PRA-PEMROSESAN DATA DAN TRANSFORMASI NAMA PROVINSI MENJADI KOORDINAT SERTA PENERAPAN HEATMAP UNTUK ANALISIS KUOTA JEMAAH HAJI DI INDONESIA (DATASET: SATUDATA KEMENTERIAN AGAMA, 2015–2023)

Muhammad Adhit Dwi Yuda*, Aliyah, Vera Nurlia dan Ghina Alya Nasution

Universitas Gunadarma
Jl. Margonda Raya No. 100, Depok, Jawa Barat, 16424
adhit@uca.ac.id, aliyah@uca.ac.id, veranurlia@uca.ac.id, ghinaalyanasution@gmail.com
*Corresponding Author

ABSTRAK

Penelitian ini menguraikan tahapan pra-pemrosesan, transformasi geospasial, dan visualisasi untuk menganalisis distribusi kuota jemaah haji di Indonesia menggunakan dataset publik SatuData Kementerian Agama (2015–2023). Metode meliputi pengunduhan data, inspeksi kualitatif untuk menjaga konsistensi, deteksi dan koreksi kesalahan numerik, serta penghapusan duplikasi. Dalam proses pembersihan teridentifikasi 24 baris dengan kesalahan desimal akibat konversi Excel yang dikoreksi sehingga nilai jumlah menjadi representatif. Peneliti memanfaatkan Anaconda Python, lingkungan Jupyter Notebook, dan library Pandas untuk transformasi data serta penambahan kolom latitude dan longitude berdasarkan koordinat pusat administratif provinsi. Nilai kuota dinormalisasi pada rentang 0–1 untuk mengurangi pengaruh nilai ekstrem sebelum divisualisasikan. Visualisasi dilakukan menggunakan library Folium dengan plugin Heatmap, dilengkapi fitur popup dan kontrol dropdown untuk analisis temporal per tahun. Hasil menunjukkan dominasi kuota pada provinsi Pulau Jawa—terutama Jawa Barat—serta pola kuota relatif lebih rendah di wilayah timur Indonesia; tahun 2021 tercatat sebagai anomali akibat pembatalan pemberangkatan pada masa pandemi COVID-19. Kontribusi penelitian ini adalah pipeline pra-pemrosesan terperinci dan metodologi pemetaan nama provinsi menjadi koordinat yang meningkatkan akurasi representasi spasial serta menyediakan dasar rekomendasi kebijakan untuk pemerataan kuota jemaah haji. Studi ini juga merekomendasikan integrasi data administrasi lokal dan evaluasi berkala untuk meningkatkan transparansi distribusi kuota serta akurasi kebijakan dan pengembangan dashboard interaktif tambahan.

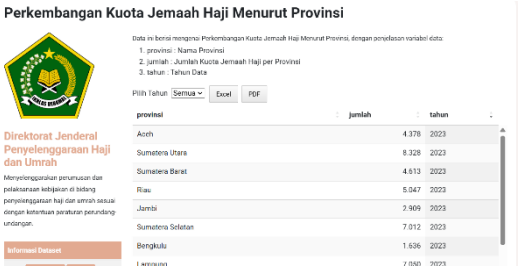
Kata Kunci: *kuota jemaah haji, pra-pemrosesan data, transformasi geospatial, heatmap (folium), satudata kementrian agama.*

PENDAHULUAN

Ibadah haji merupakan salah satu rukun islam yang wajib dilaksanakan oleh umat islam yang mampu secara fisik dan finansial [1]. Di Indonesia, negara dengan populasi muslim terbesar di dunia, pengelolaan kuota jemaah haji menjadi isu strategis yang melibatkan berbagai aspek, termasuk pendistribusian kuota per provinsi. Data jumlah jemaah haji dari setiap provinsi perlu dianalisis dan divisualisasikan agar kebijakan publik dapat lebih tepat sasaran.

Dalam era digital saat ini, data dapat diolah dan divisualisasikan dengan lebih efektif menggunakan pendekatan data science [2]. Salah satu metode yang akan digunakan yaitu visualisasi peta interaktif berbasis pemrograman Python di Jupyter Notebook. Penelitian ini bertujuan untuk menyajikan sebaran kuota jemaah haji dari tahun 2015 hingga 2023 dalam bentuk peta panas (heatmap) berdasarkan dataset publik yang diunduh dari kementrian agama. Memperoleh dataset publik dari portal web kementrian agama yang dapat diakses melalui tautan: http://satudata.kemenag.go.id/dataset/detail/perkembangan-kuota-jemaah-haji-menurut-provinsi. Dataset tersebut berisi informasi

mengenai perkembangan kuota jemaah haji berdasarkan provinsi, yang mencakup kolom-kolom seperti nama provinsi, jumlah kuota, dan tahun. Contoh isi dataset yang diambil dapat dilihat pada gambar 1.



Gambar 1. *Dataset Kuota Jemaah haji menurut provinsi*

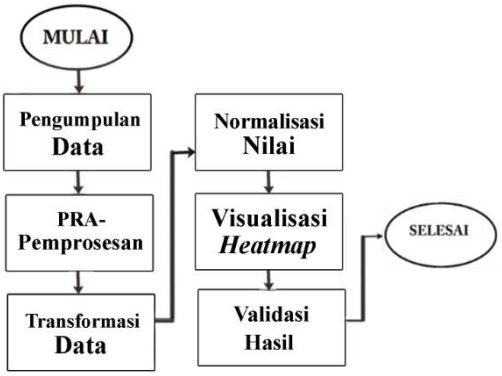
Tinjauan Pustaka

Penelitian Ajagbe et al. (2020) [3] dan Kim et al. (2024) [4] menerapkan folium untuk visualisasi spasial, namun fokus pada data kriminal dan kesehatan lokal. Kelebihan penelitian tersebut adalah kemudahan interpretasi visual, namun kelemahannya adalah terbatasnya konteks kebijakan nasional. Sementara itu, Penelitian Tu et al (2022) [5] Miftahurrahmi et al (2022) [6] menggunakan algoritma DBSCAN untuk klasterisasi, namun tidak sesuai bila tujuan utama adalah representasi distribusi kuota per tahun. Penelitian ini melengkapi kekurangan tersebut dengan pipeline *preprocessing–transformasi geospasial–heatmap interaktif* yang lebih komprehensif dan dapat mendukung kebijakan publik.

Berbeda dengan penelitian sebelumnya yang hanya menampilkan distribusi spasial secara deskriptif, penelitian ini menambahkan tahapan pra-pemrosesan data, validasi, dan normalisasi kuota. Selain itu, penelitian ini juga mengusulkan pendekatan konseptual berupa analisis kesenjangan regional serta integrasi peta interaktif sebagai alat pendukung keputusan bagi pemangku kebijakan.

METODE PENELITIAN

Metode penelitian ini di susun untuk mencapai tujuan utama, yaitu menganalisis distribusi kuota jemaah haji secara *spasial* di tingkat provinsi indonesia. Penelitian ini bersifat studi kasus menggunakan data sekunder dan *repository* publik, dan menggabungkan prosedur pra-pemrosesan data, transformasi geospasial, serta visualisasi interaktif untuk mendukung interpretasi. Secara garis besar alur penelitian ditunjukan pada gambar 2.



Gambar 2. *Metode Penelitian.*

Tahapan utama penelitian meliputi:

1. Pengumpulan Data : mengunduh file excel dari *repository* SatuData Kemenag dan menyimpan salinan mentah sebagai bukti sumber data [7].
2. Pra-pemrosesan : pemeriksaan dan pembersihan data dari memahami tipe data, deteksi duplikasi data, deteksi *missing value* dan deteksi serta koreksi kesalahan numerik [8].
3. Transformasi Data : menambahkan dua kolom baru latitude dan longitude untuk transformasi geospasial [9].
4. Normalisasi nilai : untuk menghindari dominasi nilai ekstrem [10].
5. Visualiasai *heatmap* : Pembuatan peta interaktif menggunakan *folium* dan plugin *heatmap* [11].
6. Validasi Hasil : semua perubahan penting (koreksi nilai,penghapusan duplikasi, penambahan koordinat)

divalidasi ulang terhadap file sumber [12].

Pengumpulan Data

Dataset diperoleh dari repository publik SatuData Kementrian agama Republik Indonesia (*Dataset* Perkembangan Kuota Jemaah Haji Menurut Provinsi), yang mencakup data per provinsi untuk periode 2015-2023. Pemilihan *dataset* ini dikarenakan bersifat resmi, komprehensif pada tingkat provinsi dan relevan untuk menganalisis pola sebaran kuota jemaah haji di seluruh indonesia. Salinan file sumber (format excel) disimpan sebagai arsip eksperimen penelitian ini. Isi file *dataset* dalam format excel ditunjukan pada gambar 3.

B40 | 1,999

	A	B	C
1	provinsi	jumlah	tahun
2	Aceh	4,378	2023
3	Sumatera Utara	8,328	2023
4	Sumatera Barat	4,613	2023
5	Riau	5,047	2023
6	Jambi	2,909	2023
7	Sumatera Selatan	7,012	2023
8	Bengkulu	1,636	2023
9	Lampung	7,05	2023
10	Kepulauan Bangka Belitung	1,065	2023
11	Kepulauan Riau	1,291	2023
12	DKI Jakarta	7,926	2023
13	Jawa Barat	38,723	2023
14	Jawa Tengah	30,377	2023
15	Daerah Istimewa Yogyakarta	3,147	2023
16	Jawa Timur	35,152	2023
17	Banten	9,461	2023
18	Bali	698	2023
19	Nusa Tenggara Barat	4,499	2023
20	Nusa Tenggara Timur	668	2023
21	Kalimantan Barat	2,519	2023

Sheet1

Gambar 3. *isi file dataset sumber resmi*

Pra-Pemrosesan

Dataset setelah diunduh dari *repository* resmi Kementrian Agama Republik Indonesia dalam format excel dan selanjutnya diolah menggunakan bahasa pemrograman *python* pada lingkungan *jupyter notebook (anaconda)*. Tujuan Pra-Pemrosesan adalah mempersiapkan *dataset* agar layak untuk analisis spasial dan visualisasi *heatmap* menggunakan *library folium* [13]. Struktur atribut tipe dataset ditunjukan pada gambar 4.

```
[9]: gina.dtypes

[9]: provinsi    object
     jumlah       int64
     tahun        int64
     dtype: object
```

Gambar 4. *Atribut dan tipe data utama*

Pra-Pemrosesan juga dilakukan pembersihan data secara sistematis dengan menggunakan *library pandas.* Tahapan yang di laksanakan meliputi :

1. Deteksi dan Koreksi Kesalahan Numerik : Ditemukan kasus konversi numerik akibat *entry* dataset yang keliru pada *dataset Raw* pada *file excel*, contoh nilai ditampilkan sebagai 13,6 padahal seharusnya 13.600. Setelah melakukan verifikasi manual terhadap file sumber, teridentifikasi 24 baris yang berdampak, nilai pada baris tersebut dikoreksi misal dikalikan x1000, sehingga representasi kuota menjadi sesuai dengan realitas administrasi. Ilustrasi Proses algoritma serta hasil sebelum dan sesudah koreksi disajikan pada gambar 5.

```
[10]: data = [
          ["Lampung", 7050, 2023],
          ["Kalimantan Barat", 1150, 2022],
          ["Sulawesi Selatan", 3320, 2022],
          ["Lampung", 7050, 2020],
          ["Haji Khusus",17680, 2020],
          ["Bengkulu", 1940, 2019],
          ["Jawa Tengah", 30860, 2019],
          ["Kalimantan Tengah", 1920, 2019],
          ["Sulawesi Tengah", 2250, 2019],
          ["Maluku", 1090, 2018],
          ["Maluku Utara", 1080, 2018],
          ["Papua", 1080, 2018],
          ["Jawa Timur", 35270, 2017],
          ["Maluku", 1090, 2017],
          ["Maluku Utara", 1080, 2017],
          ["Papua", 1080, 2017],
          ["Aceh", 3140, 2016],
          ["Kalimantan Tengah", 1080, 2016],
          ["Kalimantan Selatan", 3050, 2016],
          ["Haji Khusus", 13600, 2016],
          ["Aceh", 3140, 2015],
          ["Kalimantan Tengah", 1080, 2015],
          ["Kalimantan Selatan", 3050, 2015],
          ["Haji Khusus", 13600, 2015],
          ["Haji Khusus", 13600, 2015],
      ]
```

Before

Maluku Utara	853	2015
Papua	853	2015
Papua Barat	569	2015
Haji Khusus	13,6	2015

AFTER

315	Kalimantan Tengah	1080	2015
316	Kalimantan Selatan	3050	2015
317	Haji Khusus	13600	2015

Gambar 5. *algoritma proses koreksi data tidak representatif dan hasil sebelum dan sesudah dikoreksi.*

2. Deteksi dan Penghapusan Duplikat : Duplikasi baris diidentifikasi menggunakan algoritma *duplicated().sum()* dan entry duplikat dihapus dengan *drop_duplicates()* untuk mencegah bias agregasi. Pada *dataset* ini ditemukan 1 baris duplikat dan dihapus. Contoh hasil deteksi dan penghapusan disajikan pada gambar 6.

```
gina.duplicated(subset=["provinsi", "jumlah", "tahun"]).sum()
np.int64(1)
gina.drop_duplicates(subset=["provinsi", "jumlah", "tahun"], inplace=True)
```

Gambar 6. *hasi deteksi dan penghapusan data duplikasi*

3. Pemeriksaan nilai kosong : Semua kolom utama diperiksa, tidak ditemukan *missing value* sehingga tidak diperlukan imputasi atau penghapusan baris. Prosedur pemeriksaan divisualisasikan pada gambar 7.

```
[14]: gina.isnull().sum()

[14]: provinsi    0
      jumlah      0
      tahun       0
      dtype: int64
```

Gambar 7. *Algoritma mendeteksi nilai kosong*

Tranformasi Data

Setelah itu menambahkan dua kolom baru yaitu *latitude* dan *longitude*. Nama provinsi dipetakan menjadi koordinat pusat administratif provinsi menggunakan *mapping dictionary* yang bersumber pada referensi geospasial terpercaya. Implementasi *mapping* menggunakan fungsi *lambda* pada setiap baris [14]. Contoh algoritma pada gambar 8.

```
[17]: gina['Latitude'] = gina['provinsi'].map(lambda x: provinsi_coords.get(x, (None, None))[0])
      gina['Longitude'] = gina['provinsi'].map(lambda x: provinsi_coords.get(x, (None, None))[1])
```

Gambar 8. *Algoritma lambda sebagai penambah dua kolom latitude dan longitude.*

Normalisasi Nilai

Untuk menghindari dominasi nilai ekstrem pada *heatmap*, kolom jumlah dinormalisasi ke rentang [0,1] menggunakan metode min-max [15].

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

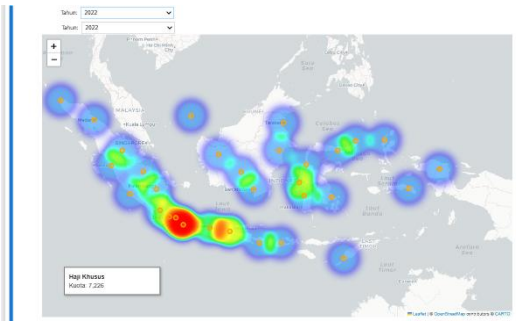
Implementasi normalisasi dilakukan menggunakan bahasa pemrograman *python* sebelum lanjut ketahap visualisasi, Lihat pada gambar 9.

```
# Normalisasi nilai jumlah
jumlah_min = data_tahun['jumlah'].min()
jumlah_max = data_tahun['jumlah'].max()
if jumlah_max != jumlah_min:
    data_tahun['jumlah_norm'] = data_tahun['jumlah'].apply(
        lambda x: (x - jumlah_min) / (jumlah_max - jumlah_min)
    )
else:
    data_tahun['jumlah_norm'] = 0.5
```

Gambar 9. *Algoritma Normalisasi nilai pada kolom data jumlah*

Visualisasi Heatmap

Pembuatan peta interaktif menggunakan *folium* dan *plugin heatmap*. Setiap titik provinsi dilengkapi popup menampilkan provinsi, jumlah, dan tahun. Fitur *ipywidgets* disertakan untuk *dropdown* pemilihan tahun sehingga analisis temporal dapat dilakukan secara interaktif [16]. Contoh tampilan pada gambar 10.



Gambar 10. *Visualisasi Peta Heatmap*

Validasi Hasil

Semua perubahan penting seperti koreksi nilai, penghapusan duplikat, dan penambahan koordinat divalidasi ulang terhadap file sumber dan validasi terhadap informasi diportal Kemenag. Perubahan di simpan atau dicatat didalam *jupyter notebook.*

HASIL DAN PEMBAHASAN

Gambaran Umum *Dataset*

Dataset yang dianalisis merupakan *dataset* publik Perkembangan Kuota Jemaah Haji Menurut Provinsi yang diunduh dari *repository* SatuData Kementrian Agama RI. Sampel mencakup periode 2015-2023 dengan atribut utama : provinsi, tahun, dan jumlah (kuota). Dari file sumber diperoleh 319 baris data mentah sebelum proses pembersihan. Ilustrasi struktur *dataset* pada gambar 11.

```
[5]: gina
```

[5]:

	provinsi	jumlah	tahun
0	Aceh	4378	2023
1	Sumatera Utara	8328	2023
2	Sumatera Barat	4613	2023
3	Riau	5047	2023
4	Jambi	2909	2023
...	...	...	...
314	Maluku Utara	8530	2015
315	Papua	8530	2015
316	Papua Barat	5690	2015
317	Haji Khusus	136	2015
318	Haji Khusus	136	2015

319 rows × 3 columns

Gambar 11. *Potongan data mentah.*

Hasil Pra-Pemrosesan (Data Cleansing)

Tahap Pra-Pemrosesan mengikuti prosedur di BAB III, yaitu mendeteksi duplikat, pengecekan *missing value* dan kesalahan koreksi numerik. Pelaksanaan tahapan ini mengikuti prinsip kualitas data (*data quality)* yang menyatakan bahwa validitas dan konsistensi input sangat menentukan hasil analisis.

Hasil Transformasi Geospasial

Setelah data numerik dan struktur dibersihkan setiap kolom provinsi dipetakan menjadi koordinat pusat administratif menambahkan dua kolom *latitude* dan *longitude* menggunakan *mapping dictionary* yang bersumber dari referensi geospasial terpercaya. Contoh hasil transformasi disajikan pada tabel 1 dan proses algoritma *mapping* terlihat pada gambar 12.

Tabel 1.*Contoh hasil Transformasi Provinsi ke koordinat*

Provinsi	Latitude	Longitude
Jawa Barat	-6.889836	107.640471
Jawa Tengah	-7.150975	110.140259
Papua	-4.269928	138.080352

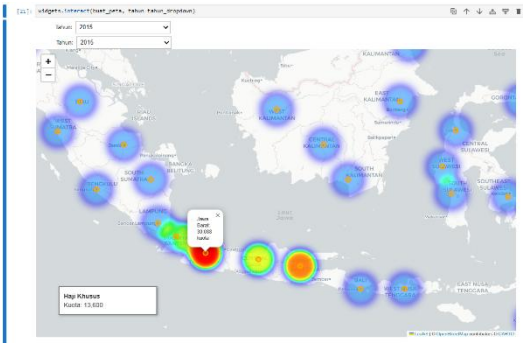
```
# --- Tambahkan Latitude dan Longitude ---
provinsi_coords = {
    'Aceh': [5.55, 95.32],
    'Sumatera Utara': [3.59, 98.67],
    'Sumatera Barat': [-0.95, 100.35],
    'Riau': [0.51, 101.45],
    'Jambi': [-1.6, 103.6],
    'Sumatera Selatan': [-3.32, 104.91],
    'Bengkulu': [-3.8, 102.26],
    'Lampung': [-5.45, 105.27],
    'Kepulauan Bangka Belitung': [-2.74, 106.44],
    'Kepulauan Riau': [3.99, 108.39],
    'DKI Jakarta': [-6.2, 106.85],
    'Jawa Barat': [-6.9, 107.6],
    'Jawa Tengah': [-7.2, 110.2],
    'Daerah Istimewa Yogyakarta': [-7.8, 110.4],
    'Jawa Timur': [-7.54, 112.23],
    'Banten': [-6.1, 106.16],
    'Bali': [-8.65, 115.2],
    'Nusa Tenggara Barat': [-8.65, 117.36],
    'Nusa Tenggara Timur': [-10.17, 123.6],
    'Kalimantan Barat': [0.13, 111.09],
    'Kalimantan Tengah': [-1.61, 113.38],
    'Kalimantan Selatan': [-3.32, 114.59],
    'Kalimantan Timur': [0.53, 117.14],
    'Kalimantan Utara': [3.3, 117.6],
    'Sulawesi Utara': [1.5, 124.8],
    'Sulawesi Tengah': [-0.89, 119.87],
    'Sulawesi Selatan': [-3.99, 119.65],
    'Sulawesi Tenggara': [-4.14, 122.43],
    'Gorontalo': [0.54, 123.06],
    'Sulawesi Barat': [-2.68, 119.2],
    'Maluku': [-3.25, 130.14],
    'Maluku Utara': [1.6, 127.7],
    'Papua Barat': [-1.33, 133.18],
    'Papua Barat Daya': [-1.15, 131.45],
}

gina['Latitude'] = gina['provinsi'].map(lambda x: provinsi_coords.get(x, [None, None])[0])
gina['Longitude'] = gina['provinsi'].map(lambda x: provinsi_coords.get(x, [None, None])[1])
```

Gambar 12. *Proses Algoritma mapping*

Visualisasi *Heatmap*

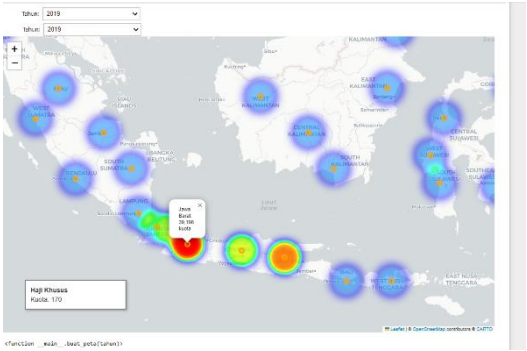
Peta interaktif dihasilkan dengan menggunakan *library folium* dan *plugin heatmap*. Beberapa keluaran penting :

1. Heatmap nasional pada tahun 2015 Memperlihatkan konsentrasi tetap dipulau jawa. Ilustrasi dapat dilihat pada gambar 13.



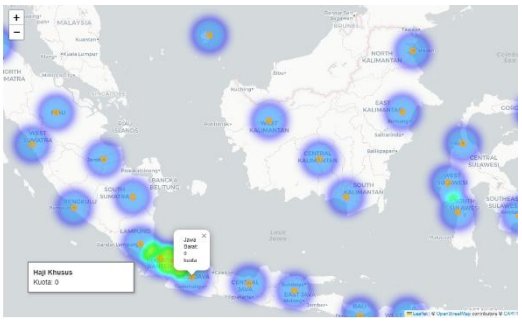
Gambar 13. *Heatmap pada tahun 2015*

2. *Heatmap* tahun 2019 menunjukan puncak alokasi dibeberapa provinsi jawa. Ilustrasi pada gambar 14.



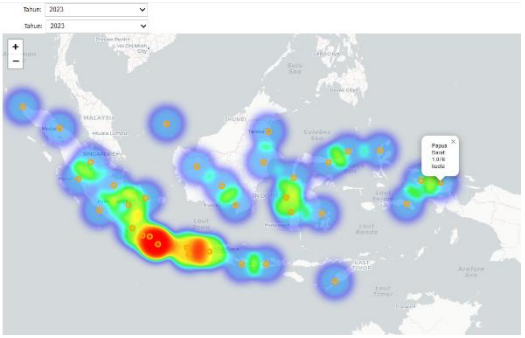
Gambar 14. *Heatmap pada tahun 2019*

3. *Heatmap pada* tahun 2021 menampilkan anomali, bahwa tidak ada alokasi akibat pembatalan pemberangkatan selama pandemi *COVID-19*. Ilustrasi pada gambar 15.



Gambar 15. *Heatmap pada tahun 2021 Pandemi COVID-19*

4. *Heatmap* pada tahun 2023 menunjukan pemulihan alokasi mendekati angka pra-pandemi *di* beberapa provinsi, misalnya jawa barat. Ilustrasi pada gambar 16



Gambar 16. *Heatmap pada tahun 2023*

Analisis dan Interpretasi Hasil

Berdasarkan hasil eksplorasi dan visualisasi yang dilakukan, menemukan beberapa temuan penting :

1. Dominasi keberangkatan jemaah haji Pulau Jawa : Provinsi-provinsi di pulau jawa, terutama jawa barat, jawa tengah dan jawa timur itu konsisten menerima kuota terbesar selama 2015-2023. Data menunjukan jawa barat sebagai provinsi dengan kuota tertinggi pada banyak tahun pengamatan. Contoh ringkasan tren jawa barat di tabel 2.

Tabel 2.*Tren Kuota Jawa Barat (2015-2023)*

Tahun	Kuota (Jawa Barat)
2015	30.088
2016	30.088

2017	38.852
2018	38.852
2019	39.198
2020	38.723
2021	(Pembatalan Penerbangan)
2022	17.679
2023	38.723

Dominasi jawa barat tidak hanya tercermin dari angka kuota, tetapi juga sejalan dengan fakta bahwa provinsi ini memiliki jumlah penduduk muslim terbesar di indonesia. Selain itu, faktor kebijakan penentuan kuota yang berbasis proporsi penduduk turut memperkuat posisi jawa barat. Hal ini mengindikasikan bahwa distribusi kuota sangat dipengaruhi oleh faktor demografi.

2. Terjadi Kesenjangan Regional : Provinsi-provinsi diwilayah timur indonesia diantaranya papua, NTT, dan maluku cenderung memperoleh kuota yang lebih rendah dan relatif stabil, terlihat sebagai area berwarna dingin pada *heatmap*. Hal ini mengindikasikan ketimpangan distribusi yang dapat berkaitan dengan proporsi penduduk, aksesibilitas pelayanan haji, serta kebijakan alokasi. Rendahnya kuota di papua, NTT, dan maluku tidak semata-mata karena jumlah penduduk, tetapi juga bisa terkait keterbatasan infrastuktur, jarak geografis, dan kapasitas embarkasi. Hal ini menunjukan adanya tantangan pemerataan yang perlu diantisipasi oleh pembuat kebijakan.
3. Dampak Pandemi *COVID-19* pada tahun 2021 : menampilkan anomali nasional yaitu terjadinya pembatalan keberangkatan yang mengakibatkan distribusi kuota praktis nihil. Visualisasi memperjelas dampak kebijakan nasional terhadap level provinsi.
4. Pemulihan Pascapandemi tahun 2022-2023 : Tahun 2022 menunjukan peningkatan tetapi belum sepenuhnya kembali ke angka pra-pandemi dan kemudian tahun 2023 menunjukan pemulihan mendekati jumlah normal untuk beberapa provinsi.

Insight Konseptual

Secara konseptual, temuan ini menegaskan pentingnya prinsip keadilan spasial dalam alokasi kuota haji. Distribusi berbasis jumlah penduduk memang efisien, tetapi berpotensi menimbulkan bias wilayah, terutama bagi provinsi dengan keterbatasan infrastruktur dan akses transportasi. Oleh karena itu, penelitian ke depan perlu mengintegrasikan indikator tambahan, misalnya rasio jamaah terhadap kapasitas embarkasi atau indeks aksesibilitas transportasi. Analisis tersebut belum dapat dilakukan pada tahap ini karena keterbatasan dataset, khususnya, data resmi Kementerian Agama mengenai kapasitas embarkasi per provinsi belum tersedia dalam dataset yang digunakan. Dengan demikian, pengembangan ini akan menjadi arah penelitian selanjutnya untuk memperkuat analisis dan mendukung kebijakan publik yang lebih adil.

Validasi dan Konsistensi Hasil

Semua koreksi dan transformasi divalidasi ulang terhadap file excel sumber. Catatan perubahan disimpan pada jupyter notebook agar bersifat reproduktif. Validasi menunjukan bahwa koreksi desimal terhadap 24 baris dan penghapusan duplikat pada satu baris tidak mengubah tren makro tetapi signifikan memperbaiki representasi numerik untuk visualisasi.

PENUTUP

Penelitian ini berhasil dilakukan eksperimen memanfaatkan pendekatan *data science* untuk menganalisis tren dan memvisualisasikan distribusi kuota jemaah haji di Indonesia periode 2015–2023 dengan menggunakan dataset resmi dari Kementerian Agama Republik Indonesia.

Melalui tahapan pra-pemrosesan data, transformasi geospasial, serta visualisasi heatmap interaktif, penelitian ini mampu memberikan gambaran spasial mengenai pola distribusi kuota antarprovinsi.

Secara metodologis, tahapan pembersihan data terbukti penting untuk menjamin validitas analisis, terutama dalam mengoreksi kesalahan numerik dan menghapus duplikasi. Integrasi koordinat geospasial serta normalisasi nilai juga memungkinkan hasil visualisasi lebih representatif.

Dengan demikian, penelitian ini tidak hanya menyajikan peta interaktif kuota haji lintas provinsi, tetapi juga memberikan kontribusi praktis sebagai alat bantu analisis bagi pemangku kebijakan, akademisi, maupun masyarakat umum dalam memahami distribusi kuota jemaah haji di Indonesia.

Kontribusi kebaruan penelitian ini adalah terletak pada integrasi *pipeline preprocessing,* transformasi geospasial, dan visualisasi interaktif yang belum dibahas secara mendalam pada penelitian sebelumnya. Selain itu, penelitian ini mengusulkan arah pengembangan lanjutan berupa penggunaan model analisis spasial kuantitatif (misalnya clustering provinsi berdasarkan rasio kuota terhadap jumlah penduduk muslim) serta pengembangan dashboard interaktif untuk mendukung transparansi kebijakan distribusi kuota.

DAFTAR PUSTAKA

[1] S. Wulandari, S. D. N. Azizi, and R. T. Hidayat, "Paradigma Ibadah Haji dan Umroh Ditinjau Berdasarkan Perspektif Hukum Islam dan Hukum Positif di Indonesia," Komparatif: Jurnal Perbandingan Hukum Dan Pemikiran Islam, vol. 3, no. 2, pp. 171–188, 2023.

[2] P. Chyan, Z. Gustiana, S. Arni, A. Yasir, H. Husain, B. A. Dermawan, and V. Afifah, Pengantar Data Science: Mengambil Keputusan Berdasarkan Data, vol. 1, no. 1. Penerbit Mifandi Mandiri Digital, 2024.

[3] S. A. Ajagbe, M. A. Oladipupo, and E. O. Balogun, "Crime belt monitoring via data visualization: A case study of Folium," International Journal of Information Security, Privacy and Digital Forensics, vol. 4, no. 2, pp. 35–45, 2020.

[4] D. Kim, B. Cánovas-Segura, M. Campos, and J. M. Juarez, "Visualization of spatial–temporal epidemiological data: A scoping review," Technologies, vol. 12, no. 3, p. 81, 2024.

[5] X. Tu, C. Fu, A. Huang, H. Chen, and X. Ding, "DBSCAN spatial clustering analysis of urban 'production–living–ecological' space based on POI data: A case study of central urban Wuhan, China," International Journal of Environmental Research and Public Health, vol. 19, no. 9, pp.1-21, 2022.

[6] S. Miftahurrahmi, Zilrahmi, N. Amalita, and T. O. Mukhti, "DBSCAN method in clustering provinces in Indonesia based on crime cases in 2022," Indonesian Journal of Data Science, vol. 2, no. 3, pp. 203–212, 2024.

[7] R. Rasmi, Repository Sebagai Sumber Informasi Proses Penyelesaian Studi Mahasiswa (Studi Kasus pada Mahasiswa IAIN Parepare), Doctoral dissertation, IAIN Parepare, 2024.

[8] C. E. Sukmawati, A. F. N. Masruriyah, A. R. Juwita, R. D. Tejayanda, and T. Nurmayanti, "Evaluasi Kinerja Algoritma AdaBoost dan XGBoost Menggunakan Dataset Penyakit Obesitas Pada Populasi Dewasa," Jambura Journal of Informatics, vol. 6, no. 2, pp. 101–111, 2024.

[9] M. A. D. Yuda, "Transformasi data solarman untuk pengungkapan informasi dan pola PLTS dengan metode semi-supervised learning," Jurnal MENTARI: Manajemen, Pendidikan dan Teknologi

Informasi, vol. 1, no. 2, pp. 100–110, 2023.

[10] I. N. Simbolon and P. D. Friskila, "Analisis dan evaluasi algoritma DBSCAN (Density-Based Spatial Clustering of Applications with Noise) pada tuberkulosis," Jurnal Informatika dan Teknik Elektro Terapan, vol. 12, no. 3S1, 2024.

[11] M. Daffa, D. S. Y. Kartika, and R. Permatasari, "Penerapan folium untuk visualisasi hasil clustering persebaran tindakan kriminal di Kota Surabaya," Jurnal Ilmiah Teknik Informatika dan Komunikasi, vol. 2, no. 3, pp. 44–53, 2022.

[12] A. G. Novarina, "Analisis tingkat kekritisan lahan daerah aliran sungai Brantas Hulu menggunakan sistem informasi geografis," Doctoral dissertation, ITN Malang, 2022.

[13] Y. F. Nurpratama, D. S. Y. Kartika, and R. Permatasari, "Implementasi folium pada hasil klaster diabetes mellitus di Puskesmas Modopuro," Jurnal Ilmiah Teknik Informatika dan Komunikasi, vol. 3, no. 3, pp. 74–85, 2023.

[14] A. Apriadsa, R. Laksono, and K. A. Pamungkas, "Pemanfaatan Arsip Peta Topografi Era Hindia Belanda pada Platform Digital Perpustakaan Universitas Leiden Sebagai Bahan Rujukan Kajian Batas Wilayah dan Toponim di Indonesia," Webinar Kearsipan, p. 54.

[15] A. M. A. Tambunan, U. Khaira, and B. F. Hutabarat, "Implementasi algoritma DBSCAN untuk klasterisasi koperasi berdasarkan indikator keuangan," Naratif: Jurnal Nasional Riset, Aplikasi dan Teknik Informatika, vol. 7, no. 1, pp. 38–55, 2025.

[16] Y. Nurdin and N. Khairun Saddami, Pengenalan Praktis Supervised Machine Learning: Dengan Jupyter Notebook. Banda Aceh, Indonesia: USK Press, 2025.

Halaman ini sengaja dikosongkan.