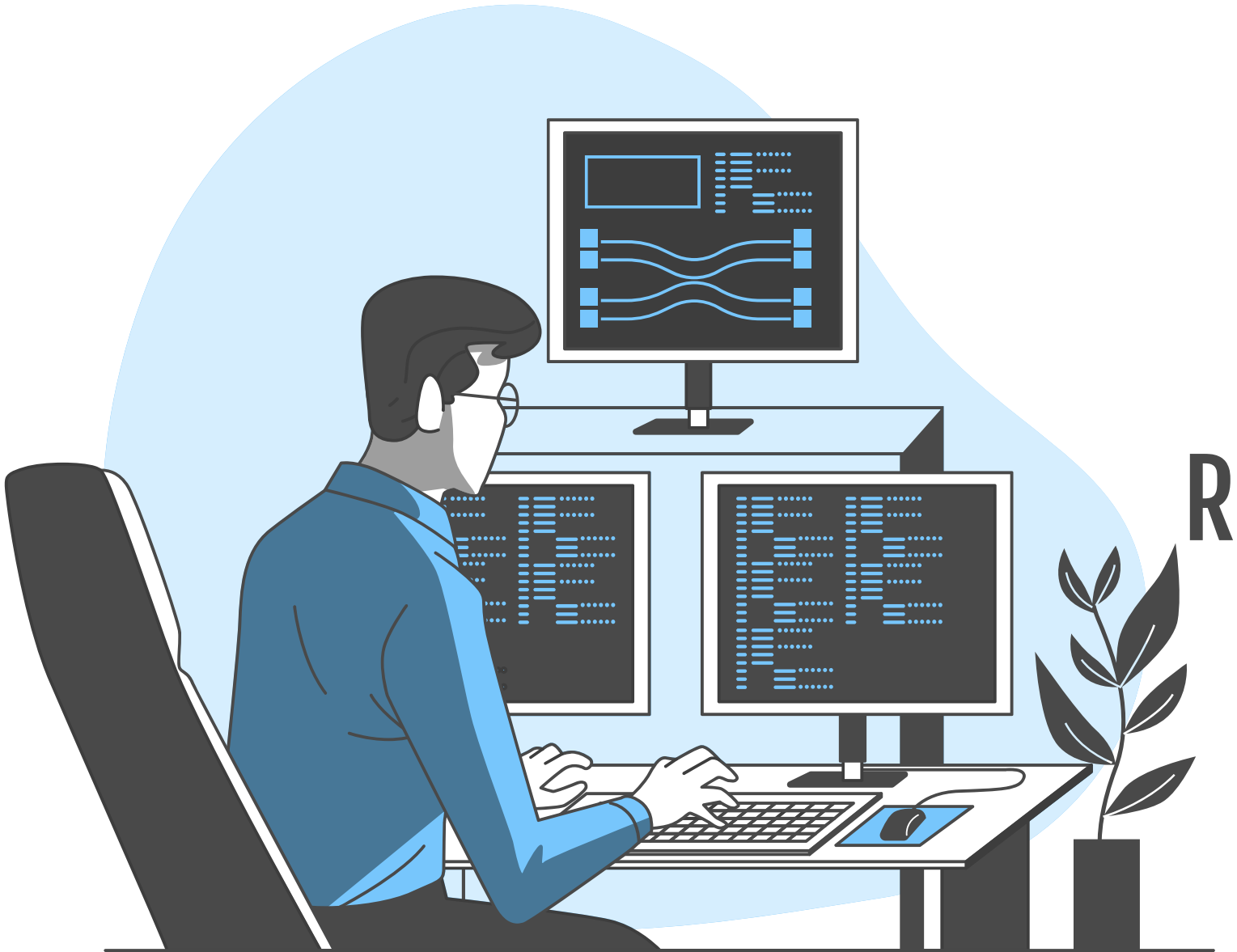
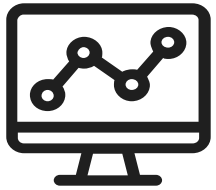


IN-VEHICLE COUPON RECOMENDATION

DEEP LEARNING 4.0 CONSULTANT

KAMI DAN KLIEN KAMI



DEEP LEARNING 4.0 CONSULTANT

Kami merupakan tim analis dari Deep Learning 4.0 Consultant. Perusahaan jasa professional di bidang konsultasi bisnis

NUDLE TECH COMPANY

Sebuah perusahaan teknologi multinasional di bidang jasa dan produk internet. Salah satu produk unggulannya yakni layanan pemetaan berbasis web dan perangkat lunak yang disebut Nmaps (Nudle Maps)

TIM ANALIS KAMI

DIKA KENCANA P.

Dika Kencana Putera | LinkedIn

HUMAIDI

https://www.linkedin.com/in/humaidi-saane-a24859168/

NORISA KURNIANI

https://www.linkedin.com/in/norisa-kurniani-33ab30231

RISA DWI R.

https://www.linkedin.com/in/risa-dwi-ratnasari-a20991a3

VIAS AULIA

Vias Aulia | LinkedIn

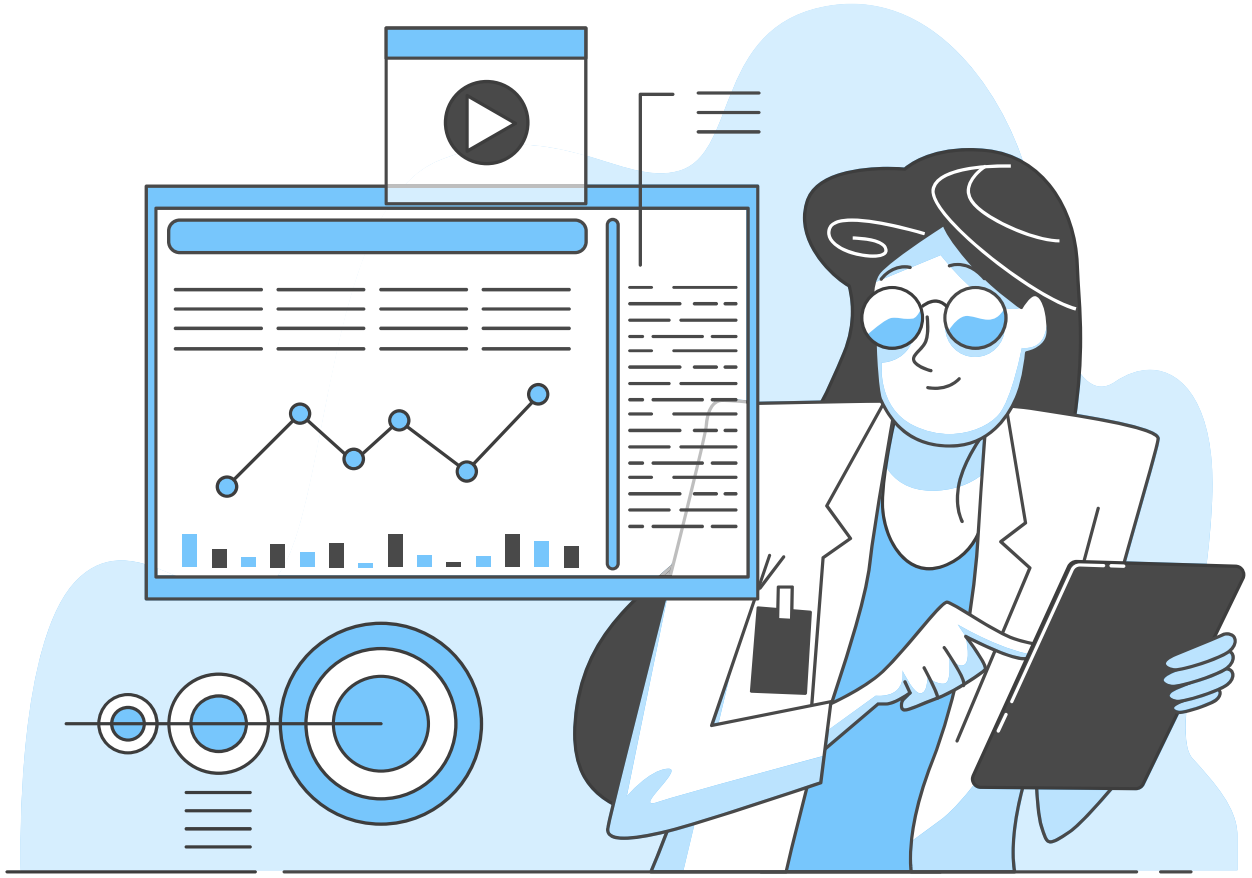
YOVAN PRADIPTA

https://www.linkedin.com/in/yovan-pradipta-121465118/

ZULIUS AKBAR A.

https://www.linkedin.com/in/zulius-akbar-amin-35469259/

DAFTAR ISI



01
LATAR BELAKANG

LATAR BELAKANG

Nudle sebagai perusahaan pihak ketiga dipercayai oleh beberapa perusahaan makanan untuk mendistribusikan kupon kepada pengemudi /pelanggan melalui aplikasi NMaps.

Berdasarkan kasus yang diberikan, NMaps hanya dapat mendistribuskan kupon kepada pelanggan dengan

COUPON ACCEPTANCE

56.84%

COUPON ACCEPTANCE

Jumlah penerima kupon dari total keseluruhan pelanggan yang diberikan kupon. Dalam kasus ini, penerima kupon pasti menggunakan kupon.

$$\text{COUPON ACCEPTANCE} = \frac{\text{Pelanggan yang menerima kupon}}{\text{Total pelanggan yang diberikan kupon}}$$

Pemberian kupon merupakan salah satu cara untuk meningkatkan revenue dan mendapatkan pelanggan baru dalam sebuah bisnis. Berdasarkan pemberian kupon, 69% mereka akan merasa menjadi pelanggan yang pintar, 60% pelanggan akan mencoba produk baru, 54% melakukan pembelian impulsif.

Sumber: Valassis Research: Coupons and Discounts Appeal Far Beyond Savings - Valassis

LATAR BELAKANG

PROBLEM STATEMENT

Rendahnya penerima kupon yakni 56.84%

GOAL

Meningkatkan jumlah penerima kupon terhadap pelanggan

OBJECTIVE

Membuat model machine learning untuk memprediksi kelas penerima kupon serta menghasilkan insight berupa business recommendation untuk meningkatkan penerima kupon

BUSINESS METRIC

Coupon Acceptance Rate
Peningkatan Benefit-Cost Ratio

02

DATASET

DATASET

destination
passanger
weather
temperature
time
coupon
expiration
gender
age
maritalStatus
has_children
education
occupation
income
car
Bar
CoffeeHouse
CarryAway
RestaurantLessThan20
Restaurant20To50
toCoupon_GEQ5min
toCoupon_GEQ15min
toCoupon_GEQ25min
direction_same
direction_opp
Y

TENTANG DATASET

Dataset ini berisi fitur yang menunjukkan karakteristik pelanggan yang diberikan kupon

SHAPE

12.684 baris data, 26 kolom (25 fitur dan 1 target 'Y')

MISSING VALUE

12.576 baris data di fitur 'car',
107 baris data di fitur 'Bar',
217 baris data di fitur 'CoffeeHouse',
151 baris data di fitur 'CarryAway',
130 baris data di fitur 'RestaurantLessThan20',
189 baris data di fitur 'Restaurant20To50'

DUPLICATED DATA

74 baris data

03

EXPLORATORY DATA ANALYSIS (EDA)

EXPLORATORY DATA ANALYSIS

DTYPE

8 Numeric, 18 Object

26 Object

Semua valuenya merupakan kategorikal

CONSTANT VALUE

- toCoupoun_GEQ5min

CORRELATION

4 features highly correlated

- ‘direction_same’ and ‘direction_opp’
- ‘time’ and ‘destination’

TOO MANY UNIQUE VALUE

- ‘occupation’

04

DATA PREPROCESSING

DATA PREPROCESSING

HANDLE MISSING VALUE

Fill with Mode (5 features)
Drop (1 feature)

- *'car'*

FEATURE TRANSFORMATION

- 'age'

Transform it from 8 unique values to 5 unique values

SPLIT DATA

70 : 30

FEATURE EXTRACTION

- 'passanger_destination'

From 'passanger' and 'destination'

- 'marital_hasChildren'

From 'maritalStatus' and 'has_children'

- 'temperature_weather'

From 'temperature' and 'weather'

FEATURE ENCODING

Label Encoding + 3 methods:

- One Hot Encoding
- Hash Encoding
- Binary Encoding

DATA PREPROCESSING

DATASET TREATMENT

Terdapat 7 jenis dataset dengan perlakuan preprocessing yang berbeda - beda

DATASET	FEATURE ENGINEERING	ENCODING METHOD			HANDLE CLASS IMBALANCE	Feature Selection	USER
	Extraction	One Hot	Hash	Binary		Manual	
1		✓				✓	Vias
2	✓	✓				✓	Norisa
3			✓			✓	Risa
4	✓		✓			✓	Zul
5	✓		✓		✓	✓	Dika
6				✓		✓	Yovan
7	✓			✓		✓	Humaidi

05

MODELING AND EVALUATION

MODELING AND EVALUATION

MODEL

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

EVALUATION METRICS

- Accuracy

Karena dataset kami memiliki fitur target dengan kelas yang balanced. Hasil prediksi model menghitung secara keseluruhan yang menerima kupon dan tidak menerima kupon.

MODELING AND EVALUATION

BEFORE TUNING HYPERPARAMETER						
DATASET	MODEL	EVALUATION METRICS				
		Accuracy	Precision	Recall	ROC-AUC Train	ROC-AUC Test
1	LogisticRegression					
	DecisionTree	0.6	0.66	0.61	0.99	0.6
	RandomForest	0.64	0.68	0.71	0.99	0.67
	XGBoost	0.61	0.63	0.79	0.66	0.64
	CatBoost					
2	LogisticRegression	0.67	0.69	0.77	0.73	0.73
	DecisionTree	0.67	0.71	0.71	0.67	0.99
	RandomForest	0.62	0.61	0.94	0.73	0.74
	XGBoost	0.73	0.73	0.81	0.79	0.83
	CatBoost	0.74	0.74	0.83	0.82	0.93
3	LogisticRegression	0,66	0,67	0,76	0,72	0,71
	DecisionTree	0,66	0,7	0,71	0,66	0,99
	RandomForest	0,74	0,75	0,8	0,8	0,99
	XGBoost	0,74	0,75	0,8	0,82	0,96
	CatBoost	0,66	0,7	0,71	0,66	0,99
4	LogisticRegression	0.67	0.69	0.77	0.72	0.72
	DecisionTree	0.67	0.71	0.71	0.66	1.00
	RandomForest	0.76	0.76	0.83	0.82	1.00
	XGBoost	0.75	0.76	0.82	0.82	0.82
	CatBoost	0.75	0.75	0.84	0.83	0.93

BEFORE TUNING HYPERPARAMETER						
DATASET	MODEL	EVALUATION METRICS				
		Accuracy	Precision	Recall	ROC-AUC Train	ROC-AUC Test
5	LogisticRegression	0.67	0.69	0.75	0.78	0.72
	DecisionTree	0.66	0.70	0.70	1.00	0.65
	RandomForest	0.74	0.76	0.78	1.00	0.81
	XGBoost	0.74	0.76	0.79	0.97	0.81
	CatBoost	0.74	0.76	0.80	0.94	0.82
6	LogisticRegression	0.68	0.69	0.78	0.74	0.73
	DecisionTree	0.70	0.72	0.77	0.78	0.75
	RandomForest	0.64	0.63	0.90	0.72	0.72
	XGBoost	0.66	0.65	0.88	0.74	0.74
	CatBoost	0.75	0.75	0.83	0.95	0.83
7	LogisticRegression	0.67	0.68	0.76	0.72	0.72
	DecisionTree	0.69	0.68	0.85	0.74	0.77
	RandomForest	0.65	0.63	0.90	0.74	0.73
	XGBoost	0.64	0.62	0.92	0.75	0.74
	CatBoost	0.75	0.75	0.82	0.82	0.93

Berdasarkan modeling sebelum tuning hyperparameters dataset 4 dengan model Random Forest menghasilkan performa model terbaik. Keakurasian dimiliki sebesar 76%

MODELING AND EVALUATION

TUNING HYPERPARAMETERS

- Dilakukan terhadap seluruh dataset dan model
- Metode yang digunakan ialah RandomizedSearchSV dan GridSearchSV

MODELING AND EVALUATION

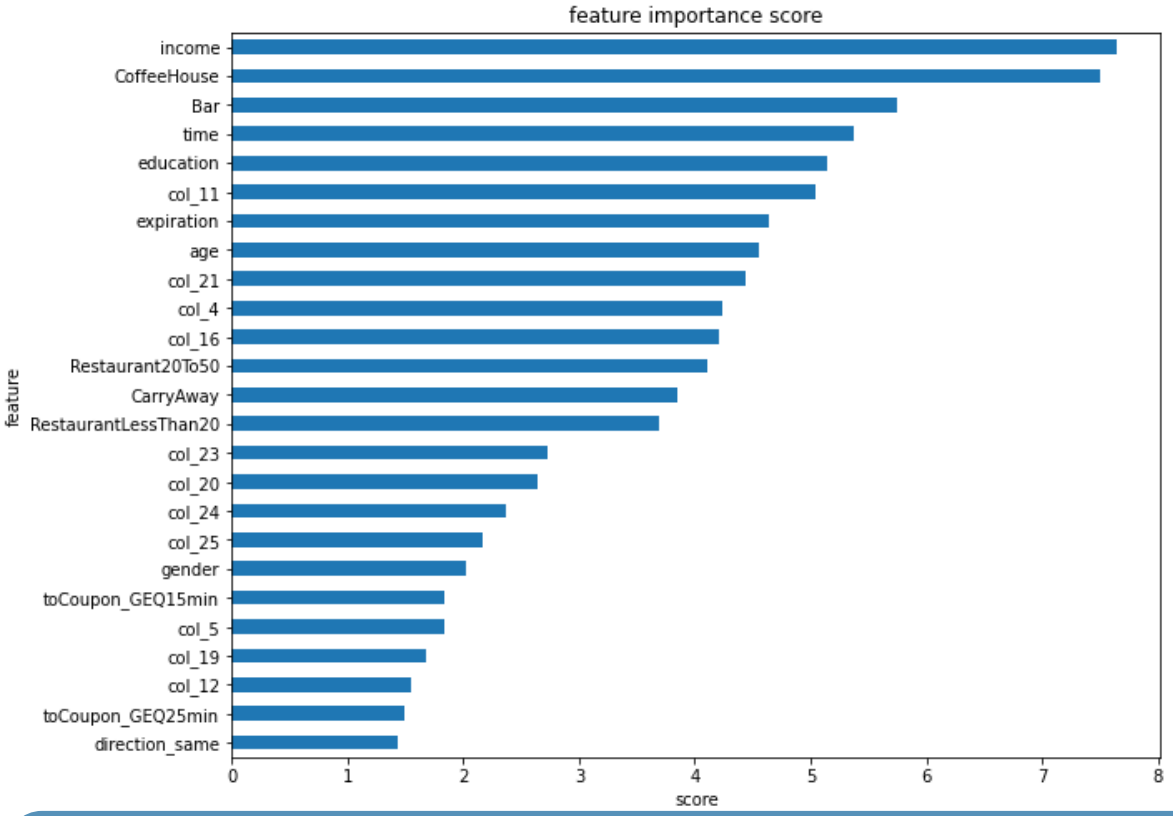
AFTER TUNING HYPERPARAMETER						
DATASET	MODEL	EVALUATION METRICS				
		Accuracy	Precision	Recall	ROC-AUC Train	ROC-AUC Test
1	LogisticRegression					
	DecisionTree	0,63	0,64	0,82	0,68	0,65
	RandomForest	0,62	0,62	0,87	0,68	0,66
	XGBoost	0,62	0,61	0,89	0,7	0,66
	CatBoost					
2	LogisticRegression	0,56	0,56	1,0	0,64	0,64
	DecisionTree	0,64	0,68	0,67	0,67	0,92
	RandomForest	0,72	0,71	0,85	0,79	0,91
	XGBoost	0,75	0,74	0,84	0,82	0,97
	CatBoost					
3	LogisticRegression	0.56	0.56	1.0	0.56	0.56
	DecisionTree	0.56	0.56	1.0	0.56	0.57
	RandomForest	0.73	0.73	0.83	0.96	0.81
	XGBoost	0.71	0.71	0.83	0.86	0.79
	CatBoost	0.73	0.73	0.84	0.84	0.79
4	LogisticRegression	0.68	0.69	0.79	0.72	0.72
	DecisionTree	0.68	0.70	0.77	0.73	0.73
	RandomForest	0.74	0.73	0.86	0.81	0.94
	XGBoost	0.73	0.73	0.85	0.80	0.87
	CatBoost	0.76	0.77	0.83	0.83	0.97

AFTER TUNING HYPERPARAMETER						
DATASET	MODEL	EVALUATION METRICS				
		Accuracy	Precision	Recall	ROC-AUC Train	ROC-AUC Test
5	LogisticRegression	0.67	0.71	0.70	0.75	0.72
	DecisionTree	0.44	0	0	0.5	0.5
	RandomForest	0.74	0.76	0.79	1.00	0.81
	XGBoost	0.74	0.76	0.80	1.00	0.82
	CatBoost	0.73	0.75	0.78	0.90	0.81
6	LogisticRegression	0.68	0.67	0.69	0.72	0.70
	DecisionTree	0.71	0.72	0.70	0.80	0.75
	RandomForest	0.70	0.75	0.76	0.85	0.76
	XGBoost	0.68	0.77	0.76	0.82	0.72
	CatBoost	0.62	0.63	0.67	0.80	0.74
7	LogisticRegression	0.66	0.67	0.76	0.72	0.72
	DecisionTree	0.66	0.70	0.71	0.99	0.66
	RandomForest	0.73	0.73	0.83	0.95	0.80
	XGBoost	0.56	0.56	0.99	0.72	0.72
	CatBoost	0.74	0.75	0.82	0.89	0.81

Berdasarkan modeling setelah tuning hyperparameters dataset 4 dengan model CatBoostmenghasilkan performa model terbaik. Keakurasian dimiliki sebesar 76%

MODELING AND EVALUATION

FEATURE SELECTION



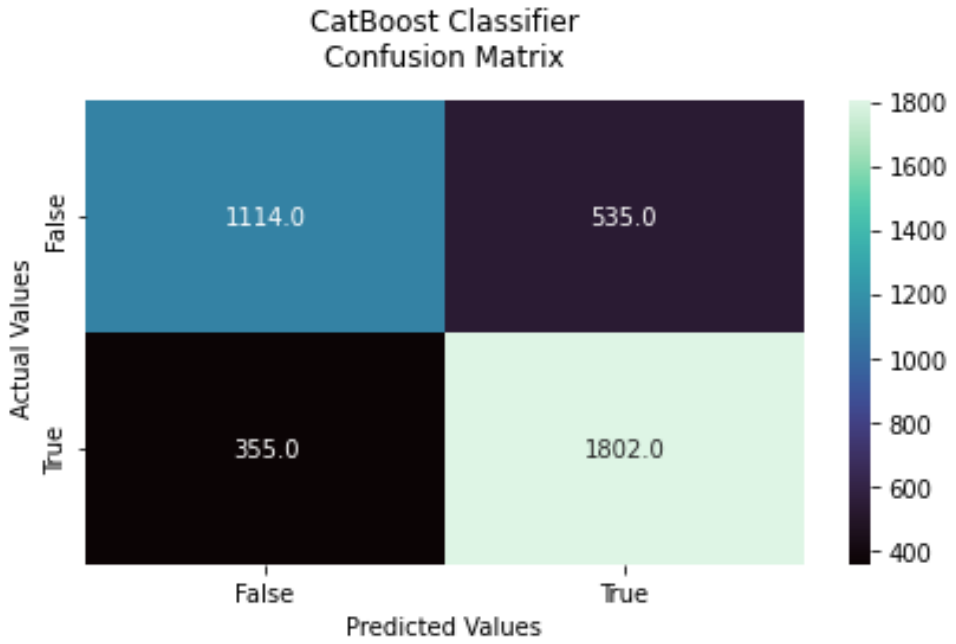
AFTER TUNING HYPERPARAMETER – BEFORE FEATURE IMPORTANCES						
DATASET	MODEL	EVALUATION METRICS				
		Accuracy	Precision	Recall	ROC-AUC Train	ROC-AUC Test
4	CatBoost	0.76	0.77	0.83	0.83	0.97

AFTER TUNING HYPERPARAMETER – AFTER FEATURE IMPORTANCES						
DATASET	MODEL	EVALUATION METRICS				
		Accuracy	Precision	Recall	ROC-AUC Train	ROC-AUC Test
4	CatBoost	0.77	0.77	0.84	0.83	0.97

Melakukan feature selection/drop 2 fitur terbawah berdasarkan hasil feature importance score. Dengan melakukan ini dapat terlihat keakurasian model meningkat menjadi 77%.

MODELING AND EVALUATION

CONFUSION MATRIX



ACCURACY = 77%
PRECISION = 77%
ROC – AUC = 97%

MODELING AND EVALUATION

SHAP FORCE PLOT

Fig a. Pelanggan Tidak Menerima Kupon (< 0.3559)

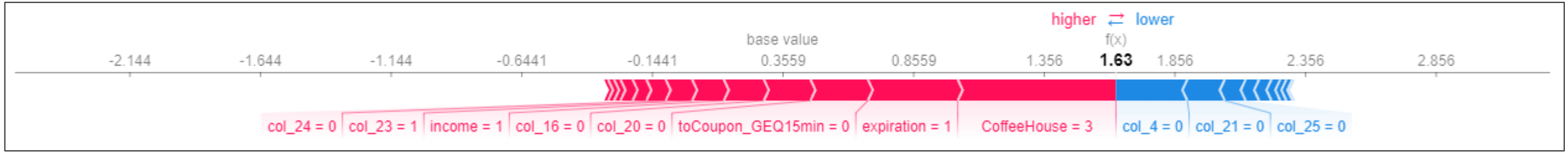


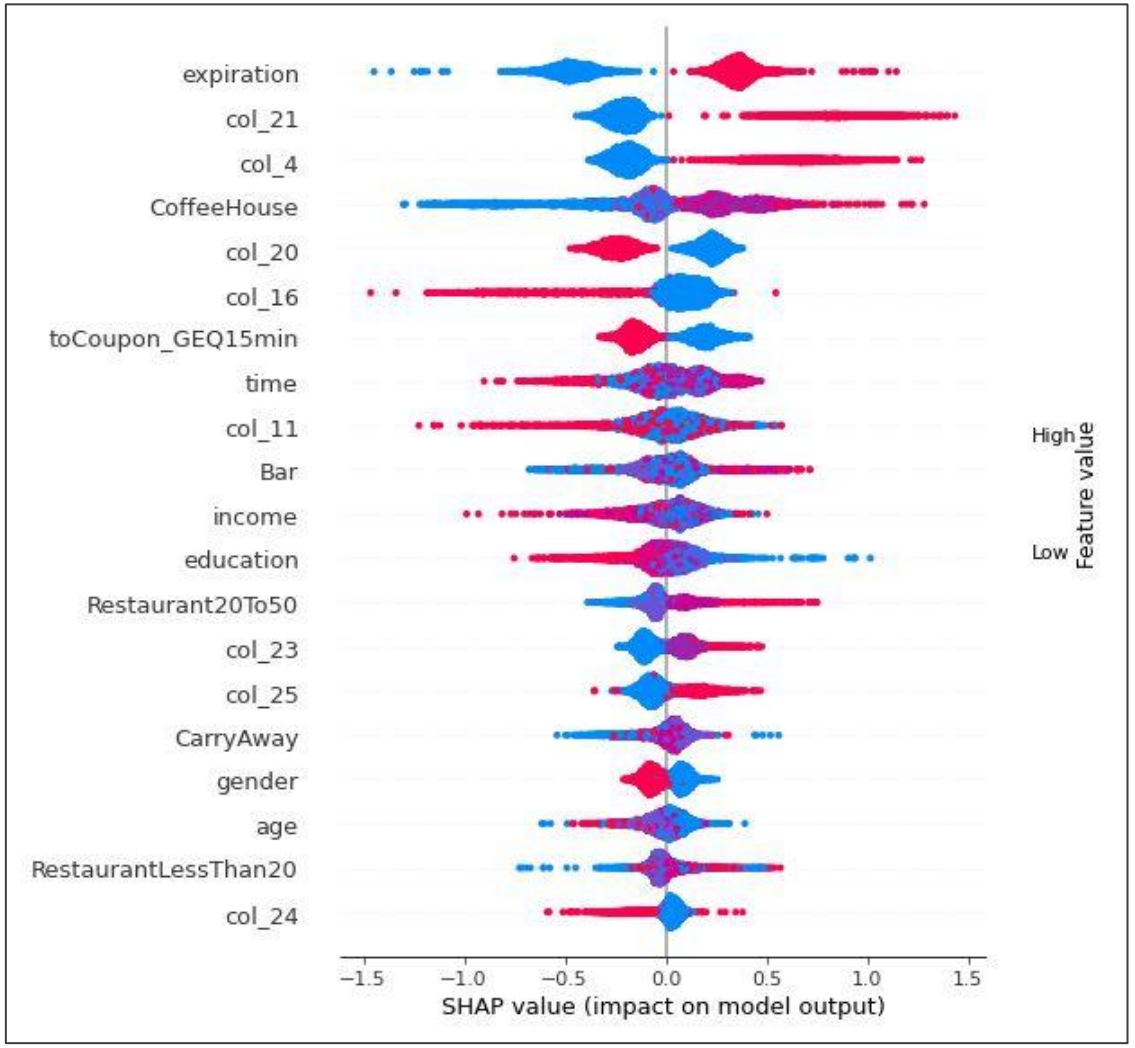
Fig b. pelanggan menerima kupon (> 0.3559)

Berdasarkan interpretasi shap force plot, dapat dilihat jika 'expiration' merupakan salah satu fitur yang Memiliki impact besar terhadap output model dalam memprediksi penerima kupon

Besarnya impact fitur ini mendeterminasikan output model apakah pelanggan Menerima/tidak menerima kupon (fig a. dan fig b.)

MODELING AND EVALUATION

SHAP VALUE



Salah satu fitur yang memiliki impact/influence terbesar terhadap output model yakni fitur ‘expiration’.

‘expiration’ akan menjadi prioritas utama menjadi fitur yang akan ditelaah lebih jauh untuk business insight and business recommendation.

06

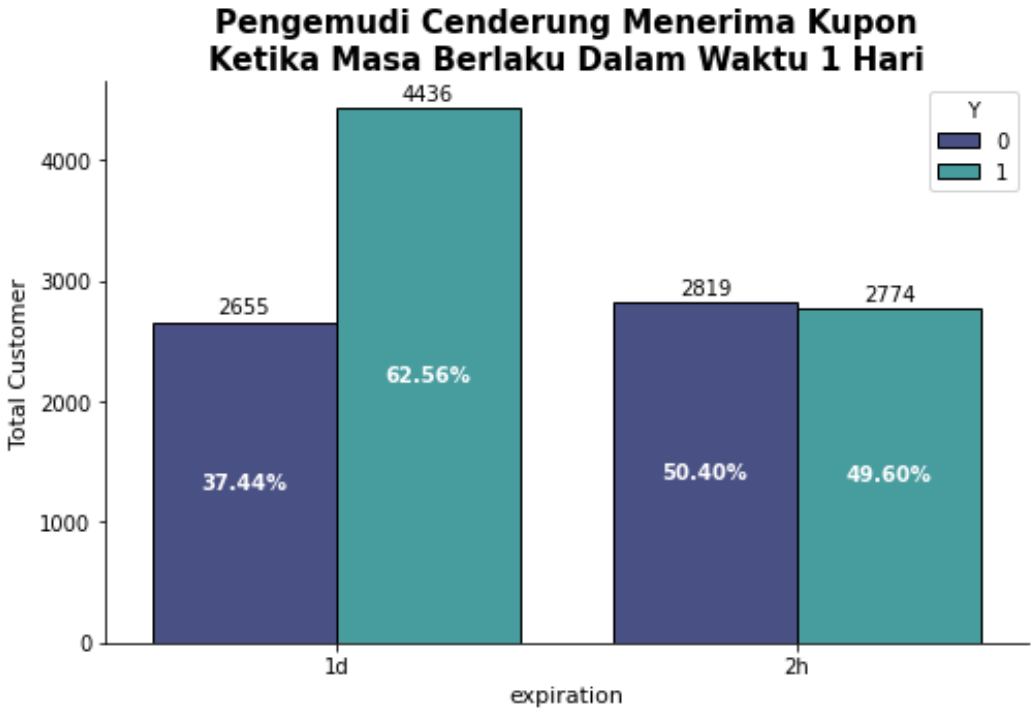
BUSINESS INSIGHT AND RECOMMENDATION

BUSINESS INSIGHT

expiration

1 DAY

62.56% Accepted



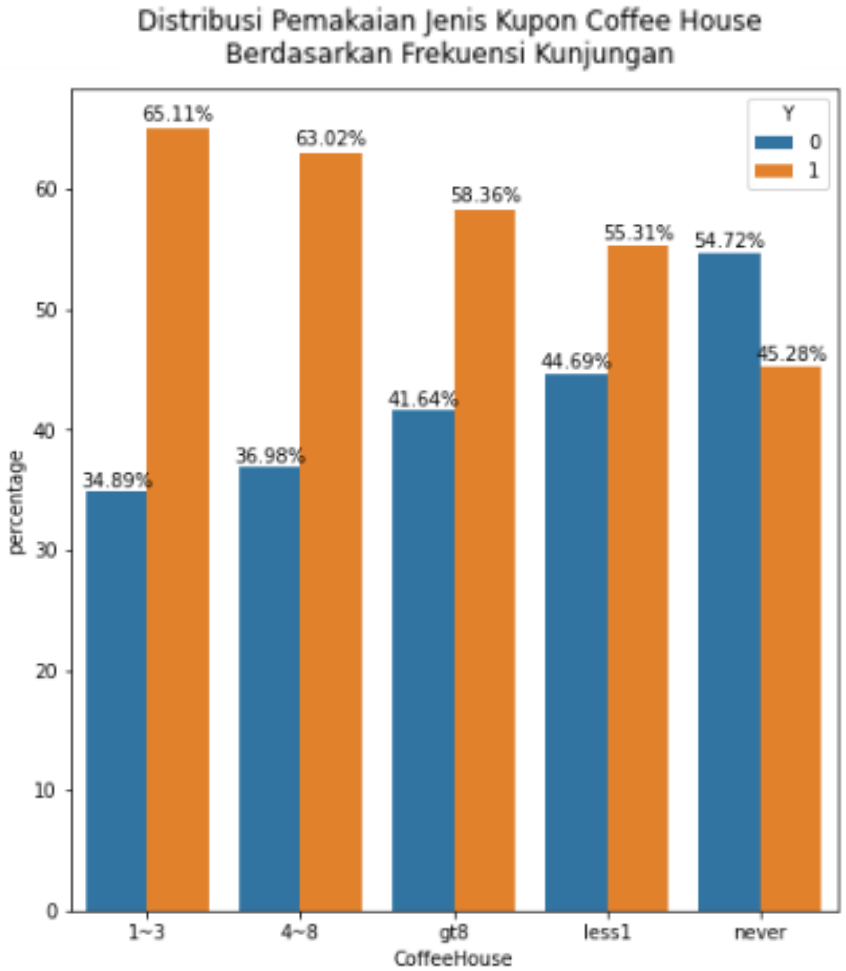
expiration

2 HOUR

49.6% Accepted

Kupon yang memiliki masa berlaku lebih lama lebih
lebih banyak diterima oleh pelanggan

BUSINESS INSIGHT

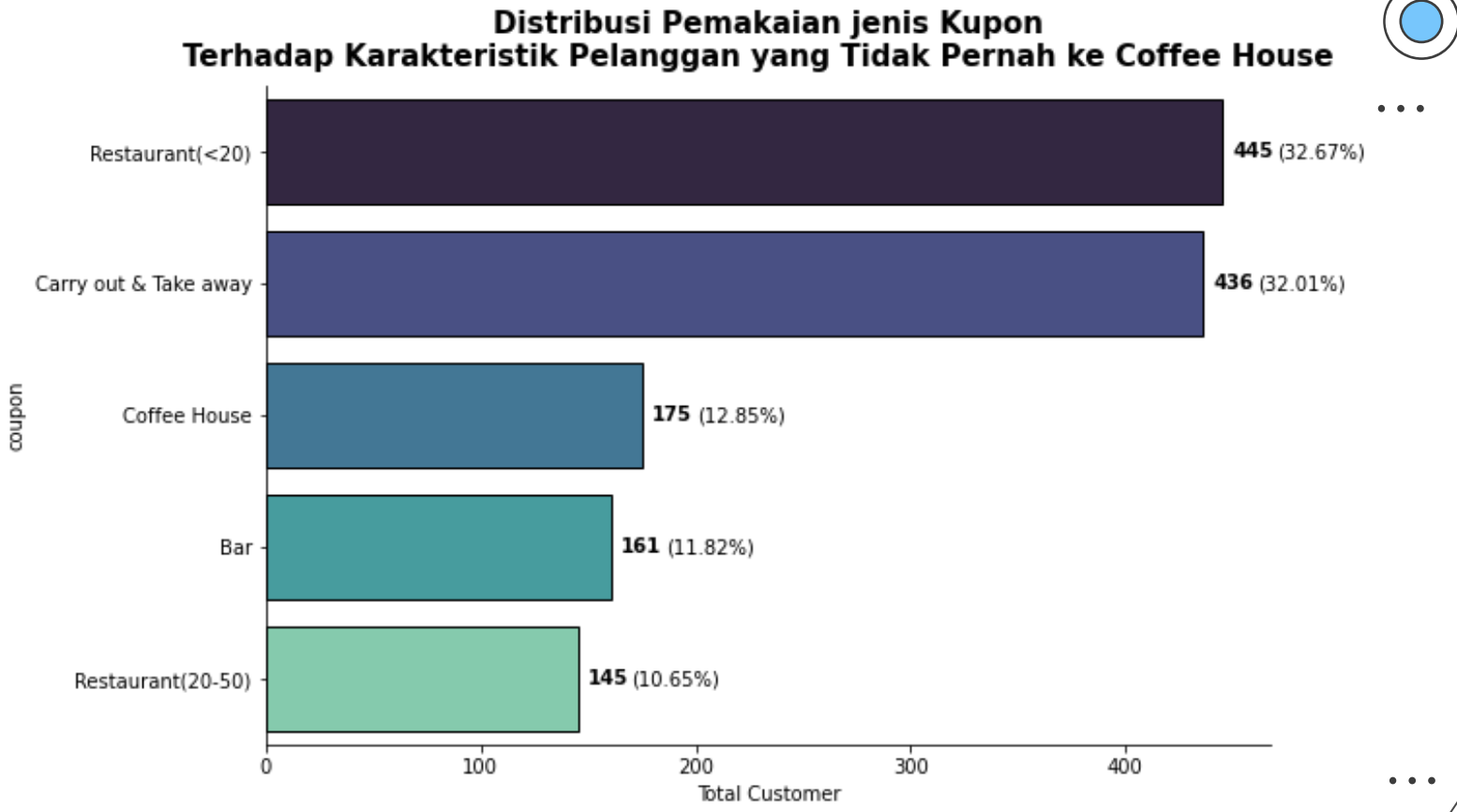


Pelanggan yang pergi ke **'CoffeeHouse' 1-3 kali** dalam sebulan memiliki presentase meerima kupon **paling tinggi**

- Pelanggan yang **memiliki histori berkunjung ke 'CoffeeHouse'** dapat diprioritaskan untuk diberikan jenis kupon CoffeeHouse
- Pelanggan yang **tidak memiliki histori berkunjung/'never' ke 'CoffeeHouse'** dapat diberikan kupon lainnya

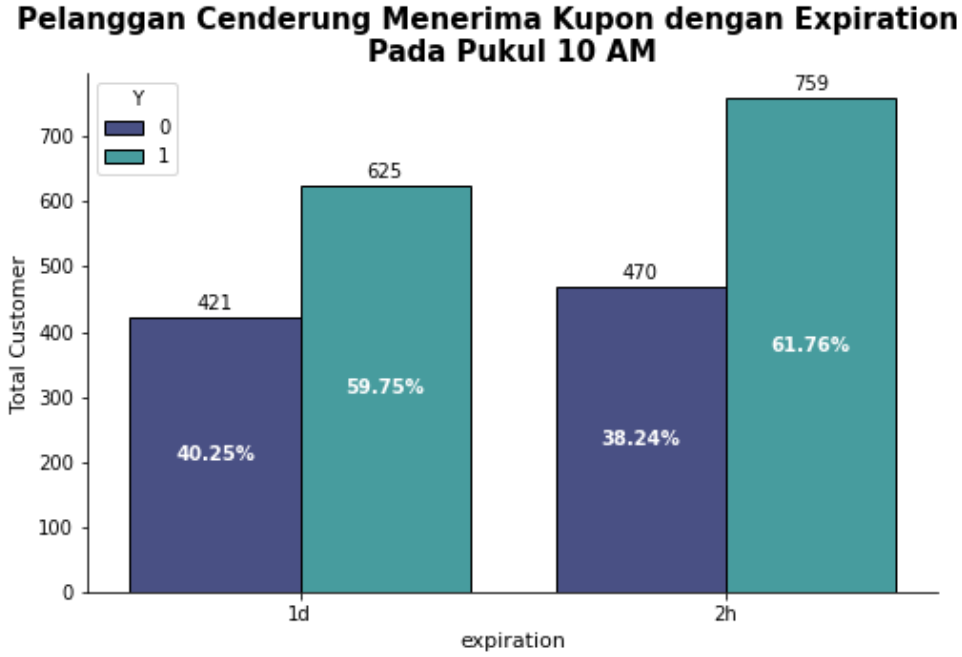
BUSINESS INSIGHT

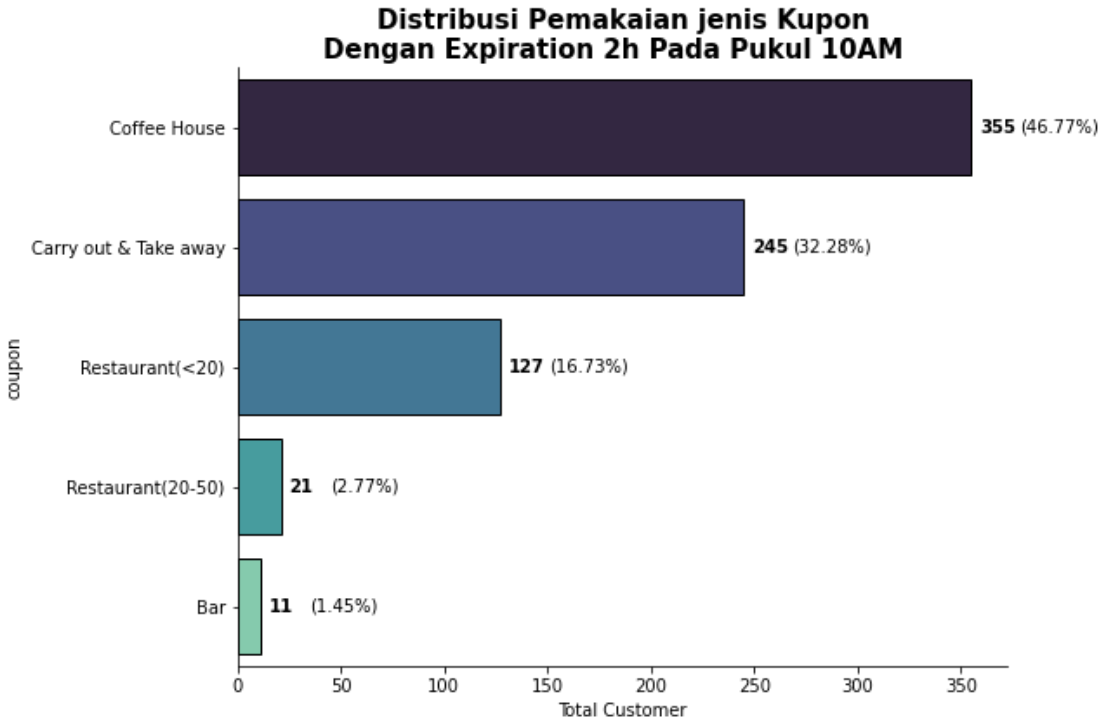
Jenis kupon tertinggi untuk pelanggan yang tidak pernah ke Coffee house adalah kupon **Restaurant Less Than $20** dan **Carry Away**



BUSINESS INSIGHT

Redeemed rate pada pukul 10AM lebih besar dengan masa kadaluarsa 2 jam dibandingkan dengan 1 hari





Jenis kupon tertinggi untuk pelanggan dengan expiration 2h adalah kupon Carry Away dan Coffee house

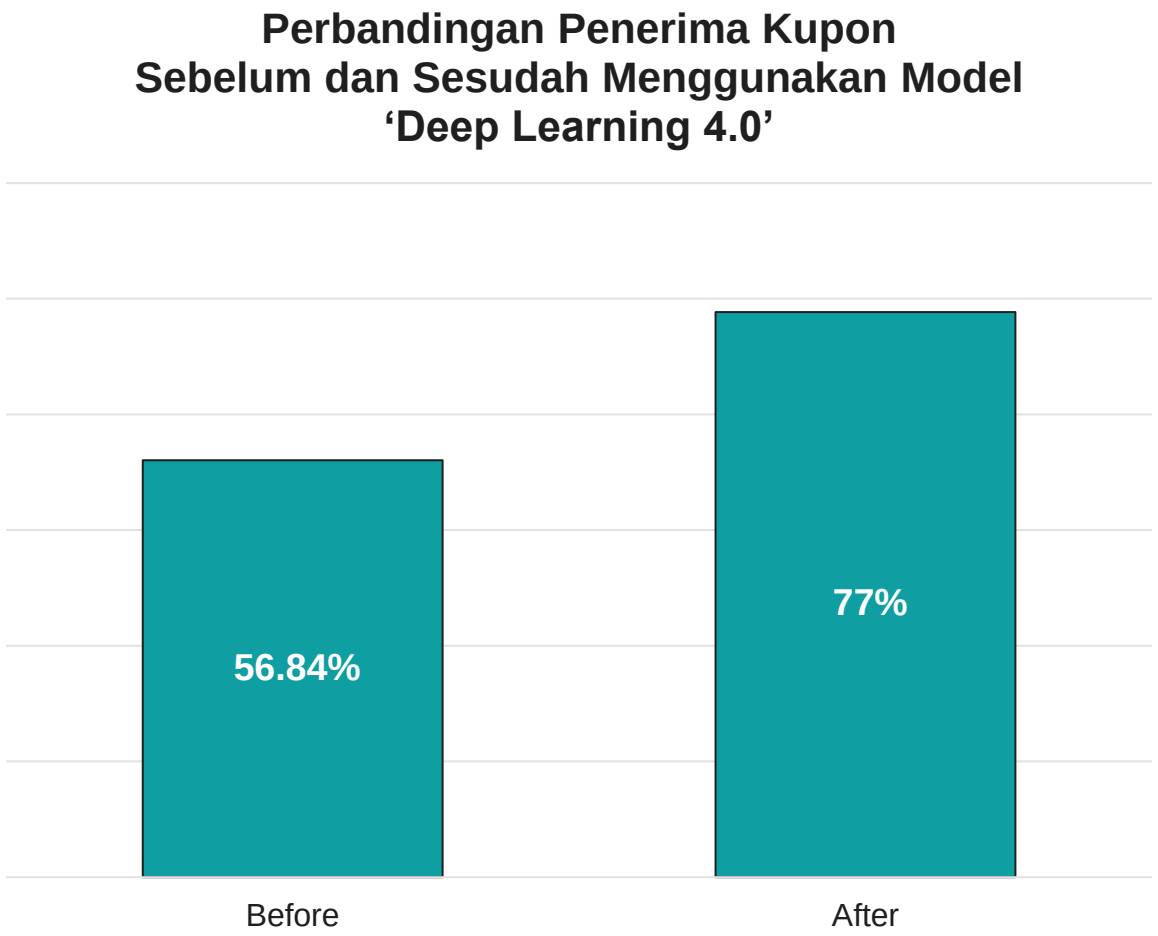
BUSINESS RECOMMENDATION

ACTIONABLE INSIGHTS

1. Pelanggan yang memiliki histori berkunjung ke Coffee House dapat diprioritaskan untuk diberikan kupon jenis Coffee House
2. Pelanggan yang tidak memiliki histori berkunjung/'never' ke Coffee House dapat direkomendasikan kupon jenis Carry Away dan Restaurant Less Than $20
3. Kupon dengan expiration 2h dapat diprioritaskan untuk diberikan pada jam 10 pagi dengan jenis kupon berupa Coffee House dan Carry Away. Karena hal ini dapat memicu perilaku impulsif terhadap pelanggan

BUSINESS RECOMMENDATION

MODEL 'DEEP LEARNING 4.0'



Terdapat *Improvement* penerima Kupon sebanyak 20.16% dengan jumlah sebelum penggunaan Model sebanyak 56.84% menjadi **77%** berdasarkan nilai presisi dari model

BUSINESS RECOMMENDATION

PERHITUNGAN B/C RATIO

$$\text{B/C Ratio} = \frac{\text{Benefit x Jumlah Coupon Acceptance}}{\text{Cost x Jumlah Coupon Acceptance}}$$

BEFORE USING OUR MODEL:

COUPON ACCEPTANCE 56.84%

Dengan asumsi Benefit percoupon = $3 dan Cost Percoupon = $1, dan penerima kupon sebanyak 7210 orang dari 12.684 orang
B/C Ratio = ($3 x 7210) / ($1 x 12.684) = 21.630/12.684

1.70x

AFTER USING OUR MODEL:

COUPON ACCEPTANCE 77%

Dengan asumsi Benefit percoupon = $3, Cost Percoupon = $1, Asumsi hasil prediksi dataset baru dengan penerima kupon sebanyak 500 orang dan nilai presisi model 77%
B/C Ratio = ($3 x (500 x 77%)) / ($1 x 500) = 1.155/500

2.31x

TERIMA KASIH