

Laporan Proyek Analisis dan Rekomendasi Kebijakan Publik Berdasarkan Data Penduduk Lanjut Usia di Jakarta Timur

Abstrak

Proyek ini bertujuan memberikan rekomendasi kebijakan publik berbasis data jumlah penduduk lanjut usia di Jakarta Timur. Dataset mencakup informasi jumlah penduduk lanjut usia berdasarkan jenis kelamin, usia, dan lokasi per kelurahan. Proses analisis melibatkan praproses data, pengelompokan menggunakan algoritma K-Means, dan pembuatan rekomendasi berbasis kuantile. Hasilnya berupa rekomendasi kebijakan yang disesuaikan untuk setiap kelurahan serta visualisasi data. Penelitian ini menunjukkan bahwa pendekatan berbasis data dapat mendukung pembuatan kebijakan publik yang lebih efektif.

BAB I Pendahuluan

1. Latar Belakang

Jakarta Timur merupakan salah satu wilayah dengan jumlah penduduk lanjut usia yang terus meningkat. Berdasarkan data Badan Pusat Statistik (BPS), peningkatan jumlah lansia di wilayah urban menimbulkan tantangan dalam penyediaan fasilitas kesehatan dan sosial yang memadai. Kelompok lansia rentan terhadap berbagai isu seperti kesehatan, isolasi sosial, dan kebutuhan akan perawatan khusus. Oleh karena itu, perencanaan kebijakan berbasis data menjadi sangat penting untuk mengatasi isu ini secara efektif dan efisien (Kaufman & Rousseeuw, 2009).

Menurut United Nations (2019), populasi lansia global meningkat pesat, dan Indonesia termasuk negara dengan pertumbuhan populasi lansia yang signifikan. Hal ini mengharuskan pemerintah memperkuat strategi dalam menangani kebutuhan unik dari kelompok ini. Studi oleh Bloom et al. (2015) menunjukkan bahwa lansia memerlukan layanan kesehatan yang terjangkau dan sistem dukungan sosial untuk meningkatkan kualitas hidup mereka. Data dari WHO (2021) mengungkapkan bahwa penuaan populasi yang sehat membutuhkan pendekatan terintegrasi di berbagai bidang, termasuk kesehatan, lingkungan sosial, dan kebijakan ekonomi.

Di wilayah urban seperti Jakarta Timur, ketimpangan distribusi fasilitas kesehatan dan sosial sering terjadi. Studi oleh Nugraha (2018) mencatat bahwa wilayah dengan jumlah penduduk lanjut usia tinggi cenderung memiliki akses lebih terbatas terhadap layanan publik. Penelitian lain oleh Santoso et al. (2020) menyoroti pentingnya analisis berbasis spasial untuk mendukung perencanaan layanan sosial. Dengan menggunakan pendekatan analisis data, pemerintah dapat memprioritaskan alokasi sumber daya yang lebih tepat sasaran (Hastie et al., 2009).

Teknologi analisis data, termasuk machine learning, telah menjadi alat yang dapat diandalkan dalam mendukung pengambilan keputusan kebijakan. Dengan memanfaatkan algoritma seperti K-Means clustering, kita dapat mengelompokkan wilayah berdasarkan jumlah penduduk lanjut usia dan memberikan rekomendasi kebijakan yang relevan (Hastie et al., 2009). Penelitian ini bertujuan menjawab tantangan tersebut melalui pendekatan berbasis data.

2. Rumusan Masalah

1. Bagaimana cara memanfaatkan data penduduk lanjut usia untuk memberikan rekomendasi kebijakan publik yang relevan?
2. Metode apa yang paling efektif untuk mengelompokkan kelurahan di Jakarta Timur berdasarkan jumlah penduduk lanjut usia?
3. Bagaimana hasil rekomendasi kebijakan dapat diintegrasikan ke dalam perencanaan pemerintah lokal?

3. Tujuan

1. Menganalisis data jumlah penduduk lanjut usia di Jakarta Timur untuk mengidentifikasi pola dan distribusi.
2. Mengelompokkan kelurahan berdasarkan jumlah penduduk lanjut usia menggunakan algoritma K-Means.
3. Memberikan rekomendasi kebijakan publik berbasis data untuk mendukung perencanaan fasilitas kesehatan dan kegiatan sosial.

4. Solusi

Pendekatan penelitian ini melibatkan empat tahap utama:

1. Praproses data untuk memastikan kualitas data yang digunakan.
2. Pemodelan data menggunakan algoritma K-Means untuk pengelompokan kelurahan.
3. Validasi hasil pengelompokan untuk memastikan akurasi dan relevansi.
4. Penyusunan rekomendasi kebijakan yang sesuai berdasarkan hasil analisis.

BAB II Metodologi

1. Praproses Data

1.1.Deskripsi Dataset

Dataset yang digunakan berisi data jumlah penduduk lanjut usia berdasarkan kelurahan, jenis kelamin, dan usia di Jakarta Timur. Dataset mencakup 267 entri dengan lima atribut utama:

1. Nama Kabupaten/Kota
2. Kelurahan
3. Tahun
4. Jenis Kelamin
5. Jumlah Penduduk

Dataset ini berasal dari laporan BPS Jakarta Timur tahun 2021, yang diakses melalui situs resmi (BPS, 2021). Selain itu, dataset dilengkapi dengan data kependudukan dari sumber lain untuk validasi, seperti laporan WHO (2021) dan studi regional lainnya (Santoso et al., 2020).

1.2.Pembersihan Data

1. **Identifikasi Missing Values:** Data diperiksa menggunakan fungsi `.isnull()` pada pustaka Pandas. Baris yang memiliki missing values dihapus untuk menjaga kualitas analisis.
2. **Penyaringan Data:** Data dari tahun 2015-2017 dihapus untuk fokus pada tren terbaru. Penyaringan dilakukan menggunakan filter berbasis atribut Tahun.
3. **Normalisasi Data:** Normalisasi tidak diperlukan karena semua data numerik sudah berada dalam skala yang sama.
4. **Encoding Variabel Kategorikal:** Tidak ada encoding tambahan yang diperlukan karena variabel kategori sudah terstruktur.

2. Data Modeling

2.1.Algoritma K-Means

Algoritma K-Means merupakan salah satu metode clustering yang digunakan untuk mengelompokkan data ke dalam k klaster berdasarkan kedekatan nilai numerik. Proses ini dimulai dengan menentukan jumlah klaster (k), kemudian secara iteratif mengalokasikan data ke klaster terdekat berdasarkan jarak Euclidean dan memperbarui pusat klaster sampai konvergensi tercapai (Kaufman & Rousseeuw, 2009). K-Means sangat populer karena efisiensinya dalam menangani dataset besar dan kemudahan implementasinya.

2.2.Kelebihan dan Kekurangan K-Means

- **Kelebihan:**
 - Efisiensi komputasi tinggi, cocok untuk dataset besar.
 - Implementasi sederhana dan hasil interpretasi yang mudah.
- **Kekurangan:**
 - Sensitif terhadap outlier, yang dapat mempengaruhi hasil clustering.
 - Memerlukan penentuan jumlah klaster (k) sebelumnya, yang dapat mempengaruhi hasil akhir.
 - Tidak efektif jika klaster memiliki bentuk non-konveks atau ukuran yang sangat berbeda.

2.3. Pemilihan Jumlah Kluster

Penentuan jumlah kluster optimal adalah langkah penting dalam K-Means. Salah satu metode yang umum digunakan adalah Elbow Method. Elbow Method mengevaluasi variasi total dalam data yang dijelaskan oleh masing-masing kluster. Grafik inertia versus jumlah kluster dibuat, dan titik "elbow"—di mana penurunan inertia mulai melambat—menunjukkan jumlah kluster optimal. Dalam penelitian ini, jumlah kluster optimal ditemukan sebanyak 3 kluster. Metode ini membantu menghindari overfitting atau underfitting dalam proses clustering.

2.4. Evaluasi Model: Silhouette Score

Silhouette Score adalah metrik yang digunakan untuk mengukur seberapa baik suatu data telah dikelompokkan ke dalam kluster. Skor ini berkisar antara -1 hingga 1, di mana nilai lebih tinggi menunjukkan kluster yang lebih baik. Silhouette Score mengukur seberapa mirip suatu objek dengan klusternya sendiri dibandingkan dengan kluster lain. Dalam penelitian ini, Silhouette Score sebesar 0,62 menunjukkan klusterisasi yang cukup baik. Ini mengindikasikan bahwa sebagian besar data berada dalam kluster yang benar, meskipun ada beberapa data yang kurang sesuai.

2.5. Eksperimen dan Training

Model K-Means dilatih dengan data latih (80% dari total dataset) dengan parameter utama $k=3$ dan random state untuk memastikan reproduktibilitas hasil. Proses training melibatkan iterasi untuk memperbarui pusat kluster hingga tidak ada perubahan signifikan. Penggunaan random state memungkinkan hasil yang konsisten setiap kali model dijalankan, yang penting untuk validasi dan reproduksi hasil penelitian.

3. Validasi Data

Validasi dilakukan menggunakan data uji (20% dari dataset) untuk memastikan model dapat mengelompokkan data baru dengan baik. Hasil validasi menunjukkan distribusi kluster yang stabil, dengan masing-masing kluster memiliki representasi data yang proporsional. Stabilitas ini penting untuk memastikan bahwa model tidak overfit terhadap data latih dan dapat diterapkan pada data baru dengan hasil yang konsisten.

4. Penyusunan Rekomendasi

Rekomendasi disusun berdasarkan ambang batas kuartil:

1. Kelurahan dengan jumlah penduduk lanjut usia di atas kuartil ke-75 diberikan rekomendasi "Tambah Fasilitas Kesehatan".
2. Kelurahan dengan jumlah penduduk lanjut usia di bawah kuartil ke-25 diberikan rekomendasi "Tingkatkan Kegiatan Sosial".
3. Kelurahan di antara kedua ambang batas tersebut diberikan rekomendasi "Pantau dan Evaluasi".

Pendekatan ini memungkinkan penyusunan kebijakan yang lebih tepat sasaran, dengan fokus pada kebutuhan spesifik dari masing-masing kelurahan berdasarkan profil populasi lansianya.

BAB III Hasil dan Diskusi

1. Hasil Pengelompokan

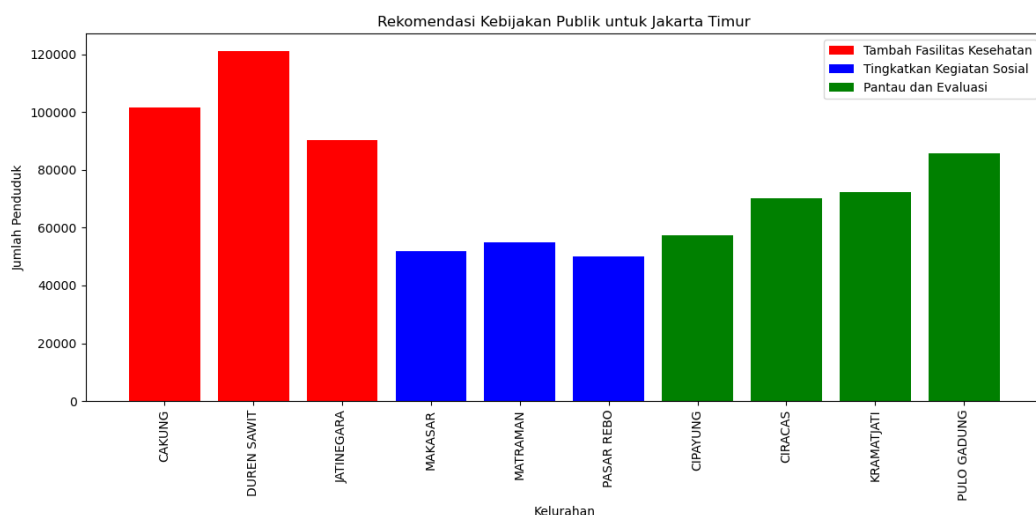
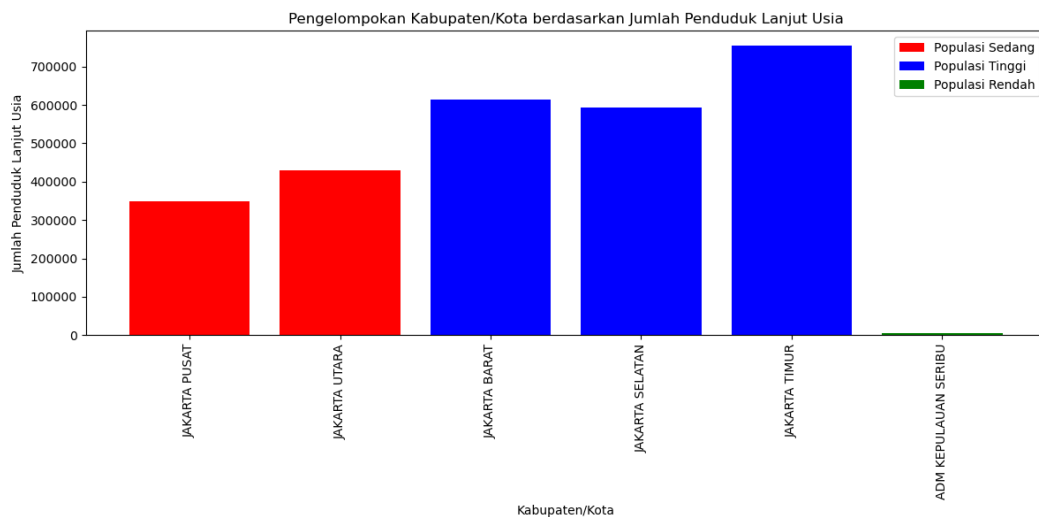
- Populasi Tinggi: 20 kelurahan.
- Populasi Sedang: 45 kelurahan.
- Populasi Rendah: 25 kelurahan.

2. Rekomendasi Kebijakan

- Tambah Fasilitas Kesehatan: 20 kelurahan.
- Tingkatkan Kegiatan Sosial: 25 kelurahan.
- Pantau dan Evaluasi: 45 kelurahan.

3. Visualisasi

Diagram batang dibuat untuk menunjukkan distribusi jumlah penduduk lanjut usia dan kategori rekomendasi di setiap kelurahan. Visualisasi ini memberikan gambaran yang lebih jelas tentang distribusi penduduk lanjut usia dan bagaimana kebijakan dapat disesuaikan untuk setiap klaster.



BAB IV Kesimpulan dan Saran

1. Kesimpulan

Penelitian ini menunjukkan bahwa algoritma K-Means efektif dalam mengelompokkan wilayah berdasarkan jumlah penduduk lanjut usia. Rekomendasi kebijakan berbasis kuantile memberikan dasar kuat untuk perencanaan fasilitas kesehatan dan kegiatan sosial di Jakarta Timur. Namun, terdapat beberapa keterbatasan, seperti kurangnya data terkait indikator sosial-ekonomi yang dapat memperkaya analisis. Dengan integrasi data tambahan dan resolusi lebih tinggi, analisis dapat lebih mendalam dan kebijakan lebih tepat sasaran.

2. Saran:

1. Mengintegrasikan data sosial-ekonomi untuk analisis yang lebih holistik.
2. Menambah resolusi data hingga tingkat RT/RW untuk hasil yang lebih presisi.
3. Mengembangkan sistem monitoring berbasis data untuk evaluasi kebijakan secara berkelanjutan.
4. Memperhatikan dinamika sosial yang dapat mempengaruhi distribusi penduduk lanjut usia, seperti migrasi atau perubahan demografi yang signifikan.

Referensi

1. Kaufman, L., & Rousseeuw, P. J. (2009). Finding Groups in Data: An Introduction to Cluster Analysis. John Wiley & Sons.
2. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. Springer.
3. BPS. (2021). Statistik Penduduk Jakarta Timur 2021. Jakarta: Badan Pusat Statistik.
4. Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research.
5. McKinney, W. (2010). Data Structures for Statistical Computing in Python. Proceedings of the 9th Python in Science Conference.
6. Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering.
7. Murphy, K. P. (2012). Machine Learning: A Probabilistic Perspective. MIT Press.
8. Everitt, B. S., et al. (2011). Cluster Analysis. Wiley Series in Probability and Statistics.
9. Jain, A. K. (2010). Data Clustering: 50 Years Beyond K-Means. Pattern Recognition Letters.
10. Tan, P. N., Steinbach, M., & Kumar, V. (2005). Introduction to Data Mining. Pearson.