

Data-Driven Insights : Optimizing Inventory and Marketing Strategies

Kalbe Nutritionals - Data Scientist

Presented by
R Firdaus Dharmawan Akbar



R Firdaus Dharmawan Akbar

Undergraduate Data Science Technology

Saya adalah seorang mahasiswa S1 di Universitas Airlangga jurusan Teknologi Sains Data. Selama studi saya dalam 2 tahun terakhir, saya telah mengembangkan pemahaman utama dalam konsep dasar ilmu data seperti statistik, pemrosesan data, visualisasi data, dan pembelajaran mesin. Terlibat dalam banyak proyek sains data secara individu maupun tim dari berbagai sumber untuk mendapatkan wawasan dan hasil yang berharga.

Study Case Experience

- JABODETABEK House Price Prediction
- Segmentasi Pengemudi Ojek Online Pontianak
- Stock Movement Analysis Technical Indicator & Historical Data
- Cryptography System Based On Vignere & Polybius Chiper
- Analysis of Purchase Patterns in Minimarket Using Apriori Association Rule Method
- Analysis of Skincare Product Recommendation Based on Customer Reviews Using Apriori Association Rule Method
- Analysis of Tiktok App Review with Topic Modeling

Outline

- Business Understanding
- Data Preparation
- Exploratory Data Analysis (EDA)
- Modeling & Evaluation
- Conclusion & Recommendations

Business Understanding

Problem Background

Bagaimana meningkatkan efektivitas promosi, penjualan, dan perlakuan dengan pendekatan yang lebih personal kepada pelanggan. Seiring dengan pertumbuhan bisnis, dapat disadari bahwa perubahan perilaku pelanggan dan kebutuhan pasar yang terus berkembang menuntut untuk menjadi lebih inovatif.

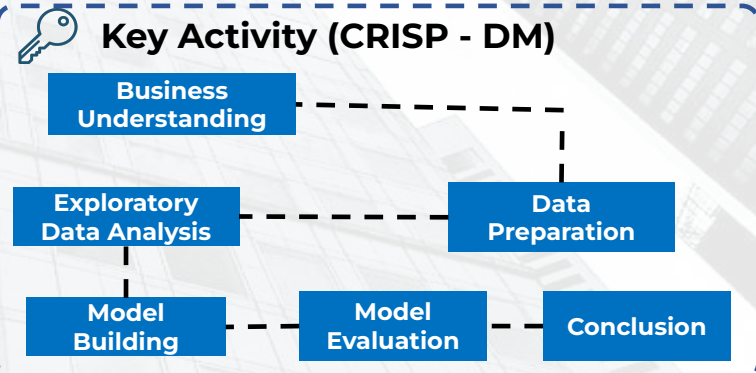


Goals

Meningkatkan efisiensi manajemen persediaan dengan memastikan stok harian dapat terjual dan meningkatkan efektivitas promosi serta penjualan dengan pendekatan personal kepada pelanggan.



Problem Statement



Data Preparation

Data Preprocessing

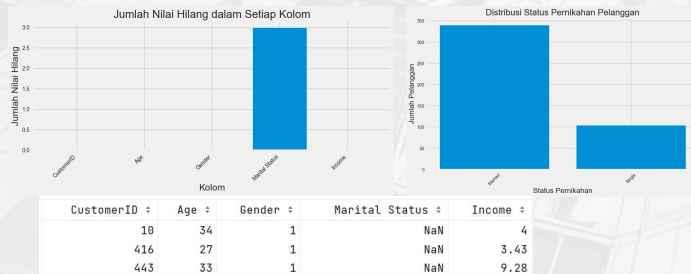
- **Handling Inconsistent Data**

Tahap 1: Mengatasi Inkonsistensi "Income":
Mengubah tanda koma (,) menjadi titik (.) pada atribut "Income" dataset customer untuk format angka yang konsisten.

Tahap 2: Mengatasi Inkonsistensi Koordinat:
Ganti koma (,) menjadi titik (.) pada atribut "Latitude" dan "Longitude" dataset store. Konversi menjadi float untuk analisis koordinat geografis.

- **Handling Missing Values**

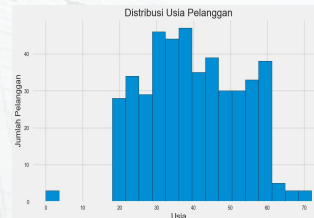
Terdapat nilai *null* pada atribut "Marital Status" dari dataset customer sesuai visualisasi berikut :



*Solve : Imputasi Modus

- **Handling Outlier**

Berdasarkan visualisasi histogram umur customer terlihat bahwa terdapat umur yang rendah (0, 2, dan 3).



CustomerID	Age	Gender	Marital Status	Income
12	2	1	Married	4.94
74	3	1	Married	5.09
128	0	1	Married	6.77

Lebih lanjut, hal ini tidak mungkin jika customer memiliki umur tersebut tetapi berstatus menikah. *Solve : Imputasi Median

- **Feature Encoding**

Menggunakan teknik *label encoding* untuk mengubah data kategori menjadi bentuk numerik berurutan agar dapat digunakan dalam pemodelan

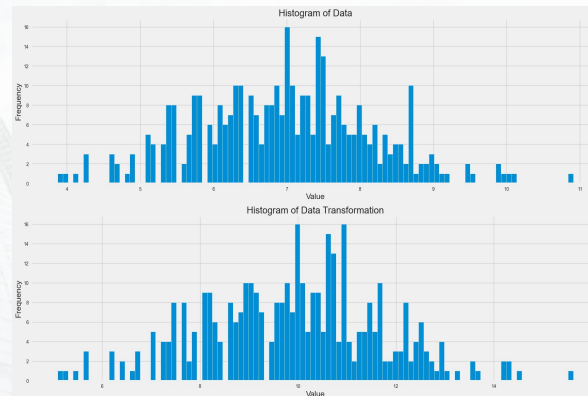
Data Preparation

Data Transformation Box- Cox

Teknik ini untuk pemodelan time series yang bertujuan untuk meminimalkan nilai margin of error. Hal ini dengan cara mengubah data yang tidak berdistribusi normal menjadi data yang lebih mendekati distribusi normal. Transformasi ini menggunakan parameter lambda (λ) untuk mentransformasi data berdasarkan tabel berikut :

λ	Transformed Data
-2	y^{-2}
-1	y^{-1}
-0.5	$1/\sqrt{y}$
0	$\ln(y)$
0.5	\sqrt{y}
1	y
2	y^2

Berdasarkan hasil temuan didapatkan nilai lambda (λ) sebesar 0.426, artinya sesuai dengan tabel berikut dapat melakukan transformasi \sqrt{y} .



Exploratory Data Analysis



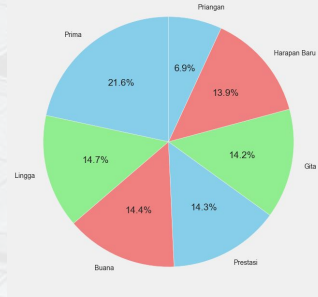
Marital Status	rata_rata_umur
Married	43.0382352941
Single	29.3846153846

gender	rata_rata_umur
Wanita	40.326446281
Pria	39.1414634146

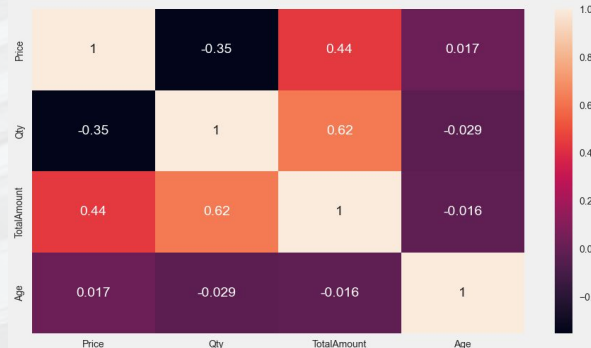
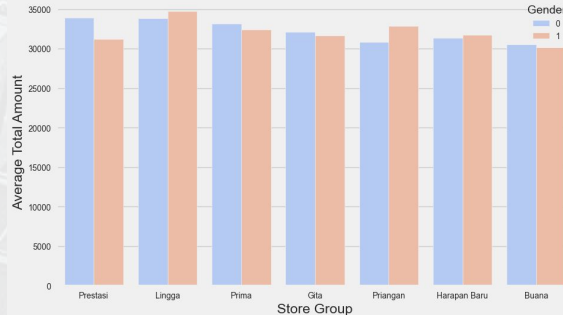
storename	qty
Lingga	2,777
Sinar Harapan	2,588
Prestasi Utama	1,395
Prima Kota	1,358
Buana	1,320

Product Name	total_amount
Cheese Stick	27,615,000
Choco Bar	21,190,400
Coffee Candy	19,711,800
Yoghurt	19,630,000
Oat	15,440,000

Percentage of Purchases by Store Group

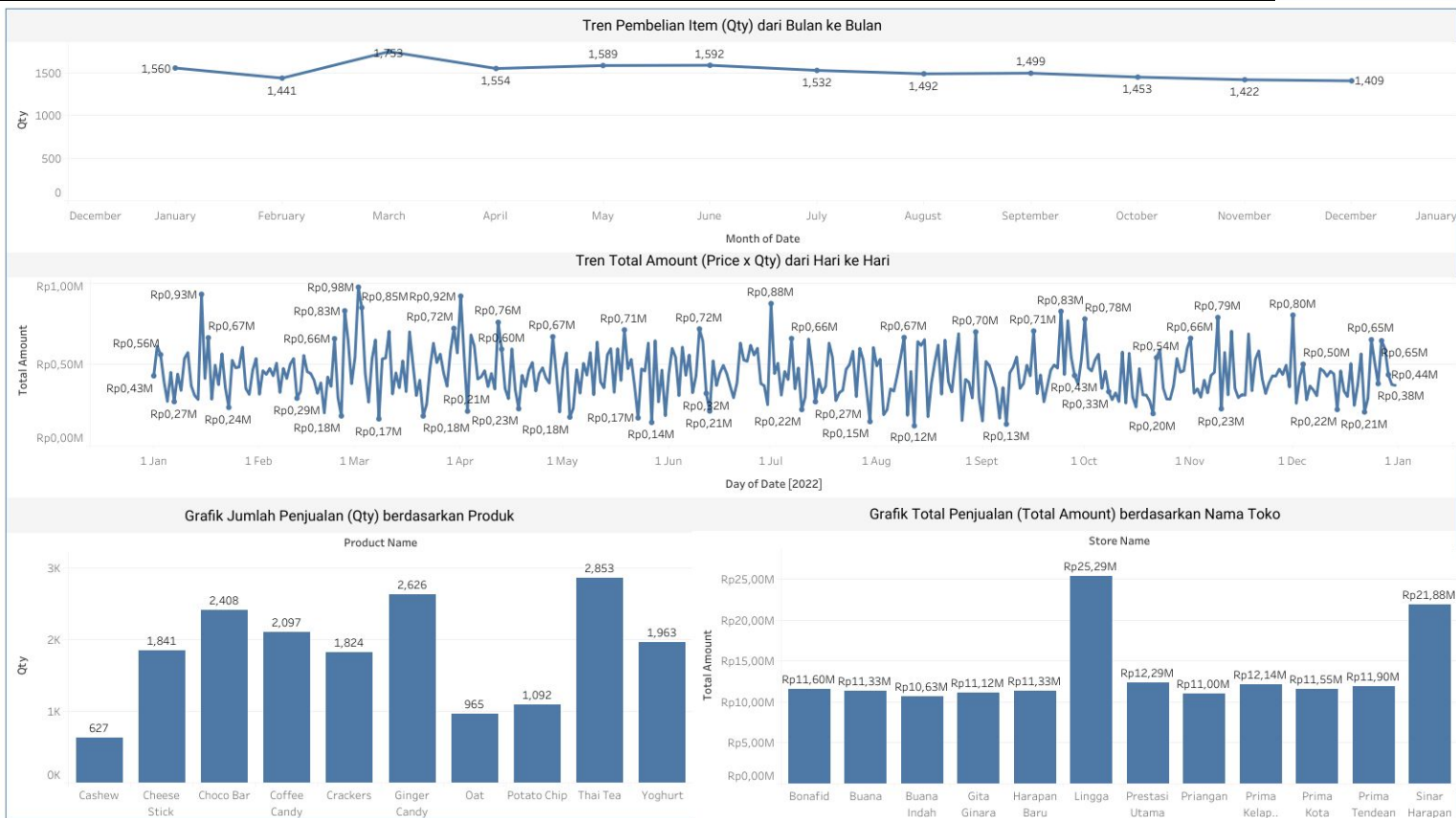


Average Total Amount by Store Group with Gender Separation



Hubungan antara "Price" dan "Qty" dengan "TotalAmount" relatif kuat, sementara "Income" dan "Age" tidak menunjukkan hubungan yang jelas dengan variabel lainnya.

Exploratory Data Analysis



Exploratory Data Analysis

Uji Asumsi Time Series

- **Uji Keberadaan Musiman**

Uji keberadaan komponen musiman dilakukan dengan menghitung perbedaan absolut antara nilai musiman dan rata-rata musiman. Berdasarkan hasil temuan dengan selang kepercayaan 95% menunjukkan **data time series bersifat musiman**.

- **Uji Stasioneritas**

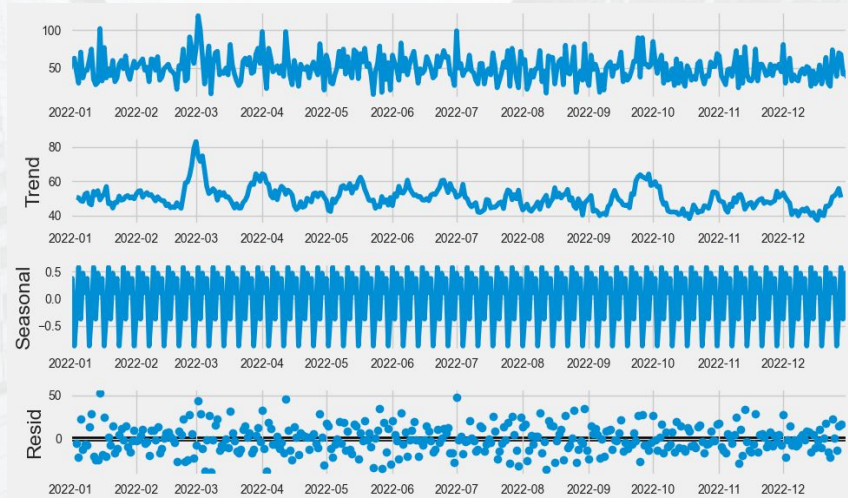
Perhitungannya menggunakan uji Augmented Dickey-Fuller (ADF).

Uji Hipotesis :

H_0 = Data residu bersifat stasioner

H_1 = Data residu tidak bersifat stasioner

Berdasarkan hasil temuan, statistik ADF = -19.448086 dan nilai P-value (0.000) > alpha (-2.87) sehingga keputusan **gagal tolak H_0** . Artinya, hasil temuan dengan selang kepercayaan 95% menunjukkan **data time series bersifat stasioner**.



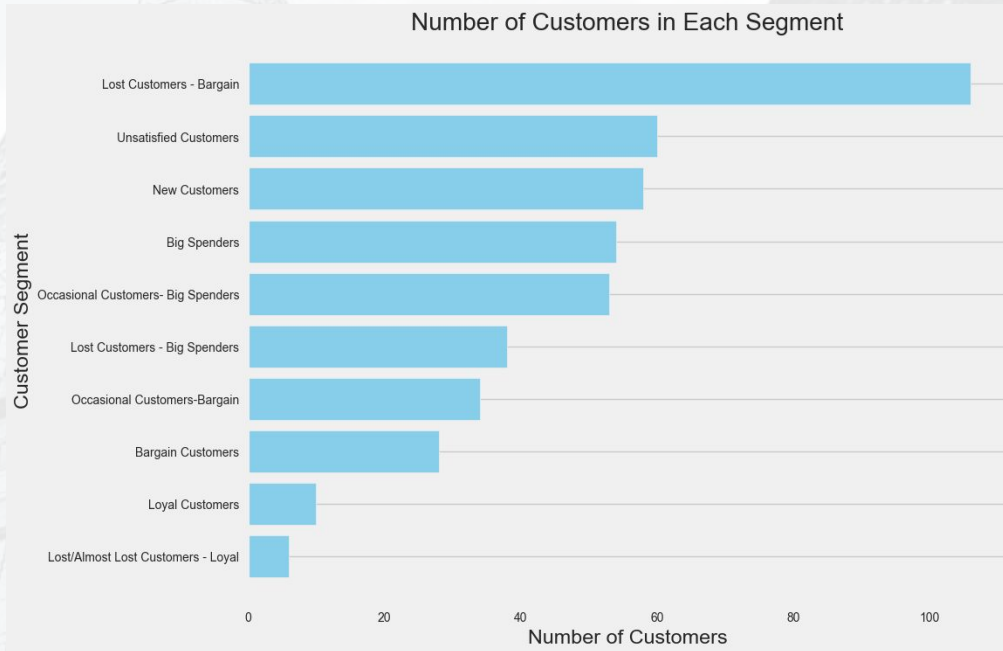
Exploratory Data Analysis

RFM Analysis

RFM adalah singkatan dari "**Recency**, **Frequency**, dan **Monetary**," yang bertujuan untuk menganalisis pelanggan dalam hal pemasaran dan manajemen hubungan pelanggan.

- **Recency** mengacu pada seberapa baru pelanggan telah melakukan transaksi terakhir dengan perusahaan.
- **Frequency** mengukur seberapa sering pelanggan bertransaksi dalam suatu periode waktu.
- **Monetary** mengukur jumlah uang yang dihabiskan oleh pelanggan dalam periode waktu.

Setelah itu, membagi data RFM dalam 3 rentang kuantil, berarti bahwa 33% pelanggan pertama memiliki jumlah total pembelian antara 1 dan 3. Pelanggan berikutnya (persentil 33-66) memiliki jumlah total pembelian antara 3 dan 5, dan seterusnya. Adapun hasil dari pengelompokkan analisis ini berdasarkan visualisasi di samping



Exploratory Data Analysis

ABC Analysis

Analisis ABC metode pengelompokan produk berdasarkan tingkat pentingnya atau kontribusinya terhadap kriteria tertentu. Metode ini digunakan untuk jenis kelompok dari sisi penawaran. Analisis ini didasarkan pada aturan Pareto yang menyatakan bahwa 80% hasil (pendapatan) berasal dari 20% penyebab (produk).

Lalu, menghitung persentase total pendapatan yang dihasilkan oleh setiap produk dan mengaturnya dalam urutan persentase yang menurun serta menghitung persentase kumulatif pendapatan untuk masing-masing produk.

- Kategori A: Produk yang menghasilkan 75% dari pendapatan
- Kategori B: Produk yang memiliki persentase kumulatif antara 75%-95%.
- Kategori C: Produk yang menghasilkan sisa 5% dari pendapatan.

ABC_Revenue	Revenue	count
A	118988700	12
B	34651100	6
C	8403200	2

Berdasarkan hasil temuan analisis ini, didapatkan kategori A, B dan C terhadap tipe store adalah berikut :

Type	Product Name	Total_Revenue	ABC_Revenue
General Trade	Cheese Stick	14475000	A
Modern Trade	Cheese Stick	13140000	A
General Trade	Coffee Candy	11392800	A
General Trade	Choco Bar	11290400	A
General Trade	Yoghurt	11070000	A
Modern Trade	Choco Bar	9900000	A
General Trade	Oat	8928000	A
Modern Trade	Yoghurt	8560000	A
Modern Trade	Coffee Candy	8319000	A
General Trade	Crackers	7747500	A
General Trade	Potato Chip	7152000	A

Type	Product Name	Total_Revenue	ABC_Revenue
Modern Trade	Oat	6512000	B
General Trade	Cashew	6138000	B
Modern Trade	Potato Chip	5952000	B
Modern Trade	Crackers	5932500	B
Modern Trade	Cashew	5148000	B
Modern Trade	Thai Tea	4968600	B

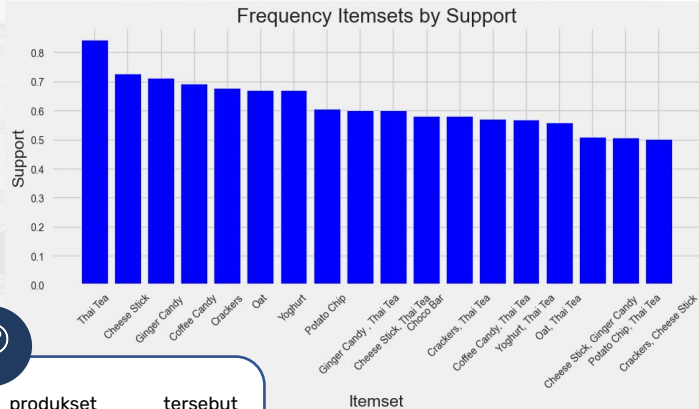
Type	Product Name	Total_Revenue	ABC_Revenue
General Trade	Ginger Candy	4963200	C
Modern Trade	Ginger Candy	3440000	C

Exploratory Data Analysis

Market Basket Analysis

• Algoritma Apriori

Adanya 5020 transaksi dan batas ambang minimum support sebesar 50%, Apriori akan menghitung berapa kali setiap produkset muncul dalam transaksi dan kemudian menyaring produkset yang memiliki tingkat support setidaknya 50%.



produkset tersebut memiliki tingkat support lebih tinggi dari 50%, sehingga dianggap signifikan dalam dataset.

• Association Rule

Sebuah aturan asosiasi adalah pola yang menyatakan bahwa ketika X terjadi, Y terjadi dengan probabilitas tertentu. Berikut hasil temuan berdasarkan nilai minimum confidence 80% :

antecedents	consequents	support	confidence	lift
(Ginger Candy)	(Thai Tea)	0.599553	0.842767	1.001907
(Cheese Stick)	(Thai Tea)	0.599553	0.824615	0.980327
(Crackers)	(Thai Tea)	0.579418	0.854785	1.016194
(Coffee Candy)	(Thai Tea)	0.570470	0.825243	0.981073
(Yoghurt)	(Thai Tea)	0.568233	0.849498	1.009909
(Oat)	(Thai Tea)	0.557047	0.832776	0.990029
(Potato Chip)	(Thai Tea)	0.505593	0.833948	0.991423

Jika pelanggan membeli Thai Tea, peluang mereka juga membeli produk seperti Ginger Candy, Cheese Stick, Crackers, Coffee Candy, Yoghurt, Oat, dan Potato Chip relatif tinggi (confidence tinggi).

Support : Aturan asosiasi dianggap berlaku dengan nilai support (sup) pada himpunan data transaksi (T) jika persentase sup transaksi mengandung X U Y adalah setidaknya sup% dalam T.

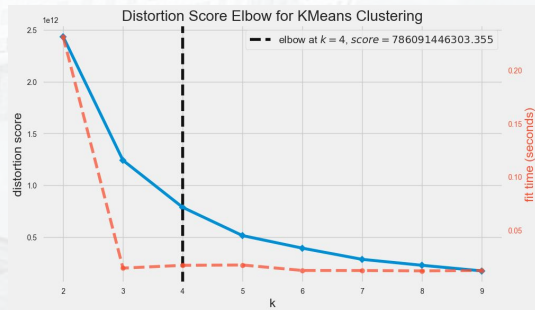
$$*sup = Pr(X \cup Y)$$

Confidence : Aturan asosiasi dianggap berlaku dalam T dengan nilai confidence (conf) jika persentase conf transaksi yang mengandung X juga mengandung Y adalah setidaknya conf%.

$$*conf = Pr(Y | X)$$

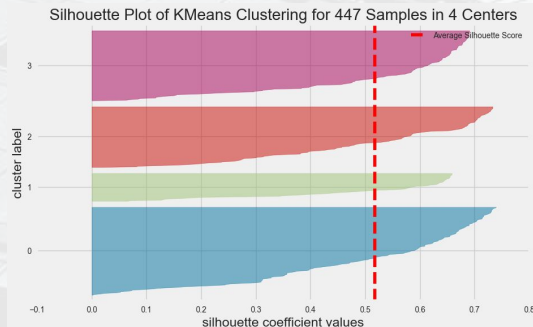
Modeling & Evaluation

Elbow Method



Metode elbow digunakan untuk menemukan jumlah optimal cluster dalam analisis klasterisasi. Pada titik elbow, distorsi mulai menunjukkan penurunan yang lebih lambat setelahnya, sehingga hasil k yang didapatkan sebesar 4.

Silhouette Score



Visualisasi menampilkan koefisien siluet untuk setiap sampel pada basis per cluster, memvisualisasikan cluster mana yang padat dan mana yang tidak. Hal ini sangat berguna untuk menentukan ketidakseimbangan klaster. Rata-rata hasil score yang didapatkan sebesar 0.52 menandakan partisi data relatif baik.

Result

cluster	TransactionID	Qty	TotalAmount	Income	Marital Status	Gender
0	11	39	340749.685535	8.544906	Married	Wanita
1	16	61	591511.764706	8.186078	Married	Wanita
2	13	49	450864.545455	8.450545	Married	Wanita
3	8	27	221272.440945	8.936850	Married	Pria

- Terdapat empat kelompok pelanggan (kluster) berdasarkan pola pembelian dan karakteristik mereka.
- Setiap kluster memiliki rata-rata jumlah transaksi, jumlah barang yang dibeli (Qty), total jumlah pembelian (TotalAmount), dan tingkat pendapatan pelanggan (Income Customer) yang berbeda-beda.
- Semua kluster didominasi oleh pelanggan yang status pernikahannya adalah "Married".
- Kluster 0, 1, dan 2 didominasi oleh pelanggan dengan jenis kelamin "Wanita".
- Kluster 3 didominasi oleh pelanggan dengan jenis kelamin "Pria".

Modeling & Evaluation

Auto-ARIMA adalah algoritma yang digunakan untuk secara otomatis menentukan model ARIMA yang paling sesuai yang menggabungkan pendekatan grid search dengan seleksi model berbasis statistik untuk memilih parameter terbaik (p, d, q) dari model ARIMA. Ukuran statistik untuk mengukur evaluasi model ini menggunakan nilai AIC dengan semakin rendah nilainya, semakin baik modelnya. Berikut tabel hasil pemodelan :

Model	AIC	Time (sec)
ARIMA(0,1,0)(0,1,0)[12]	1601.433	0.03
ARIMA(1,1,0)(1,1,0)[12]	1401.802	0.10
ARIMA(1,1,0)(0,1,0)[12]	1471.078	0.03
ARIMA(1,1,0)(2,1,0)[12]	1373.291	0.28
ARIMA(0,1,0)(2,1,0)[12]	1482.214	0.18
ARIMA(2,1,0)(2,1,0)[12]	1339.784	0.37
ARIMA(2,1,0)(1,1,0)[12]	1378.534	0.14
ARIMA(3,1,0)(2,1,0)[12]	1302.938	0.46
ARIMA(3,1,0)(1,1,0)[12]	1341.777	0.26
ARIMA(3,1,0)(2,1,0)[12]		
intercept	1304.934	1.61



Berdasarkan hasil tersebut, didapatkan model terbaik dengan nilai AIC terkecil, yaitu **ARIMA(3,1,0)(2,1,0)[12]**.

Modeling & Evaluation

Uji Asumsi Hasil Pemodelan Terbaik

1

Uji Independensi Residu (White-Noise Residual)

Hasil uji Ljung-Box (Q) statistik (1.34) dengan Prob(Q) (0.25), menunjukkan bahwa model memenuhi asumsi tentang autokorelasi dalam residuals.

2

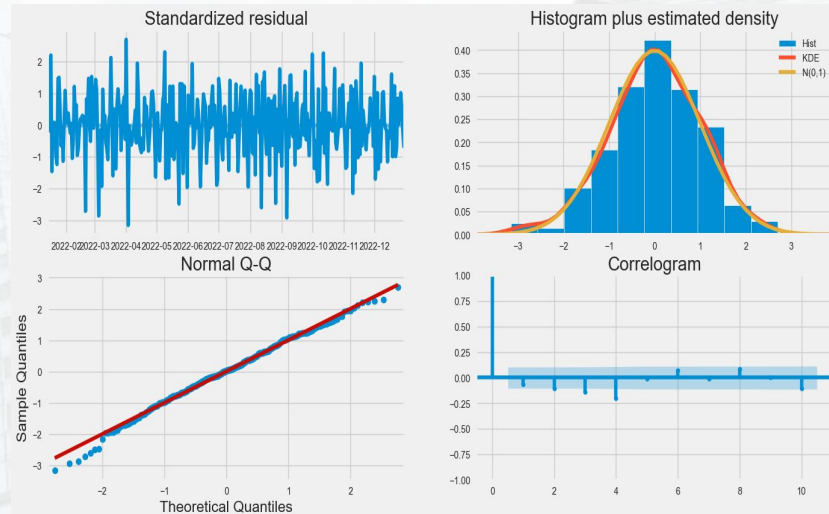
Uji Distribusi Residual

Hasil uji Jarque-Bera (JB) menunjukkan nilai 3.72 dengan Prob(JB) sebesar 0.16, menunjukkan bahwa model memenuhi asumsi tentang distribusi normal dalam residuals.

3

Uji Heteroskedastisitas

Hasil uji Heteroskedastisitas menunjukkan nilai 0.91 dengan Prob(H) sebesar 0.6, menunjukkan bahwa model memenuhi asumsi tentang homoskedastisitas dalam residuals.



Ljung-Box (L1) (Q):	1.34	Jarque-Bera (JB):	3.72
Prob(Q):	0.25	Prob(JB):	0.16
Heteroskedasticity (H):	0.91	Skew:	-0.24
Prob(H) (two-sided):	0.60	Kurtosis:	3.15

Modeling & Evaluation

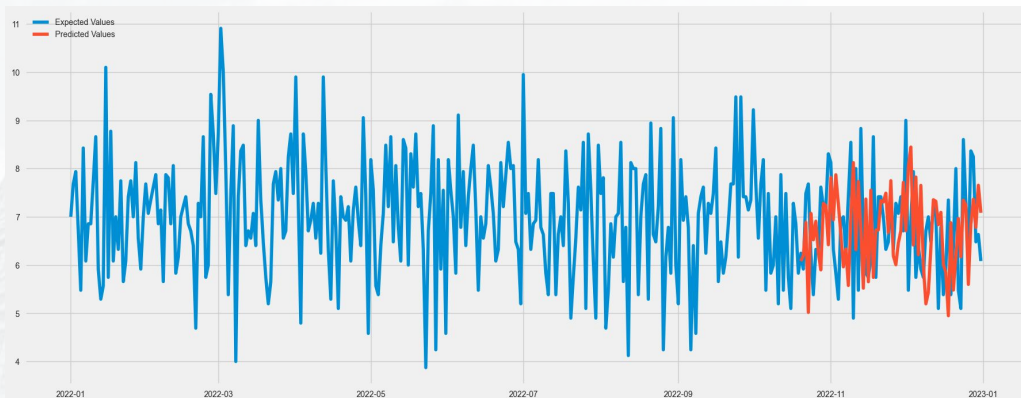
Prediksi Data Test

Memisahkan data time series menjadi dua subset: subset pelatihan (train) dan subset pengujian (test) dengan perbandingan 80% untuk pelatihan dan 20% untuk pengujian. Lalu, melakukan pemodelan sehingga diperoleh berdasarkan visualisasi berikut dengan nilai data asli vs data prediksi model.

Evaluasi Model

Error Metric	MAE	MAPE	MSE	RMSE	MAD
Value	1.219	18.519	2.484	1.576	1.474

Semakin rendah nilai MAE, MSE, RMSE, dan MAD, semakin baik performa model dalam memprediksi data. Sedangkan, nilai MAPE menunjukkan persentase kesalahan rata-rata dari prediksi model terhadap nilai sebenarnya.

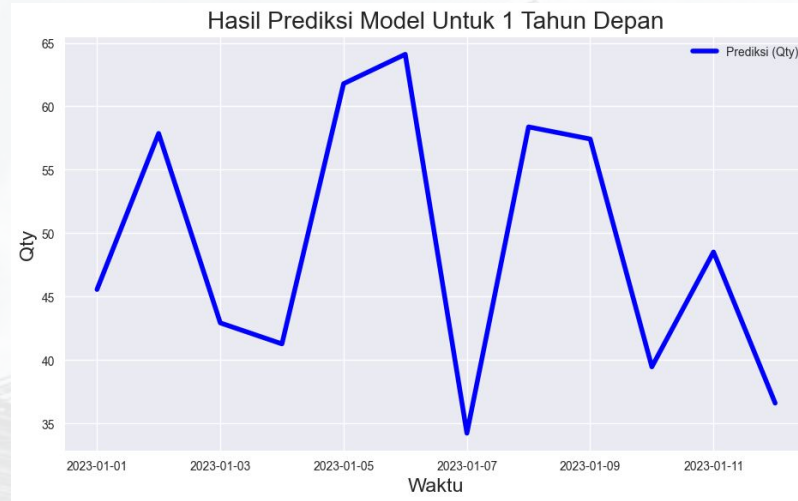


Berdasarkan hasil berikut, variabel-variabel dalam model dinyatakan signifikan secara statistik dengan nilai p-value. Jika nilai p-value < alpha (0.05), kita dapat menolak hipotesis nol dan menyimpulkan bahwa koefisien tersebut signifikan dalam model.

	coef	std err	z	P> z	[0.025	0.975]
intercept	0.0053	0.085	0.062	0.950	-0.162	0.173
ar.L1	-0.7841	0.055	-14.315	0.000	-0.891	-0.677
ar.L2	-0.5360	0.067	-7.970	0.000	-0.668	-0.404
ar.L3	-0.3263	0.058	-5.671	0.000	-0.439	-0.214
ar.S.L12	-0.6192	0.054	-11.378	0.000	-0.726	-0.513
ar.S.L24	-0.3412	0.053	-6.416	0.000	-0.445	-0.237
sigma2	2.2492	0.167	13.458	0.000	1.922	2.577

Modeling & Evaluation

Prediksi Data 1 Tahun Berikutnya

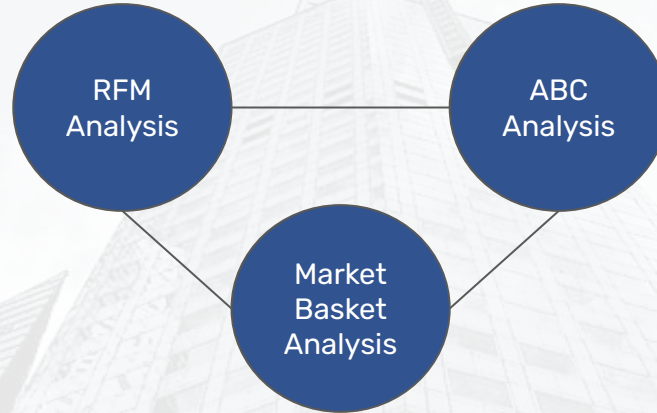


Berdasarkan prediksi model, nilai penjualan produk atau data time series tersebut diperkirakan akan bervariasi dalam rentang yang cukup lebar selama 1 tahun ke depan dari tanggal 01 Januari 2023 hingga 01 Desember 2023. Terdapat fluktuasi nilai dari tanggal ke tanggal, prediksi nilai tertinggi terjadi pada 6 Januari 2023 dengan nilai 64.10, sementara nilai terendah terjadi pada 7 Januari 2023 dengan nilai 34.24.

Conclusion & Recommendations

Strategi Pemasaran

Pada hasil analisis RFM menunjukkan jumlah pelanggan loyal yang rendah, kelompok "Unsatisfied Customers" cukup besar, dan kelompok "Lost Customers - Bargain" memiliki jumlah tertinggi. Untuk meningkatkan retensi pelanggan, perusahaan perlu memberikan diskon, tawaran promosi, meningkatkan pelayanan, dan menawarkan produk dengan harga yang lebih kompetitif."



Hasil analisis ABC menunjukkan Ginger Candy sebagai produk C di Modern Trade dan General Trade, dengan kontribusi pendapatan kecil. Sementara itu, Cheese Stick tergolong dalam kategori produk A di kedua tipe toko, menjanjikan potensi pendapatan. Informasi ini berguna untuk manajemen stok barang dan menargetkan kelompok hasil analisis RFM seperti "Lost Customers - Bargain", "Big Spenders", dan "Occasional Customers - Big Spenders".

Dari analisis market basket, produk Thai Tea mayoritas direkomendasikan meskipun berada dalam kategori B. Ginger Candy memiliki nilai support, confidence, dan lift tertinggi karena sering dibeli oleh pelanggan, meskipun termasuk dalam kategori C. Cheese Stick, yang termasuk dalam kategori A dan sering dibeli, dapat meningkatkan pendapatan. Hasil analisis ini relevan untuk semua kelompok pelanggan, terutama "New Customer", "Big Spenders", dan "Lost Customer".

Conclusion & Recommendations

Rekomendasi Bisnis Segmentasi Pelanggan

Cluster 0

Menunjukkan tingkat transaksi dan pembelian produk (Qty) yang tinggi secara rata-rata. Meskipun total transaksi dan jumlah produk yang dibeli lebih rendah dari cluster lain, rata-rata TotalAmount atau total pendapatan dari cluster ini masih cukup besar. Mayoritas pelanggan dalam cluster ini adalah wanita yang sudah menikah. Fokus bisnis harus pada mempertahankan loyalitas dan meningkatkan interaksi dengan pelanggan wanita yang sudah menikah, mendorong mereka untuk bertransaksi lebih sering guna meningkatkan pendapatan.

Cluster 2

Memiliki rata-rata total pendapatan yang tinggi. Dalam hal jumlah transaksi dan jumlah produk yang dibeli (Qty), cluster ini berada di antara Cluster 0 dan Cluster 1. Bisnis dapat fokus untuk meningkatkan interaksi dengan pelanggan dalam cluster ini dan mendorong mereka untuk meningkatkan frekuensi pembelian.

Cluster 1

Memiliki rata-rata tinggi dalam hal jumlah transaksi dan jumlah produk yang dibeli (Qty). Rata-rata TotalAmount atau total pendapatan dari cluster ini lebih tinggi dibandingkan dengan cluster lainnya. Pelanggan dalam cluster ini mayoritas adalah wanita yang sudah menikah. Bisnis dapat fokus untuk meningkatkan frekuensi pembelian dan mempertahankan loyalitas pelanggan dalam cluster ini.

Cluster 3

Memiliki rata-rata jumlah transaksi, jumlah produk yang dibeli (Qty), dan total pendapatan yang lebih rendah dibandingkan dengan cluster lainnya. Mayoritas pelanggan dalam cluster ini adalah pria yang sudah menikah. Meskipun tingkat total pendapatan relatif lebih rendah, bisnis dapat mencari cara untuk meningkatkan nilai transaksi pelanggan dalam cluster ini dengan strategi seperti penawaran khusus atau produk yang sesuai dengan preferensi mereka.

Conclusion & Recommendations

Prediksi Penjualan Produk (Qty)

Hasil analisis ini berhasil menganalisis faktor-faktor runtun waktu pada penjualan produk melalui asumsi-asumsi yang mendalam dan pemodelan yang optimal.

01



Keberhasilan menganalisis faktor-faktor



02

ARIMA dapat memprediksi jumlah penjualan produk dengan baik

Model ARIMA dengan parameter $ARIMA(3,1,0)(2,1,0)[12]$ dapat secara baik memprediksi penjualan produk dengan nilai AIC rendah dan margin of error yang kecil.

Variabel-variabel dalam model runtun waktu dinyatakan signifikan secara statistik melalui hasil pemodelan runtun waktu, yaitu $ar.L1$, $ar.L2$, $ar.L3$, $ar.S.L12$, $ar.S.L24$, dan σ^2 .

03



Variabel berpengaruh signifikan terhadap jumlah penjualan produk



04

Rekomendasi model ARIMA sebagai alat prediksi

Model Arima direkomendasikan sebagai alat prediksi penjualan produk untuk membuat estimasi jumlah produk yang dibeli oleh pelanggan dengan mempertimbangkan faktor-faktor yang mempengaruhi tersebut.

Link Folder :

[FinalTask_Kalbe_DS_RFirdausDharmawanAkbar - Google Drive](#)

Link Presentasi :

[\(1\) FinalTask_Kalbe_DS_Presentasi_RFirdausDharmawanAkbar - YouTube](#)

Link Dashboard Tableau :

public.tableau.com/app/profile/r.firdaus.dharmawan.akbar/viz/ExploratoryDataAnalysis-Kalbe/Dashboard1

Link Github :

[rfdharma/FinalTask_Kalbe_DS \(github.com\)](https://github.com/rfdharma/FinalTask_Kalbe_DS)

Thank You



Rakamin
Academy



KALBE
Nutritional