



Team Alnnovate

Membangun Strategi Pemasaran yang Tepat untuk Segmen Pengguna Berdasarkan Karakteristik Finansial menggunakan Clustering dengan K-Means

Team



Mentor
Wiranda Aditya Fidin



Fasilitator
Dara Aprita



Agatha Paskaline
Universitas Hasanudin



Erlangga Dwi Atha
Universitas Riau



Mitta Permatasari
Universitas Sriwijaya



Muhammad Ilyas Arsyad
Universitas Bina Nusantara



Nurita Fitriana Luthfa
Universitas Negeri Semarang



Suristhia Rahmaidta
Universitas Sriwijaya

Work Flow



1



**Bussines
Understanding**

2



**Data
Understanding**

3



**Data
Preparation**

4



**Data
Modeling**

5



**Model
Evaluation**

6



**Business
Solution**

Work Flow

Goals

Data Preprocessing

Feature Selection

Business Understanding



Business Understanding

Super App merupakan perusahaan yang mendominasi industri retail di Indonesia dengan berbagai layanan dan fitur yang terintegrasi dalam satu platform. Perusahaan, Super App, aplikasi yang menawarkan berbagai layanan yang beragam dalam satu platform, mulai dari pembelian barang dan jasa, hingga layanan transportasi, keuangan, dan hiburan.

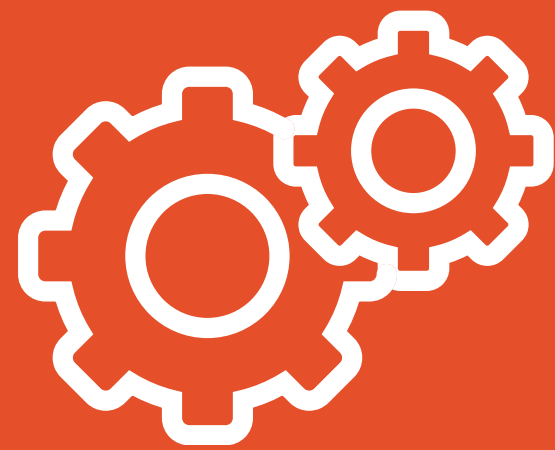
Problem Statement

Super App menghadapi tantangan kompleks dalam mengidentifikasi dan memahami tren serta pola penggunaan aplikasinya yang berbeda-beda, sehingga perlu memahami perilaku dan karakteristik pengguna untuk membantu perusahaan menyusun strategi pemasaran yang lebih tepat sasaran.



Untuk menyusun strategi pemasaran yang tepat bagi **Super App**, **tim kami** melakukan analisis segmentasi pengguna berdasarkan karakteristik finansial mereka.

Focus Target



**Mengidentifikasi
Segmen Pengguna**



**Menganalisis
Karakteristik Setiap
Segmen**



**Memberikan Strategi
Pemasaran yang
Efektif**

Data Understanding



Data Describe

Data User	Berisi informasi terperinci mengenai pengguna, meliputi data demografis dan riwayat transaksi yang membantu memahami perilaku pengguna.
TRX	Catatan historis transaksi yang mendetail memungkinkan analisis tren pembelian dan pengaruh promosi.
Location Reference	Menyediakan rincian lokasi geografis yang penting untuk analisis regional.
Product Reference	Menawarkan gambaran lengkap tentang jajaran produk yang vital untuk menghubungkan preferensi regional dengan tawaran spesifik.

18

Total Unique
Products yang
ditawarkan

143M

Total income selama periode 01
Januari 2024 sampai 31
Desember 2024

81.6K

Total pelanggan
melakukan transaksi pada
Super App

Data Understanding

2.5JT



74.5%

25.5%



Total transaksi pelanggan
selama periode 01 Januari 2024
sampai 31 Desember 2024

Transaksi pada Super App didominasi oleh
pelanggan dengan jenis kelamin perempuan,
dengan perbandingan sebesar 3/4

Data Preparation



Data Preparation

Missing Values

- Drop missing values 7% pengguna yang tidak memiliki transaksi
- Imputasi nilai 0 pada kolom discounts

Outlier

- Mengganti data dengan nominal kuadriliun pada kolom money_spent, refund, wallet_balance menjadi nominal jutaan.
- Drop minus values pada kolom money_spent, refund, wallet_balance
- Handling outlier values (tahun kelahiran di atas 2006) pada kolom birth dengan median, mean, dan manual.

Duplicated Values

- 100% unique rows

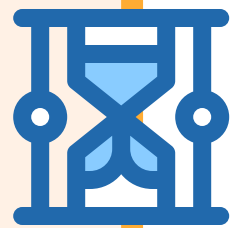
Correcting Data Values

- Kolom join_date, birth, dan transaction_date yang bertipe data object, diubah menjadi tipe data datetime

Convert Data Types

- Memperbaiki nilai pada kolom money_spent dengan sum gross_amount-discounts

Feature Engineering



Age

Dibuat kolom usia pengguna dari kolom birth agar lebih mudah diinterpretasikan dibandingkan dengan tanggal lahir.



Recency

Membuat kolom rfm_recency untuk menunjukkan berapa lama waktu yang sudah berlalu sejak pengguna terakhir melakukan transaksi.



Refund Ratio

Membuat kolom refund_ratio untuk menunjukkan berapa persentase uang yang telah dikembalikan kepada pelanggan dalam bentuk refund.



Frequency

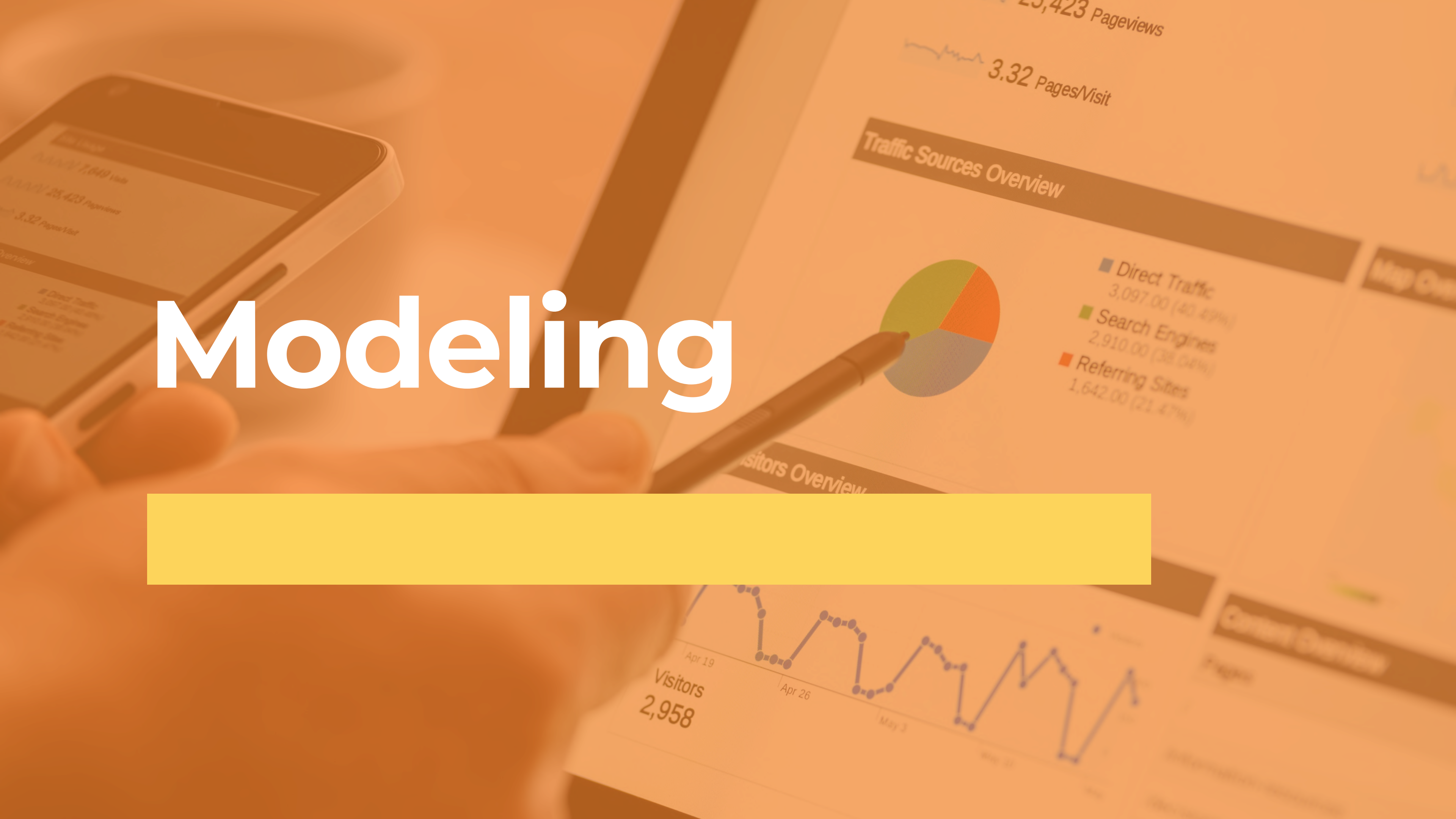
Membuat kolom rfm_frequency untuk menunjukkan seberapa sering pelanggan melakukan transaksi.



Monetary Value

Membuat kolom rfm_monetary_value untuk menunjukkan berapa banyak uang yang telah dibelanjakan oleh pengguna.

Modeling



Relevant Feature for Cluster

- Money Spent
- Balance
- Total Spent
- Gross Amount
- Discounts
- RFM Recency
- RFM Frequency
- Refund
- Refund Ratio

Pemilihan fitur ini penting untuk meningkatkan akurasi pada saat clustering. Fitur yang dipilih berhubungan dengan karakteristik finansial pengguna.

Normalization & Invert Feature

Normalisasi fitur money_spent, balance, total_spent, gross_amount, rfm_recency, 'discounts, refund, refund_ratio menggunakan Min Max Scaler agar nilai pada kolom-kolom tersebut berada dalam rentang yang sama. Setelah normalisasi dilakukan **invert** untuk fitur refund, rfm_recency, dan discounts.

	money_spent	balance	total_spent	gross_amount	rfm_recency	discounts	refund	refund_ratio
0	0.563758	0.089947	0.464646	0.051847	0.901408	0.853686	0.763359	0.017280
1	0.724832	0.108466	0.585859	0.064266	0.515493	0.845549	0.671756	0.019010
2	0.546980	0.063492	0.481481	0.050605	0.971831	0.927821	0.839695	0.011296
3	0.674497	0.097884	0.569024	0.061472	0.816901	0.828474	0.748092	0.015020
4	0.697987	0.103175	0.572391	0.063024	0.946479	0.845445	0.702290	0.017647



Model Clustering

K-Means

K-Means Clustering adalah algoritma yang mengelompokkan data ke dalam K cluster berdasarkan jarak terdekat ke titik pusat (centroid) masing-masing cluster.

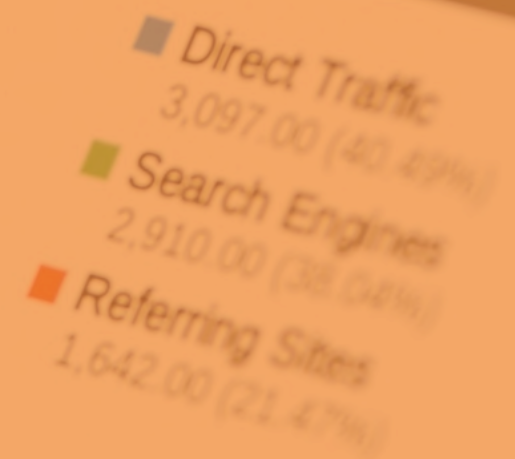
DBScan

DBSCAN mengelompokkan data berdasarkan kepadatan titik-titik data, mengidentifikasi cluster sebagai area dengan kepadatan tinggi dan noise sebagai titik data yang terisolasi.

Dari beberapa model clustering yang ada, tim kami mencoba K-Means dan DBScan.

Model Evaluation

Traffic Sources Overview



Visitors Overview





Metric Evaluation

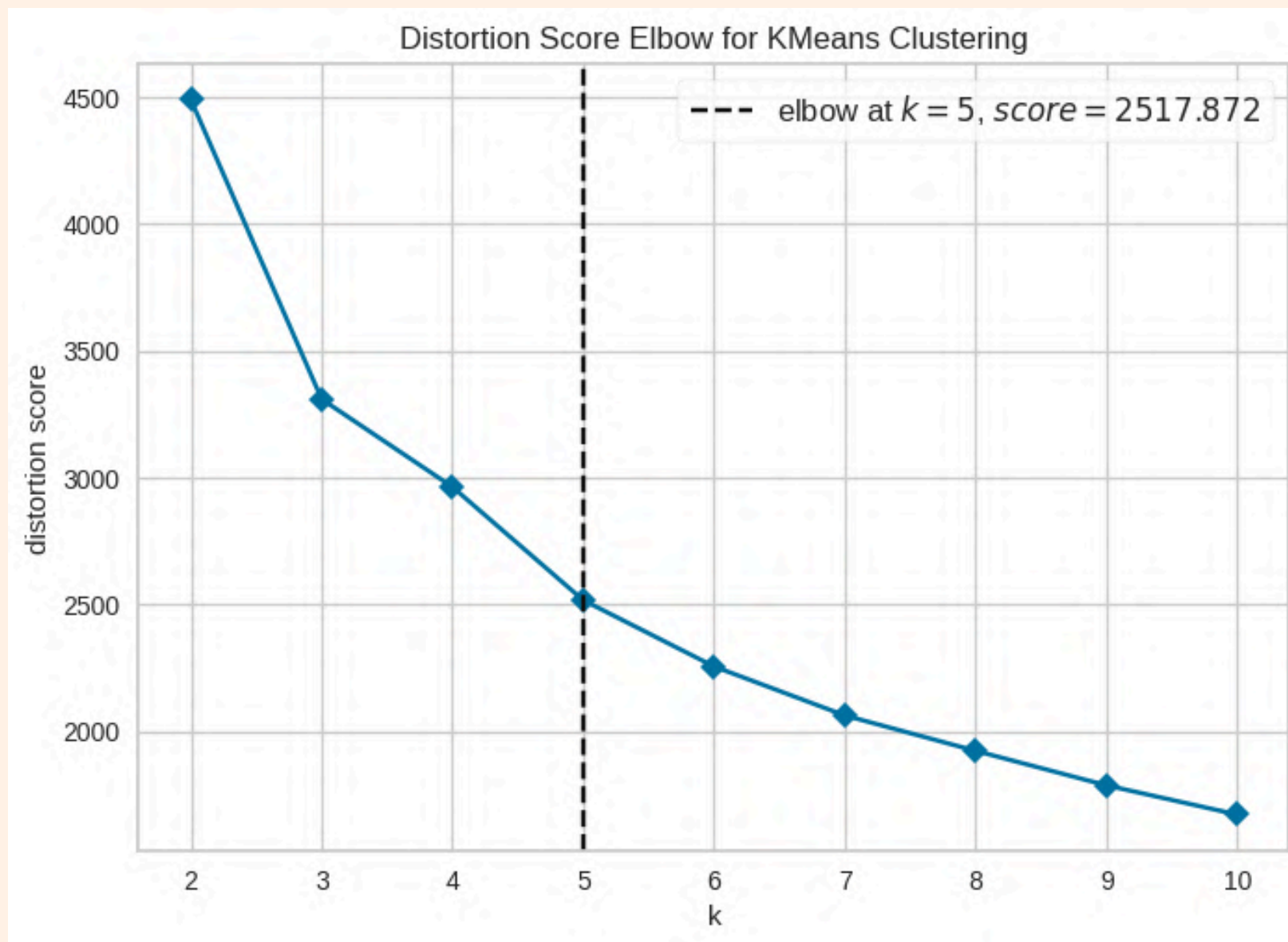
Perbandingan nilai metric evaluation antara K-Means dan DBScan

Algoritma	Silhouette	Calinski-Harabasz	Davies-Bouldin
K-Means	0.2721	62954.2993	1.2126
DBScan	0.7074	17.2518	1.8030

Secara umum, clustering yang baik ditunjukkan dengan nilai Silhouette Score mendekati 1, Calinski-Harabasz Index tinggi, dan Davies-Bouldin Index rendah, namun perlu mempertimbangkan konteks dan visualisasi untuk interpretasi yang lebih komprehensif.

Meskipun DBSCAN menunjukkan skor Silhouette yang lebih tinggi dibandingkan K-Means, analisis menyeluruh terhadap metrik lainnya dan pola cluster yang terbentuk menunjukkan bahwa K-Means merupakan model yang terbaik.

Elbow Method



Walaupun hasil dari elbow method menunjukkan 5, kami lebih memilih 4 klaster karena ini memungkinkan hasil analisis yang lebih mudah diinterpretasikan dan diimplementasikan dalam strategi bisnis. Dengan lebih sedikit klaster yang perlu dikelola, fokus pada perbedaan utama antar segmen pengguna menjadi lebih jelas, serta efisiensi dalam pengembangan strategi dan operasional dapat tercapai.

Cluster Results



**Cluster 1: High
Spenders with
High Refunds**



**Cluster 3: Higher
Spending and
Discounts**



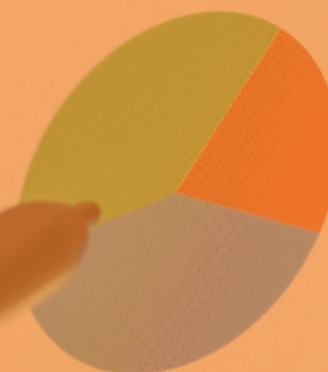
**Cluster 2:
Consistent High
Spenders**



**Cluster 4: Balanced
Spenders**

Business Solution

Traffic Sources Overview



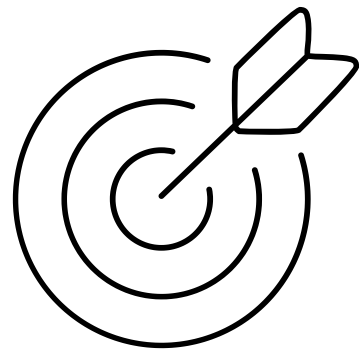
Direct Traffic	3,097.00 (40.49%)
Search Engines	2,910.00 (38.04%)
Referring Sites	1,642.00 (21.47%)

Visitors Overview



Cluster 1: High Spenders with High Refunds

Pengguna dalam cluster ini cenderung memiliki pengeluaran tinggi dan sering melakukan refund. Ini mungkin menandakan bahwa mereka sangat memperhatikan kualitas produk atau layanan, dan akan melakukan refund jika tidak memenuhi standar mereka. Selain itu, kemungkinan besar mereka pernah mengalami pengalaman buruk dengan layanan super app sebelumnya, sehingga lebih mudah bagi mereka untuk melakukan refund. Mereka juga terlihat sebagai pencari diskon yang aktif, namun tidak segan untuk meminta pengembalian dana jika produk atau layanan yang mereka dapatkan tidak sesuai dengan harapan.

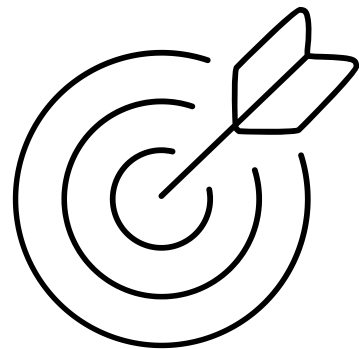


Strategi:

Perusahaan dapat meningkatkan pengalaman pelanggan dengan fokus pada peningkatan kualitas produk dan layanan serta menyederhanakan proses refund. Selain itu, menawarkan program loyalitas khusus bagi pengguna yang sering melakukan refund dan memberikan garansi pada produk tertentu juga dapat membantu membangun kepercayaan dan meningkatkan loyalitas pelanggan.

Cluster 2: Consistent High Spenders

Pengguna dalam cluster ini dikenal sebagai pengeluar konsisten yang tinggi namun jarang meminta pengembalian dana. Ini menandakan kesetiaan dan kepuasan mereka terhadap layanan super app. Mereka mungkin telah memiliki pengalaman positif yang kuat dengan platform tersebut, yang mengurangi keinginan mereka untuk melakukan refund. Selain itu, kemungkinan besar mereka memiliki pemahaman yang baik tentang produk dan layanan yang mereka gunakan, sehingga kesalahan yang memerlukan pengembalian dana menjadi jarang terjadi.

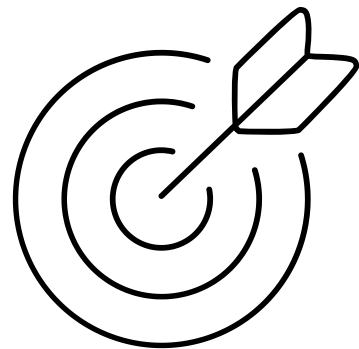
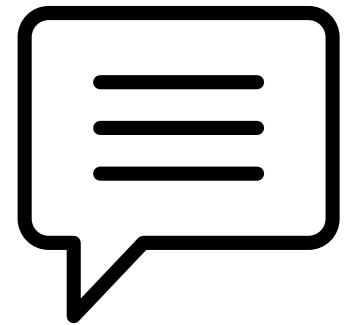


Strategi:

Perusahaan dapat mempertahankan loyalitas pengguna dengan program loyalitas eksklusif, promo personal yang sesuai preferensi, komunikasi melalui email dan push notification, serta layanan pelanggan premium dengan respon cepat dan solusi efektif.

Cluster 3: Higher Spending and Discounts

Pengguna dalam cluster ini dikenali sebagai individu dengan pengeluaran tinggi yang secara aktif memanfaatkan diskon. Ini menunjukkan bahwa mereka adalah pencari promo yang sangat aktif dan responsif terhadap penawaran diskon. Kemungkinan besar, mereka memiliki pengalaman positif dengan penggunaan diskon dan promo di layanan super app, yang mendorong mereka untuk terus memanfaatkannya.

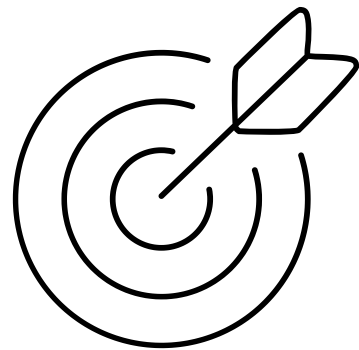


Strategi:

Untuk mempertahankan dan meningkatkan loyalitas pengguna, perusahaan dapat meningkatkan frekuensi promo, menyesuaikan penawaran dengan preferensi pengguna, memperkenalkan program VIP dengan keuntungan eksklusif, dan memperkuat brand loyalty melalui pengalaman belanja yang positif dan memorable.

Cluster 4: Balanced Spenders

Pengguna dalam cluster ini memiliki pengeluaran tinggi namun seimbang antara pengeluaran dan pengembalian dana. Ini menunjukkan bahwa mereka mungkin merupakan pengguna yang stabil dengan pola pengeluaran yang konsisten. Kemungkinan besar, mereka puas dengan layanan super app dan telah memiliki pengalaman yang positif dengan produk dan layanan yang ditawarkan. Mereka mungkin menggunakan layanan super app dengan terkontrol, tidak terlalu impulsif dalam pembelian mereka.



Strategi:

Perusahaan perlu mengadopsi beberapa strategi penting: memberikan layanan konsisten dengan kualitas tinggi, membangun kepercayaan melalui informasi yang transparan, menawarkan program reward yang menarik, dan memperkuat brand image dengan pengalaman positif. Strategi-strategi ini akan memperkuat hubungan dengan pengguna dan meningkatkan penggunaan layanan super app.

Thank You

By : TEAM AINNOVATE