

**PERBANDINGAN KINERJA ALGORITMA K-MEANS DAN
AGGLOMERATIVE CLUSTERING UNTUK SEGMENTASI
PENJUALAN ONLINE PADA CUSTOMER RETAIL**

**SKRIPSI NON REGULER
SKEMA SCIENTIST PUBLIKASI**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh
GHANIM RAMADHAN
20.11.3412

Kepada
**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2024**

**PERBANDINGAN KINERJA ALGORITMA K-MEANS DAN
AGGLOMERATIVE CLUSTERING UNTUK SEGMENTASI
PENJUALAN ONLINE PADA CUSTOMER RETAIL**

**SKRIPSI NON REGULER
SKEMA SCIENTIST PUBLIKASI**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

GHANIM RAMADHAN

20.11.3412

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

HALAMAN PERSETUJUAN

NASKAH PUBLIKASI

Perbandingan Kinerja Algoritma K-means dan Agglomerative Clustering Untuk Segmentasi Penjualan Online Pada Customer Retail

yang disusun dan diajukan oleh

Ghanim Ramadhan

20.11.3412

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 27 Juni 2024

Dosen Pembimbing,



Yuli Astuti, M.Kom

NIK. 190302146

HALAMAN PENGESAHAN

NASKAH PUBLIKASI

Perbandingan Kinerja Algoritma K-means dan Agglomerative Clustering Untuk Segmentasi Penjualan Online Pada Customer Retail

yang disusun dan diajukan oleh

Ghanim Ramadhan

20.11.3412

Telah dipertabakan di depan Dewan Penguji pada tanggal 27 Juni 2024

Susunan Dewan Penguji

Nama Penguji

Nur'aini, M.Kom
NIK. 190302066

Acihmah Sidauruk, M.Kom
NIK. 190302238

Yuli Astuti, M.Kom
NIK. 190302146

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan untuk memperoleh gelar Sarjana Komputer Tanggal 27 Juni 2024

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom., Ph.D.
NIK. 190302096

HALAMAN PERNYATAAN KONTRIBUSI KARYA

Yang bertandatangan di bawah ini,

Nama mahasiswa : Ghanim Ramadhan
NIM : 20.11.3412

Menyatakan bahwa Skripsi dengan judul berikut:

Perbandingan Kinerja Algoritma K-means dan Agglomerative Clustering Untuk Segmentasi Penjualan Online Pada Customer Retail

Dosen Pembimbing : Yuli Astuti, M.Kom

1. Karya tulis ini adalah benar-benar **ASLI** dan **BELUM PERNAH** diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian **SAYA** sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab **SAYA**, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini **SAYA** buat dengan **sesungguhnya**, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka **SAYA** bersedia menerima **SANKSI AKADEMIK** dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 27 Juni 2024

Yang Menyatakan,



Ghanim Ramadhan

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa. Atas rahmat dan karunia-Nya, penulis dapat menyelesaikan Skripsi Non-Reguler dengan judul “Perbandingan Kinerja Algoritma K-means dan Agglomerative Clustering Untuk Segmentasi Penjualan Online Pada Customer Retail” disusun sebagai salah satu syarat utama dalam menyelesaikan program sarjana pada Universitas AMIKOM Yogyakarta. Penyelesaian skripsi non-reguler ini juga tidak lepas dari bantuan berbagai pihak, oleh karena ini pada kesempatan kali ini penulis menyampaikan rasa hormat dan terimakasih kepada:

1. Prof. Dr. M. Suyanto, M.M. selaku Rektor Universitas AMIKOM Yogyakarta.
2. Bapak Hanif Al Fatta, M.Kom. selaku Dekan Fakultas Ilmu Komputer.
3. Ibu Windha Mega Pradnya D, M.Kom. selaku Ketua Program Studi Informatika Universitas AMIKOM Yogyakarta.
4. Ibu Yuli Astuti, M.Kom. selaku dosen pembimbing yang telah memberikan banyak bimbingan, saran, serta waktunya selama penyusunan skripsi non-reguler ini.
5. Bapak Suryatno dan Ibu Sukarti selaku kedua orang tua saya yang selalu memberi dukungan penuh serta doa bagi penulis.
6. Pihak-pihak lain yang telah memberikan kontribusi baik secara langsung maupun tidak langsung terhadap penelitian dan penyusunan skripsi non-reguler ini.

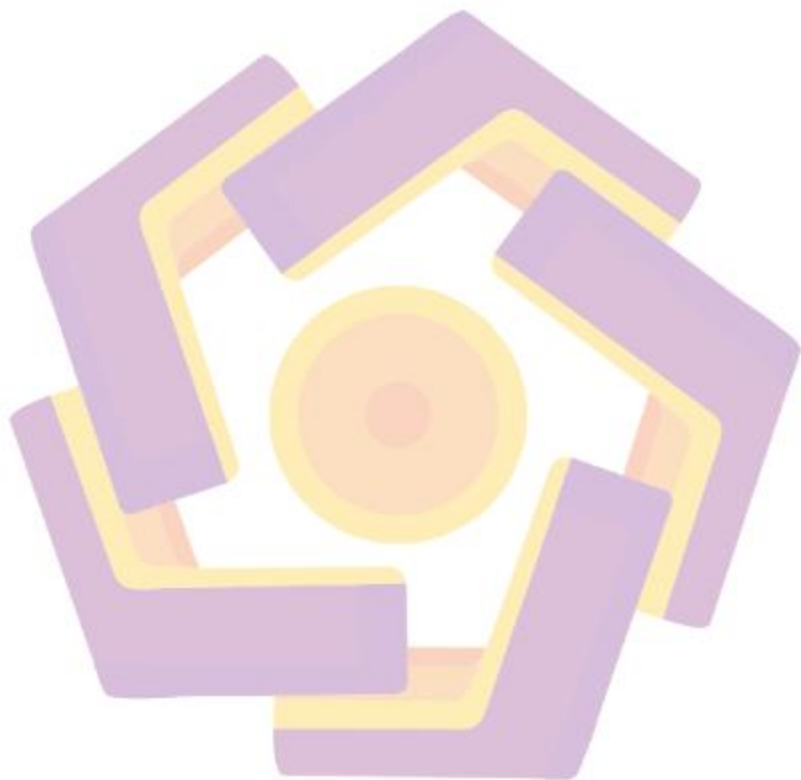
Yogyakarta, 27 Juni 2024

Ghanim Ramadhan

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
KATA PENGANTAR	v
DAFTAR ISI	vi
DAFTAR TABEL	viii
DAFTAR GAMBAR	ix
DAFTAR LAMPIRAN	x
INTISARI	xi
<i>ABSTRACT</i>	xii
BAB I PENDAHULUAN	1
BAB II TINJAUAN PUSTAKA	3
BAB III METODE PENELITIAN	4
3.1 Alur Penelitian	4
3.2 Data Analisis	4
3.3 Data Cleansing	5
3.3 Data Preparation	6
3.4 Modelling	7
3.5 Analisis Akhir	7
BAB IV HASIL DAN PEMBAHASAN	9
BAB V PENUTUP	14
5.1 Kesimpulan	14

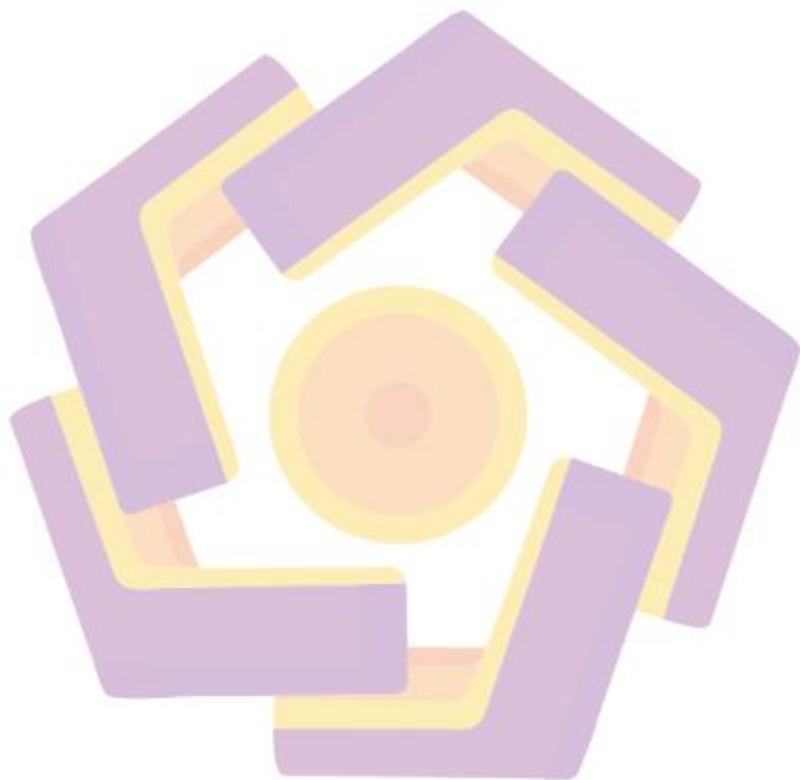
5.2	Saran	14
	REFERENSI	15
	LAMPIRAN	17



DAFTAR TABEL

Tabel 1 Perbandingan metode

5



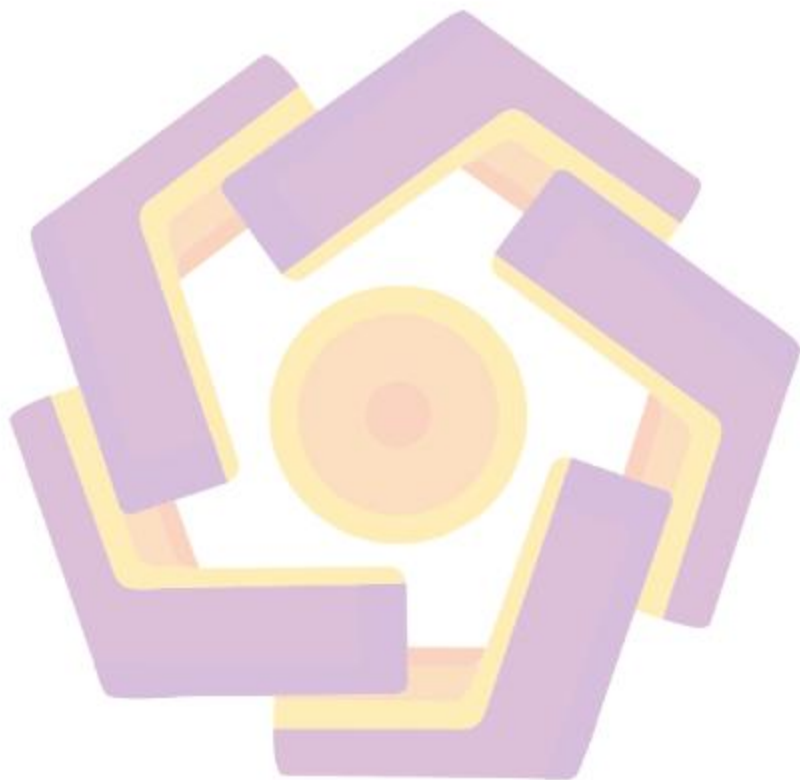
DAFTAR GAMBAR

Gambar. 1 Alur Penelitian	4
Gambar. 2 Data Pembelian Pelanggan	5
Gambar. 3 Jumlah Data Null Dataset	6
Gambar. 4 Boxplot cluster dengan variabel amount	9
Gambar. 5 Boxplot cluster dengan variabel frequency	10
Gambar. 6 Boxplot cluster dengan variabel recency	10
Gambar. 7 Dendogram cluster	11
Gambar. 8 Boxplot cluster dengan variabel amount	12
Gambar. 9 Boxplot cluster dengan variabel frequency	12
Gambar. 10 Boxplot cluster dengan variabel recency	13



DAFTAR LAMPIRAN

Lampiran 1. Lembar Review Dan Acceptance	17
Lampiran 2. Bukti Index Penerbit JPIT	18



INTISARI

Penelitian ini berfokus pada perbandingan antara dua algoritma populer dalam bidang data science, yaitu algoritma K-means dan Agglomerative Clustering. Konteks utama dari penelitian ini adalah segmentasi data pelanggan, suatu proses yang sangat penting dalam dunia bisnis untuk memahami dan melayani pelanggan dengan lebih baik. Tujuan utama dari penelitian ini adalah untuk mengevaluasi dan membandingkan kinerja kedua algoritma tersebut dalam menghasilkan segmen pelanggan yang efektif dan efisien. Dalam penelitian ini, dataset yang digunakan adalah dataset pelanggan retail. Dataset ini mencakup berbagai atribut yang mencerminkan karakteristik dan perilaku pelanggan. Untuk mengukur kinerja kedua algoritma, penelitian ini menggunakan metode pembobotan RFM (*Recency, Frequency, Monetary*). Metode ini adalah metode yang umum digunakan dalam analisis pelanggan untuk mengidentifikasi pelanggan yang paling berharga berdasarkan seberapa baru mereka bertransaksi (*Recency*), seberapa sering mereka bertransaksi (*Frequency*), dan seberapa besar nilai transaksi mereka (*Monetary*). Selain itu, penelitian ini juga menggunakan metrik evaluasi yang dikenal sebagai silhouette score. Metrik ini digunakan untuk mengukur seberapa baik suatu objek cocok dengan klusternya sendiri dibandingkan dengan kluster lain. Hasil penelitian ini memberikan wawasan yang berharga tentang kualitas kedua algoritma dalam melakukan segmentasi data pelanggan. Ditemukan bahwa algoritma K-Means menghasilkan nilai silhouette score sebesar 0.5087, sedangkan Agglomerative Clustering menghasilkan nilai yang lebih tinggi, yaitu 0.6363. Ini menunjukkan bahwa, dalam konteks dataset ini, Agglomerative Clustering mungkin lebih efektif dibandingkan dengan K-Means. Namun, penelitian lebih lanjut tentunya diperlukan untuk memvalidasi temuan ini dan untuk mengeksplorasi lebih lanjut tentang bagaimana kedua algoritma ini dapat dioptimalkan untuk segmentasi data pelanggan.

Kata kunci: Analisis algoritma, K-means, Agglomerative Clustering, RFM, Silhouette Score.

ABSTRACT

This research focuses on the comparison between two popular algorithms in data science, namely K-means and Agglomerative Clustering algorithms. The main context of this research is customer data segmentation, a very important process in the business world to understand and serve customers better. The main objective of this research is to evaluate and compare the performance of the two algorithms in generating effective and efficient customer segments. In this research, the dataset used is a retail customer dataset. This dataset includes various attributes that reflect customer characteristics and behavior. To measure the performance of both algorithms, this research uses the RFM (Recency, Frequency, Monetary) weighting method. This method is a commonly used method in customer analysis to identify the most valuable customers based on how recently they transacted (Recency), how often they transact (Frequency), and how much their transactions are worth (Monetary). In addition, this research also uses an evaluation metric known as silhouette score. This metric is used to measure how well an object fits into its own cluster compared to other clusters. The results of this study provide valuable insights into the quality of both algorithms in segmenting customer data. It was found that the K-Means algorithm produced a silhouette score value of 0.5087, while Agglomerative Clustering produced a higher value of 0.6363. This suggests that, in the context of this dataset, Agglomerative Clustering may be more effective compared to K-Means. However, further research is certainly needed to validate these findings and to further explore how these two algorithms can be optimized for customer data segmentation.

Keyword: Analysis algorithm, K-means, Agglomerative Clustering, RFM, Silhouette Score.