Predict Clicked Ads Customer Classification by using Machine Learning

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

Rakamin Academy

Created by:
Muhammad Iqbal Mudzakky
Email : iqbalmudzakky@gmail.com
LinkedIn : [(4) Muhammad Iqbal Mudzakky | LinkedIn]
Github : [iqbalmudzakky (github.com)]

"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. "

PROJECT BACKGROUND

Company 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.

Rakamin Academy

The Main Issue/Problem?

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

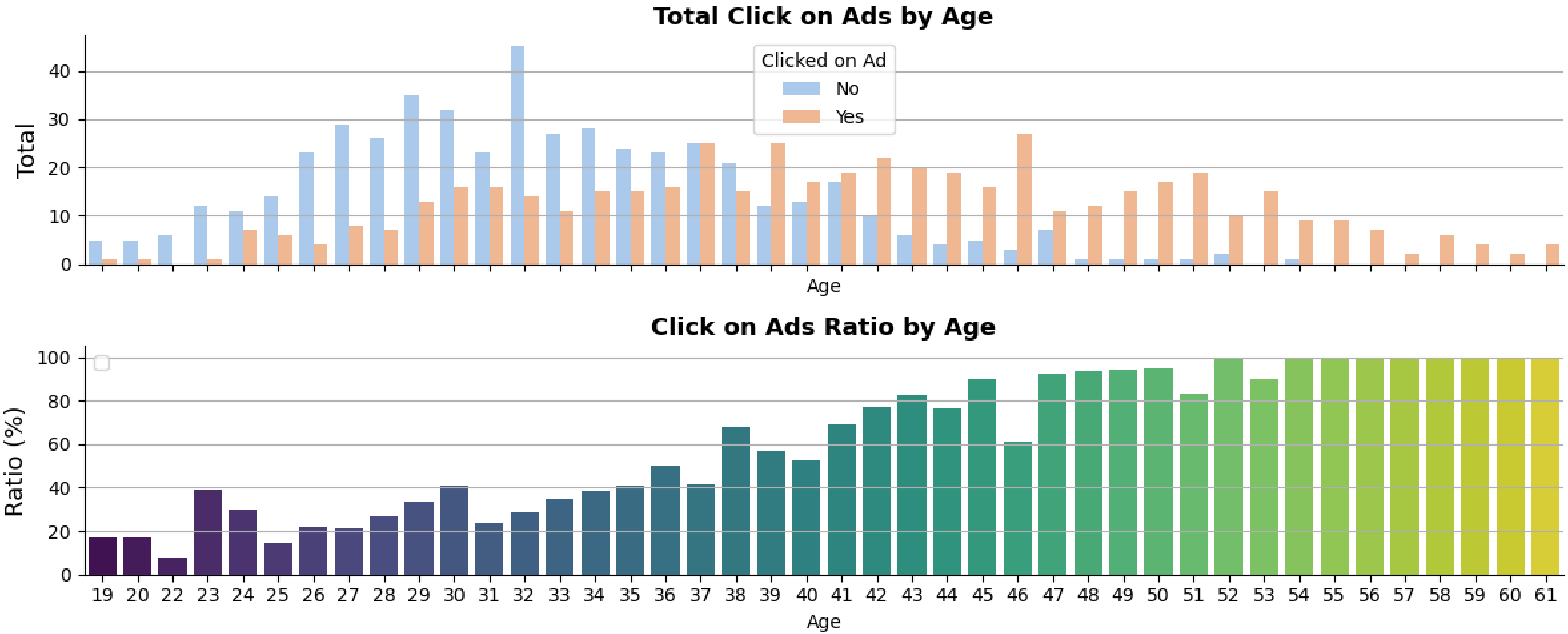
Workflow

Rakamin Academy

1. Exploratory Data Analysis
2. Data Cleaning & Preprocessing
3. Data Modeling
4. Business Recommendation and Simulation

EXPLORATORY DATA ANALYSIS

Exploratory Data Analysis

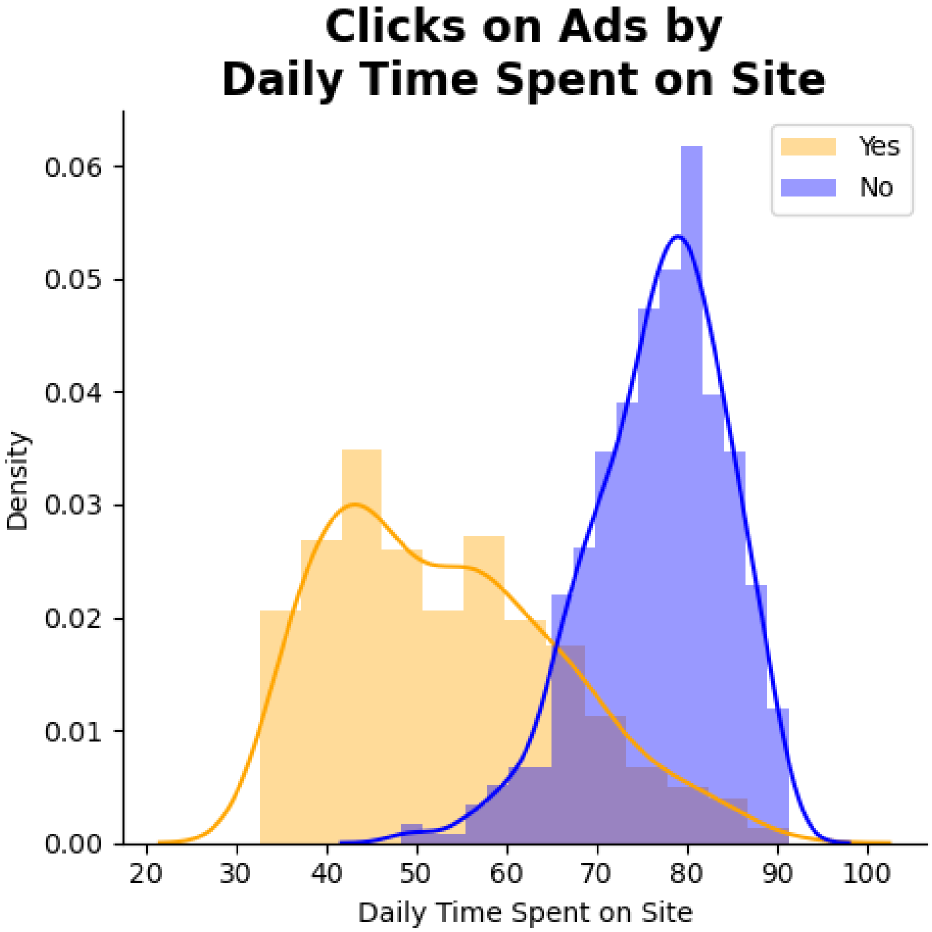


Total User
1000

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.

Exploratory Data Analysis

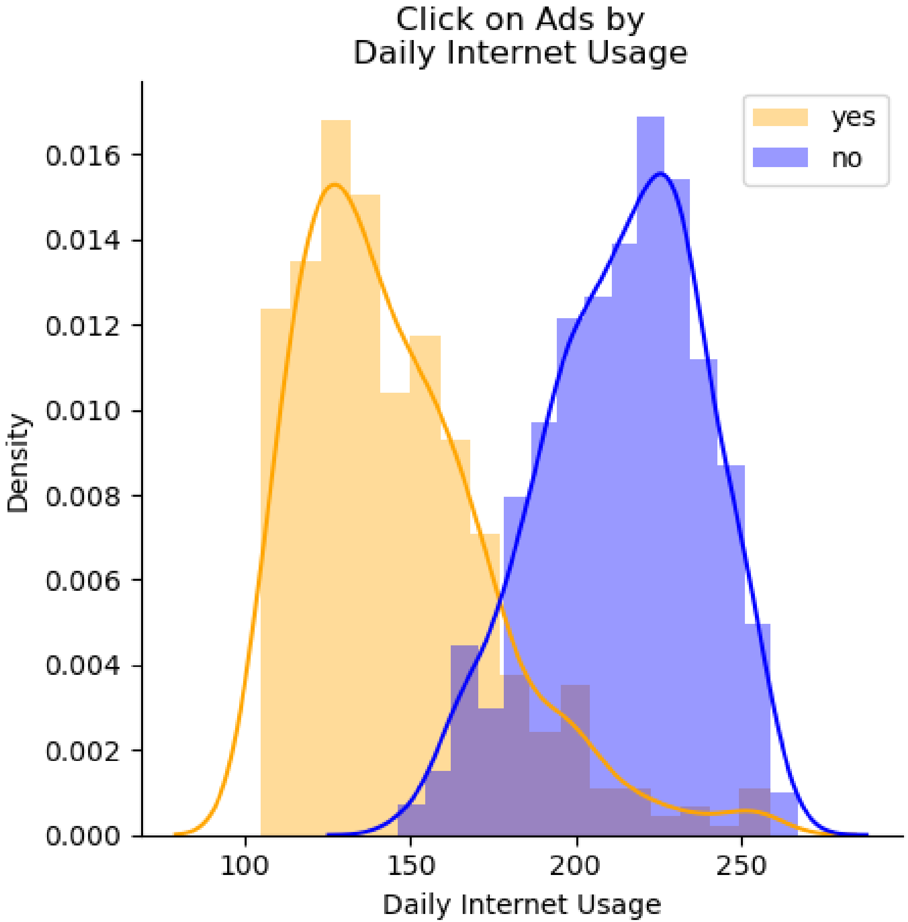
Rakamin Academy



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.

Exploratory Data Analysis



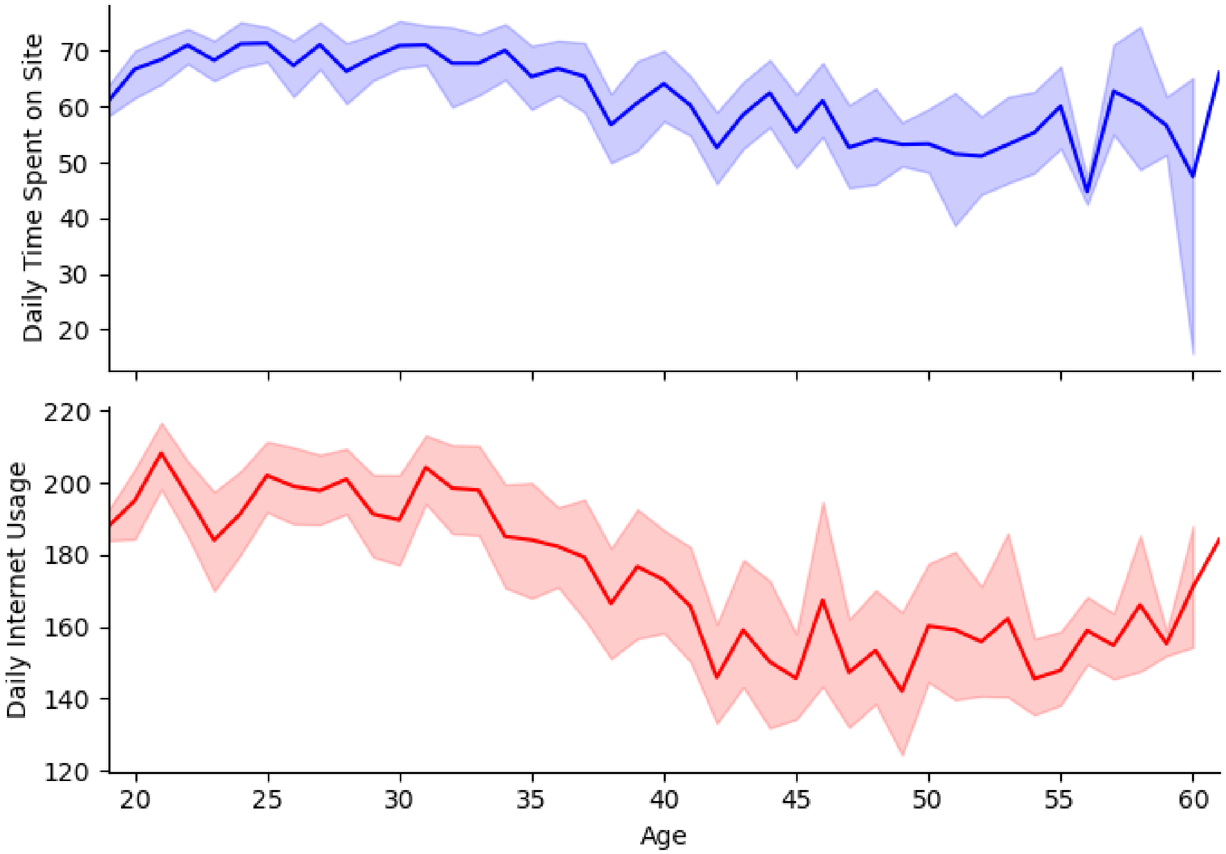
Total User
1000

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.

Exploratory Data Analysis

Rakamin Academy

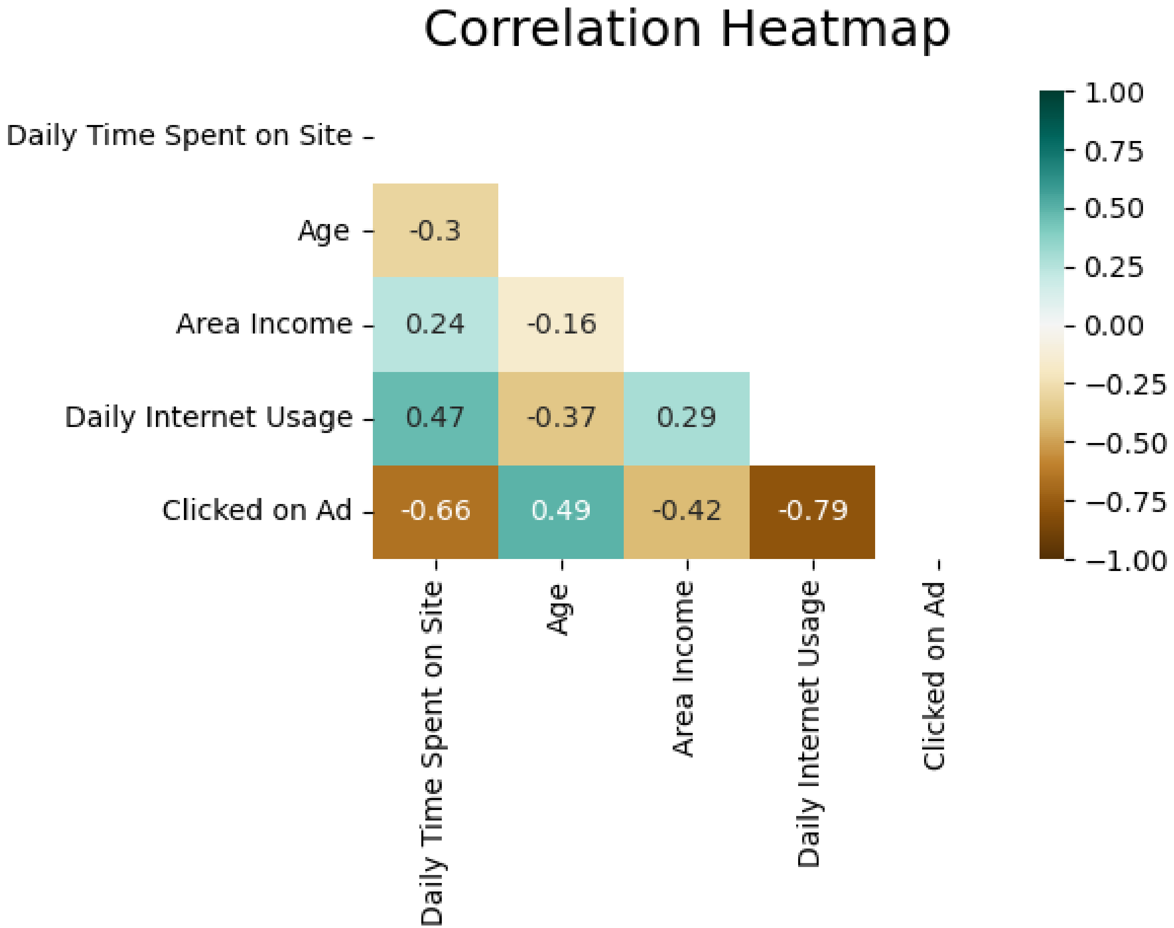
Bivariate Analysis: Age, Time Spent, Internet Usage



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.

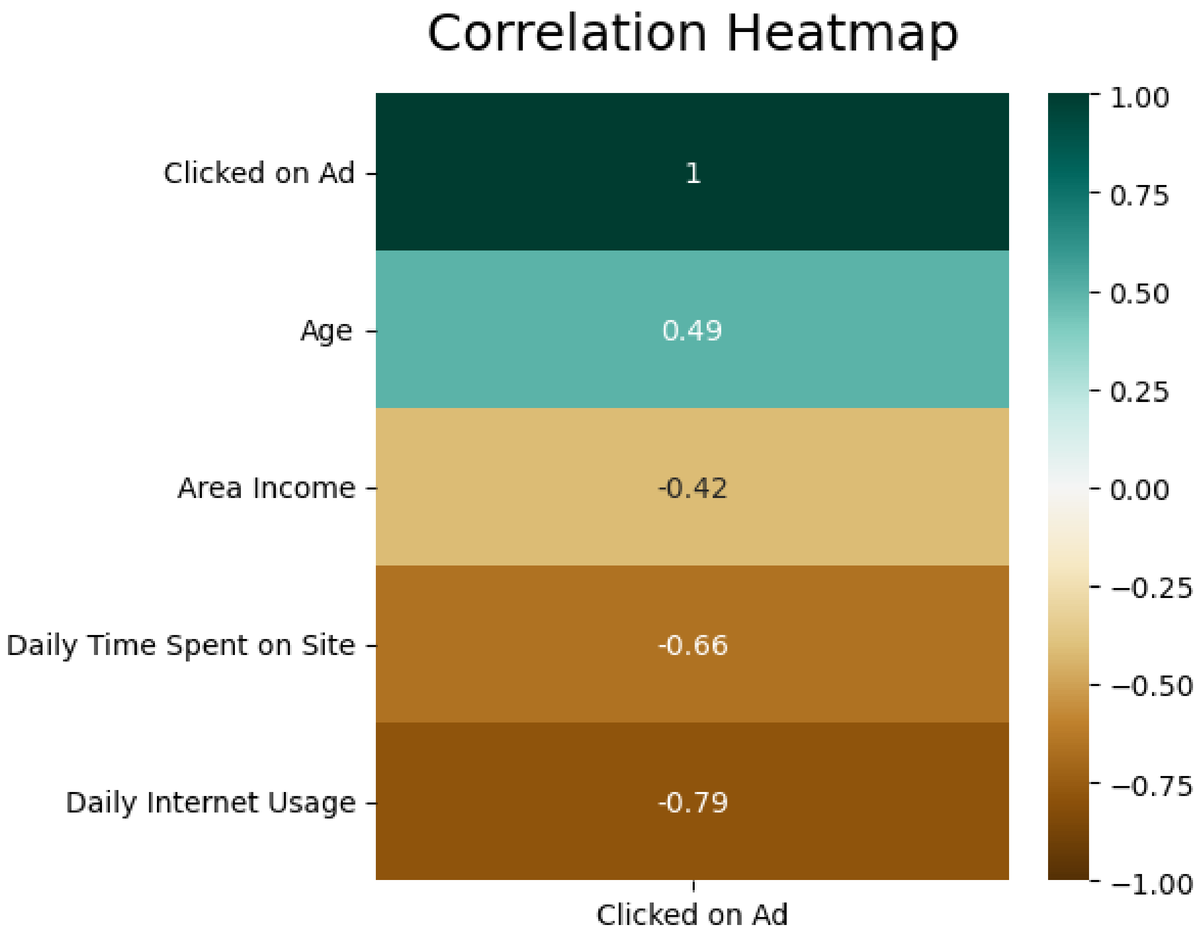
Exploratory Data Analysis

Rakamin Academy



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.

Exploratory Data Analysis



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.

DATA CLEANING AND PREPROCESSING

Data Cleansing and Preprocessing

Rakamin Academy

Handle Missing Value

- Time Spent
- Income
- Internet Usage
- Gender

Handle Data Duplicate

Feature Engineering

Extract Timestamp :

- years
- months
- weeks
- days

Feature Encoding

- Gender
- Clicked Ads
- City
- Province
- Category

Split Feature And Target

1. **X (Feature)** :
Time Spent, Age, Income, Internet Usage, Gender, City, Province, Category, Bulan, Pekan, Tanggal.

2. **y (Target) :**
Click ads

DATA MODELING & EVALUATION

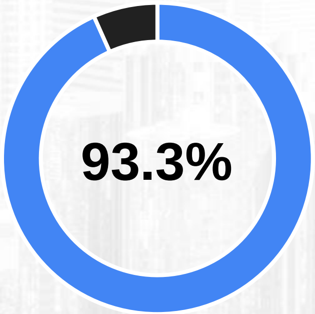
Experiment 1 : Without Normalization

Rakamin Academy

Data Train : Data Test
80 : 20

Decision Tree

Precision Score :

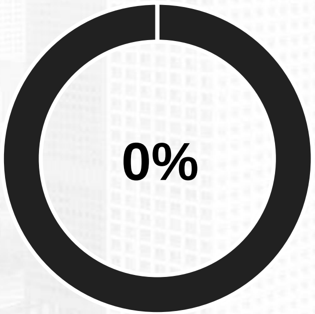


Evaluation Model:

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

Logistic Regression

Precision Score :

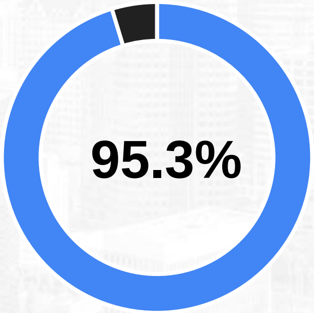


Evaluation Model:

- 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%

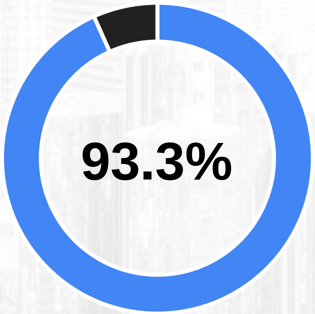
Experiment 2 : Standard Scaler

Rakamin Academy

Data Train : Data Test
80 : 20

Decision Tree

Precision Score :

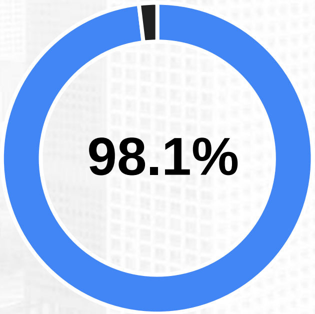


Evaluation Model:

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

Logistic Regression

Precision Score :

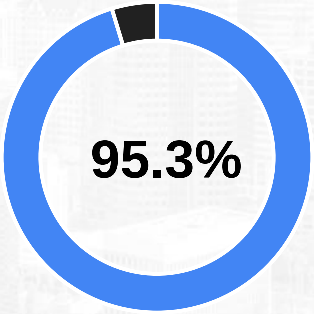


Evaluation Model:

- 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%

Hyperparameter Tuning

Logistic Regression

Precision Score :

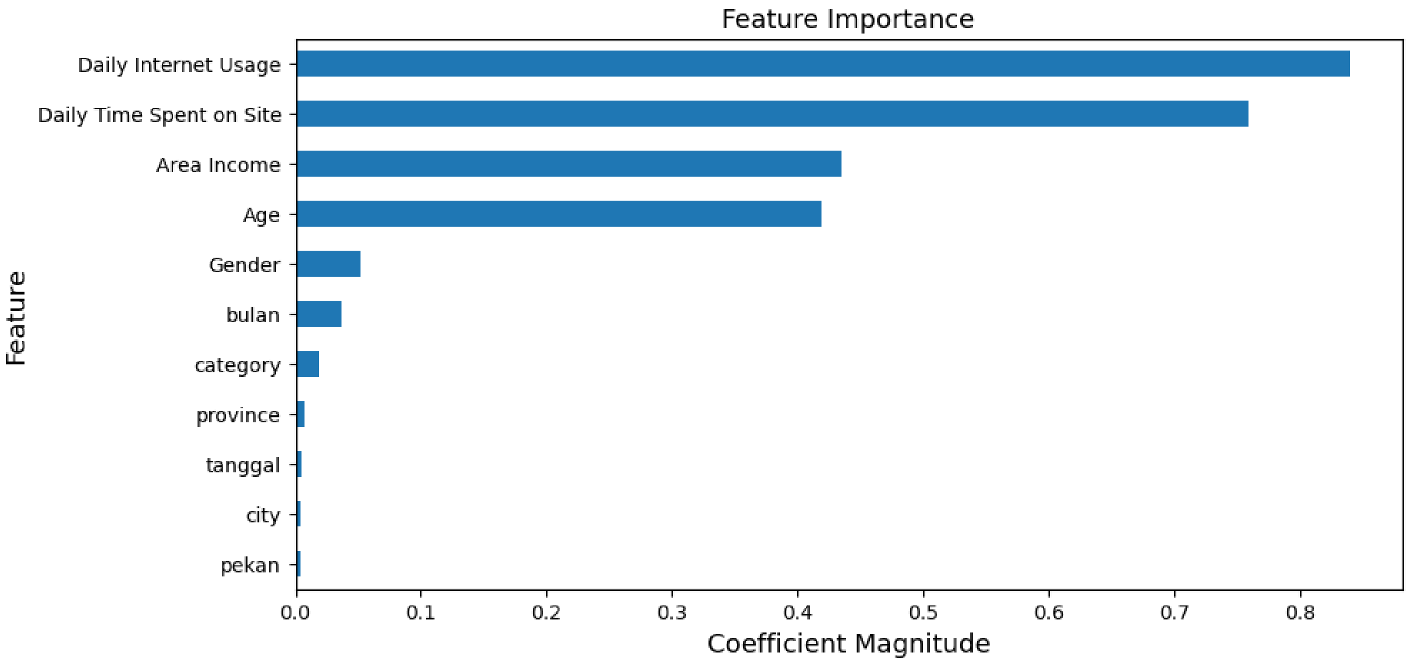
Evaluation Model:

- 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 Identfied :

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

THANK YOU

Click icon for the code :