

# Predict Clicked Ads Customer Classification by using Machine Learning

Created By : Ammar Khaq Baasir



# About Me

Saya merupakan seorang freshgraduate dari universitas muhammadiyah malang program studi teknik elektro. saya berhasil menyelesaikan pendidikan S1 saya dengan meraih penghargaan sebagai lulusan terbaik ke-1 periode 1 2024 dengan raihan IPK 3,76. sebagai seorang freshgraduate yang memiliki ketertarikan berkarir sebagai seorang data scientist maupun data analyst. saya memiliki beberapa kemampuan seperti pemahaman tentang statistik dan matematika, python programming, SQL, penerapan algorithms machine learning maupun deep learning.



# Table of contents

**01**

**About  
Project**

**02**

**Demo &  
Business  
Insight**

**03**

**Simulation  
Bisnis**

**04**

**Finalisasi  
Model**





# 01

## About Project

# Business Problem

Sebagai Perusahaan consulting digital marketing yang terkemuka, kami dihadapkan dengan sebuah project yang menarik oleh klien kami, sebuah Perusahaan barang – barang rumah tangga telah melakukan promosi digital melalui ads dalam beberapa waktu tahun terakhir. Perusahaan tersebut merasa bahwa yang mereka lakukan mengalami penurunan dalam konversi. Perusahaan tersebut ingin mengetahui dan mengungkap customers dengan karakteristi seperti apa yang cocok mendapatkan penawaran konten ads dari perusahaan. Mereka ingin mendapatkan beberapa wawasan atau insight dari history iklan yang telah mereka lakukan beberapa tahun belakangan, dengan insight yang didapatkan Perusahaan berharap dapat meningkatkan profit mereka dan mengurangi biaya cost ads yang telah dilakukan.



# Predict Clicked Ads Customer Classification Using Machine Learning

Merupakan sebuah proses menggunakan teknik machine learning untuk mengklasifikasikan (mengelompokkan) pelanggan berdasarkan kemungkinan mereka untuk melakukan clicked on ads. Tujuan utama dari project ini adalah memprediksi apakah seorang customers akan mengklik iklan tertentu atau tidak. Proses ini melibatkan beberapa tahap, mulai dari pembersihan dan pengolahan data, pemilihan fitur yang relevan, pemilihan model machine learning yang sesuai, hingga evaluasi kinerja model. Hasil dari pembuatan model klasifikasi diharapkan dapat membantu pemasar atau perusahaan untuk mengoptimalkan strategi pemasaran mereka dengan menargetkan iklan kepada customers yang memiliki kemungkinan lebih tinggi untuk melakukan click, sehingga dapat meningkatkan efektivitas dan efisiensi ads mereka.



# Machine Learning

Machine Learning (ML) adalah cabang dari kecerdasan buatan (AI) yang memungkinkan sistem komputer untuk belajar dari data yang ada, mengidentifikasi pola, membuat keputusan, dan melakukan prediksi tanpa harus secara eksplisit diprogram. Dengan kata lain, alih-alih mengandalkan instruksi yang diprogram secara kaku, sistem machine learning mampu belajar dan memperbaiki kinerjanya seiring waktu sebagaimana mereka terpapar dengan data baru. Terdapat beberapa jenis pendekatan dalam machine learning, diantaranya : **Supervised Learning, Unsupervised Learning, dan Reinforcement Learning.**

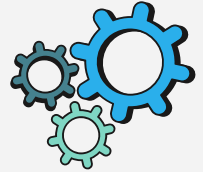
Manfaat dari machine learning sangat luas dan terus berkembang seiring dengan penemuan Teknik baru dan kemajuan dalam komputasi. Berikut beberapa manfaat utama penggunaan machine learning :

1. **Prediksi yang akurat** : Dengan memanfaatkan algoritma machine learning yang sesuai, sistem dapat memberikan prediksi yang akurat tentang hasil masa depan berdasarkan data historis.
2. **Analisis data yang kompleks** : Dengan kemampuannya mengekstrak pola yang kompleks dari data besar, machine learning memungkinkan organisasi untuk mendapatkan wawasan berharga yang tidak akan terdeteksi melalui metode analisis tradisional.
3. **Pengoptimalan Proses Bisnis** : Dengan memanfaatkan machine learning pada bisnis, dapat meningkatkan efisiensi operasional, meningkatkan kualitas produk atau layanan, dan mengidentifikasi peluang untuk inovasi.



# Type Machine Learning

## Supervised Learning



Supervised Learning merupakan jenis pembelajaran mesin (Machine Learning) di mana algoritma belajar dari data yang sudah dilabeli. Artinya setiap contoh data masukan (input) dikaitkan dengan label keluaran (output) yang sesuai. Tujuan dari supervised learning adalah untuk menghasilkan model yang dapat memetakan input ke output yang benar.

Terdapat beberapa jenis supervised learning diantaranya :

1. **Regresi** : Dalam regresi, tujuannya adalah untuk memprediksi nilai kontinu. Model mempelajari hubungan antara variabel input dan variabel output kontinu.
2. **Klasifikasi** : Dalam klasifikasi, tujuannya adalah untuk memprediksi kelas atau label. Model mempelajari keputusan dari contoh-contoh yang diberikan untuk memprediksi kelas dari data yang diberikan.



# Environment & Tools For The Project

Language :



Environment :



Tools :



Library :

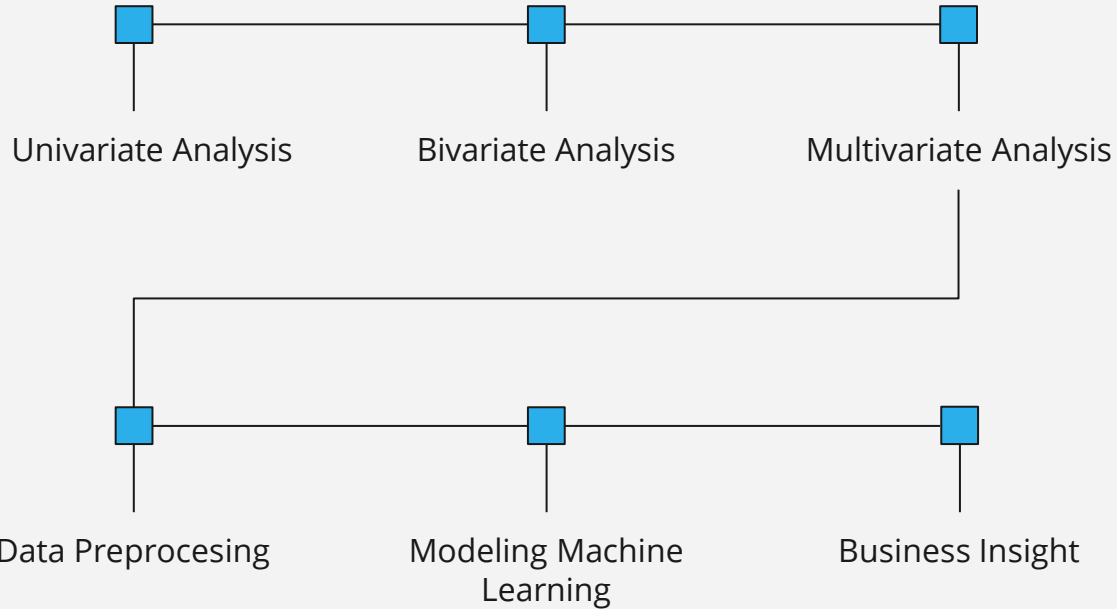
The matplotlib logo, featuring the word "matplotlib" in a blue, sans-serif font, with a small circular icon containing a colorful star-like shape.The seaborn logo, featuring the word "seaborn" in a blue, sans-serif font, with a small circular icon containing a colorful bar chart.The pandas logo, featuring a blue icon of a stylized 'P' with a white 'X' inside, and the word "pandas" in a blue, sans-serif font to its right.



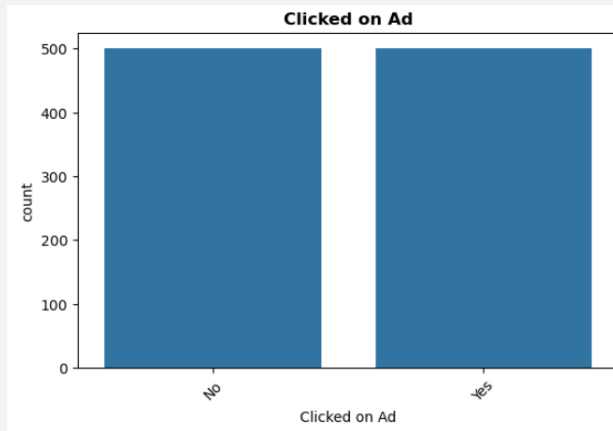
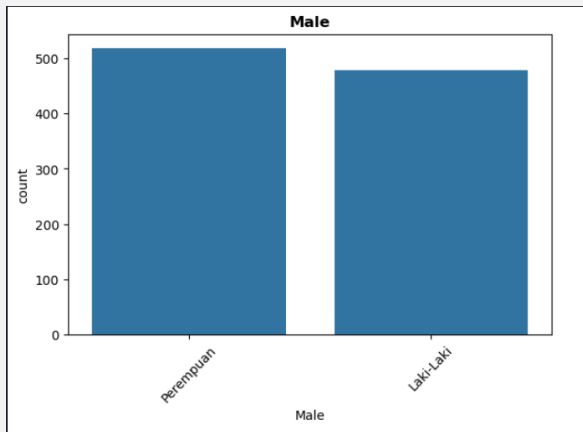
# 02

## DEMO PROJECT

# Demo Project Timeline

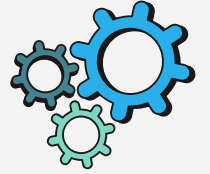


# Univariate Analysis

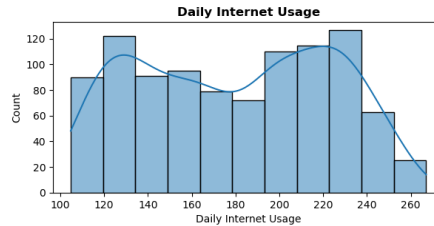
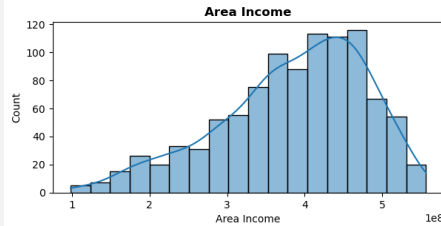
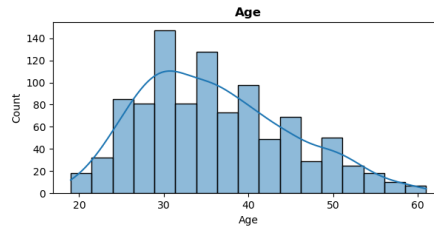
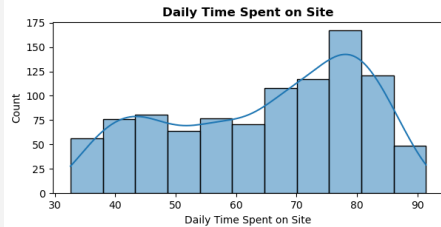


## Data Insight :

1. Persebaran **Gender Customers** yang melakukan **clicked on ads** maupun **tidak**, cenderung **memiliki proporsi** yang **hampir sama**. Ini menandakan **tidak adanya dominasi gender** dalam feature male, feature ini masuk dalam kriteria **feature yang dapat digunakan sebagai feature prediksi**. Akan tetapi perlu dilakukan analisis lebih lanjut mengenai feature male ini, apakah memiliki pengaruh yang kuat terhadap kemungkinan customers melakukan clicked on ads.
2. Kolom target (**Clicked on Ad**) memiliki **proporsi** atau **jumlah** customers **yang sama** antara customers **"yes"** atau **"no"** dalam **clicked on ads**. Ini menandakan bahwa data yang digunakan **merupakan data** yang memiliki jumlah **class seimbang** atau **balance**, data yang **balance** akan **memudahkan machine learning** mempelajari semua feature yang diberikan pada **saat pembelajaran** dilakukan, hal ini karena **machine learning** akan **bersifat adil** dan **tidak fokus** hanya pada **class yang memiliki jumlah lebih banyak**.



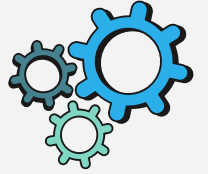
# Univariate Analysis



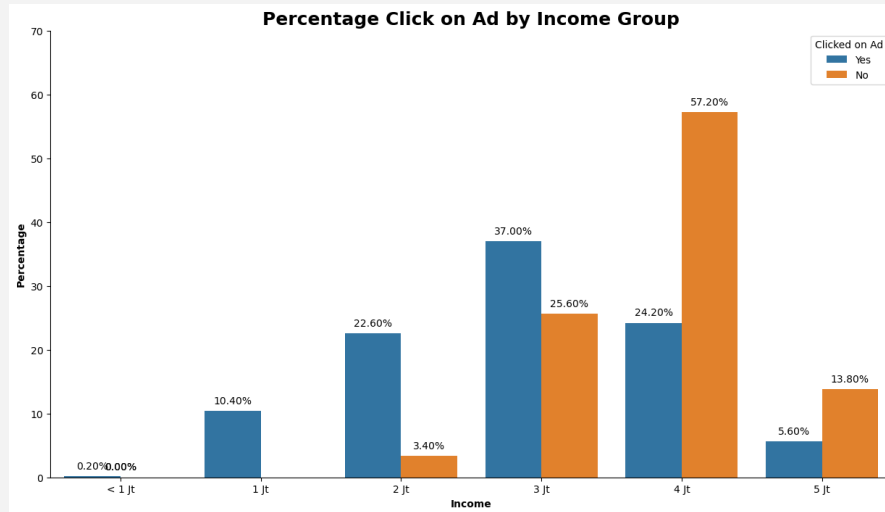
## Data Insight :

1. Feature **Daily Time Spent on Site** memiliki **distribusi binomial** atau memiliki **2 jenis kelompok** yakni kelompok **pertama customers** yang memiliki **rata - rata daily time spent on site 45 menit** dan kelompok **kedua** memiliki **rata - rata daily time on site 80 menit**.
2. Feature **Age** memiliki **distribusi data skewness positif** atau **miring kekanan**, Dimana **rata - rata umur** customer di angka **30 tahunan**.
3. Feature **Area Income** memiliki **distribusi data skewness negative** atau **miring kekiri**, Dimana **rata - rata income** customer di angka **4.5 juta**.
4. Feature **Daily Internet Usage** memiliki **distribusi binomial**, Dimana customer **terbagi** menjadi **2 kelompok** yakni kelompok **daily internet usage di 130 menit** dan **daily internet usage di 220 menit**.

File Code [Disini](#)



# Bivariate Analysis



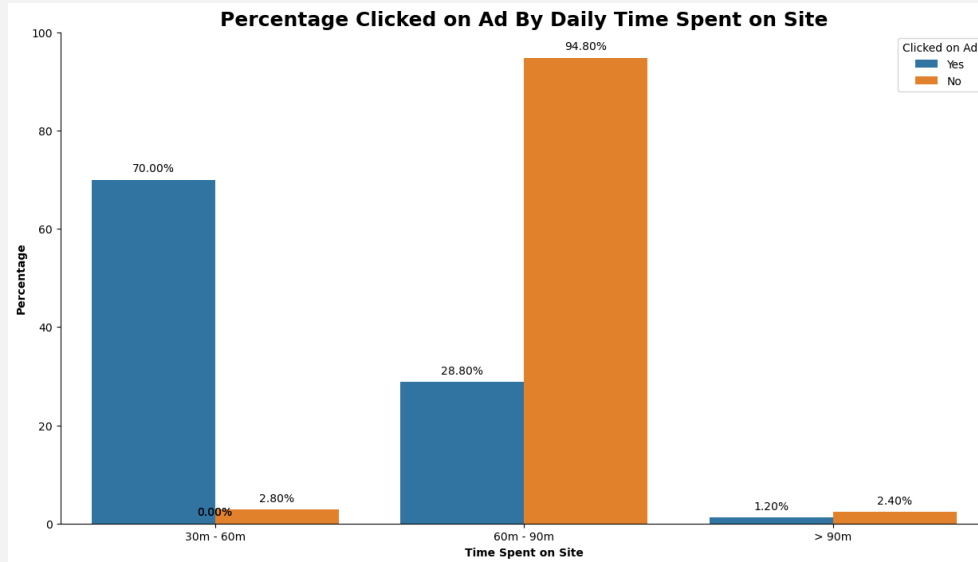
## Business Insight :

Terdapat 2 kelompok customers berdasarkan analisis pengaruh **income grub** terhadap **persentase** customers **clicked on ads** yakni :

1. Kedua kelompok ini terbagi berdasarkan **batas income** di angka **3 juta**.
2. Customer yang memiliki **income dibawah** atau **maksimal** diangka **3 juta**, maka **persentase** customer **melakukan clicked on ads** sangat besar dibandingkan tidak melakukan clicked on ads.
3. Apablia customers memiliki **income diatas 3 juta**, maka **persentase** customers tersebut **melakukan clicked on ads** sangat kecil (cenderung **tidak** melakukan clicked on ads).

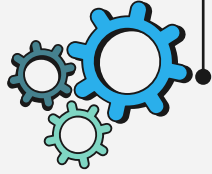


# Bivariate Analysis

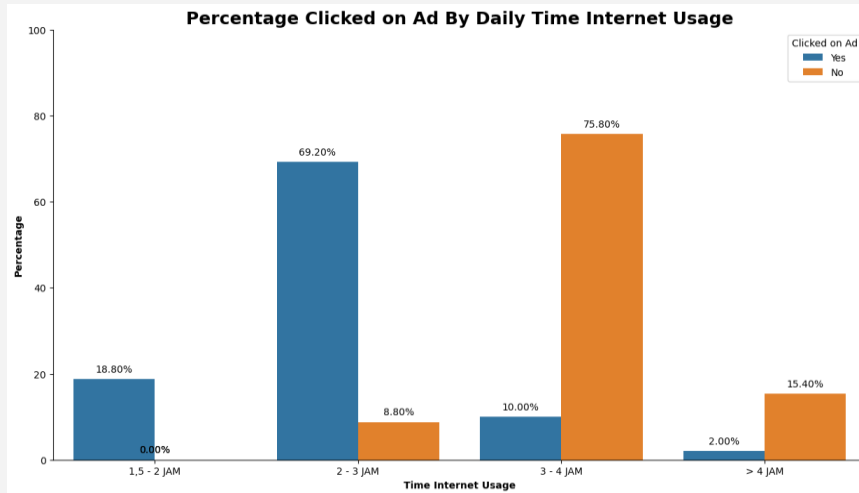


## Business Insight :

Terdapat pola atau trend negative customers **Daily Time Spent on Site** terhadap persentase clicked on ads, dimana **semakin banyak waktu yang dihabiskan** customers untuk melakukan aktivitas **didalam site**, maka **persentase** customers tersebut dalam **melakukan clicked on ads akan semakin kecil** atau customers cenderung **tidak melakukan clicked on ads**.



# Bivariate Analysis



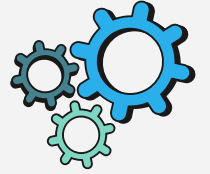
## Business Insight :

Terdapat 2 jenis kelompok dalam hasil analisis pengaruh Time Internet Usage terhadap persentase customers clicked on ads, berikut detail kelompok tersebut :

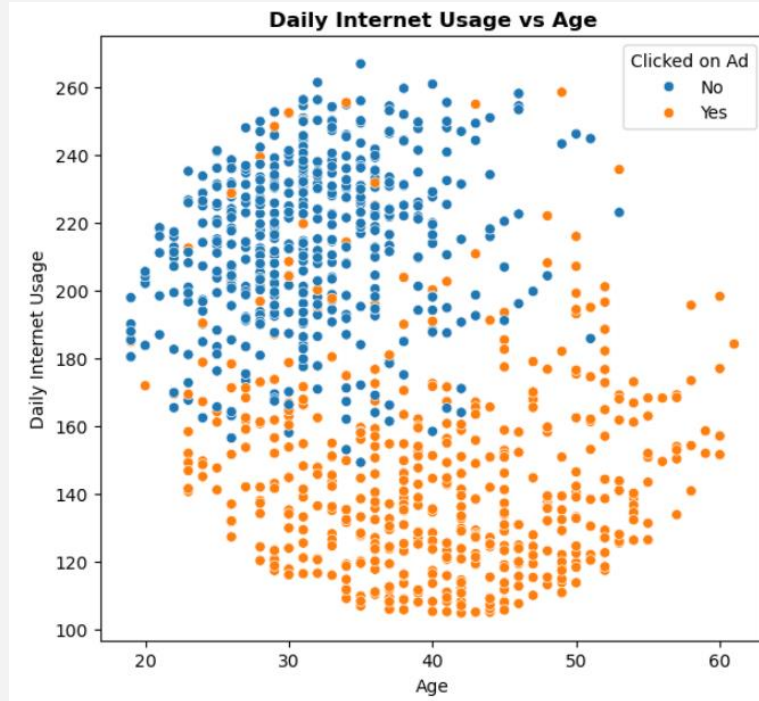
1. Kedua kelompok ini dibatasi oleh Time Internet Usage diangka 2 - 3 jam.
2. Kelompok pertama merupakan kelompok customers yang memiliki persentase clicked on ads lebih tinggi dibandingkan tidak melakukan clicked on ads, kelompok ini memiliki karakteristik yakni Time Internet Usage dibawah atau maksimal diangka 2-3 jam.
3. Kelompok kedua merupakan kelompok customers yang cenderung tidak melakukan clicked on ads, customers pada kelompok ini memiliki karakteristik Time Internet Usage diatas 3 jam.

File Code [Disini](#)





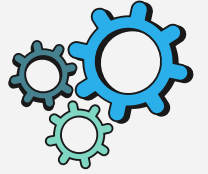
# Multivariate Analysis



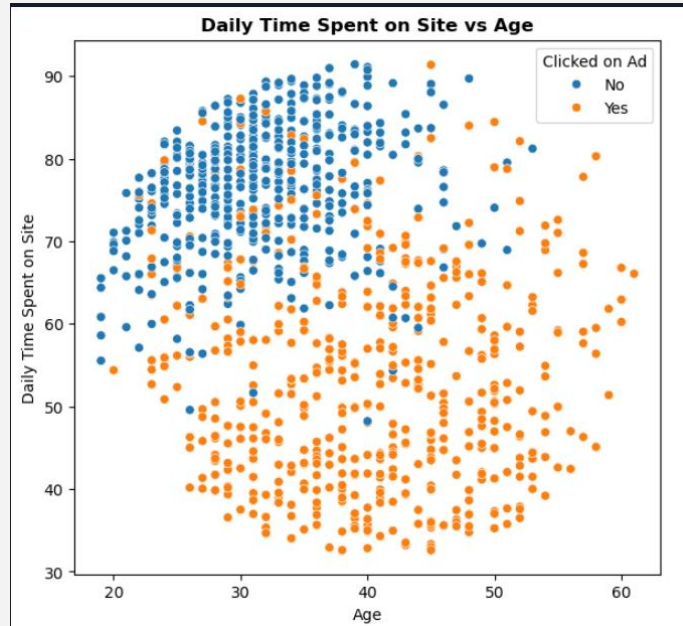
## Business Insight :

Analisis selanjutnya adalah mengetahui **pengaruh Daily Internet Usage** dan **Age** dari customers terhadap **kemungkinan customers** tersebut melakukan **clicked on ads**. Didapatkan beberapa point yakni :

1. Terdapat **pola customers** yang cenderung melakukan **clicked on ads** memiliki **rentang umur** antara **25 hingga 60 tahun** dan **memiliki** kebiasaan **Daily Internet Usage** dibawah **180 menit**.
2. Sebaliknya **customers** yang cenderung **tidak** melakukan **clicked on ads** memiliki **rentang umur** antara **20 hingga 45 tahun** dan **memiliki** kebiasaan **daily internet usage** diatas **180 menit**.



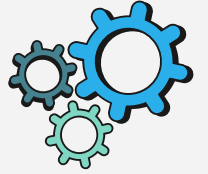
# Multivariate Analysis



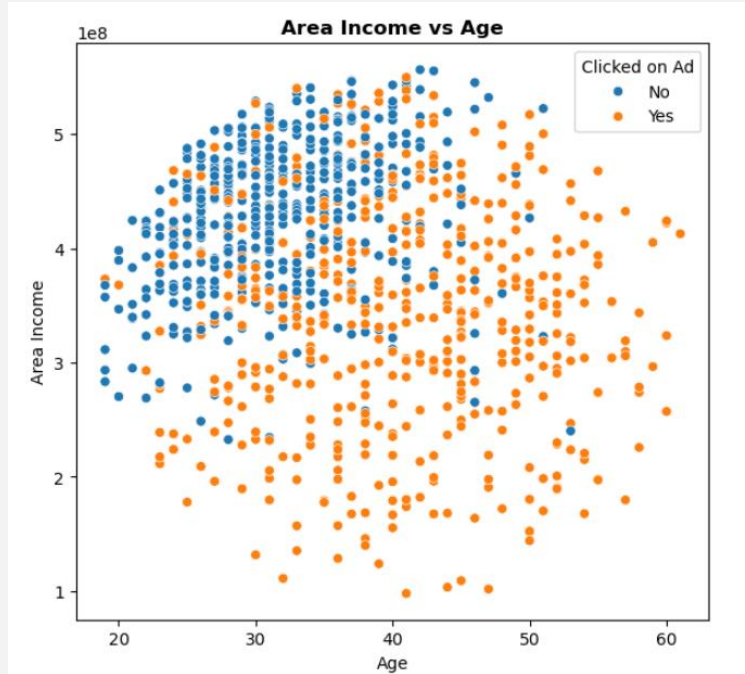
## Business Insight :

Dilakukan sebuah analisis mengenai **pengaruh age** dan **daily time spent on site** terhadap kemungkinan **customers** melakukan **clicked on ads**. Didapatkan kesimpulan sebagai berikut :

1. Terdapat **Pola customers** yang cenderung melakukan **clicked on ads** memiliki **umur** diantara **25 hingga 60 tahun** dan **daily time spent on site** dibawah **70 menit**.
2. Sebaliknya **customers** yang cenderung **tidak** melakukan **clicked on ads** memiliki **umur** antara **20 hingga 45 tahun** dan **daily time spent on site** diatas **70 menit**.



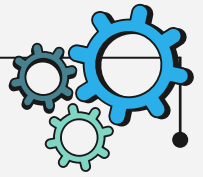
# Multivariate Analysis



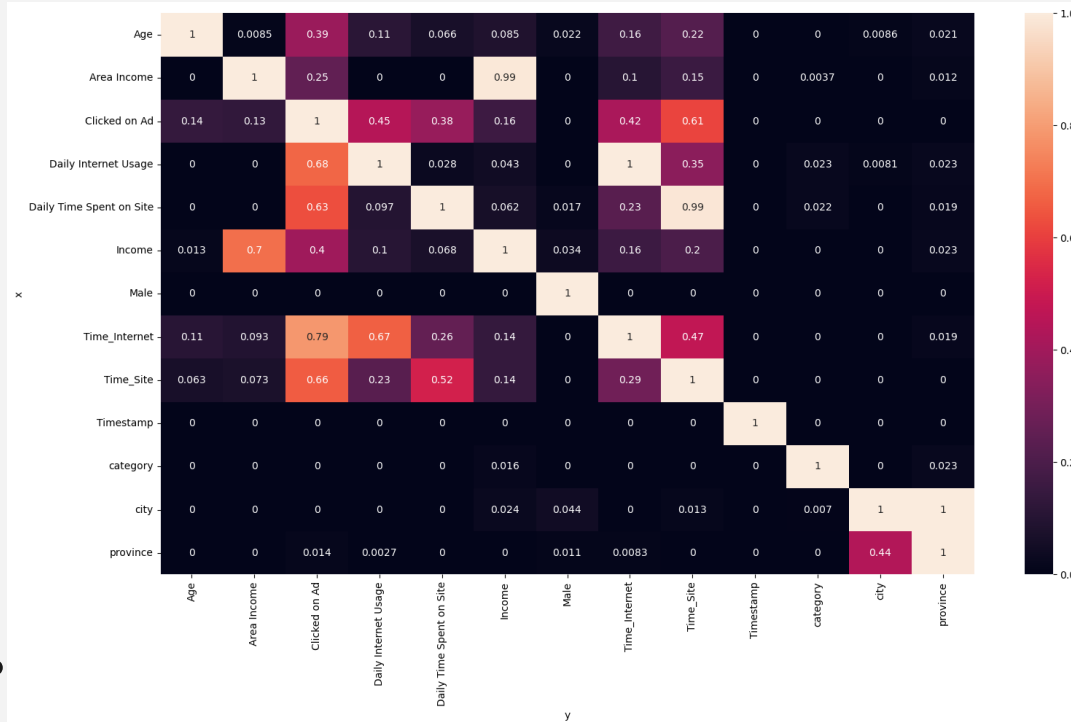
## Business Insight :

Dilakukan **analisis** menggunakan visualisasi **Heatmap** mengenai **pengaruh Area income** dan **Age** terhadap kemungkinan **customers** melakukan **clicked on ads**. Pada analisis ini, **customer** terbagi menjadi **dua kelompok** utama yakni :

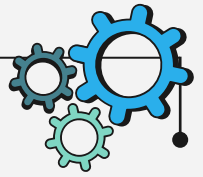
1. Terdapat **pola customers** yang cenderung melakukan **clicked on ads** memiliki **area income dibawah 3.5 Juta** dan memiliki rentang **umur 25 hingga 60 tahun**.
2. Sedangkan **customers** yang **tidak** melakukan **clicked on ads** memiliki **area income diatas 3.5 Juta** dan memiliki rentang **umur** lebih pendek yakni antara **20 hingga 45 tahun**.



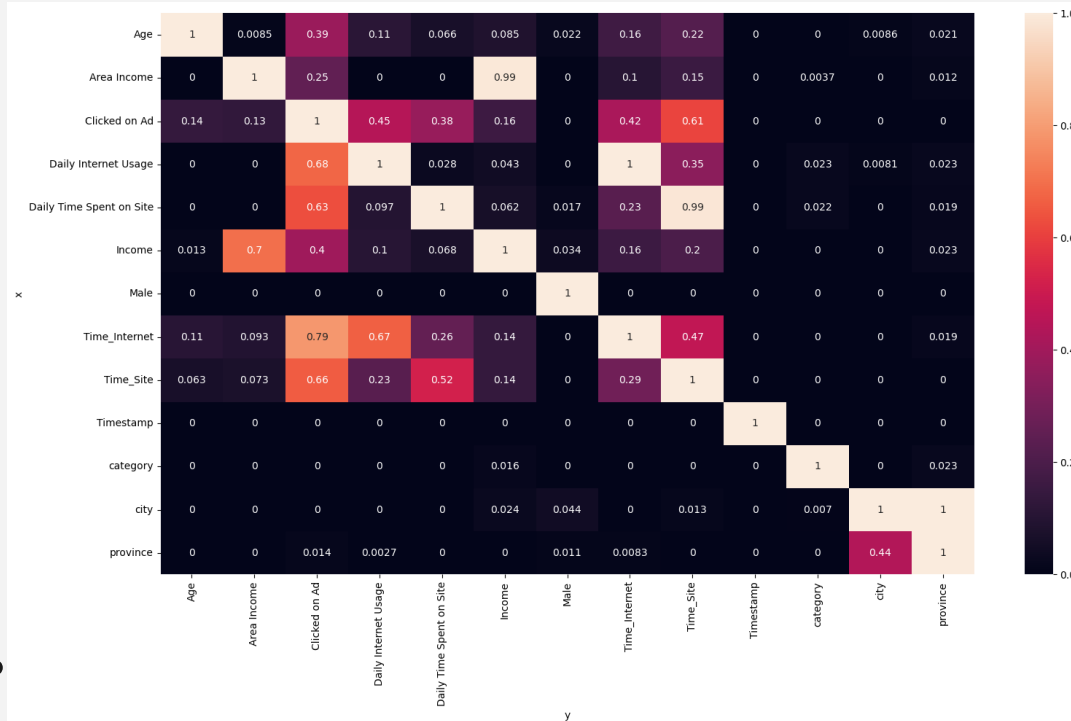
# Multivariate Analysis



Dalam Multivariate Analysis dilakukan perhitungan **korelasi** antara sebuah **feature** dengan **targetnya (clicked on ad)**, teknik yang digunakan pada tahapan ini adalah **PPS (Predictive Power Score)**. **PPS** dipilih karena **memilik** beberapa **keunggulan** seperti dapat **memproses data numerical** dan **categorical** secara bersamaan dan dapat **digunakan** apabila **data bersifat non-linear**.



# Multivariate Analysis

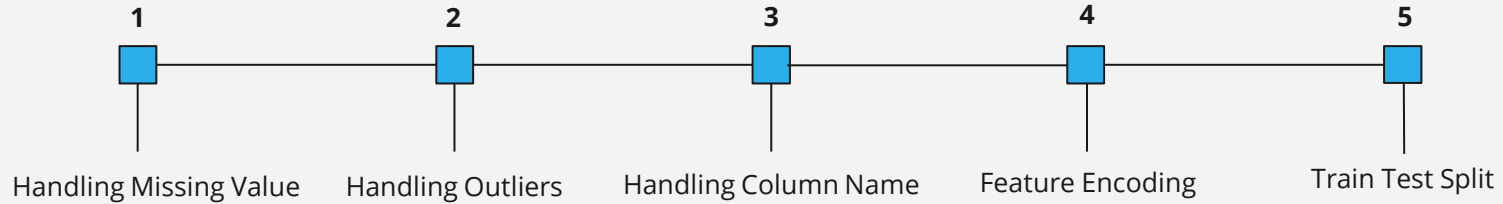
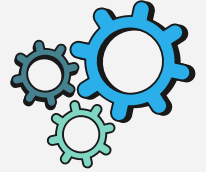


PPS memiliki nilai atau skor antara 0 hingga 1, dimana nilai yang lebih tinggi menunjukkan bahwa sebuah feature memiliki hubungan yang lebih kuat dengan target.

Terdapat 4 feature yang memiliki skor mendekati 1 atau memiliki hubungan kuat terhadap kolom target, feature tersebut yakni :

1. Daily Internet Usage.
2. Daily Time Spent on Site.
3. Area Income.
4. Age

# Data Preprocessing Timeline



# Data Preprocessing

```
#Handling Feature that have Missing Value
df['Daily Time Spent on Site'].fillna(df['Daily Time Spent on Site'].mode()[0], inplace=True)
df['Area Income'].fillna(df['Area Income'].median(), inplace=True)
df['Daily Internet Usage'].fillna(df['Daily Internet Usage'].mode()[0], inplace=True)

#Handling Outliers in Feature Area Income
outlier_lambda = df['Area Income']
Q1 = df['Area Income'].quantile(0.25)
Q3 = df['Area Income'].quantile(0.75)
IQR = Q3 - Q1
lower_thres = Q1 - 1.5 * IQR
upper_thres = Q3 + 1.5 * IQR

df = df[(df['Area Income'] >= lower_thres) & (df['Area Income'] <= upper_thres)]

#Rename Columns Target (Clicked on Ad)
df.rename(columns={'Clicked on Ad': 'Label'}, inplace=True)

#Feature Encoding With label Encoder
df['Label'] = df['Label'].replace({'No': 0,
                                  'Yes': 1})

from sklearn.model_selection import train_test_split

#Split Feature and Target
x = df.drop(columns='Label')
y = df['Label']

#Split Data to Data Train and Data Test
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=42)
```

**Data Preprocessing** merupakan **proses persiapan** data sebelum data tersebut digunakan **untuk analisis** atau **pemodelan** lebih lanjut. **Tujuan** dari data preprocessing adalah **memastikan kualitas dari data** tersebut **telah sesuai** atau cocok **untuk analisis** atau **pemodelan**.

Data preprocessing yang dilakukan terbagi menjadi 5 tahapan yakni :

## 1. Handling Missing Value

Merupakan teknik untuk mengatasi data – data yang kosong atau hilang.

## 2. Handling Outliers

Merupakan teknik untuk mengatasi data – data yang memiliki nilai diluar range data pada umumnya (outliers).

## 3. Handling Columns Name

Merupakan teknik mengubah nama column menjadi standart tertentu.

## 4. Feature Encoding

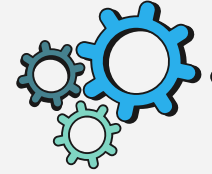
Merupakan teknik mengubah format value dari categorical menjadi numerical.

## 5. Splitting Data

Merupakan teknik membagi data menjadi beberapa bagian, data dibagi memiliki tujuan yakni untuk bahan pembelah=jaran model dan bahan test atau evaluasi model.

**File Code [Disini](#)**

# Modeling Machine Learning



Logistic Regression



Support Vector Machine



Decision Tree



Random Forest



AdaBoost  
Algorithm

**XGBoost**

XGBoost Classifier



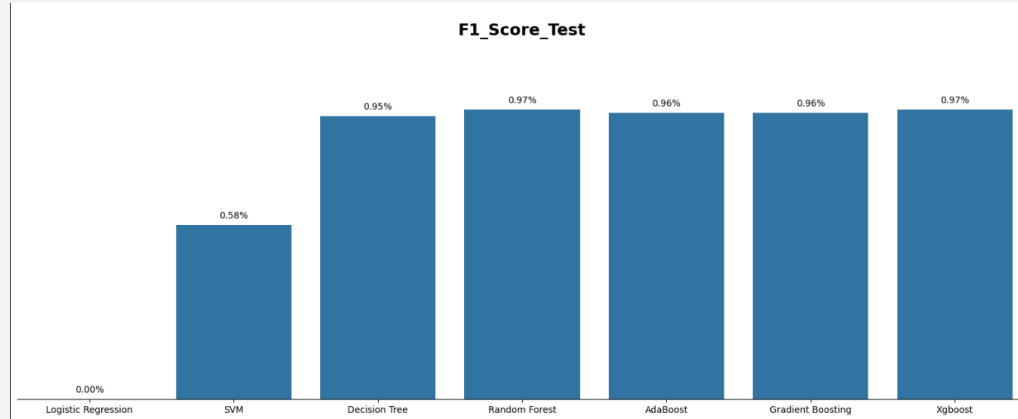
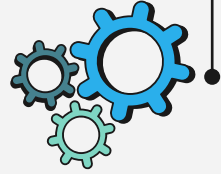
Gradient  
Boosting  
Algorithm

**Modeling machine learning** dilakukan dengan percobaan menggunakan **7 algoritma classification**, algoritma – algoritma tersebut **dipilih karena** ke tujuh **algoritma ini paling populer** dan **cukup powerfull** digunakan **untuk pemodelan** machine learning berbasis **classification**. Algoritma yang digunakan yakni :

1. **Logistic Regression**
2. **Support Vector Machine**
3. **Decision Tree**
4. **Random Forest**
5. **AdaBoost Algorithm**
6. **Gradient Boosting Algorithm**
7. **XGBoost Classifier**



# Modeling Machine Learning



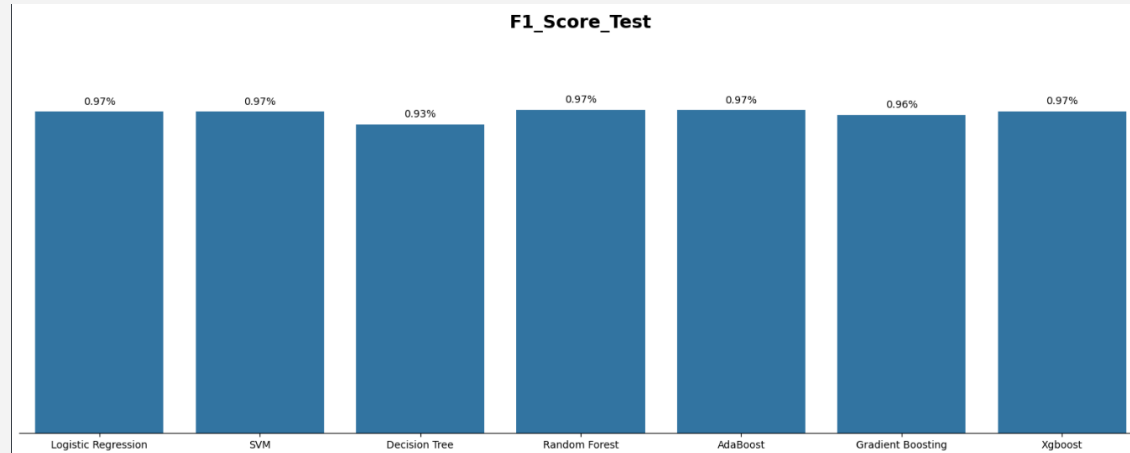
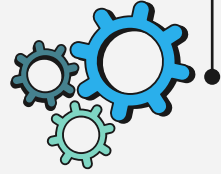
Pada tahap percobaan pertama, **7 algoritma classification dilatih** dengan **data** yang **tidak dilakukan normalisasi** terlebih dahulu. **Data normalisasi** sendiri merupakan **pengubahan range data** dari yang **tidak seragam** menjadi data yang **memiliki range seragam**. **Normalisasi data** hanya **mengubah range nilainya** saja **tanpa mengubah informasi** yang terkandung **didalamnya**, jadi normalisasi data cukup aman digunakan untuk meningkatkan kualitas data.

**Hasil yang didapatkan** pada modelling machine learning tanpa normalisasi data yakni algoritma **Random Forest memiliki F1-Score paling tinggi** dibandingkan algoritma yang lain, dan **Logistic Regression memiliki F1-Score paling rendah** karena algoritma ini **berbasis jarak** dan **sangat sensitive** terhadap feature yang memiliki **range data tidak seragam**.

**F1-Score digunakan** sebagai metric evaluation karena pada kasus ini kami **ingin mendapatkan keseimbangan nilai** antara **Precision** dan **Recall**.

File Code [Disini](#)

# Modeling Machine Learning

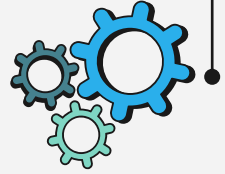


Pada tahap percobaan kedua, **7 algoritma classification** dilatih dengan **data** yang sudah **dilakukan normalisasi**. Hasil yang **didapatkan** pada modeling machine learning dengan normalisasi data yakni algoritma **Random Forest** tetap memiliki **F1-Score paling tinggi** dibandingkan algoritma yang lain, dan **Logistic Regression** memiliki **F1-Score hampir sama dengan Random Forest**, karena algoritma ini **berbasis jarak** jadi **sangat penting dilakukan normalisasi data** terlebih dahulu **sebelum** dilakukan **modelling machine learning**.

Karena **nilai F1-Score** pada algoritma **Random Forest paling tinggi** dari pada algoritma yang lain, **baik tanpa normalisasi data maupun dengan normalisasi data**, maka **algoritma ini** yang **dipilih sebagai model utama** dalam memprediksi customers clicked on ads.

File Code [Disini](#)

# Modeling Machine Learning



	Nama Model	F1_Score_Test
0	Logistic Regression	0.000000
1	SVM	0.583851
2	Decision Tree	0.937500
3	Random Forest	0.968750
4	AdaBoost	0.958333
5	Gradient Boosting	0.958763
6	Xgboost	0.968750

Tanpa Normalisasi Data

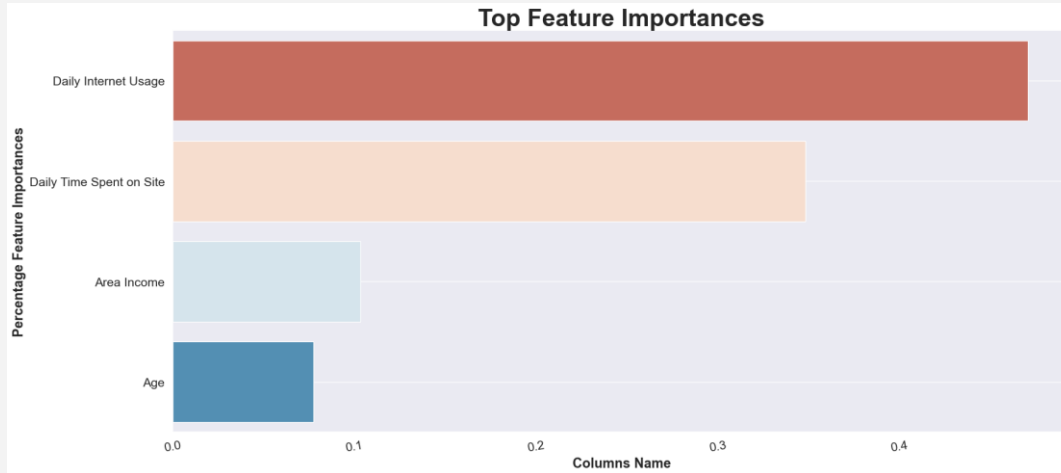
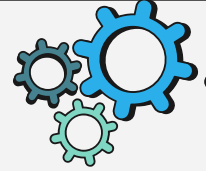
	Nama Model	F1_Score_Test
0	Logistic Regression	0.966184
1	SVM	0.966184
2	Decision Tree	0.947368
3	Random Forest	0.970874
4	AdaBoost	0.970588
5	Gradient Boosting	0.960784
6	Xgboost	0.966507

Dengan Normalisasi Data

Kedua **DataFrame** diatas merupakan **hasil** lebih **detailnya** nilai **F1-Score** yang didapatkan oleh setiap algortima yang diujikan **baik** menggunakan **normalisasi data** maupun **tidak melakukan normalisasi data**. Dapat **ditarik kesimpulan** bahwa **normalisasi data** **sangat penting dilakukan untuk meningkatkan performa** dari **model machine learning** yang dibuat, dilihat dari algoritma **Random Forest** terdapat **kenaikan nilai F1-Score** dari **tanpa** menggunakan **normalisasi data** **menjadi** menggunakan **normalisasi data**. Meskipun kenaikan nilai F1-Score sangat minim tetapi **ini membuktikan** bahwa **normalisasi data** merupakan **hal yang sangat penting dilakukan sebelum modeling machine learning**.

File Code [Disini](#)

# Business Insight



## Business Insight :

Hasil modeling machine learning menghasilkan beberapa **feature penting** atau **feature importance** yang memiliki **pengaruh** sangat **besar** terhadap kemungkinan **customers** melakukan **clicked on ads** atau **tidak clicked on ads**, feature tersebut yakni :

1. **Daily Internet Usage**
2. **Daily Time Spent on Site**
3. **Area Income**
4. **Age**

# Business Insight

## Segment Customers

Customers yang ada terbagi menjadi 2 kelompok utama yakni :

### **1. Kelompok Clicked on Ads.**

Customer pada kelompok ini merupakan customers yang memiliki karakteristik yakni :

- Menggunakan internet dibawah 3 jam (180 menit) dalam sehari.
- Lama waktu mengunjungi situs dibawah 1 jam dalam sehari
- Rata – rata berumur lebih dari 40 tahun
- Memiliki income dibawah 3 juta.

### **2. Kelompok Tidak Clicked on Ads.**

Customer pada kelompok ini merupakan customers yang memiliki karakteristik yakni :

- Menggunakan internet diatas 3 jam (180 menit) dalam sehari.
- Lama waktu mengunjungi situs diatas 1 jam dalam sehari.
- Rata – rata berumur lebih muda dibawah 40 tahun.
- Memiliki income diatas 3 juta.

# Business Recommendation

1. Melakukan Ads yang lebih tertarget kepada customers yang memiliki umur lebih dari 40 tahun, karena customer dengan umur tersebut mudah tergiur terhadap iklan yang ditawarkan.
2. Dilakukan improvement pada kualitas ads yang ditayangkan, lalu dilakukan sebuah uji hipotesis kepada customers yang memiliki income diatas 3 juta, apakah dengan improvement tersebut customers menjadi tertarik dengan ads atau tidak. Apabila customers dengan income diatas 3 juta tetap tidak tertarik dengan ads yang ditawarkan, kemungkinan besar customers tidak tertarik karena produknya bukan karena ads yang ditawarkan.
3. Pemberian konten yang singkat dan menarik, hal ini perlu dilakukan karena hasil analisis customers yang melakukan clicked on ads cenderung tidak punya banyak waktu untuk menggunakan internet maupun mengunjungi site, kemungkinan customers dengan karakteristik ini cenderung mudah bosan dalam menggunakan internet. maka dari itu pemberian konten yang singkat dan menarik diharapkan dapat menarik customers untuk clicked ads.
4. Menggunakan Ads berbasis Influencer, customers yang cenderung tidak terlalu lama menggunakan internet perlu dilakukan penawaran menggunakan influencer. Hal ini karena sebagian besar customers akan menggunakan internet untuk melihat konten – konten orang terdekat maupun beberapa tokoh yang mereka percayai, jadi ads melalui influencer kemungkinan besar akan membuat customer tersebut tertarik dan percaya atas iklan yang ditawarkan.



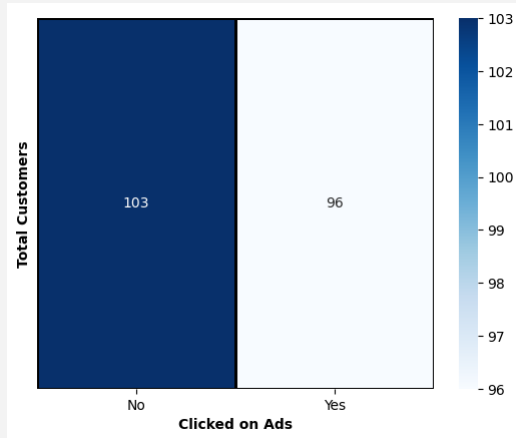
# 03

## Business Simulation

# Business Simulation

Simulasi Bisnis akan dilakukan dengan **membandingkan** antara **tanpa** penggunaan **machine learning** dengan **penggunaan machine learning**. Simulasi dilakukan dengan menggunakan **data testing** yang berjumlah **199 customers**, Dimana data aslinya customers yang **melakukan click ads** sebanyak **96 customers** dan sebanyak **103 customers tidak melakukan click ads**. Pada tahapan ini diberikan **asumsi** bahwa **biaya iklan 1 customers** sebesar **Rp 20.000** dan **profit** yang didapatkan apabila **customers click ads** adalah **Rp 50.000**

## Tanpa Machine Learning



Tanpa menggunakan machine learning, akan dilakukan perhitungan profit yang didapatkan oleh perusahaan.

**Profit = Revenue - Cost**

**Profit = (Customers Click \* Profit Click) - (Total Customers \* Cost Ads)**

**Profit = (96 \* Rp 10.000) - (199 \* Rp 2.000)**

**Profit = (Rp 960.000) - (Rp 398.000)**

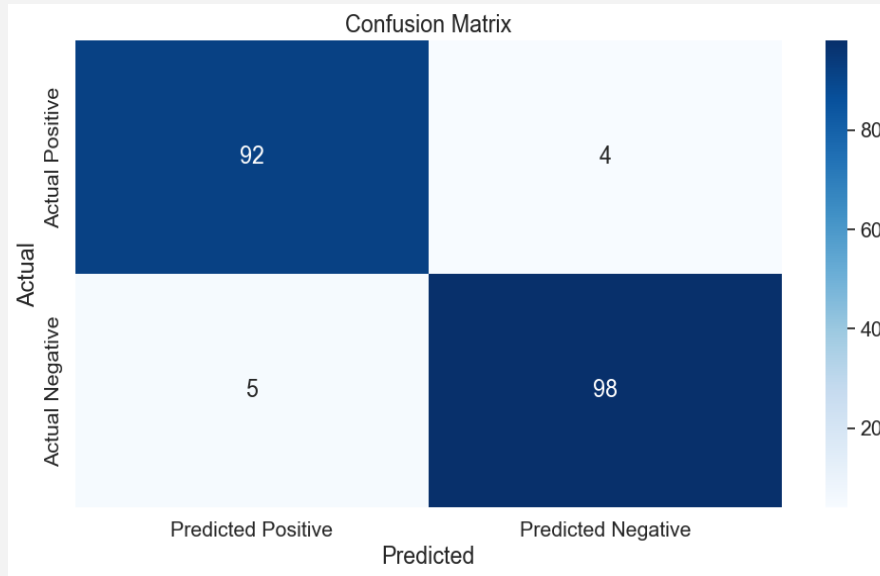
**Profit = Rp 562.000**

File Code [Disini](#)



# Business Simulation

## Dengan Machine Learning



Dengan menggunakan machine learning, akan dilakukan perhitungan profit yang didapatkan oleh perusahaan.

**Profit = Revenue - Cost**

**Profit = (Customers Click \* Profit Click) - (Total Customers \* Cost Ads)**

**Profit = (92 \* Rp 10.000) - (92 \* Rp 2000)**

**Profit = (Rp 920.000) - (Rp 184.000)**

**Profit = Rp 736.000**



Peningkatan Profit :

# Rp 174.000

Dengan menggunakan Machine Learning profit yang didapatkan **meningkat** sebesar **30,9%** dibandingkan tanpa menggunakan machine learning.



# 04

## Finalisasi Model

# Saved Model Machine Learning

```
import pickle

file_name = "Models_ads.sav"
pickle.dump(model_final, open(file_name, 'wb'))
```

**Tahapan Finalisasi Model** merupakan tahapan Dimana model yang sudah dibuat dan dilatih pada tahapan sebelumnya disimpan dalam sebuah format khusus. Tahapan ini dilakukan agar model dapat digunakan Kembali untuk melakukan prediksi pada data – data yang lain. Tentunya jika ingin menggunakan Kembali model yang telah dilatih atau disimpan ini, hanya perlu melakukan load pada modelnya saja, tanpa perlu melakukan end to end analisis dari awal.



# Thanks!

**Do you have any questions?**

[ammarbaasir19@gmail.com](mailto:ammarbaasir19@gmail.com)

<https://www.linkedin.com/in/ammar-baasir-973873234/>