

**SEGMENTASI PELANGGAN TRANSJAKARTA
MENGUNAKAN METODE *CLUSTERING* DENGAN
MODEL RFM (*RECENCY, FREQUENCY, MONETARY*)
DAN ALGORITMA *K-MEANS***

SKRIPSI SARJANA SISTEM INFORMASI

Oleh :

Trie Widiarti Ningsih

207006516033



**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNOLOGI KOMUNIKASI DAN
INFORMATIKA
UNIVERSITAS NASIONAL**

2024

**SEGMENTASI PELANGGAN TRANSJAKARTA
MENGUNAKAN METODE *CLUSTERING* DENGAN
MODEL RFM (*RECENCY, FREQUENCY, MONETARY*)
DAN ALGORITMA *K-MEANS***

Karya ilmiah sebagai salah satu syarat untuk memperoleh gelar
Sarjana Sistem Informasi dari Fakultas Teknologi Komunikasi dan Informatika

Oleh

Trie Widiarti Ningsih

207006516033



**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNOLOGI KOMUNIKASI DAN
INFORMATIKA
UNIVERSITAS NASIONAL**

2024

HALAMAN PENGESAHAN

HALAMAN PENGESAHAN

TUGAS AKHIR

SEGMENTASI PELANGGAN TRANSJAKARTA MENGGUNAKAN METODE
CLUSTERING DENGAN MODEL RFM (*REGENCY, FREQUENCY, MONETARY*)
DAN ALGORITMA K-MEANS



Trie Widiarti Ningsih

207006516033

Dosen Pembimbing 1

Sari Ningsih, S.Si., M.M.

NIDN. 0302066701

HALAMAN KEASLIAN TUGAS AKHIR

PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa Tugas Akhir dengan judul :

**SEGMENTASI PELANGGAN TRANSJAKARTA MENGGUNAKAN METODE
CLUSTERING DENGAN MODEL RFM (*RECENCY, FREQUENCY, MONETARY*)
DAN ALGORITMA K-MEANS**

Yang dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional, sebagaimana yang saya ketahui adalah bukan merupakan tiruan atau publikasi dari Tugas Akhir yang pernah diajukan atau dipakai untuk mendapatkan gelar di lingkungan Universitas Nasional maupun perguruan tinggi atau instansi lainnya, kecuali pada bagian – bagian tertentu yang menjadi sumber Informasi atau acuan yang dicantumkan sebagaimana mestinya.

Jakarta, 26 Agustus 2024



Trie Widiarti Ningsih

NPM : 207006516033

LEMBAR PERSETUJUAN TUGAS AKHIR

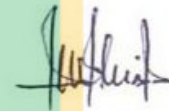
LEMBAR PERSETUJUAN TUGAS AKHIR

Tugas Akhir dengan judul :

SEGMENTASI PELANGGAN TRANSJAKARTA MENGGUNAKAN METODE *CLUSTERING* DENGAN MODEL RFM (*RECENCY, FREQUENCY, MONETARY*) DAN ALGORITMA K-MEANS

Dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional. Tugas Akhir ini diujikan pada Sidang Akhir Semester Genap 2023-2024 pada tanggal 22 Agustus 2024

Dosen Pembimbing 1



Sari Ningsih, S.Si., M.M.

NIDN. 0302066701

Ketua Program Studi



Dr. Andrianingsih, S. Kom., MMSI.

NIDN. 0303097902

HALAMAN PENGESAHAN
TUGAS SARJANA
SEGMENTASI PELANGGAN TRANSJAKARTA
MENGGUNAKAN METODE *CLUSTERING* DENGAN MODEL
RFM (*RECENCY, FREQUENCY, MONETARY*) DAN
ALGORITMA *K-MEANS*



Penguji 1

Dr. Septi Andryana, S.Kom, MMSI
NIDN. 0317097209

Penguji 2

Agus Iskandar, S.Kom, M.Kom
NIDN. 0310087503

LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI

Nama : Trie Widiarti Ningsih
NPM : 207006516033
Fakultas/Akademi : Fakultas Teknologi Komunikasi dan Informatika
Program Studi : Sistem Informasi
Tanggal Sidang : 22 Agustus 2024

JUDUL DALAM BAHASA INDONESIA :

SEGMENTASI PELANGGAN TRANSJAKARTA MENGGUNAKAN
METODE *CLUSTERING* DENGAN MODEL RFM (*RECENCY*,
FREQUENCY, *MONETARY*) DAN ALGORITMA K-MEANS

JUDUL DALAM BAHASA INGGRIS :

CUSTOMER SEGMENTATION OF TRANSJAKARTA USING
CLUSTERING METHOD WITH RFM (RECENCY, FREQUENCY,
MONETARY) MODEL AND K-MEANS ALGORITHM

TANDA TANGAN DAN TANGGAL		
Pembimbing 1	Ka. Prodi	Mahasiswa
TGL : 26-08-2024	TGL : 26-08-2024	TGL : 26-08-2024
 Sari Ningsih, SS, MM		

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kepada Allah SWT atas segala rahmat, nikmat, barakah, dan kemudahan yang telah di berikan, sehinggann penulis dapat menyelesaikan penyusunan skripsi yang berjudul : **SEGMENTASI PELANGGAN TRANSJAKARTA MENGGUNAKAN METODE CLUSTERING DENGAN MODEL RFM (RECENCY, FREQUENCY, MONETARY) DAN ALGORITMA K-MEANS**, skripsi ini merupakan salah satu syarat kelulusan untuk Program Studi Sistem Informasi di Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional.

Penyusunan skripsi ini tidak akan terlaksana tanpa dukungan dan bantuan dari berbagai pihak yang telah mengorbankan waktu, tenaga, pemikiran dan sumber daya. Oleh karena itu, penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Allah SWT yang telah melimpahkan rahmat dan barakah-Nya sehingga skripsi ini dapat diselesaikan.
2. Kedua orang tua tercinta yang selalu memberikan dukungan materi dan doa yang tiada henti kepada penulis.
3. Kakak-kakak tercinta, Atika, Aisyah, Laras dan Muadz, yang telah memberikan dukungan dan dorongan yang sangat berarti selama proses penyusunan skripsi ini.
4. Ibu Sari Ningsih, S.Si., M.M., dosen pembimbing di prodi Sistem Informasi Universitas Nasional.
5. Seluruh dosen pengajar Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional, terimakasih atas segala pengetahuan dan wawasan yang telah diberikan selama masa perkuliahan.
6. Yusri, Adit dan Raka, sahabat sejak awal perkuliahan yang senantiasa memberikan semangat, motivasi, serta masukan dan saran berharga yang sangat membantu penulis dalam menyelesaikan skripsi ini.

7. Afifah dan Octa, sahabat yang selalu memberikan dukungan semangat, motivasi, dan mendengarkan segala keluh kesah penulis, sehingga proses penyusunan skripsi ini dapat berjalan lancar.
8. Diri sendiri, atas kerja keras dan keteguhan dalam menghadapi proses penyusunan skripsi ini, yang memungkinkan penulis menyelesaikan tugas ini dengan sebaik mungkin.

Penulis menyadari bahwa skripsi ini masih memiliki kekurangan dan tidak sempurna. Oleh karena itu, penulis sangat mengharapkan masukan, saran, dan kritik yang membangun untuk perbaikan di masa depan. Penulis juga berharap skripsi ini dapat memberikan manfaat bagi semua pihak yang terlibat.



Depok, 13 Agustus 2024

Trie Widiarti Ningsih

Trie Widiarti Ningsih

ABSTRAK

Penelitian ini bertujuan untuk melakukan segmentasi pelanggan Transjakarta menggunakan metode *clustering* dengan model RFM (*Recency, Frequency, Monetary*) dan algoritma *K-Means*. Sebagai kota metropolitan yang padat penduduk, Jakarta menghadapi tantangan besar dalam menyediakan layanan transportasi umum yang efektif. Dalam konteks ini, Transjakarta sebagai angkutan massal berbasis sistem transit menjadi solusi utama untuk memenuhi kebutuhan mobilitas penduduk Jakarta. Pemahaman yang mendalam terhadap perilaku pengguna Transjakarta menjadi hal krusial dalam meningkatkan kualitas pelayanan dan efisiensi operasional. Segmentasi pelanggan, sebagai pendekatan strategis memungkinkan pihak manajemen untuk mengidentifikasi kelompok pelanggan dengan karakteristik serupa, memahami kebutuhan, dan merancang strategi pelayanan yang lebih terfokus. Metode yang diterapkan dalam penelitian ini melibatkan model RFM, yang mempertimbangkan faktor waktu dan nilai transaksi, serta algoritma *K-Means*, yang membantu pembagian pelanggan ke dalam kelompok-kelompok dengan kemiripan tertentu. Penentuan jumlah *cluster* optimal dilakukan menggunakan metode *Elbow*, *silhouette*, dan *Davies-Bouldin Index*. Hasil penelitian menunjukkan skor *Silhouette* untuk model dengan tujuh cluster adalah 0.789, menunjukkan bahwa cluster tersebut memiliki pemisahan yang baik dan kepadatan yang memadai. Nilai *Davies-Bouldin Index* untuk tujuh cluster adalah 0.303, menunjukkan bahwa model dengan jumlah cluster ini memiliki rasio jarak antar cluster dan jarak dalam cluster yang optimal. Hasil analisis mengidentifikasi tujuh segmen pelanggan, yaitu Aktif, Rutin, Jarang, Sering Terjadi, Terbatas, Hemat, dan Minim. Setiap segmen memiliki karakteristik unik dalam frekuensi penggunaan, nilai transaksi, dan waktu terakhir transaksi, serta rekomendasi strategi pemasaran yang spesifik untuk meningkatkan kepuasan pelanggan dan efisiensi operasional Transjakarta. Misalnya, segmen Aktif disarankan untuk diberi program loyalitas dan promosi digital, sementara segmen Jarang dapat diaktifkan kembali melalui kampanye kesadaran dan penawaran khusus. Dengan merinci kelompok-kelompok pelanggan berdasarkan model RFM dan algoritma *K-Means*, Transjakarta dapat meningkatkan kepuasan pelanggan, mengoptimalkan strategi pelayanan, dan menciptakan sistem transportasi umum yang lebih adaptif serta berkelanjutan.

Kata Kunci: Algoritma *K-Means*, *Clustering*, *Davies-Bouldin Index*, Metode *Elbow*, Model RFM, Segmentasi Pelanggan, *Silhouette*, Transjakarta.

ABSTRACT

This research aims to conduct customer segmentation for Transjakarta using clustering methods with the RFM (Recency, Frequency, Monetary) model and the K-Means algorithm. As a densely populated metropolitan city, Jakarta faces significant challenges in providing effective public transportation services. In this context, Transjakarta, as a mass transit system, serves as the primary solution to meet the mobility needs of the capital city's residents. A profound understanding of Transjakarta user behavior is crucial for enhancing service quality and operational efficiency. Customer segmentation, as a strategic approach, enables management to identify customer groups with similar characteristics, understand their needs, and design more focused service strategies. The methods employed in this research involve the RFM model, which considers time and transaction value factors, and the K-Means algorithm, which facilitates the classification of customers into groups with similar characteristics. The optimal number of clusters was determined using the Elbow method, silhouette score, and Davies-Bouldin Index. The research's results showed that the Silhouette score for the model with seven clusters was 0.789, indicating that these clusters have good separation and adequate density. The Davies-Bouldin Index value for the seven clusters was 0.303, suggesting that the model with this number of clusters has an optimal ratio of inter-cluster distance to intra-cluster distance. The analysis identified seven customer segments: Active, Routine, Rare, Frequent, Limited, Frugal, and Minimal. Each segment has unique characteristics in terms of usage frequency, transaction value, and recency of transactions, along with specific marketing strategies to enhance customer satisfaction and operational efficiency. For instance, the Active segment is recommended to be targeted with loyalty programs and digital promotions, while the Rare segment could be reactivated through awareness campaigns and special offers. By detailing customer groups based on the RFM model and K-Means algorithm, Transjakarta can improve customer satisfaction, optimize service strategies, and create a more adaptive and sustainable public transportation system.

Keyword: Clustering, Customer Segmentation, Davies-Bouldin Index, Elbow Method, K-Means Algorithm, RFM Model, Silhouette, Transjakarta.

DAFTAR ISI

HALAMAN PENGESAHAN.....	ii
KATA PENGANTAR	vii
ABSTRAK	ix
ABSTRACT	x
DAFTAR ISI.....	xi
DAFTAR GAMBAR	xiii
DAFTAR TABEL.....	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	3
1.3 Tujuan Penelitian.....	4
1.4 Batasan Masalah.....	4
1.5 Manfaat Penelitian.....	5
1.6 Sistematika Penulisan.....	5
BAB II TINJAUAN PUSTAKA.....	6
2.1 Penelitian Terdahulu.....	6
2.2 Gap Analisis	11
2.3 Studi Literatur.....	11
2.4 Landasan Teori	12
2.3.1 Segmentasi Pelanggan	12
2.3.2 Data Mining	12
2.3.3 <i>Clustering</i>	12
2.3.4 RFM Model.....	13
2.3.5 Algoritma <i>K-Means</i>	13
2.3.6 Metode <i>Elbow</i>	14
2.3.7 <i>Davies-Bouldin Index</i>	14
2.3.8 <i>Silhouette</i>	14
2.3.9 <i>Google Colab</i>	15
2.3.10 <i>Rapid Miner</i>	15
2.3.11 <i>Streamlit</i>	15
BAB III METODOLOGI PENELITIAN.....	16
3.1 Kerangka Pemikiran	16
3.2 Tahapan Penelitian	17
3.2.1 <i>View Data</i>	18
3.2.2 <i>Data Understanding</i>	19
3.2.3 <i>Data Preparation</i>	19
3.2.3.1 <i>Data Cleaning</i>	19
3.2.3.2 <i>Data Selection</i>	19
3.2.3.3 <i>Data Transformation</i>	20
3.2.4 <i>Modeling</i>	21
3.2.4.1 <i>RFM Model</i>	21
3.2.4.2 <i>K-Means Clustering</i>	22

3.2.5 <i>Evaluation</i>	27
3.2.6 <i>Deployment</i>	27
BAB IV HASIL DAN PEMBAHASAN	28
4.1 Implementasi Tahapan Penelitian	28
4.1.1 <i>View Data</i>	28
4.1.2 <i>Data Understanding</i>	29
4.1.3 <i>Data Preparation</i>	29
4.1.3.1 <i>Data Cleaning</i>	29
4.1.3.2 <i>Data Selection</i>	30
4.1.3.3 <i>Data Transformation</i>	35
4.1.4 <i>Modeling</i>	36
4.1.4.1 <i>RFM Model</i>	36
4.1.4.2 <i>K-Means Clustering</i>	39
4.1.5 <i>Evaluation</i>	42
4.1.6 <i>Deployment</i>	43
4.2 Perhitungan Manual	49
4.3 Pengujian <i>Clustering K-Means</i>	52
4.4 Analisa Hasil	53
4.4.1 Perbandingan Hasil <i>Clustering</i>	53
4.4.2 Visualisasi	54
4.4.3 Karakteristik RFM	56
4.4.4 Strategi	59
BAB V KESIMPULAN DAN SARAN	65
5.1 Kesimpulan	65
5.2 Saran	67
DAFTAR PUSTAKA	xv
LAMPIRAN	xviii



DAFTAR GAMBAR

Gambar 3. 1 Kerangka Penelitian	17
Gambar 3. 2 Tahapan Penelitian	18
Gambar 4.1 1 Kaggle data Transjakarta.....	28
Gambar 4.1 2 Dataset Transaksi Transjakarta	29
Gambar 4.1 3 <i>Elbow Method</i>	40
Gambar 4.1 4 <i>Silhouette</i>	41
Gambar 4.1 5 <i>Davies-Bouldin Index</i>	42
Gambar 4.1 6 <i>Home</i>	43
Gambar 4.1 7 <i>Data Overview 1</i>	44
Gambar 4.1 8 <i>Data Overview 2</i>	44
Gambar 4.1 9 <i>RFM Analyst 1</i>	45
Gambar 4.1 10 <i>RFM Analyst 2</i>	45
Gambar 4.1 11 <i>RFM Analyst 3</i>	46
Gambar 4.1 12 <i>Clustering 1</i>	46
Gambar 4.1 13 <i>Clustering 2</i>	47
Gambar 4.1 14 <i>Clustering 3</i>	47
Gambar 4.1 15 <i>Strategy 1</i>	48
Gambar 4.1 16 <i>Strategy 2</i>	48
Gambar 4.1 17 <i>Strategy 3</i>	49
Gambar 4.3 1 Hasil <i>Clustering RapidMiner</i>	52
Gambar 4.4 1 Visualisasi Distribusi <i>Recency</i>	54
Gambar 4.4 2 Visualisasi Distribusi <i>Frequency</i>	55
Gambar 4.4 3 Visualisasi Distribusi <i>Monetary</i>	56
Gambar 4.4 4 Jumlah Anggota.....	60
Gambar 4.4 5 Rata-rata Umur.....	60
Gambar 4.4 6 Jenis Kelamin.....	61

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	6
Tabel 3.2 1 Atribut data Transjakarta	19
Tabel 3.2 2 Atribut data RFM	21
Tabel 3.2 3 Atribut <i>K-Means Clustering</i>	22
Tabel 3.2 4 Data <i>Centroid</i> 1	24
Tabel 3.2 5 Hasil <i>Iteration</i> 1	26
Tabel 3.2 6 Hasil Jumlah <i>Cluster</i>	26
Tabel 4.1 1 Tipe data <i>dataset</i>	29
Tabel 4.1 2 <i>Missing Value</i>	29
Tabel 4.1 3 Setelah <i>Missing Value</i>	30
Tabel 4.1 4 Tabel Pelanggan	31
Tabel 4.1 5 Tabel Corridor	32
Tabel 4.1 6 Tabel StopsIn	32
Tabel 4.1 7 Tabel StopsOut	33
Tabel 4.1 8 Tabel Transactions	34
Tabel 4.1 9 Tabel setelah Normalisasi data	35
Tabel 4.1 10 <i>Data Transformation</i>	36
Tabel 4.1 11 <i>Recency</i>	37
Tabel 4.1 12 <i>Frequency</i>	38
Tabel 4.1 13 <i>Monetary</i>	39
Tabel 4.1 14 <i>RFM</i> Model	39
Tabel 4.1 15 <i>RFM</i> Standarisasi	40
Tabel 4.1 16 <i>Silhouette Score</i>	41
Tabel 4.1 17 <i>Davies-bouldin Score</i>	42
Tabel 4.2 1 Data <i>Centorid</i> 2	49
Tabel 4.2 2 Hasil <i>Iteration</i> 2	51
Tabel 4.2 3 Hasil Jumlah Anggota <i>Cluster</i>	51
Tabel 4.3 1 Jumlah Anggota <i>Cluster RapidMiner</i>	52
Tabel 4.3 2 Hasil <i>Performance Metric Clustering</i>	53
Tabel 4.4 1 Perbandingan Hasil <i>Clustering</i>	54
Tabel 4.4 2 Rata-rata Pola Perjalanan Karakteristik <i>RFM</i>	57
Tabel 4.4 3 Keterangan Karakteristik	58
Tabel 4.4 4 Rekomendasi Strategi	62