

# Implementasi Algoritma K-Means untuk Clustering Wilayah Jamaah Haji KBIH Jabal Nur

Oleh:

Deva Martha Shella

Ade Eviyanti

Progam Studi Informatika

Universitas Muhammadiyah Sidoarjo

Januari, 2025

# Pendahuluan

Pelaksanaan ibadah haji adalah kewajiban bagi umat Islam yang mampu secara fisik dan finansial. Haji dilakukan di Mekkah, Arab Saudi, dengan tujuan mendekatkan diri kepada Allah dan memperkuat iman.

KBIH Jabal Nur, lembaga konseling haji di bawah Muhammadiyah Sidoarjo, menghadapi tantangan dalam pengelolaan data pendaftar yang terus meningkat setiap tahun. Saat ini, pengelompokan data masih dilakukan secara manual, yang menghambat efisiensi strategi pemasaran.

Penelitian ini bertujuan untuk mengelompokkan wilayah pendaftar jamaah haji menggunakan algoritma K-Means dengan evaluasi melalui metode Elbow untuk meningkatkan efektivitas promosi dan pengelolaan data jamaah.

# Pertanyaan Penelitian (Rumusan Masalah)

- Bagaimana cara mengimplementasikan algoritma K-Means untuk melakukan clustering wilayah jamaah haji pada KBIH Jabal Nur?
- Bagaimana hasil clustering dengan algoritma K-Means dapat digunakan untuk mendukung pengambilan keputusan dalam pengelompokan wilayah pada bimbingan ibadah haji (KBIH) Jabal Nur?
- Bagaimana cara mengukur seberapa efektif algoritma K-Means dalam mengelompokkan wilayah jamaah haji berdasarkan parameter tertentu, seperti alamat, usia, dan tahun keberangkatan?

# Metode

## 1. Studi Literatur

- Studi Literatur digunakan untuk mendukung analisis

## 2. Pengumpulan Data

- Data diambil dari pendaftar haji KBIH Jabal Nur pada keberangkatan 2011-2050.

## 3. Data Preparation

- Seleksi atribut penting.
- Data Cleaning
- Data Transformation

## 4. Implementasi K-Means Clustering

- Penentuan 2 kluster

## 5. Visualisasi dan Analisis

- Hasil clustering divisualisasikan menggunakan scatter plot.
- Penggunaan metode Elbow untuk menentukan k terbaik

# Hasil

hasil klasterisasi menggunakan algoritma K-Means pada penelitian ini, menunjukkan bahwa alamat yang paling terbanyak adalah wilayah 5 atau Sidoarjo yang terletak pada cluster 1, dimana Sidoarjo memiliki 1261 jumlah data pendaftar. Sedangkan wilayah yang paling sedikit dengan jumlah 6 data dimiliki oleh wilayah 6,4,3,2,0 dan 1 yaitu Surabaya, Pasuruan, Mojokerto, Lamongan, Blitar dan Jember.

# Pembahasan

- Dataset

Data yang diperoleh dari KBIH Jabal Nur merupakan pendaftar jamaah haji pada keberangkatan tahun 2011-2050 dapat dilihat pada gambar dibawah ini. Terdapat 1442 dataset

Tahun Berangkat	Nama	Jenis Kelamin	Usia	Alamat
2011	Bina Retina	W	64	J. H. P. Blok AC-16 RT. 5/4 Gedangan- Sidoarjo
2011	Nur Madiah Bt Chusairi	W	66	Jl. Wijaya Kusuma 31 RT. 31 RW. 04 Bligo- Candi
2012	Suwito BN Moch Akiyak	P	50	Kepatihan RT. 4/2 Tulangan- Sidoarjo
2012	Nafi'ah BT Kasmun	W	60	Kepatihan RT. 4/2 Tulangan- Sidoarjo
2012	Siti Fatimah BT Soenjoen Fatah	W	70	Penambangan RT. 02 RW. 01 Balongbendo- Sidoarjo
2012	Susanto BN Sumarlan	P	60	Penambangan RT. 02 RW. 01 Balongbendo- Sidoarjo
2012	Innamah BT Mudakir	W	62	Sambisari KAV Utara F-6 RT 35 RW 06 Sambibulu- Taman
2012	Broto Mulyoko BN Sudewo Benowo	P	62	Sambisari KAV Utara F-6 RT 35 RW 06 Sambibulu- Taman
2012	Titi Surtiningsih BT Sunarja	W	67	Simomulyo RT. 15 RW. 04 Kesambi- Porong
2012	Yuli Sudarmanto BT Munaji	W	58	Simomulyo RT. 15 RW. 04 Kesambi- Porong
2012	Siti Muawanah BT H. Affandi	W	70	Trosobo No. 8 RT 3/4 Taman- Sidoarjo

# Pembahasan

- Data Cleaning

Setelah tahap seleksi data nilai selesai dilakukan, langkah selanjutnya adalah proses data cleaning dari hasil seleksi sebelumnya. Proses ini bertujuan untuk membersihkan data dari komponen yang tidak diperlukan. Berikut ini adalah tahapan processing pada datasheet pendaftar jamaah haji.



	Tahun	Nama	Gender	Usia	Alamat
0	2011	Heru Santoso BN Madsaridi	P	84	SIDOARJO
1	2011	Bina Retina	W	56	SIDOARJO
2	2012	Nur Madiah BT Chusairi	W	60	SIDOARJO
3	2012	SuWito BN Moch Akiyak	P	61	SIDOARJO
4	2012	Nafi'ah BT Kasmun	W	64	SIDOARJO
...	...	...	...	...	...
1418	2050	Zainuri Bn Adam SariPah	P	65	SIDOARJO
1419	2050	Mariyam Bt Qasim/Tasrifah	W	58	SIDOARJO
1420	2050	Senifah ( Haji Plus)	W	67	SIDOARJO
1421	2050	Soffyah Bt Kadiran Miun	W	59	SIDOARJO
1422	2050	Suliamah	W	67	SIDOARJO

1423 rows x 5 columns

# Pembahasan

- Transformasi Data

Setelah proses data cleaning selesai, langkah berikutnya adalah melakukan transformasi data pada hasil tahapan sebelumnya. Transformasi data adalah langkah penting dalam proses analisis, di mana data diubah menjadi format yang sesuai atau tepat untuk diolah. Berikut ini adalah tahapan transformasi data pada dataset pendaftar haji

```
from sklearn.preprocessing import StandardScaler
#transformasi data
X = df[['Alamat', 'Gender', 'Usia']]
sc = StandardScaler()
sc.fit(X)
x = sc.transform(X)
print(X)
```

	Alamat	Gender	Usia
0	5	0	84
1	5	1	56
2	5	1	60
3	5	0	61
4	5	1	64
...	...	...	...
1418	5	0	65
1419	5	1	58
1420	5	1	67
1421	5	1	59
1422	5	1	67

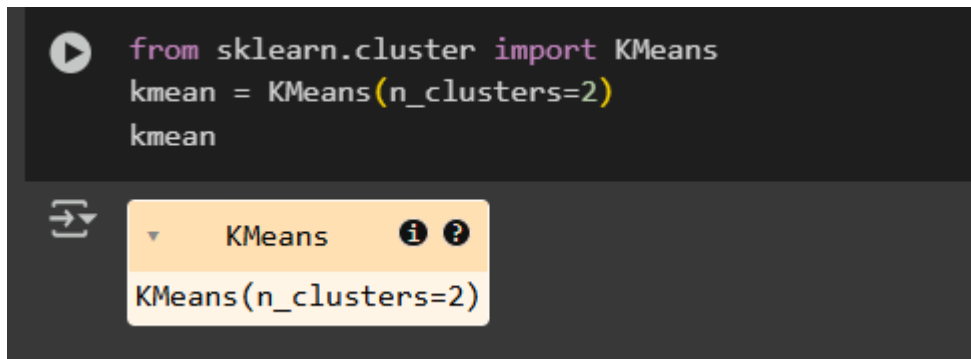
[1423 rows x 3 columns]

# Pembahasan

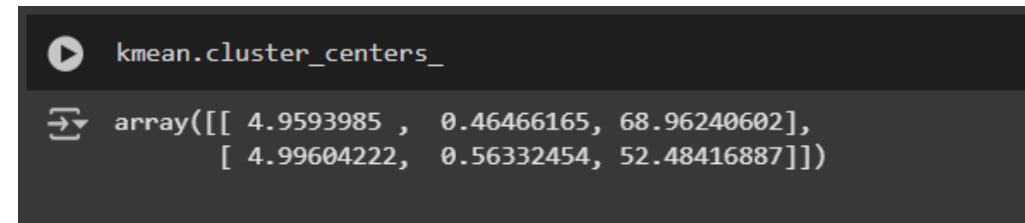
- Penerapan K-Means Clustering

Setelah tahap transformasi data selesai, langkah selanjutnya adalah menerapkan proses K-Means Clustering pada hasil transformasi tersebut. Dan juga menentukan titik centroid dari masing-masing kluster yang terbentuk dari K-Means Clustering. Centroid merupakan titik pusat setiap kluster, yang dihitung berdasarkan rata-rata posisi data dalam cluster tersebut.

```
from sklearn.cluster import KMeans
kmean = KMeans(n_clusters=2)
kmean
```



```
kmean.cluster_centers_
```



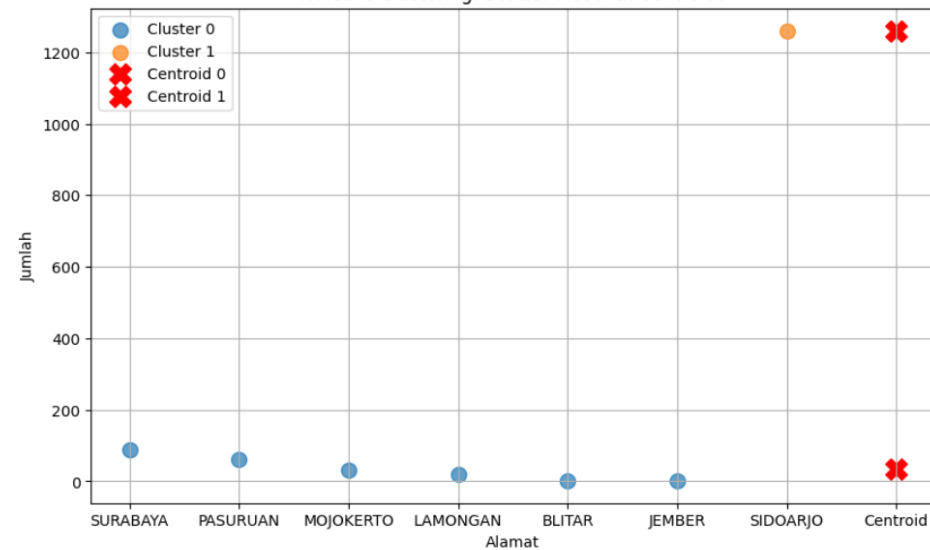
```
array([[ 4.9593985 ,  0.46466165, 68.96240602],
       [ 4.99604222,  0.56332454, 52.48416887]])
```

# Pembahasan

- Hasil Visualisasi dan Analisa

Hasil pengelompokan data pada KBIH Jabal Nur menggunakan dua kelompok, yang disajikan dalam tampilan bar. Setelah mengumpulkan data tentang masing-masing centroid, peneliti melakukan inisialisasi k pusat cluster (centroid) secara acak. Visualisasi hasil clustering ditampilkan dalam grafik scatter plot

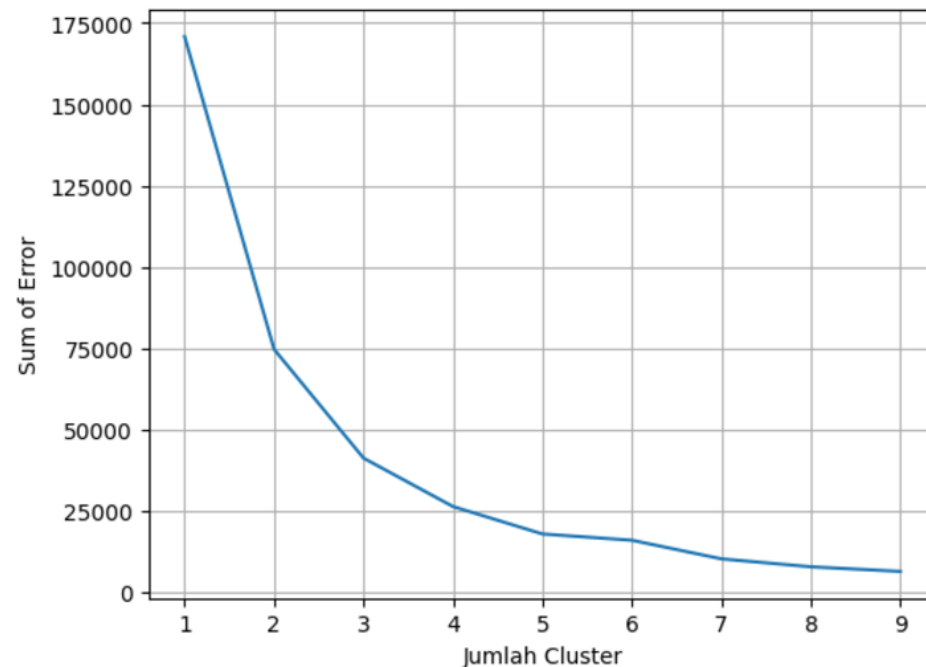
Index	Alamat	Frequency	Cluster	Cluster_Label
0	5	1261	1	Wilayah Terbanyak
1	6	89	0	Wilayah Sedikit
2	4	40	0	Wilayah Sedikit
3	3	26	0	Wilayah Sedikit
4	2	3	0	Wilayah Sedikit
5	0	2	0	Wilayah Sedikit
6	1	2	0	Wilayah Sedikit



# Pembahasan

- Evaluasi Hasil Clustering menggunakan Metode Elbow

Pada bagian ini disampaikan mekanisme pengukuran k terbaik. Metode uji dari penelitian ini menggunakan metode Elbow. Pada gambar 4 berikut merupakan gambar yang menunjukkan grafik Elbow.



# Temuan Penting Penelitian

Penelitian ini menunjukkan bahwa banyak wilayah yang paling sedikit pendaftaran ibadah haji di Kbih Jabal Nur untuk lebih ditingkatkan promosinya pada wilayah wilayah tertentu.

# Manfaat Penelitian

Penelitian ini bermanfaat bagi KBIH Jabal Nur untuk menyusun strategi pemasaran yang lebih efektif dengan mengidentifikasi wilayah dengan jumlah jamaah haji terbanyak dan tersedikit. Selain itu, penelitian ini memberikan panduan praktis bagi praktisi hukum dan data dalam menerapkan algoritma K-Means untuk pengelompokan data secara efisien, yang dapat diadaptasi untuk pengelolaan data di sektor lain. Bagi akademisi, hasil penelitian ini menambah referensi dalam penerapan data mining khususnya algoritma K-Means, serta menjadi landasan bagi penelitian lanjutan. Penelitian ini juga mendukung pengembangan sistem informasi berbasis data mining untuk membantu pengambilan keputusan strategis berbasis data.

# Referensi

- [1] Ali, A. (2019). Klasterisasi Data Rekam Medis Pasien Menggunakan Metode K-Means *Clustering* di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo. *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, 186-195. <https://doi.org/10.30812/matrik.v19i1.529>
- [2] Djaka Permana, M., Lia Hananto, A., Novalia, E., Huda, B., & Paryono, T. (2023). Klasterisasi Data Jamaah Umrah pada Tanurmutmainah Tour Menggunakan Algoritma K-Means. *Jurnal KomtekInfo*, 15–20) <https://doi.org/10.35134/komtekinfo.v10i1.332>
- [3] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk *Clustering* Data Obat-Obatan," *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.
- [4] M. Iqbal, "Klasterisasi Data Jamaah Umroh Pada Auliya Tour & Travel Menggunakan Metode K-Means *Clustering*," *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 5, no. 2, pp. 97–104, 2019, doi: 10.33330/jurteksi.v5i2.352.
- [5] M.A Albaihaqi, "STRATEGI PEMASARAN KELOMPOK BIMBINGAN IBADAH HAJI (KBIH) AT TAQWA UJUNG HARAPAN BEKASI DALAM MENINGKATKAN JUMLAH JAMA'AH."
- [6] Muhammad Yamin Nurzaman, & Nurina Sari, B. (2023). Implementasi K-Means *Clustering* Dalam Pengelompokan Banyaknya jumlah petani. *Jurnal Teknik Informatika Dan Sistem Informasi*, 10(3). <http://jurnal.mdp.ac.id>
- [7] Naufal Hayyu Triwardana. (n.d.). IMPLEMENTASI METODE FUZZY C-MEANS UNTUK CLUSTERING TRAVEL UMROH BERDASARKAN PRINCIPAL COMPONENT ANALYSIS
- [8] S. A. Azis, S. Defit, and Y. Yunus, "Jurnal Informatika Ekonomi Bisnis Klasterisasi Dana Bantuan Pada Program Keluarga Harapan ( PKH )," vol. 3, pp. 53–59, 2021, doi: 10.37034/infeb.v3i2.66.
- [9] T. Hidayat, "Klasifikasi Data Jamaah Umroh Menggunakan Metode K-Means *Clustering*," *Jurnal Sistim Informasi dan Teknologi*, pp. 19–24, Feb. 2022, doi: 10.37034/jsisfotek.v4i1.115.
- [10] Yunitasari, Melinda, Tamara Maharani, and Bagus Hikmahwan Akademi Komunitas Negeri Pacitan Jl Walanda Maramis. 2022. "IMPLEMENTASI METODE K-MEANS UNTUK PENGELOMPOKAN DATA JAMAAH" 09 (1).

# Referensi

- [11] I. Manajemen Promosi KBH Dalam Meningkatkan Minat Bimbingan Jamaah Haji Saepurrahmat, A. Iwan Setiawan, J. Manajemen Dakwah, and U. Sunan Gunung Djati, "Anida (Aktualisasi Nuansa Ilmu Dakwah)," 2017.
- [12] I. Sanela, A. Nazir, F. Syafria, E. Haerani, and L. Oktavia, "Penerapan Metode Clustering Dengan K-Means Untuk Memetakan Potensi Tanaman Padi di Sumatera," *Journal of Computer System and Informatics (JoSYC)*, vol. 5, no. 1, pp. 82–92, Nov. 2023, doi: 10.47065/josyc.v5i1.4523.
- [13] A. A. Hussein, "Improve The Performance of K-means by using Genetic Algorithm for Classification Heart Attack," *Int. J. Electr. Comput. Eng.*, vol. 8, no. 2, p. 1256, 2018, doi: 10.11591/ijece.v8i2.pp1256-1261.
- [14] R. K. Dinata, H. Novriando, N. Hasdyna, and S. Retno, "Reduksi Atribut Menggunakan Information Gain," vol. 6, no. 1, pp. 48–53, 2020.
- [15] I. Kamila, U. Khairunnisa, and M. Mustakim, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 5, no. 1, p. 119, 2019, doi: 10.24014/rmsi.v5i1.7381.

