

# Predict Clicked Ads Customer Classification by using Machine Learning



**Created by:**

**Muhammad Iqbal Mudzakky**

Email : [iqbalmudzakky@gmail.com](mailto:iqbalmudzakky@gmail.com)

LinkedIn : [\(4\) Muhammad Iqbal Mudzakky | LinkedIn](#)

Github : [iqbalmudzakky \(github.com\)](https://github.com/iqbalmudzakky)

*"Iqbal is a junior data scientist with a robust background in exploratory data analysis, preprocessing, modeling, visualization, and providing actionable insights. He possesses extensive experience in handling diverse data types, particularly in the Fintech and E-Commerce industries. His expertise is evident through successful completion of numerous supervised and unsupervised learning projects, which demonstrate his ability to extract valuable information from data, develop accurate models, and effectively communicate findings. With a proven track record in multiple facets of data science, Iqbal serves as a highly valuable asset to any team or organization seeking expertise in data-driven decision making and problem-solving."*

A faded, light grey background image of a city skyline with various skyscrapers and buildings.

# PROJECT BACKGROUND

“In the world of digital marketing, online ads are essential for promoting products and services. To improve ad effectiveness, companies use machine learning to understand customer behavior and preferences.

By analyzing customer behavior, companies can figure out which ads are more likely to catch their attention and interest, enabling them to show more relevant ads.”



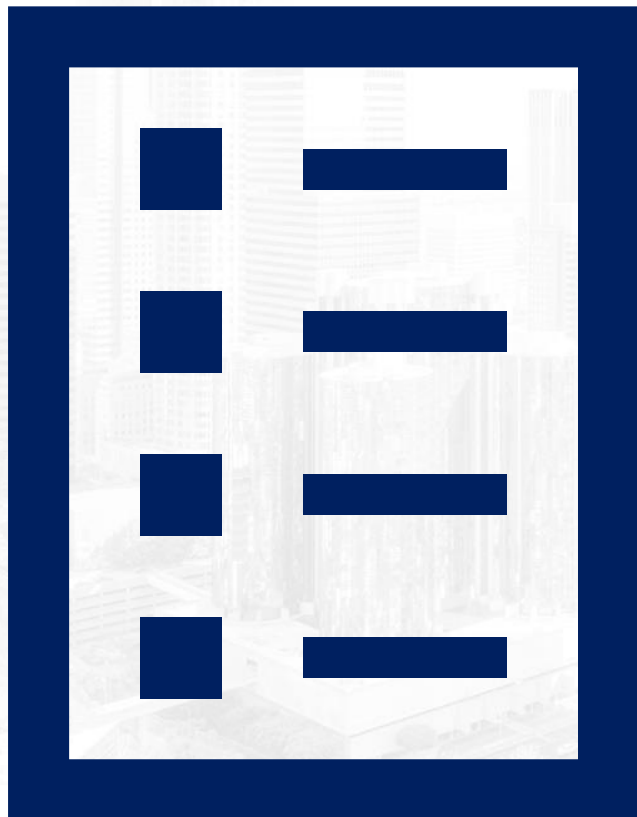
## Goals?

is to create a machine learning model that sorts customers based on their chances of clicking on specific ads. This way, the project aims to gain valuable insights into different customer types and their responses to ads.

## The Main Issue/Problem?



“How to identify and classify customers based on their likelihood to click on specific ads.”



1. Exploratory Data Analysis

2. Data Cleaning & Preprocessing

3. Data Modeling

4. Business Recommendation and Simulation

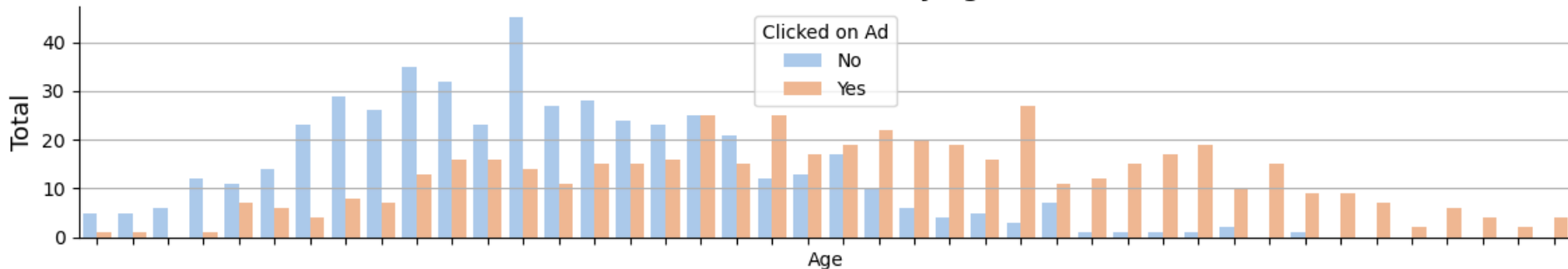




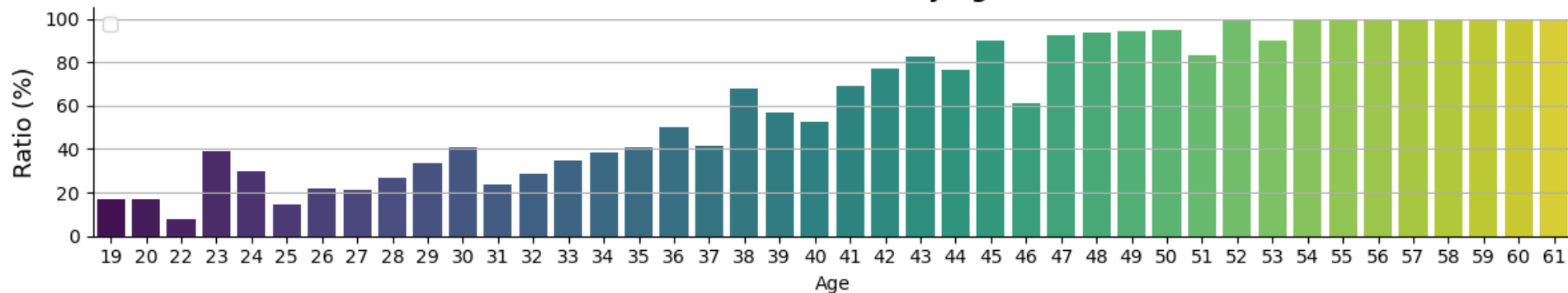
# EXPLORATORY DATA ANALYSIS

# Exploratory Data Analysis

## Total Click on Ads by Age



## Click on Ads Ratio by Age

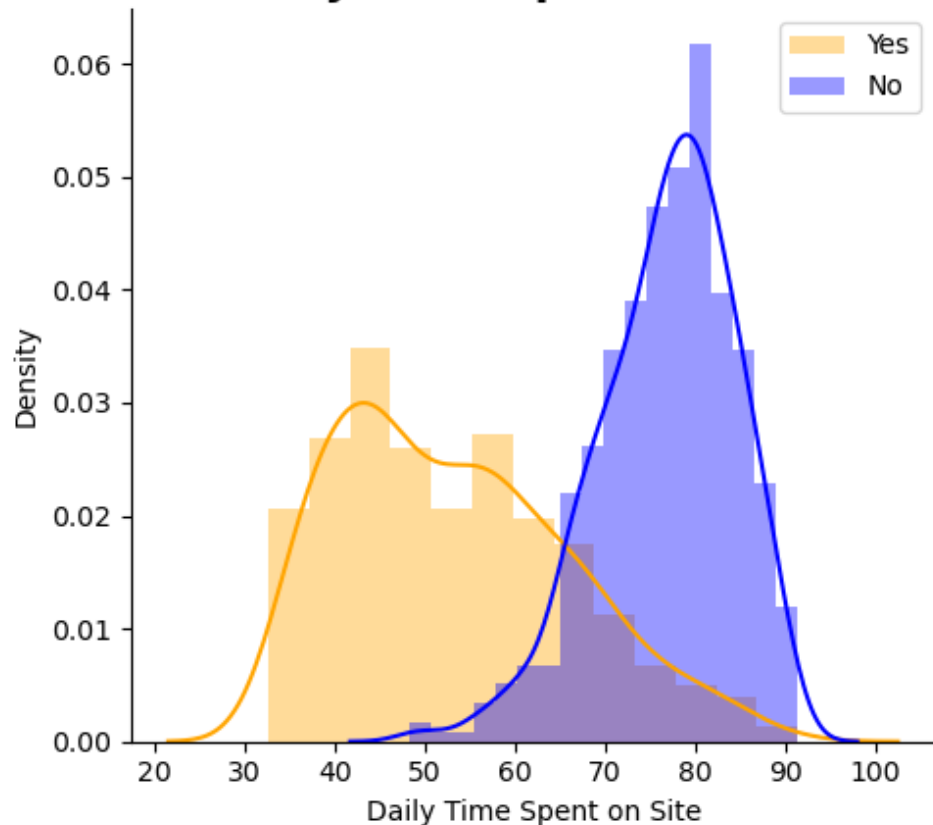


Total User  
1000

The chart shows that users under 35 years old are less likely to click on ads, while those above 35 are more likely to click. The most significant user group falls into the 32-year age range.



## Clicks on Ads by Daily Time Spent on Site



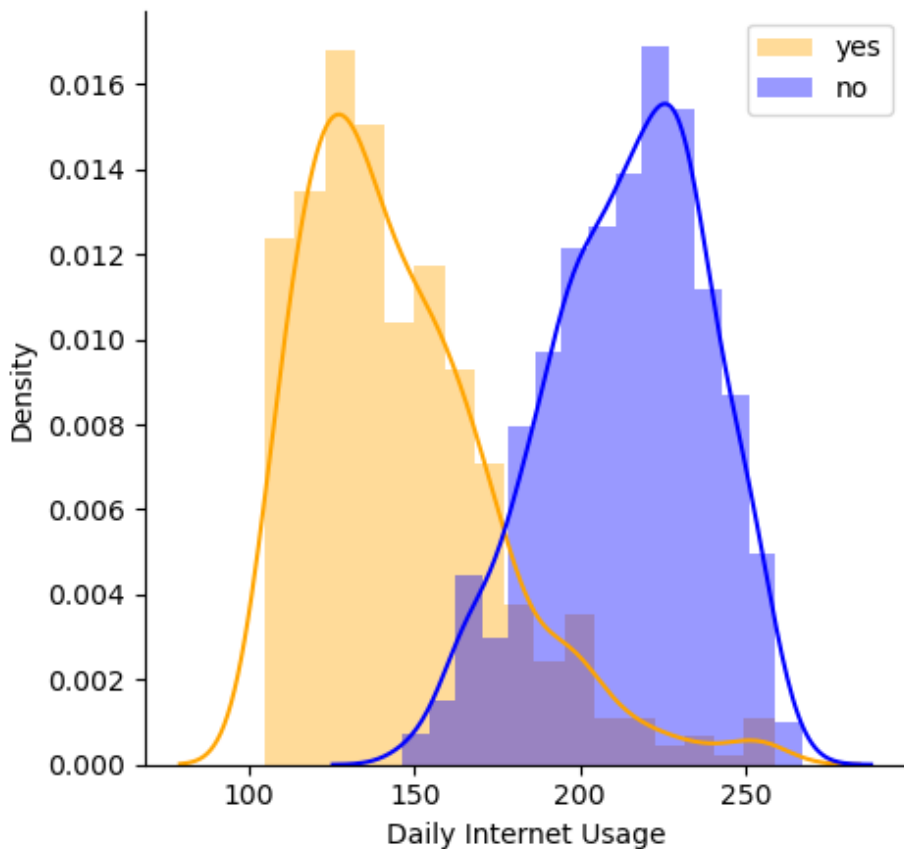
Total User  
1000

The chart shows that fewer users click on ads compared to those who don't. Users who don't click on ads spend more time on the website, around 60 to 90 seconds on average, while those who click on ads spend around 30 to 80 seconds on average.



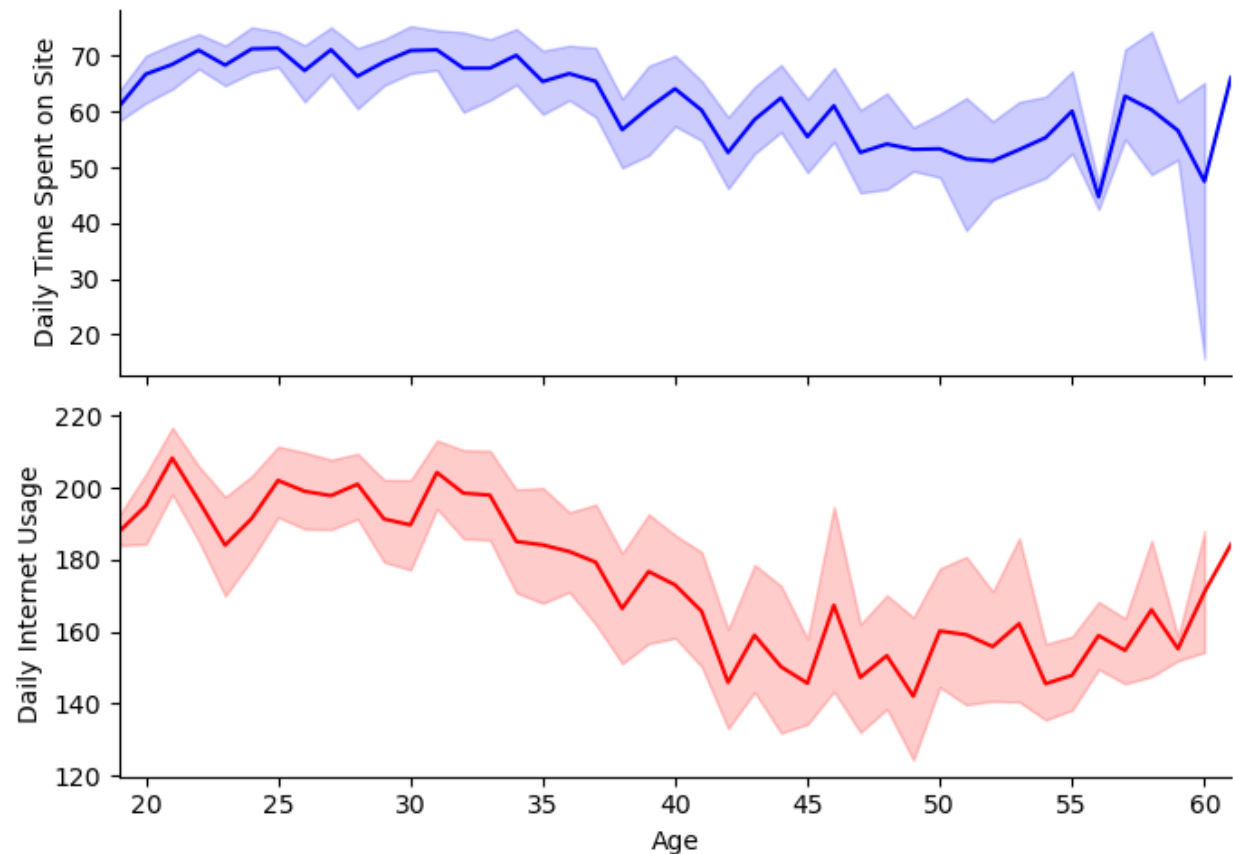
Total User  
1000

Click on Ads by  
Daily Internet Usage



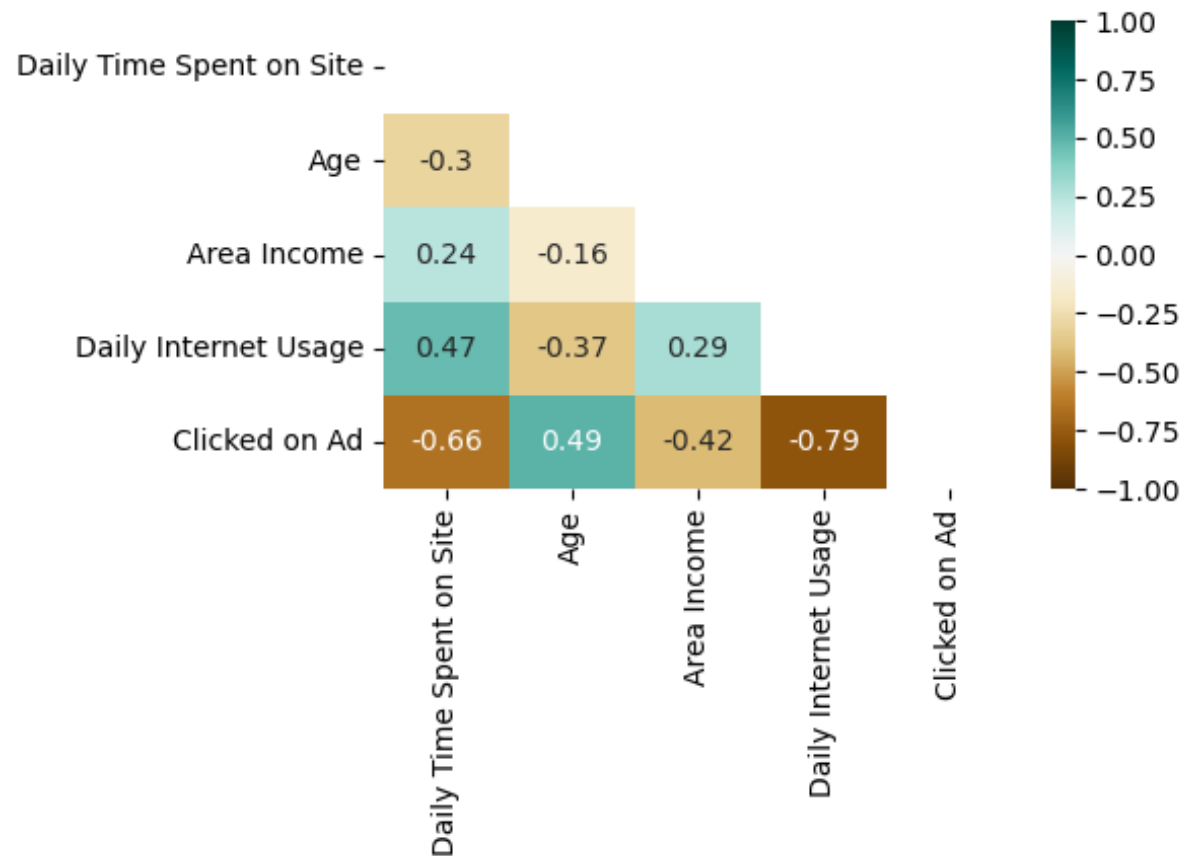
The chart shows that users who don't click on ads use more internet data, mostly between 200 to 250 MB, while users who click on ads use less data, typically between 100 to 150 MB.

## Bivariate Analysis: Age, Time Spent, Internet Usage



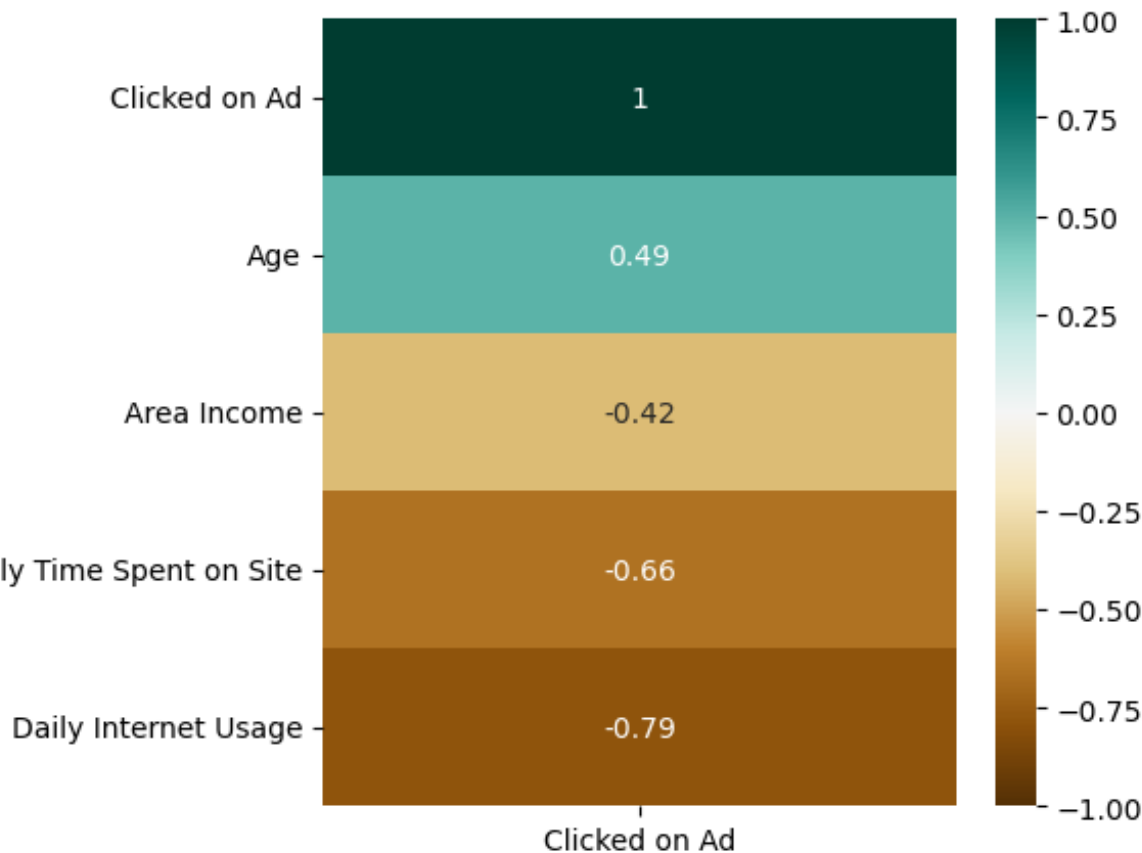
This chart represents the correlation between the age column and internet usage, as well as the time spent. From both plots, we can conclude that there is a negative correlation between age and internet usage, as well as between age and time spent. In other words, as age increases, both internet usage and time spent tend to decrease.

## Correlation Heatmap



Based on the chart, each feature has a relatively strong correlation with the target feature. Additionally, there are some features that show strong correlations with each other, such as internet usage with time spent, internet usage with age, and age with time spent. In other words, these features have a significant relationship with each other, and their values tend to change together.

## Correlation Heatmap



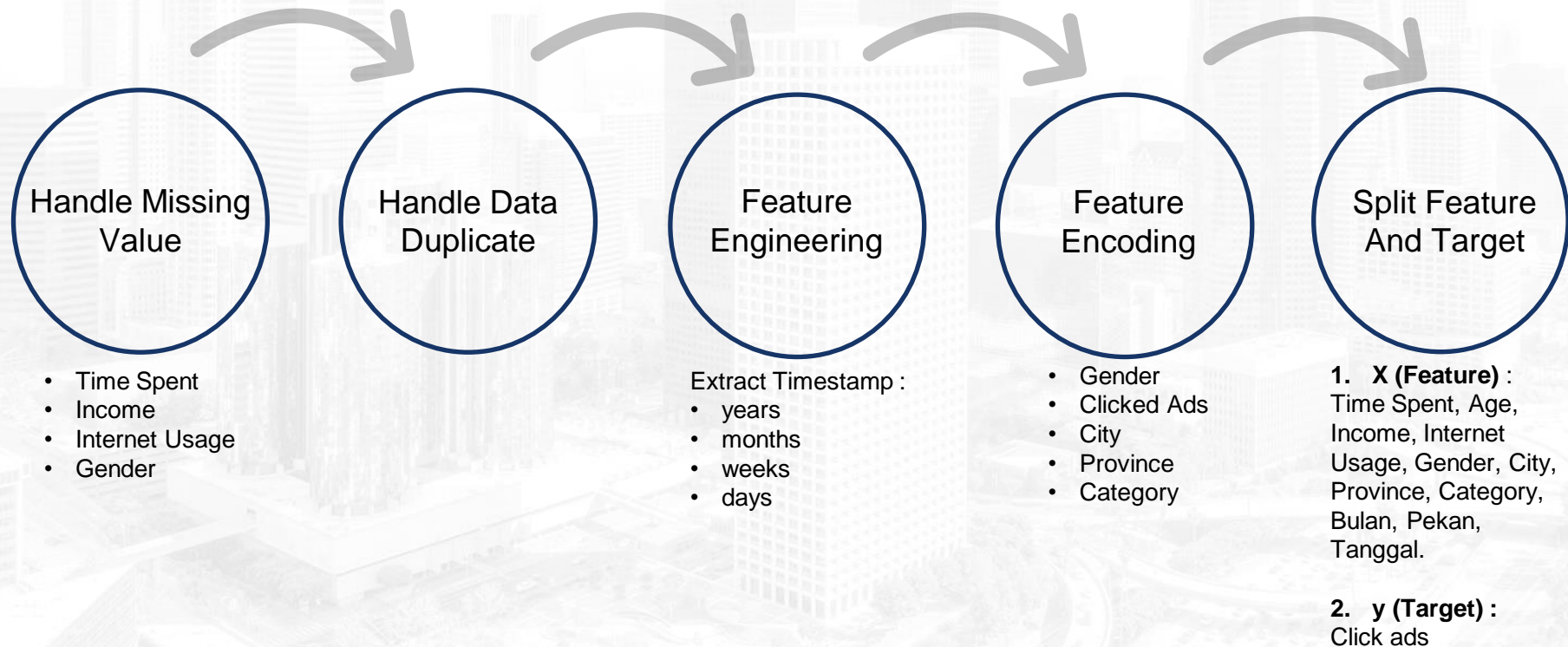
The correlation between independent features and the dependent feature (target feature = 'Clicked on Ad') can be seen more clearly in the adjacent chart. Among the four features, only the "age" feature has a positive correlation, which means that as the user's age increases, the likelihood of clicking on the ad also increases. On the other hand, for the other features, as their values increase, the probability of clicking on the ad decreases.

A faded, grayscale background image of a city skyline with various skyscrapers and buildings.

# **DATA CLEANING AND PREPROCESSING**



# Data Cleansing and Preprocessing







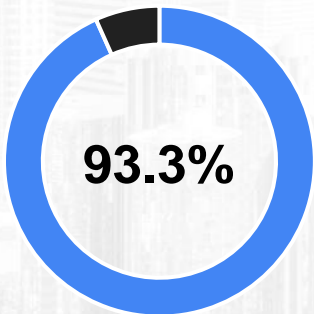
# DATA MODELING & EVALUATION

# Experiment 1 : Without Normalization

Data Train : Data Test  
80 : 20

## Decision Tree

Precision Score :

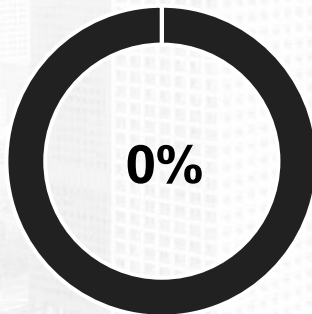


### Evaluation Model:

- Accuracy : 92.5%
- Recall : 92.4%
- F1 Score : 92.8%
- ROC AUC : 92.5%

## Logistic Regression

Precision Score :

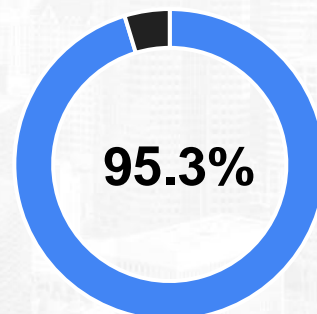


### Evaluation Model:

- Accuracy : 47.5%
- Recall : 0%
- F1 Score : 0%
- ROC AUC : 75%

## Random Forest

Precision Score :



### Evaluation Model:

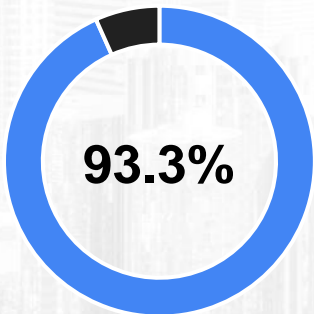
- Accuracy : 95.5%
- Recall : 96.2%
- F1 Score : 95.7%
- ROC AUC : 98.6%

# Experiment 2 : Standard Scaler

Data Train : Data Test  
80 : 20

## Decision Tree

Precision Score :

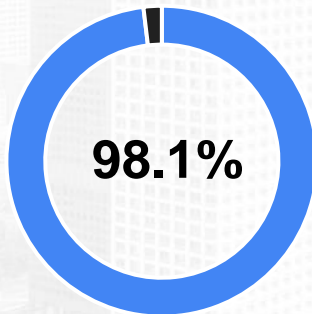


### Evaluation Model:

- Accuracy : 92.5%
- Recall : 92.4%
- F1 Score : 92.8%
- ROC AUC : 92.5%

## Logistic Regression

Precision Score :

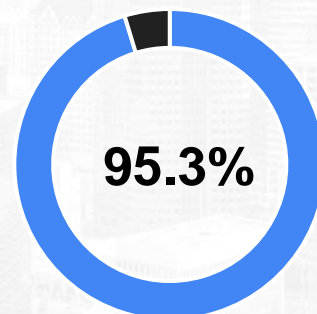


### Evaluation Model:

- Accuracy : 97.5%
- Recall : 97.1%
- F1 Score : 97.6%
- ROC AUC : 99.1%

## Random Forest

Precision Score :

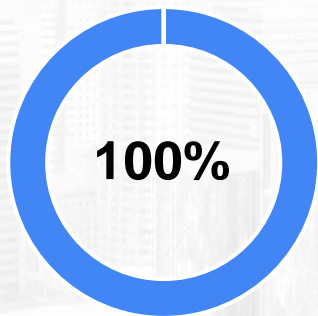


### Evaluation Model:

- Accuracy : 95.5%
- Recall : 96.2%
- F1 Score : 95.7%
- ROC AUC : 98.6%

## Logistic Regression

Precision Score :



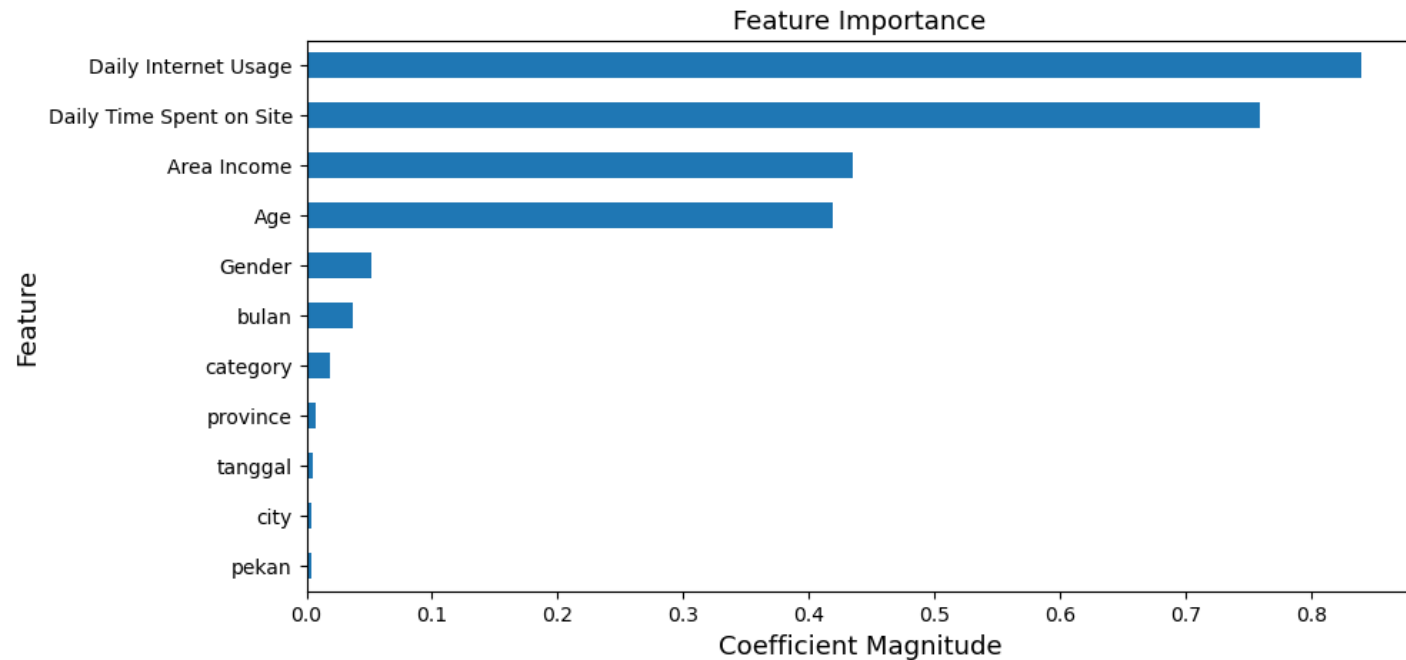
### Evaluation Model:

- Accuracy : 97%
- Recall : 94.3%
- F1 Score : 97.1%
- ROC AUC : 99.2%

## Setting Parameter :

- C : 0.01, 0.1, 1, 10
- solver : lbfgs, liblinear, sag, saga

# Feature Importance



## Influential Features Identified :

1. Internet Usage
2. Time Spent
3. Income
4. Age



# **BUSINESS RECOMMENDATION**



## **Target Market by Age:**

- 1 Focus ad campaigns on users above 35, as they are more likely to click on ads. Also, cater ads and products to the preferences of the significant age group of 30 to 40 years old.

## **Optimize Ads Timing:**

- 2 For users who don't click on ads and spend more time on the website, improve ad appearance and placement to grab their attention. For users who click on ads and spend less time, use attractive and concise offers to engage them effectively.

## **Segment Users by Internet Usage:**

- 3 Users who don't click ads use more internet data, so offer engaging content for them. For users who click ads and use less data, focus on products/services that meet their needs quickly.

## **Focus on Marketing Cost:**

- 4 Daily internet usage and time spent on site are crucial factors in campaign success. Allocate budget wisely based on these features for effective campaigns. Further analysis can optimize costs by targeting specific user segments.

## **A/B Testing:**

- 5 Always conduct A/B tests to try different marketing strategies, ad content, and placements. This way, we can identify the most effective approach to capture user interest and clicks.

## **Personalize Content:**

- 6 Utilize the available data to present more personalized and relevant content to users. By understanding their preferences and habits, we can deliver more persuasive messages and increase the chances of ad clicks.



A faded, light grey background image of a city skyline with various skyscrapers and buildings.

# **BUSINESS SIMULATION**

# Without Machine Learning Model

## Assuming :

1. Target Reach (TR) = 1.000 users
2. Ads Cost (AC) = Rp 500/reach
3. Conversion Rate (CR) = 1.5%
4. Revenue per Transaction (RpT) = Rp 100.000

## Simulation :

1. Total Cost =  $AC * TR$   
=  $Rp\ 500 * 1.000\ user$   
=  $Rp\ 500.000$
2. Total Revenue =  $TR * CTR * CR * RpT$   
=  $1.000\ user * 50\% * 1.5\% * Rp\ 100.000$   
=  $Rp\ 750.000$

3. Profit = Total Revenue – Total Cost

$$= Rp\ 750.000 - Rp\ 500.000$$

=>

## Based On Dataset :

1. CTR = 50%

**Rp 250.000**

## Assuming :

1. Target Reach (TR) = 1.000 users
2. Ads Cost (AC) = Rp 500/reach
3. Conversion Rate (CR) = 1.5%
4. Revenue per Transaction (RpT) = Rp 100.000

## Simulation :

1. Total Cost =  $AC * TR$   
 $= Rp\ 500 * 1.000\ user$   
 $= Rp\ 500.000$
2. Total Revenue =  $TR * Accuracy * CR * RpT$   
 $= 1.000\ user * 97\% * 1.5\% * Rp\ 100.000$   
 $= Rp\ 1.455.000$

3. Profit = Total Revenue – Total Cost  
 $= Rp\ 1.455.000 - Rp\ 500.000$

=>

## Evaluation Metrics Score (Optimized CTR) :

1. Accuracy = 97%
2. Recall = 95.3%
3. Precision = 100%
4. F1 Score = 97.1%

**\*UP TO 382%!**

**Rp 955.000**

# THANK YOU

Click icon for the code : 