

PURCHASING INTENTION PREDICTION ON ONLINE SHOPPER

GROUP 8 - MAU BELI??



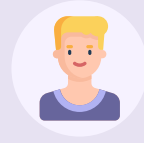
OUR TEAM



ARIF B



DESY N



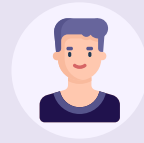
DANANG R



INGGRIANI



MAYA M



M DONNY F

TABLE OF CONTENTS

01 Business Understanding

Memahami peran, masalah, dan goal yang ingin dicapai dalam project

02 Exploratory Data Analysis (EDA)

Memeriksa dan memahami dataset untuk mendapatkan informasi yang dapat digunakan untuk meningkatkan bisnis

03 Data Pre-Processing

Membersihkan data yang kurang lengkap dan memilih feature yang akan digunakan untuk memperkecil error

04 Modeling & Evaluation

Membuat machine learning model yang sesuai dengan tujuan kita

05 Business Recommendation

Memberikan rekomendasi bisnis berdasarkan hasil modeling

01

BUSINESS UNDERSTANDING



ROLE & PROBLEM STATEMENT

ROLE

Tim Data Scientist di PT Beli Dong bertugas untuk mencari **solusi** dalam mengatasi masalah Tim Bisnis melalui dataset URL Information yang tersedia

PROBLEM STATEMENT

E-Commerce PT Beli Dong
terdapat 12.330 visitor



14% New Visitor
86% Returning Visitor



$CvR \text{ New Visitor} > CvR \text{ Returning Visitor}$

GOAL & BUSINESS METRICS

Goal



Meningkatkan Revenue (Conversion Rate) dengan memfollow up Returning visitor yang diprediksi membeli dengan call to action dan pemberian promo

Objective



Membuat pemodelan yang mampu memprediksi user yang berkunjung mempunyai intensi membeli atau tidak

Metrics



Revenue (Conversion Rate)





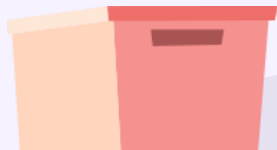
02 EXPLORATORY DATA ANALYSIS (EDA)

ABOUT DATASET

Online Shoppers Intention

18 Features

12330 bars
0 null
125 duplicates



10 Numerical

- **Administrative**
- **Administrative Duration**
- **Informational**
- **Informational Duration**
- **ProductRelated**
- **ProductRelated Duration**
- **BounceRates**
- **ExitRates**
- **PageValues**
- **SpecialDay**

8 Categorical

- **Month**
- **OperatingSystems**
- **Browser**
- **Region**
- **TrafficType**
- **VisitorType**
- **Weekend**
- **Revenue**

ABOUT DATASET

REVENUE

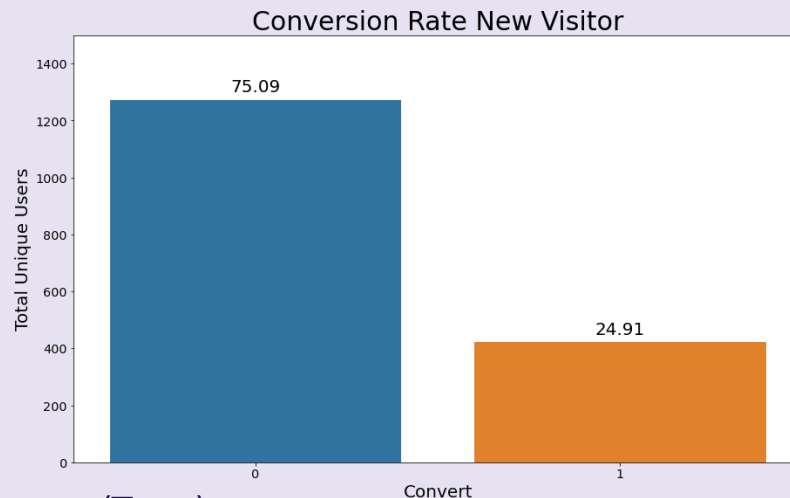
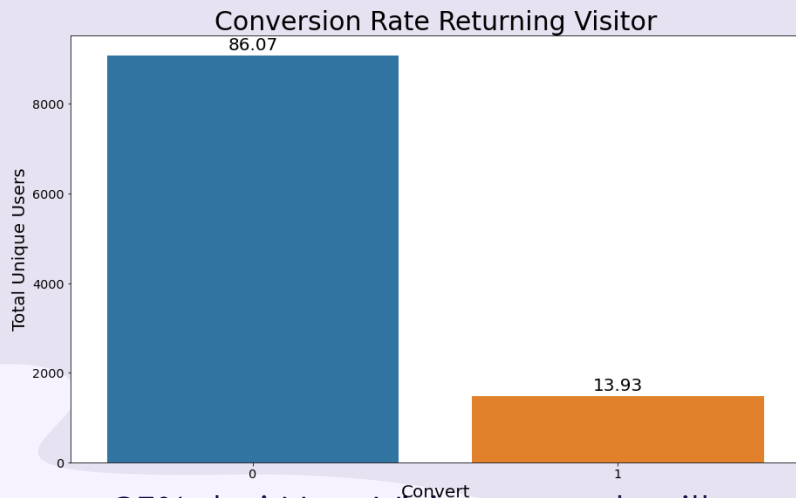


- 86% visitor tidak melakukan transaksi (False)
- Hanya 14% visitor melakukan transaksi
- Perlu diberikan treatment untuk meningkatkan conversion rate



ABOUT DATASET

VISITOR TYPE



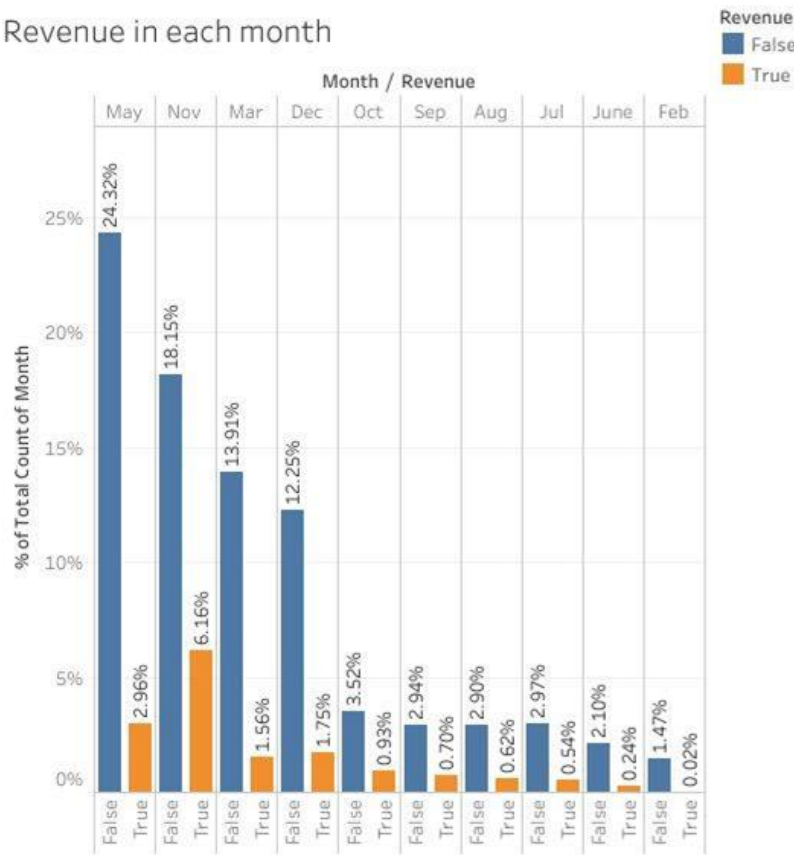
- 25% dari New Visitor menghasilkan revenue (True)
- Conversion Rate Returning visitor hanya 14%
- Conversion Rate Other 19%
- Nilai Other perlu ditreatment di tahap Preprocessing

ABOUT DATASET

MONTH

- Top 3 pengunjung terbanyak ada pada bulan Maret, Mei dan November
- Top 3 conversion rate tertinggi ada pada bulan November, Mei dan Desember

Revenue in each month



ABOUT DATASET

DATA STATISTICAL

1. Rata-rata durasi kunjungan website adalah 30 menit, dimana 70% pada halaman Produk
2. User secara rata-rata mengalami 25 page baru melakukan exit
3. 86% pengunjung merupakan ReturningVisitor, yang mana seharusnya memiliki niat beli lebih karena sengaja melakukan kunjungan kembali.
4. Hanya ada 0,2% user yang tidak menyelesaikan scrolling paganya

Melihat dari data ini, dapat disimpulkan bahwa **engagement pada web sudah baik, namun mengapa user tidak convert?**

ABOUT DATASET

EXTERNAL INSIGHT

anderson-review.ucla.edu

- Hypothesa : User mengalami *search fatigue*
- Sebuah penelitian menemukan bahwa lebih dari 40% nya memilih untuk 'istirahat' sebelum kembali berbelanja **ini yang menyebabkan jumlah returning visitor lebih banyak dari new visitor**

High Level Approach: How Might We Improve This?

- 1. Berfokus pada segmen Returning Visitor, karena CvR nya jauh dibawah New Visitor**
- 2. Mencari faktor yang paling berpengaruh terhadap pembelian konsumen**
- 3. Membuat model machine learning untuk memprediksi niat beli user**
4. Memberikan treatment khusus pada user yang tidak berniat membeli dengan melihat history dalam appsnya lebih lanjut
5. Treatment bisa dicluster berdasarkan probabilitas pembelian,
 - a. User dengan minat tinggi beli (90%) -> Treatment A
 - b. User dengan minat beli sedang (50%) -> Treatment B
 - c. User dengan minat beli rendah (<50%) -> Treatment C
6. Mempercepat time to checkout customer agar tidak mengalami search fatigue sebelum pembelian

03

DATA PREPROCESSING



DATA CLEANSING



HANDLING DUPLICATED DATA

125 duplicated data namun tidak didrop karena diasumsikan setiap baris pada dataset ini adalah sesi kunjungan user yang kebetulan memiliki pola sama



HANDLING VISITOR TYPE

- Drop semua baris nilai New_Visitor
- Replace nilai Other menjadi Returning_Visitor
- Jumlah baris setelah di drop 10636 baris

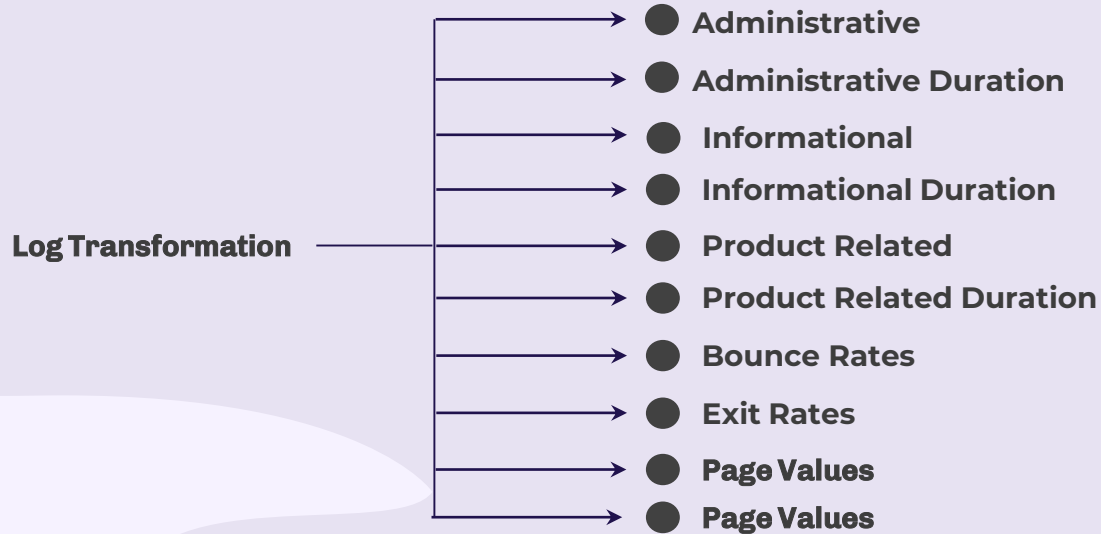


HANDLING OUTLIER

Menggunakan z-score dengan jumlah baris data setelah memfilter outlier menjadi 9883

DATA CLEANSING

FEATURE TRANSFORMATION



DATA CLEANSING

FEATURE ENCODING

Label Encoding



One Hot Encoder

- **Month**
- **Operating Systems**
- **Browser**
- **Region**
- **Traffic Type**
- **Visitor Type**

FEATURE SELECTION

8 Features

- log_Administrative
- log_Informational
- log_ProductRelated
- log_ExitRates
- log_PageValues
- agg_Month_Nov
- agg_OperatingSystems_2
- agg_TrafficType_2

Target

- **Revenue**

SPLIT DATA & CLASS IMBALANCE

SPLIT DATA
Train 70% (**6918**) : Test 30% (**2965**)



HANDLE CLASS IMBALANCE
SMOTE data train (**1 : 1**)





04

MODELING & EVALUATION

MODELING



5 Model

- Logistic Regression
- Decision Tree
- Random Forest
- Adaboost
- XGBoost

MODELING

Model	Accuracy (%)	Precision (%)	Recall (%)	AUC(%)	F1(%)
Logistic Regression	86	51	80	83	62
Decision Tree	84	47	62	74	53
Random Forest	86	53	70	79	60
Adaboost	87	53	79	83	64
XGB	87	56	63	77	59

TUNING HYPERPARAMETERS

Solver	Newton-cg, lbfgs, liblinear
Penalty	L2, l1, elasticnet, none
C	100, 10, 1.0, 0.1, 0.01, 0.001, 0.0001

Inisiasi model

```
logres = LogisticRegression(random_state=42) # Init Logres dengan Gridsearch,  
cross validation = 5
```

```
lr_tuned = RandomizedSearchCV(logres, hyperparameters, cv=5, random_state=42,  
scoring='recall')
```


BEST MODEL

AFTER TUNING HYPERPARAMETER

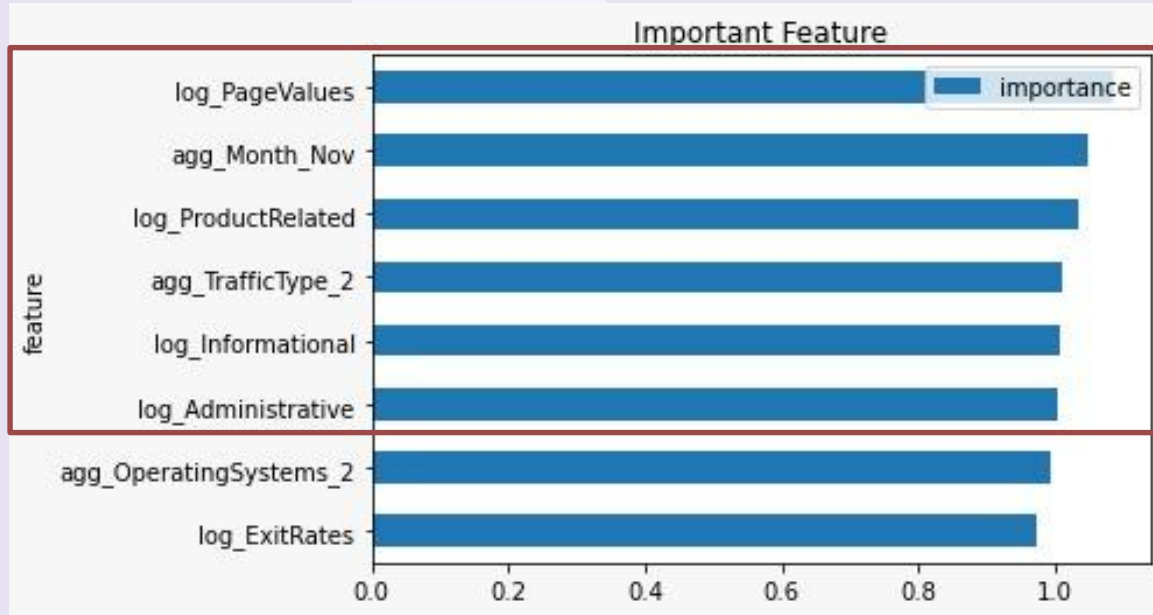
Hasil tuning
model LOGISTIC
REGRESSION.

Metrics	Before Tuning (%)	After Tuning (%)
Accuracy	86	86
Precision	51	51
Recall	80	80
AUC	83	83
F1	62	62

Best Params

Random State = 42, C = 100, Solver = newton-cg

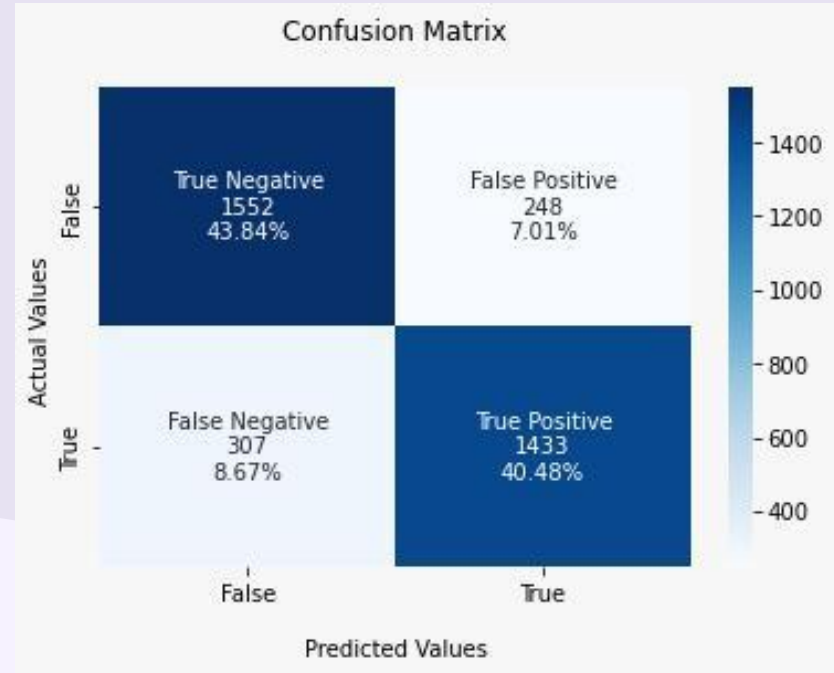
FEATURE IMPORTANCE



- **log_PageValues** merupakan feature tertinggi dalam model
- 7 feature lainnya memiliki korelasi yang cukup tinggi

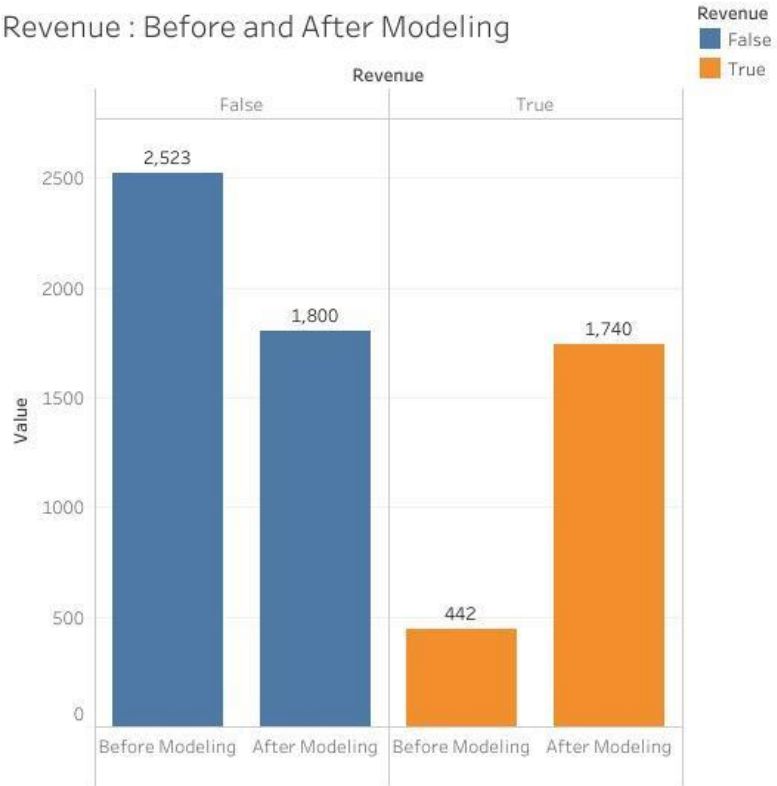
FEATURE SELECTION

Metrics	Tuned Logres Before Iteration	Tuned Logres After Iteration
Accuracy	86	84
Precision	51	85
Recall	80	82



RESULT MODELLING

Revenue : Before and After Modeling



- jumlah visitor yang convert mengalami kenaikan sebesar 1298 (75%)

Asumsi **profit** e-Commerce :
Rp. 500 M/bulan

	Profit/bulan	CvR Returning Visitor
Before	Rp.500 M	15%
After	Rp.875 M	26%



05

BUSINESS RECOMENDATION

EXTERNAL INSIGHT

- Berdasarkan anderson-review.ucla.edu banyak nya pilihan pencarian pada suatu produk dapat menyebabkan *search fatigue* pada online shopper yang mengunjungi website.
- Sebuah penelitian menemukan bahwa lebih dari 40% nya memilih untuk 'istirahat' sebelum kembali berbelanja ini yang menyebabkan jumlah returning visitor lebih banyak dari new visitor.

BUSINESS RECOMENDATION

1. Memberikan **treatment khusus pada user yang tidak berniat membeli** dengan **melihat history** dalam appsnya lebih lanjut
2. Treatment bisa **dicluster berdasarkan probabilitas pembelian**,
 - a. User dengan minat tinggi beli (90%) -> Treatment A
 - b. User dengan minat beli sedang (50%) -> Treatment B
 - c. User dengan minat beli rendah (<50%) -> Treatment C
3. **Mempercepat time to checkout** customer agar tidak mengalami search fatigue sebelum pembelian

BUSINESS RECOMENDATION

- Berdasarkan karakteristik hasil pemodelan, perusahaan perlu melakukan **treatment** lebih terhadap visitor yang yang **diprediksi tidak membeli** seperti diberi promo gratis ongkir/potongan harga sebelum mengalami *search fatigue* dan memutuskan untuk istirahat dengan harapan segera melakukan checkout.
- Perusahaan dapat **melakukan prediksi probabilitas membeli dan mengelompokkan user** sesuai niat membelinya (Agak Tinggi, Sedang, Rendah) lalu memberikan A/B test untuk memaksimalkan niat membelinya
- Jika **asumsi** profit e-Commerce Rp. 500M/bulan, model dapat **meningkatkan** Conversion Rate dari **15%** menjadi **26%** atau setara Rp. 875M/bulan

ADDED INSIGHT

Untuk membuat A/B Test, kita perlu mempelajari:

- Bagaimana konten page dengan PageValue yang tinggi
- Traceback metode **marketing pada bulan November dan May**, apa yang membuat serbuan pengunjung dan konversi.
- Mengetahui **traffic pembeli yang kurang produktif**, yaitu selain dari Operating System 2 dan TrafficType 2.
- Mengecek **payment method pada page administrative**, serta information targeting yang tepat pada information page (sesuai minat customer).
- **Mengetahui history produk** yang dilihat oleh user (Product Type, Price Avg)

EXECUTION GUIDELINE

- Perusahaan bisa melakukan A/B testing kepada segmen customer yang **“True Negative” namun melakukan kunjungan ulang pada website**
- A/B testing merupakan outcomes dari formulasi fitur-fitur penting dalam melakukan pembelian, misal :
 - A. Memberikan stimulus page, serupa dengan page dengan pageValue yang tinggi
 - B. Memberikan kampanye serupa yang dilakukan pada bulan November/May
 - C. **Stimulus diberikan pada page 24**, mempertimbangkan exit rate di halaman 25



**TERIMA
KASIH**