





Analisis Produk Terlaris dari Data Perilaku Customer Toko Online



By: Ilham Rofi'i, N.



Overview



Objective 1	Memprediksi pembelian produk
Objective 2	Seperti apa produk yang laris ?



Dataset

Open CDP

Oktober 2019 – Februari 2020

event_time

waktu tiap terjadinya event/perilaku dalam UTC

event_type

perilaku yang dilakukan, ada 3 jenis yaitu view, cart, dan purchase

product_id

ID dari produk

category_id

ID untuk kategori produk

category_code

Kategori produk

brand

nama brand produk

price

harga produk

user_id

ID permanen pengguna

user_session

ID pengguna tiap yang berubah tiap sesi

Exploratory Data Analysis

Analisis kolom category_code level dasar

Analisis kolom category_code lengkap

Analisis kolom brand

Analisis brand produk top 10 kategori produk

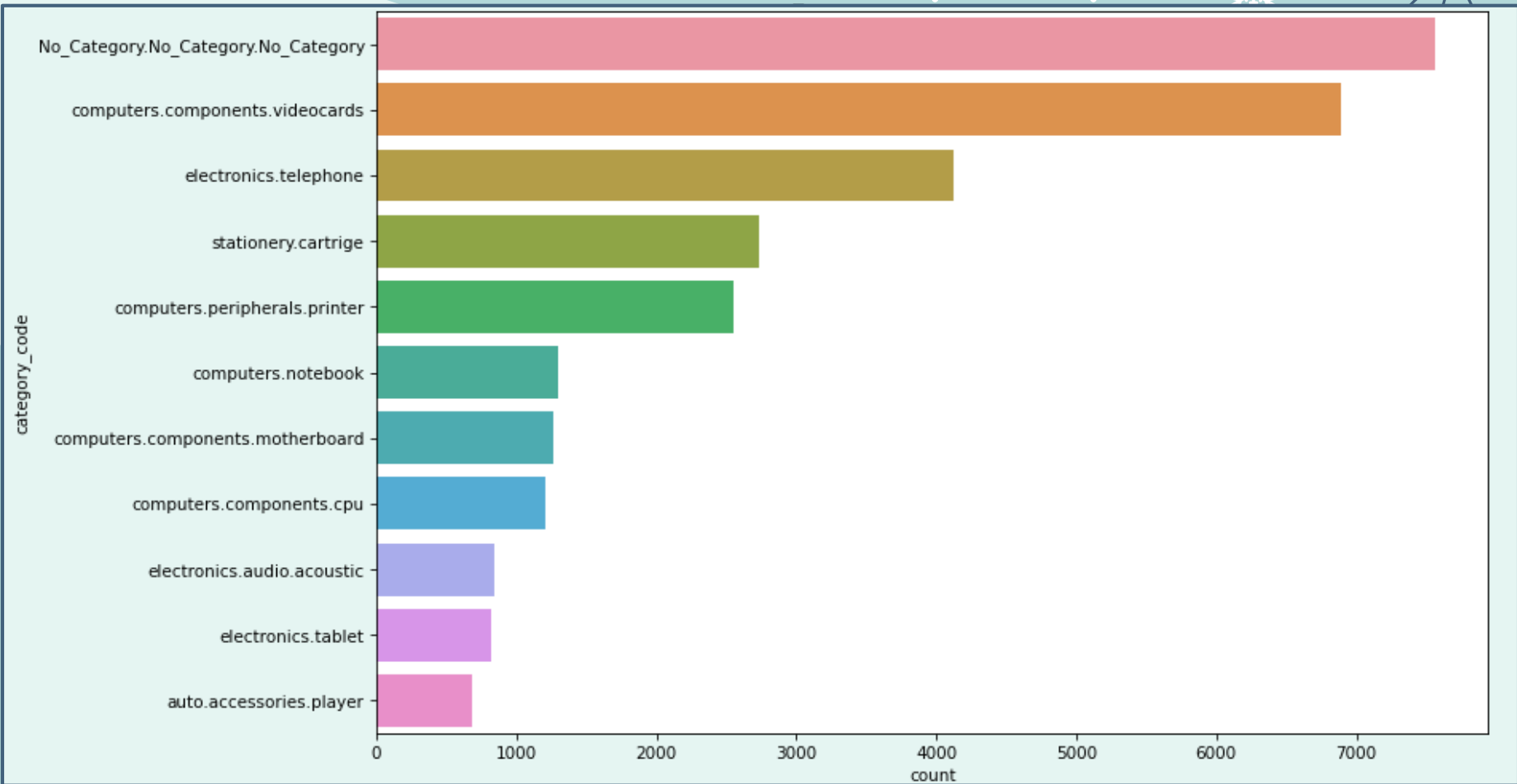
Analisis kolom price

Analisis price produk top 10 kategori produk

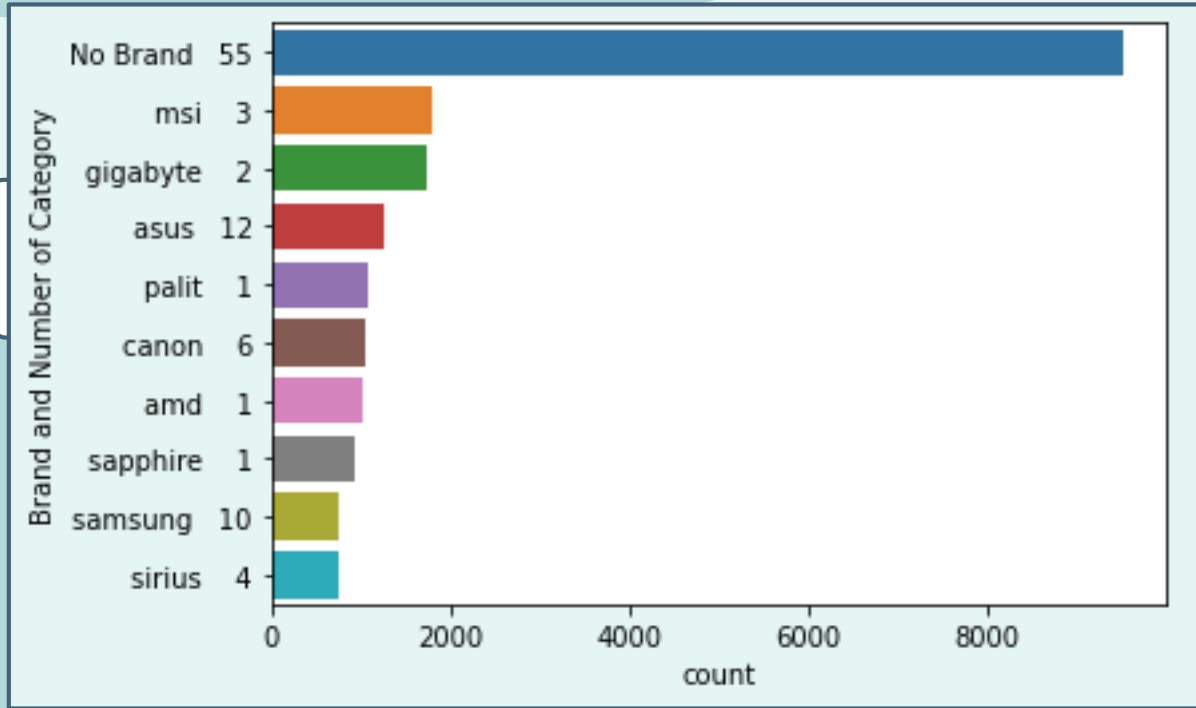
Analisis kolom activity

Analisis kolom weekday

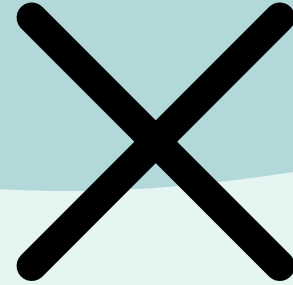
Analisis kolom category_code



Analisis kolom brand



Analisis brand produk top 10 kategori produk



	Top 3 Brand	Persentase No Brand
videocards	gigabyte, No Brand, sapphire	10%
telephone	No Brand, sirius, No Brand	61%
cartridge	hp, canon, nv-print	33%
printer	canon, pantum, canon	22%
notebook	mobilepc, No Brand, topon	28%
motherboard	asrock, asus, asrock	0%
cpu	amd, amd, amd	0%
acoustic audio	edifer, edifer, edifer	4%
tablet	samsung, xiaomi, irbis	35%
player (auto accessories)	pioneer, kenwood, pioneer	0%

Analisis price produk top 10 kategori produk

	product_id	category_code	brand	price	count	cheaper_percentage
0	4099645	computers.components.videocards	gigabyte	292.83	564	22.24%
1	1821813	computers.components.videocards	No Brand	397.48	538	51.04%
2	893196	computers.components.videocards	sapphire	214.10	384	8.92%
0	3829355	electronics.telephone	No Brand	32.22	543	38.44%
1	1821557	electronics.telephone	sirius	16.19	121	6.23%
2	3632668	electronics.telephone	No Brand	34.44	116	82.81%
0	3642540	stationery.cartridge	hp	26.19	345	21.91%
1	1785245	stationery.cartridge	canon	43.11	207	69.79%
2	471387	stationery.cartridge	nv-print	13.86	110	0.00%
0	799067	computers.peripherals.printer	canon	82.97	233	8.69%
1	3829912	computers.peripherals.printer	pantum	122.86	122	41.58%
2	3790736	computers.peripherals.printer	canon	163.52	94	55.01%
0	1044616	computers.notebook	mobilepc	22.22	95	9.29%
1	3627467	computers.notebook	No Brand	31.11	46	45.36%
2	1006967	computers.notebook	topon	55.87	46	87.43%

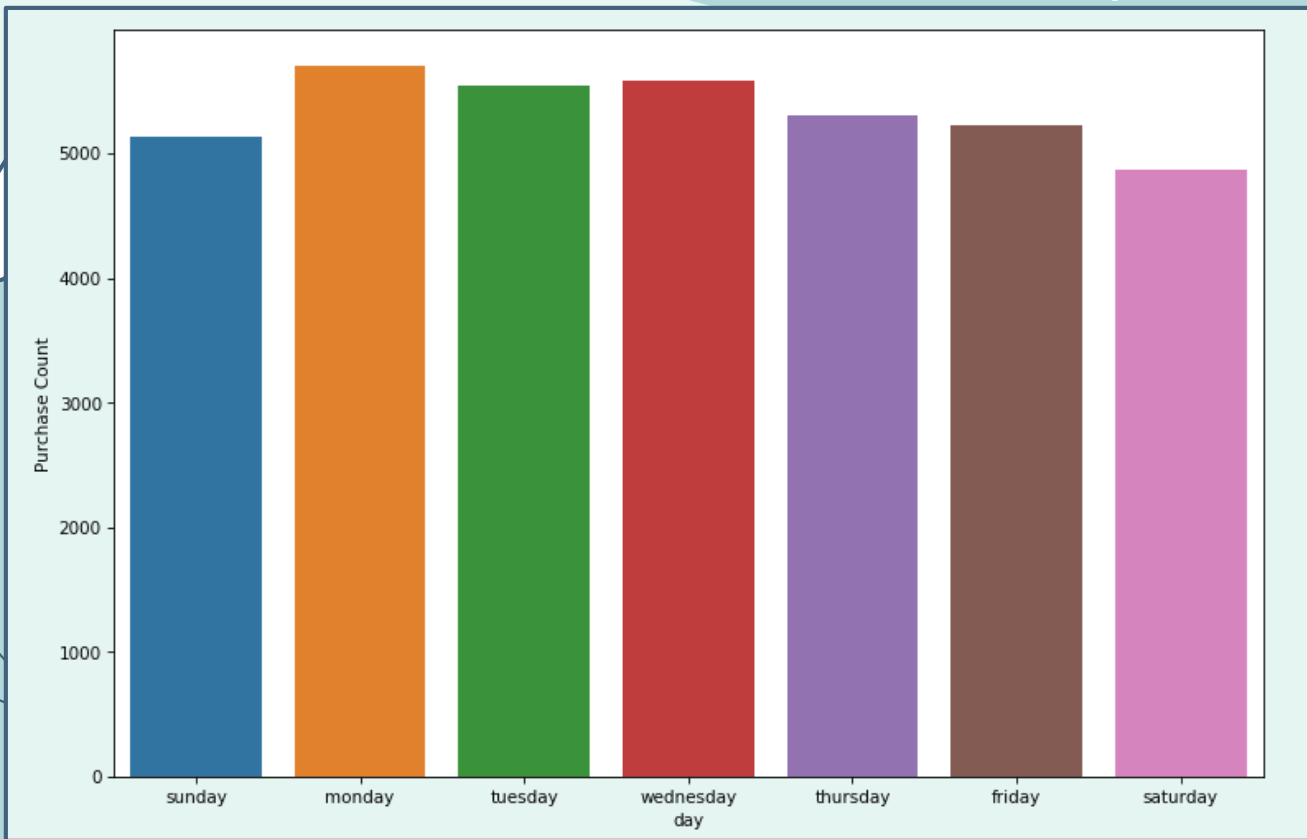
Analisis price produk top 10 kategori produk

	product_id	category_code	brand	price	count	cheaper_percentage
0	136700	computers.components.motherboard	asrock	77.73	194	69.05%
1	809948	computers.components.motherboard	asus	64.48	187	26.73%
2	523117	computers.components.motherboard	asrock	73.81	125	54.40%
0	3791351	computers.components.cpu	amd	204.29	423	40.22%
1	942339	computers.components.cpu	amd	56.27	96	0.00%
2	3804514	computers.components.cpu	amd	140.86	79	27.82%
0	1271550	electronics.audio.acoustic	edifier	84.89	76	14.97%
1	124712	electronics.audio.acoustic	edifier	204.79	61	67.07%
2	471287	electronics.audio.acoustic	edifier	117.75	43	48.50%
0	3829572	electronics.tablet	samsung	57.62	67	30.58%
1	1578612	electronics.tablet	xiaomi	284.30	58	82.26%
2	1804316	electronics.tablet	irbis	127.08	40	70.03%
0	4099764	auto.accessories.player	pioneer	126.68	180	36.09%
1	4100427	auto.accessories.player	kenwood	104.86	68	11.09%
2	1843507	auto.accessories.player	pioneer	153.83	50	84.57%

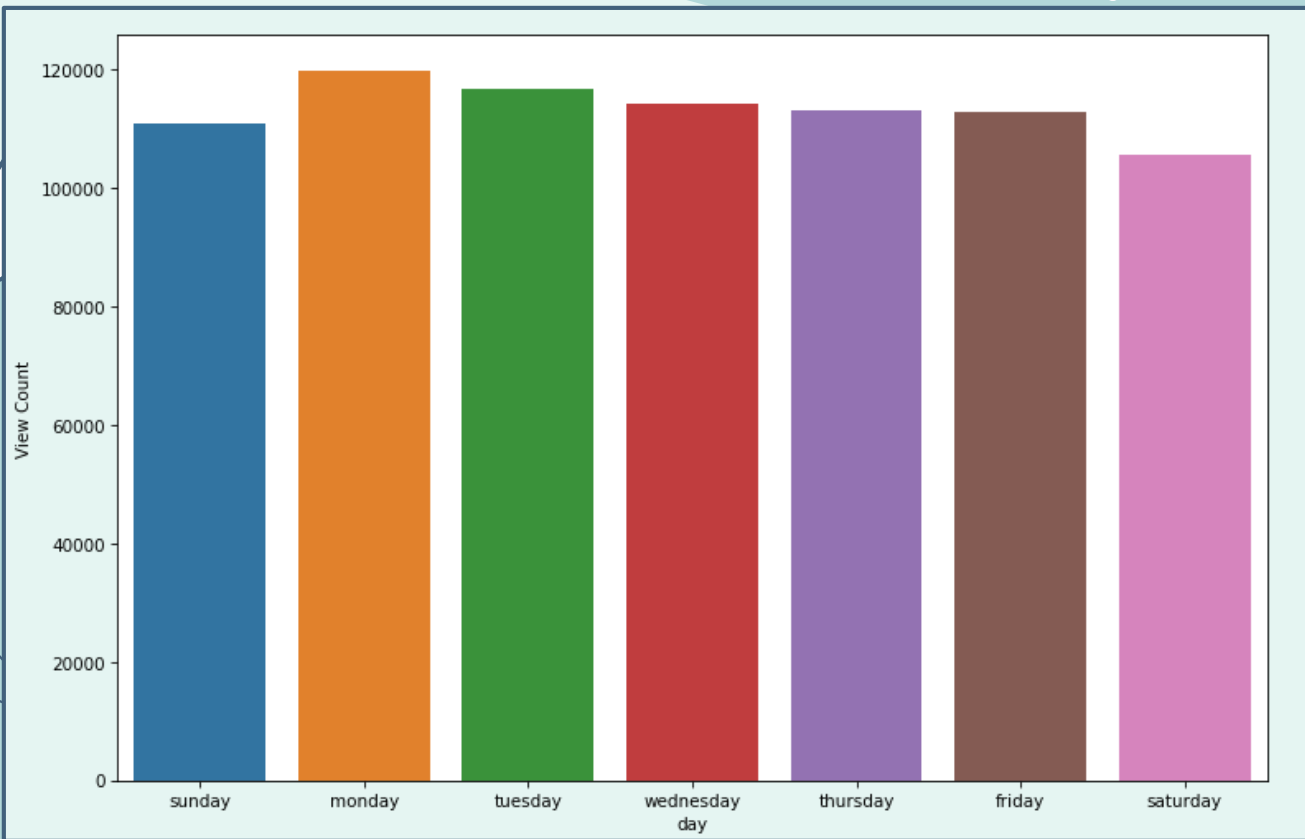
Analisis kolom activity

	product_id	category_code	view_count	rank_view	purchase_count	rank_purchase	cart_count	rank_cart
0	1821813	computers.components.videocards	12794	1	564	1	1220	1
1	3791351	computers.components.cpu	5717	2	543	2	1033	2
2	4099645	computers.components.videocards	5489	3	538	3	761	3
3	1785245	stationery.cartrige	4630	4	423	4	664	4
4	3829355	electronics.telephone	4552	5	384	5	662	5
5	809948	computers.components.motherboard	3830	6	257	9	492	6
6	1830099	computers.components.videocards	3660	7	207	12	450	7
7	3791509	computers.components.videocards	3150	8	187	16	318	10
8	4154414	computers.components.videocards	3000	9	169	19	294	13
9	893196	computers.components.videocards	2865	10	149	22	239	19

Analisis kolom weekday



Analisis kolom weekday









Analisis Machine Learning



Objective 1	Apakah feature dalam dataset mempengaruhi target variable ?
Objective 2	Feature apa yang paling penting ?
Data Preparation	feature engineering, data cleansing, overcoming imbalances, categorical variable encoding
Data Modeling	Xgboost Machine Learning Algorithm
Feature Importance	xgboost built-in feature, permutation method, SHAP







Data Preparation



feature engineering	'brand', 'price', 'event_weekday', 'category_code_level1', 'category_code_level2', 'activity_count'
data cleansing	Drop duplicated data Fill missing data
encoding	Using LabelEncoder() from sklearn
data imbalances	45% Imbalances



Data Modeling

F1-Score

0.63



Sumber: <https://www.educba.com/xgboost-algorithm/>

Feature Importance

SHAP

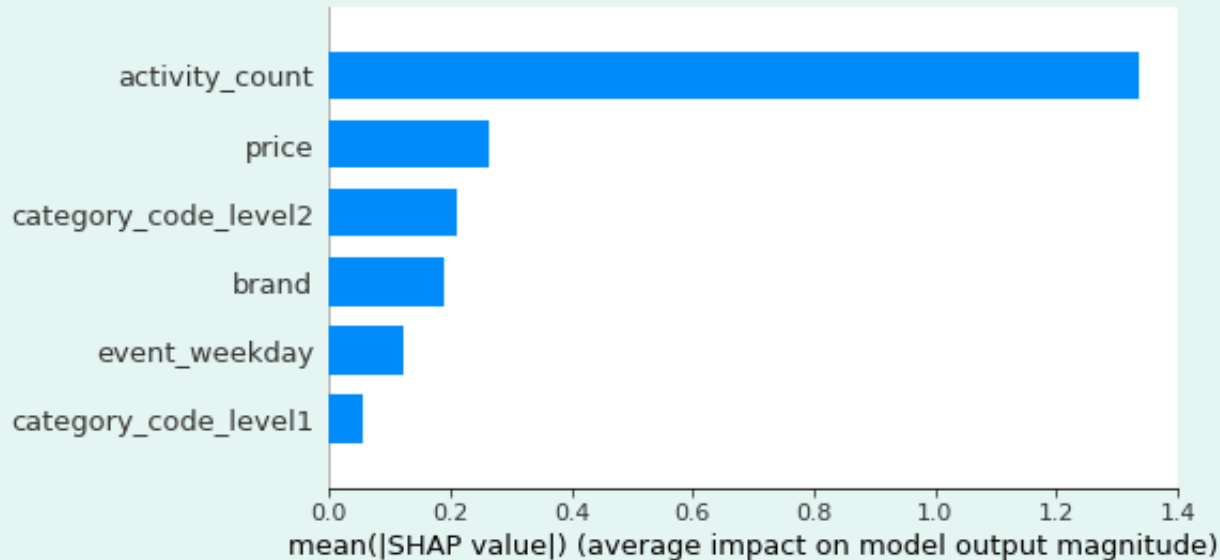
Metode feature importance utama

Permutaion Methode

Sama persis dengan hasil SHAP

XGBoost Built-in Function

Sedikit berbeda dengan hasil SHAP



Rekomendasi

1. Menjual barang yang populer

- a. Barang dengan brand terkenal
- b. Melakukan kegiatan pengiklanan

3. Menjual produk yang termasuk 10 kategori terlaris

videocards, telephone, cartrige, printer, notebook, motherboard, cpu, acoustic audio, tablet, player (auto accessories)

2. Menjual produk dengan harga murah hingga mahal

Produk paling laris harganya dibawah 50% barang terjual di kategori yang sama.

4. Menjual produk dengan brand terkenal

msi, gigabyte, asus, palit, canon, amd, sapphire, samsung, sirius



Terima Kasih atas Perhatiannya

For more info:

SLIDESGO | BLOG | FAQs

You can visit our sister projects:

FREEPIK | FLATICON | STORYSET | WEPIK | VIDEVO

