0.08
7. Province 0.07
8. Category 0.00
9. Analysis
10. Relationship between Age and Daily Time on Site
11. Relationship between Internet Usage and Age



1. Relationship between Internet Usage and Daily Time Spent on Site

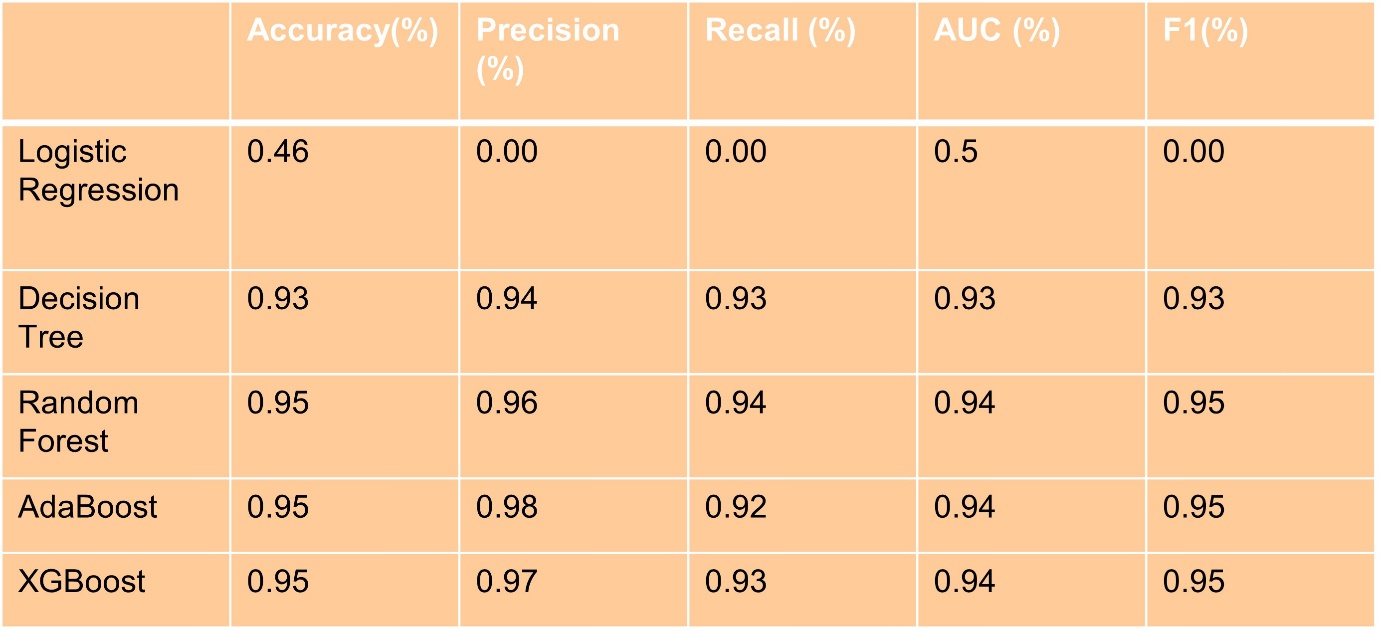


**Data Pre processing**

* For null data, numeric using mean and categoric use modus.
* There’s no duplicated data.
* For Timestamp column, there’s a change datatype to date using pandas (pd.to\_datetime). After that, the new date column are extracted according to their respective function such as year,month, day, and week.
* Future Encoding mapping for male and clicked on Ad. One hot encoding for city, category and province
* Split data 70:30 with data from columns : Daily Time Spent on Site, Age ,Area Income ,Daily Internet Usage, city\_Balikpapan, city\_Cimahi, city\_Jakarta Pusat, city\_Malang, city\_Serang, city\_Tangerang Selatan, province\_Banten, province\_Jawa Barat, province\_Kalimantan Timur, Clikked on ad. Clicked on ad column for the target.

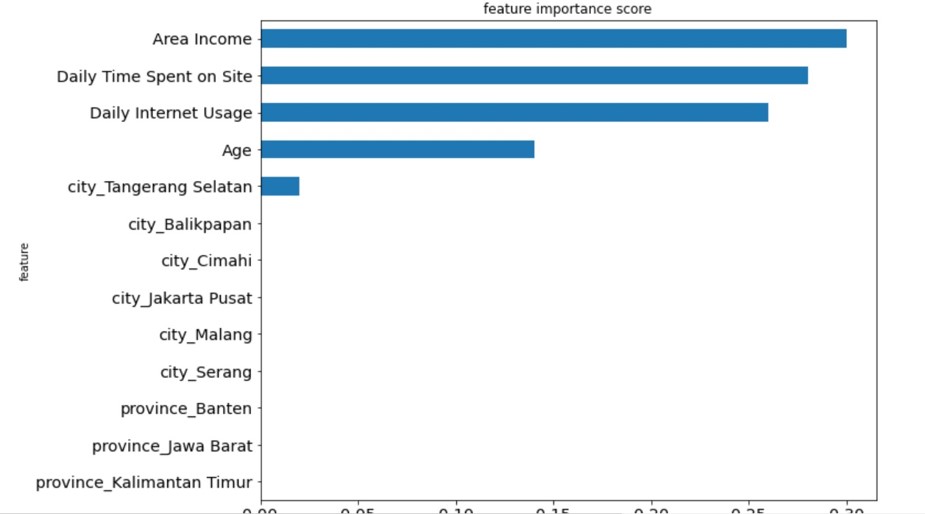
**Modelling**

This dataset using 5 model, Logistic Regression; Decision Tree; Random Forest; Adaboost; and XGboost. With two different treatment, before and after standardization.

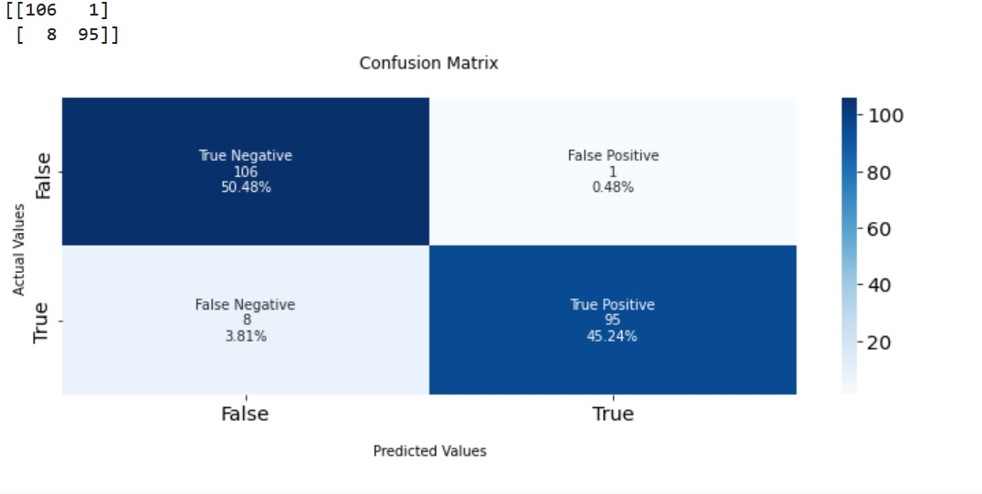
1. Before Standardization
2. After Standardization



From the result, we used modelling without standardization and for the best modeeling is Adaboost. This is the future Importance from the modelling.

The future importance are Area Income, Daily Time Spent on Site. Daily Internet Usage, Age, and City Tanggerang Selatan.

Form This modelling, we can calculate the matrix.



**Results**

The business recommendation are:

1. Give a easier access to click for people who in adults until elderly.
2. Giving a colour for the clicked so that can make an eye catching and easy.
3. Simple instruction for people so they can follow convenient.
4. Offer a promotion for province that have a lower consuming surfing internet.

Result from the Modelling:

* Assumption:

The company conducts a survey to click ads that appear on the website. The sales department conducts a campaign with a cost 30000IDR per customer. If a customner clicks, the company gets 100000 IDR per customer. So the company earns a profit:

Cost : 30000IDR

Revenue : 100000IDR

Profit = Revenue-Cost

= 100000IDR-30000IDR

= 70000IDR

If we don’t using the modelling:

Profit = (Revenuex1000)-(Costx1000)

= (100000IDRx500)-(30000IDRx1000)

= 20000000IDR

If we using the modelling:

Profit = (Revenuex1000)-(Costx1000)

= (100000IDRx95)-(30000IDRx96)

= 6620000IDR

With using a machine learning, the profit will increase 3.31%.

For more detail, you can visit the link below:

<https://github.com/inggridpris/Predict-Clicked-Ads-Customer-Classification-by-using-Machine-Learming>

<https://drive.google.com/drive/folders/1bn4pdGtVZAwFSYAYMvSp__vC4BXDzBro?usp=sharing>

or you can connect with me with linkedin

<https://www.linkedin.com/in/inggriani-priscilia-69779b179/>