

# Final Project

Online Shoppers  
Purchasing Intention

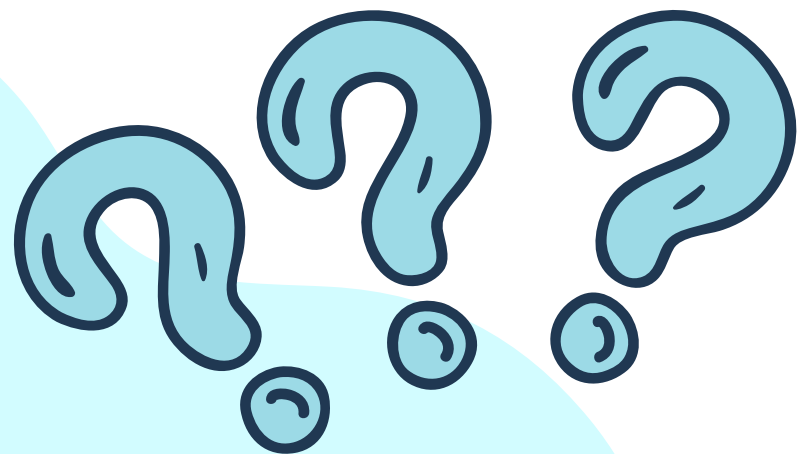
INFINITY - Data Scientist Team in Leslar E-Commerce

Anggota Kelompok:

Abdurrahman Bachsin  
Annisa Pratiwi  
Ida Parwati  
Jad Al Maula  
Kyagus Maula  
Neneng Murnasih  
Putu Arya Saputrawan  
Reza Aulia



# Table of Content



Background  
Problem

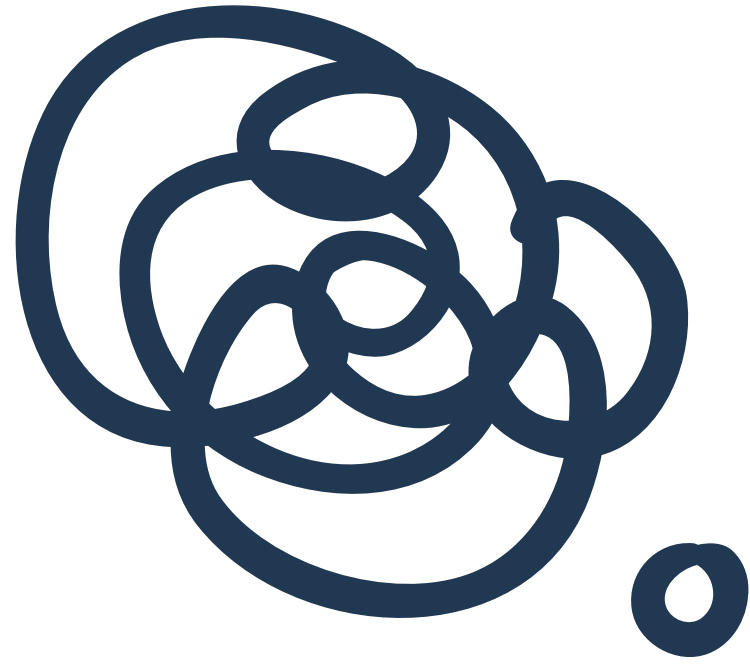


Analyze &  
Modeling

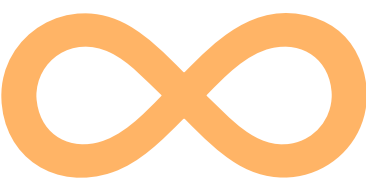


Result &  
Recommendation



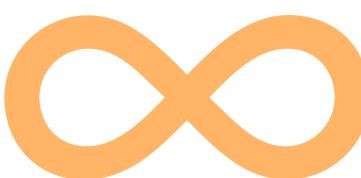
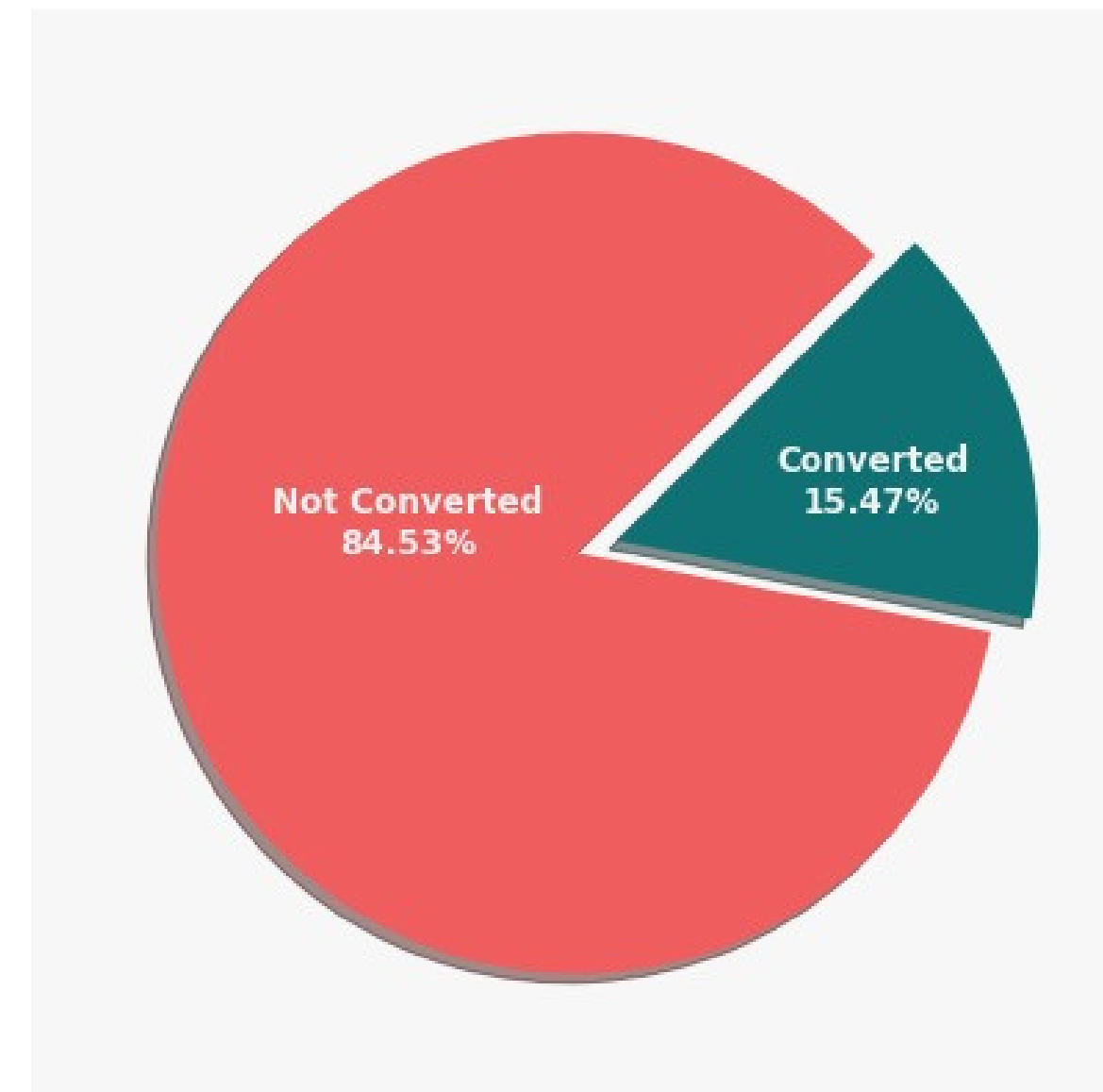


# Background Problem



## Problem Statement

Perusahaan Leslar memiliki conversion rate diangka 15% pada tahun pertama. Pada tahun kedua perusahaan menargetkan peningkatan conversion rate sebesar 10%. Tim Data Scientist diminta membuat model untuk membantu mencapai target tersebut.



## Goals

- Meningkatkan pertumbuhan conversion rate pada tahun kedua sebesar 10%

## Objectives

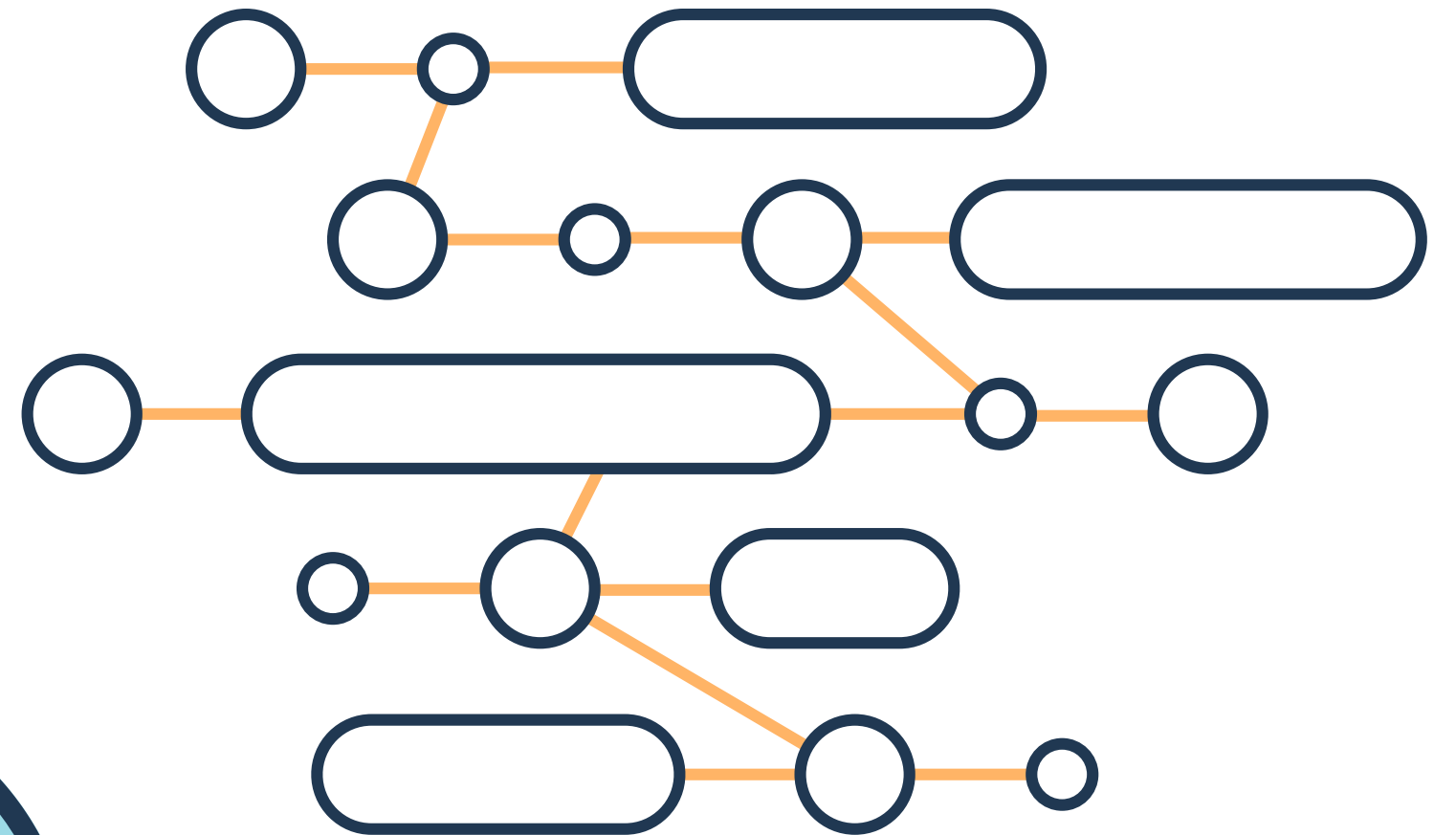
- Menganalisis korelasi antar atribut terhadap pembelian.
- Membuat model machine learning yang dapat memprediksi apakah seorang visitor akan purchase atau tidak.

## Metrics

- Conversion Rate



# Analyze & Modeling



# Dataset Overview

1 Year Historic data with 18 Columns & 12,330 Rows

## Columns list

- Administrative
- Administrative\_Duration
- Informational
- Informational\_Duration
- ProductRelated
- ProductRelated\_Duration
- BounceRates
- ExitRates
- PageValues
- SpecialDay
- Month
- OperatingSystems
- Browser
- Region
- Traffic\_Type
- Visitor\_Type
- Weekend
- Revenue

## Information

Dataset Terdiri dari 12330 Baris dan 18 Kolom.

Tidak Memiliki Nilai Null.

Revenue Merupakan Kolom Target

Seluruh Data Pada Setiap Kolom Sudah Mewakili Informasinya masing-masing



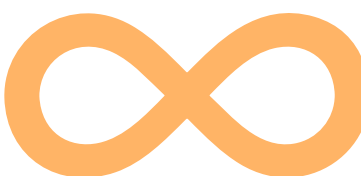
# Exploratory Data Analysis



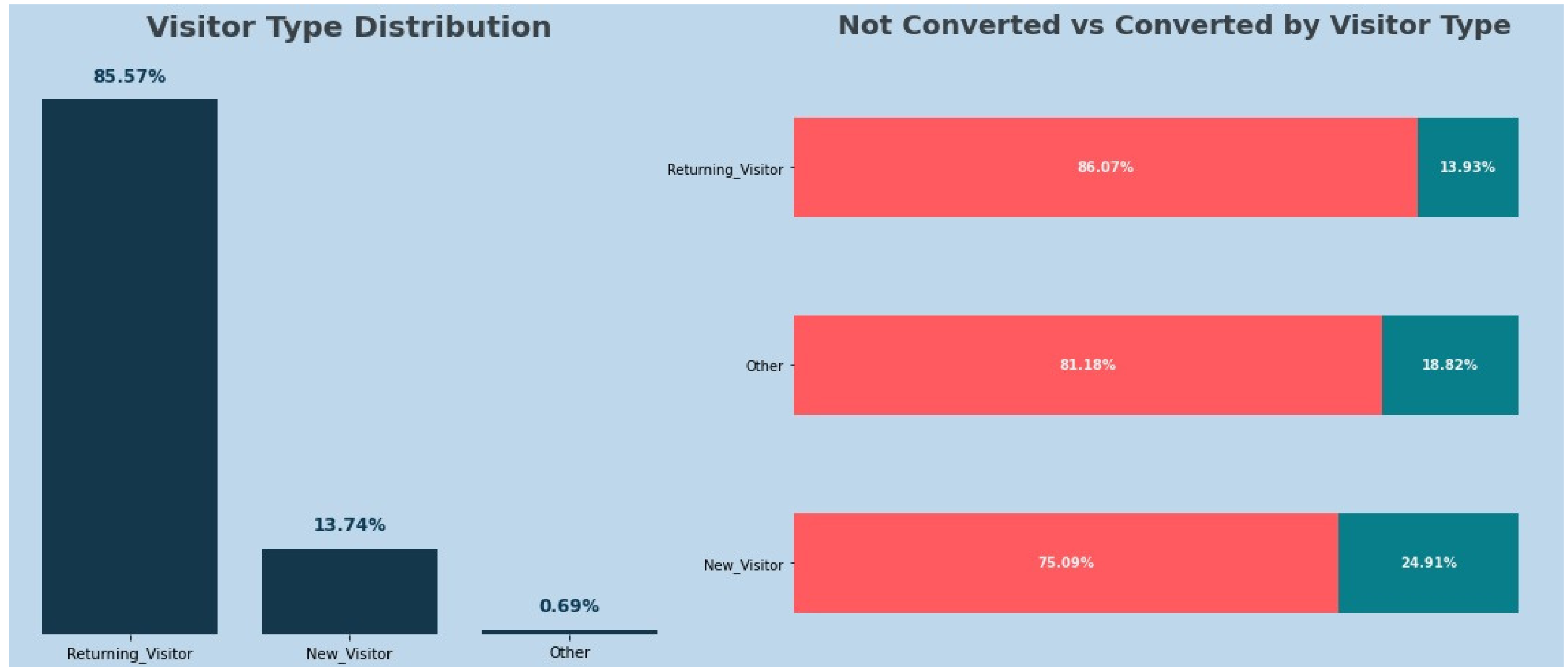
Pertumbuhan penjualan produk di tahun pertama masih fluktuatif atau tidak stabil.

- Pada bulan April, website mengalami server down
- Pada bulan Jun - Oct, jumlah kunjungan dan penjualan relatif rendah dan stagnan.
- Pada bulan Nov merupakan jumlah penjualan yang paling tinggi.

Sebagian besar penjualan produk per bulan masih berada di bawah rata-rata (rata-rata 191 transaksi per bulan).







Persentase returning visitor yang membeli produk sebesar 14%. Lebih rendah jika dibandingkan new visitor dengan persentase sebesar 25%.



# Data Pre Processing

Missing Value	Duplicate	Outlier	Encoding	Transform	Feature Selection	Imbalance
Tidak Ada Missing Value	Drop 125 Duplicate	Tidak Menghilangkan Outlier	<ul style="list-style-type: none"><li>• One Hot Encoding</li><li>• Label Encoder</li></ul>	<ul style="list-style-type: none"><li>• Yeo Johnson</li><li>• Standard Scaler</li></ul>	<ul style="list-style-type: none"><li>• Fitur berkorelasi tinggi</li><li>• Drop fitur kurang relevan</li></ul>	SMOTE 0.5



# Modeling

## XGB Classifier

*default hyperparameters*

```
Accuracy (train): 0.9
Accuracy (test): 0.88
=====
Precision (train): 0.93
Precision (test): 0.95
=====
Recall (train): 0.92
Recall (test): 0.91
=====
F1 (train): 0.92
F1 (test): 0.93
=====
AUC (train): 0.95
AUC (test): 0.91
=====
CV AUC (train): 0.95
CV AUC (test): 0.94
```

Actual \ Predicted	Not Purchase	Purchase
Not Purchase	1900	179
Purchase	102	260

\*Positive = 0



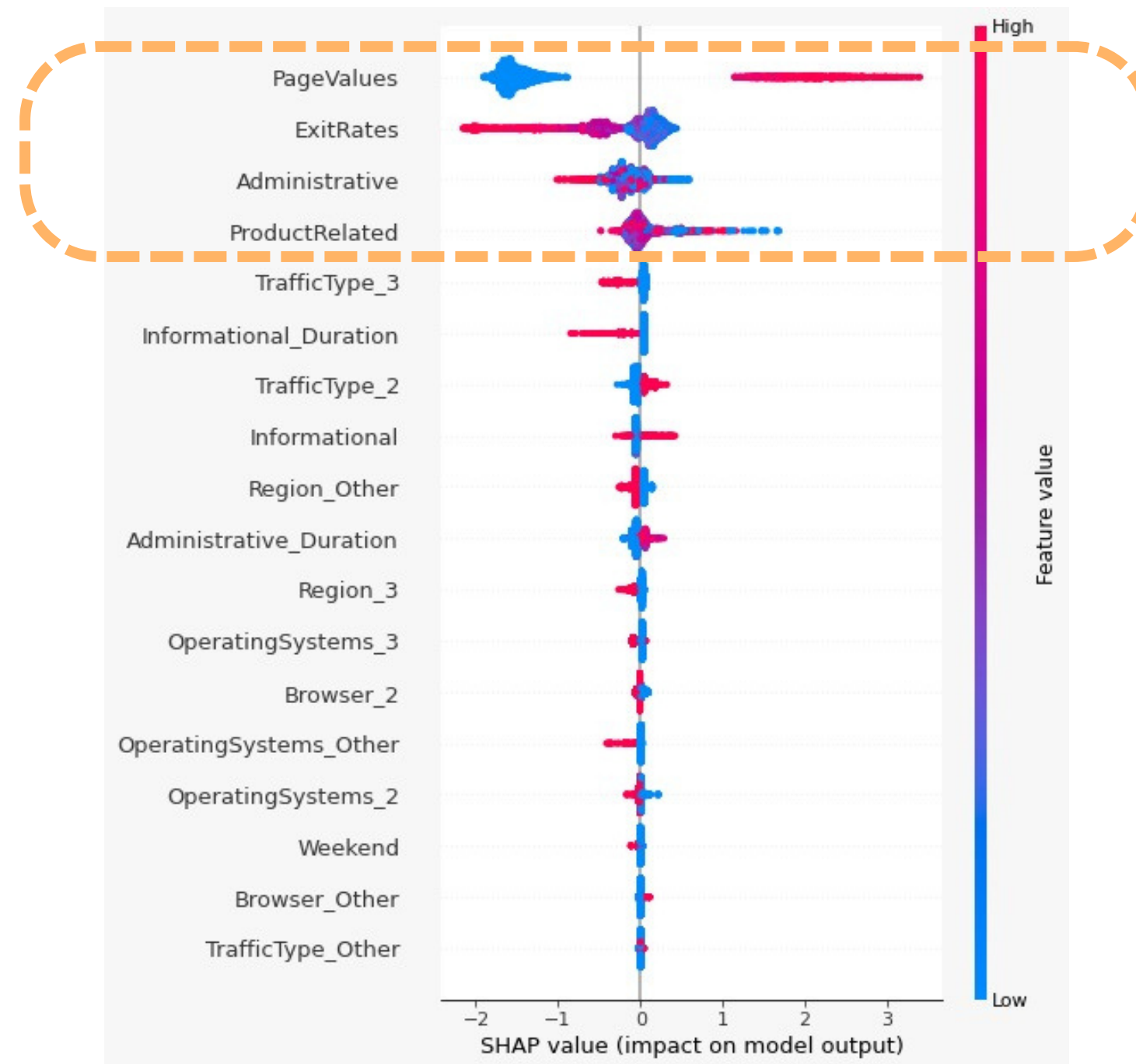
Perusahaan tidak bisa hanya fokus pada visitors yang terdeteksi akan membeli saja.



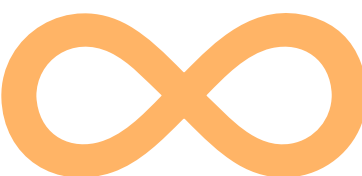
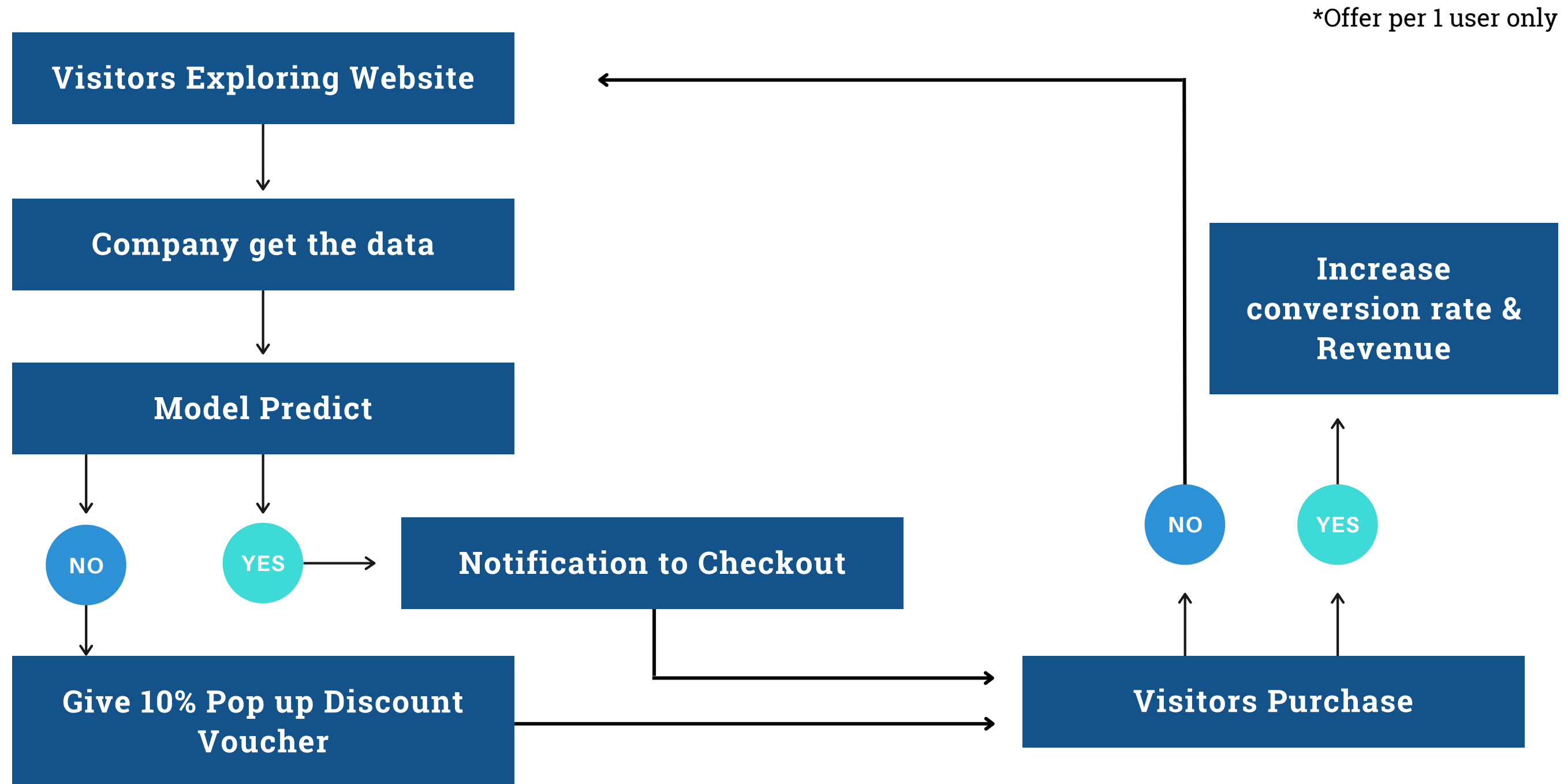
Kita perlu memperhatikan visitors yang terdeteksi tidak membeli agar terkonversi.



# Feature Importance



# Model Flow





# Result & Recommendation



# Model Simulation

Dengan model tersebut 91% dari visitors yang tidak membeli dapat terdeteksi (data test).

Jika kita memberikan campaign berupa voucher discount sebesar 10% kepada visitor yang terdeteksi tidak membeli, dengan asumsi visitor terkonversi sebesar 11%, maka simulasi kenaikan conversion rate adalah sebagai berikut.

Converted Before Model (CBM) = 1908

Correctly Predicted (CP) =  $0.91 * 10422 = 9484$

Pop-up Converted (PC) =  $0.11 * CP = 0.11 * 9484$   
= 1043

Converted After Model (CAM) =  $CBM + PC = 2951$

**Before Model**

**15.47%**

CR Increased 54.66%

**After Model**

**23.93%**



# Model Simulation

Berdasarkan hasil perhitungan banyaknya converted visitors sebelum dan sesudah model, dan dengan asumsi besar Gross Profit Margin (GPM) 42.53% dan rata-rata value per order sebesar \$102/order, maka peningkatan GPM dapat disimulasikan sebagai berikut.

**Before Model:**

$$\text{Revenue} = 1908 * \$102 = \$194,616$$

$$\text{GPM} = 0.4253 * \$194,616 = \$82,770.19$$

**After Model:**

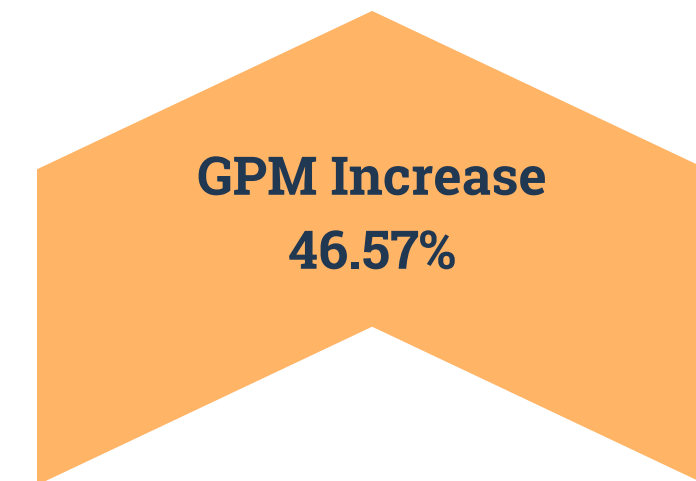
$$\text{PC} = 1043, \text{TN} = 1409, \text{FP} = 499$$

$$\begin{aligned} \text{Revenue} &= \$102 * 1409 + \$102 * 0.9 * (1043+499) \\ &= \$285,273.6 \end{aligned}$$

$$\text{GPM} = 0.4253 * \$285,273.6 = \$121,326.86$$

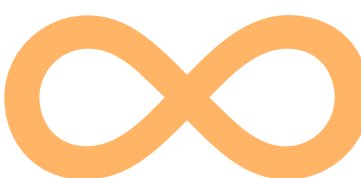
**After Model**

**\$121,326.86**



**Before Model**

**\$82,770.19**





# Business Recommendation

Problem	Root Cause	Action Items
<ul style="list-style-type: none"><li>• Exit Rates</li></ul>	<ul style="list-style-type: none"><li>• Page Load Time</li></ul>	<ul style="list-style-type: none"><li>• Image and page optimization</li></ul>
<ul style="list-style-type: none"><li>• Page Value</li></ul>	<ul style="list-style-type: none"><li>• Too much abandoned cart ( High Page Values but Not Purchase)</li></ul>	<ul style="list-style-type: none"><li>• Push Email Reminder to user</li></ul>
<ul style="list-style-type: none"><li>• Product Related</li></ul>	<ul style="list-style-type: none"><li>• Bad product Recommendation</li></ul>	<ul style="list-style-type: none"><li>• Optimizing system to giving more relevant product</li></ul>
<ul style="list-style-type: none"><li>• Administrative</li></ul>	<ul style="list-style-type: none"><li>• Difficult / Error transaction process</li></ul>	<ul style="list-style-type: none"><li>• Fixing transaction error issue (ex: can't input card) and making transaction process easy.</li></ul>



**TERIMA KASIH**