



Surabaya, 6 Juli 2023

SEMINAR NASIONAL HASIL RISET DAN PENGABDIAN

“Peran Riset, Inovasi dan Pengabdian Kepada Masyarakat Bagi Pembangunan Indonesia Berkelanjutan”



Klasterisasi Kabupaten Kota di Jawa Timur Berdasarkan Kualitas Ketenagakerjaan Menggunakan K Means

Novi Rahmawati^{1*}, Faldianus Karno², Rezky Bryan Jonfris Purba³, Muhammad Athoillah⁴

^{1,2,4}Program Studi Statistika, Universitas PGRI Adi Buana Surabaya, Indonesia

³Enterprise Data Management, Bank Rakyat Indonesia, Indonesia

*Email: novi.rahmawati541@gmail.com

Abstrak

Ketenagakerjaan merupakan bagian integral dari kehidupan ekonomi dan sosial suatu negara, dengan tujuan menciptakan lingkungan kerja yang adil, produktif, aman, dan berkelanjutan bagi pekerja dan perusahaan. mengklasifikasikan kabupaten dan kota ke dalam kelompok-kelompok berdasarkan kualitas ketenagakerjaan, pemerintah dapat mengidentifikasi daerah-daerah yang memerlukan perhatian khusus, seperti daerah dengan tingkat pengangguran yang tinggi atau rendahnya akses pendidikan. Metode yang cocok digunakan adalah Algoritma *K-means*. sasi adalah proses pembagian satu set objek data ke dalam kelompok-kelompok yang disebut kluster, objek yang berada dalam satu kluster memiliki karakteristik yang serupa satu sama lain dan berbeda dengan kluster lainnya. Dalam penelitian ini, analisis data dilakukan menggunakan *software Python*. Penelitian ini menggunakan variabel pencari kerja terdaftar, TPT, Angkatan Kerja, dan gaji pegawai formal untuk menganalisis kualitas ketenagakerjaan di Jawa Timur. Metode klasterisasi K-Means digunakan untuk mengelompokkan observasi berdasarkan variabel-variabel tersebut. Melalui metode elbow dan metode silhouette, jumlah kluster optimal ditentukan menjadi 2. Hasil klasterisasi menunjukkan bahwa cluster 2 berisi provinsi-provinsi dengan kualitas ketenagakerjaan yang kurang baik, ditandai dengan tingginya jumlah pencari kerja, TPT, dan Angkatan Kerja meskipun rata-rata upah pekerja formal lebih tinggi. Cluster 1, di sisi lain, berisi provinsi-provinsi dengan kualitas ketenagakerjaan yang baik, dengan tingkat pengangguran yang rendah meskipun rata-rata upah pekerja formal lebih rendah. Cluster 2 dapat berpotensi mengancam perekonomian Jawa Timur di masa bonus demografi karena tingginya tingkat pengangguran, sedangkan cluster 1 memiliki potensi lebih baik karena tingkat pengangguran yang rendah.

Kata kunci: Ketenagakerjaan; K-Means; Python

PENDAHULUAN

Bonus demografi terjadi ketika terjadi peningkatan jumlah penduduk usia produktif dalam suatu negara atau wilayah. Peningkatan tersebut diikuti pula dengan menurunnya angka kelahiran serta kematian (Yusri, 2022). Fenomena ini memberikan peluang besar bagi pertumbuhan ekonomi dan sosial. Dengan populasi usia produktif yang besar, jika tenaga kerja

tersebut dapat dilatih dan terlibat dalam sektor ekonomi yang berkembang, maka dapat tercipta pertumbuhan ekonomi yang signifikan, peningkatan produktivitas, dan kemajuan sosial.

Ketenagakerjaan merupakan istilah yang menggambarkan segala hal yang terkait dengan tenaga kerja dalam suatu negara atau organisasi. Ini meliputi peraturan ketenagakerjaan, hubungan industrial, kebijakan ketenagakerjaan, pasar tenaga kerja, dan isu-isu yang melibatkan pekerja dan perusahaan. Ketenagakerjaan penting dalam menciptakan lingkungan kerja yang adil, produktif, aman, dan berkelanjutan bagi pekerja dan perusahaan. Kualitas tenaga kerja menjadi faktor kunci dalam penyediaan sumber daya manusia yang berkualitas.

Indonesia memiliki populasi yang besar dan banyak angkatan kerja. Namun, tingkat pengangguran dan underemployment masih menjadi masalah di beberapa daerah. Indonesia saat ini mengalami bonus demografi dengan sekitar 69,3% penduduk usia produktif (15-64 tahun) dari total 275,36 juta jiwa pada Juni 2022 (Kusnandar, 2020). Jawa Timur, provinsi kaya di Indonesia, memiliki populasi usia kerja yang besar. Meskipun proporsi penduduk usia produktif masih tinggi, kualitas ketenagakerjaannya masih perlu diperhatikan. Bonus demografi memberikan potensi tenaga kerja untuk mendukung pertumbuhan ekonomi. Dengan menggunakan metode K-means, variasi potensi ekonomi, peluang kerja, dan keterampilan tenaga kerja di Jawa Timur dapat diidentifikasi. Hasil klasterisasi ini membantu pemerintah merencanakan kebijakan pembangunan yang lebih efektif dan efisien dengan mengalokasikan sumber daya yang tepat untuk meningkatkan kualitas ketenagakerjaan di setiap klaster.

Beberapa penelitian dan kajian mengenai tenaga kerja telah dilakukan, diantaranya penelitian yang membahas mengenai implementasi metode K-Means dalam perbandingan tingkat pengangguran dengan tenaga kerja berdasarkan provinsi (Damanik et al., 2020). Dari penelitian tersebut, hasil klasterisasi menggunakan metode k-means, terdapat lebih banyak provinsi dengan tingkat pengangguran melebihi 57% dibandingkan dengan jumlah provinsi yang memiliki persentase tenaga kerja sebesar 43%.

Berdasarkan uraian permasalahan diatas, penelitian ini bertujuan untuk mengelompokkan kabupaten/kota di Jawa timur berdasarkan kualitas ketenagakerjaan, sehingga dapat diketahui bagaimana kualitas ketenagakerjaan di Jawa timur pada tahun 2022. Metode yang cocok digunakan adalah Algoritma K-means. Algoritma K-means dapat mengkategorikan atau pengelompokan sekelompok objek sesuai dengan atribut yang sama atau karakteristik ke dalam sejumlah grup yang mewakili *mean* dari *cluster*.

METODE

Penelitian ini menggunakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) Jawa Timur. Data yang digunakan dalam penelitian ini adalah data Jumlah Pencari Kerja Terdaftar, TPT, Angkatan Kerja, dan Gaji Pekerja Formal 38 kabupaten / kota di Jawa Timur.

Tabel 1. Variabel Penelitian

| No. | Variabel | Deskripsi | Tipe Data | Sumber Data |
|-----|-------------------------|---|-----------|---------------|
| 1. | Pencari Kerja Terdaftar | Jumlah Pencari Kerja Yang Terdaftar Pada Tahun 2022 | Numerik | www.bps.go.id |
| 2. | TPT | Tingkat Pengangguran Terbuka Seluruh Kabupaten / Kota di Jawa Timur Tahun 2022 | Numerik | www.bps.go.id |
| 3. | Angkatan Kerja | Seluruh penduduk Kabupaten / Kota di Jawa Timur Tahun 2022 yang bisa dan bersedia untuk melakukan pekerjaan | Numerik | www.bps.go.id |
| 4. | Gaji Pekerja Formal | Kompensasi yang diterima oleh individu di seluruh Kabupaten / Kota di Jawa Timur Tahun 2022 yang bekerja dalam pekerjaan formal | Numerik | www.bps.go.id |

Dalam penelitian ini, klasterisasi K-Means akan dilakukan menggunakan software Python. Klasterisasi adalah proses pembagian objek data ke dalam kelompok-kelompok berdasarkan karakteristik serupa menggunakan algoritma klasterisasi (IRWANSYAH, n.d.). *K-means* merupakan algoritma yang bersifat *unsupervised learning* untuk mengelompokkan observasi ke dalam *cluster*. Pada algoritma *k-means* setiap objek akan masuk ke kelompok tertentu dalam suatu proses tertentu sehingga data yang memiliki karakteristik berbeda dikelompokkan ke dalam *cluster* yang lain dan data yang mempunyai karakteristik sama dikelompokkan ke dalam kelompok yang sama (Yunita, 2018). Dalam melakukan analisis *K-Means* yang dilakukan antara lain (Izzadin, 2020), menentukan jumlah cluster yang akan dibentuk, secara acak mendistribusikan data cluster atau menginisialisasikan data (*centroid*), menghitung jarak data input terhadap masing-masing *centroid* dengan menggunakan jarak euclidean (*d*).

Dalam penelitian langkah-langkah penelitian yang dilakukan yakni,

Algoritma K means

Algoritmanya sebagai berikut (Rahman, 2017); menentukan jumlah k-cluster, membangkitkan k-centroid (titik pusat cluster) secara acak, menghitung jarak setiap data terhadap masing-masing centroid. Rumus yang digunakan adalah rumus jarak Euclidean (Euclidean Distance) (Rahman, 2017) dengan rumus sebagai berikut :

$$d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}, i = 1, 2, 3, \dots, n \quad (1)$$

Dimana $d(x_i, \mu_i)$ adalah jarak antara cluster n x dengan pusat cluster μ pada kata ke- i , x_i adalah bobot kata ke- i pada cluster yang ingin dicari jaraknya, μ_i bobot kata ke- i pada pusat cluster.

Mengelompokkan data berdasarkan jarak terdekat antara data dengan centroid. Menentukan nilai centroid yang baru dengan cara menghitung rata-rata dari cluster yang bersangkutan. Lakukan perulangan dari langkah 3 - 5 hingga anggota tiap cluster tidak ada yang berubah.

Metode Elbow

Algoritmanya sebagai berikut (Rahman, 2017); Menginisialisasi awal nilai k , menaikkan nilai k , menghitung hasil sum of square error dari tiap nilai k , analisis hasil sum of square error dari nilai k yang mengalami penurunan secara drastis, cari dan tentukan nilai k yang berbentuk siku. Pada metode Elbow nilai cluster terbaik yang akan diambil dari nilai Sum of Square Error (SSE) yang mengalami penurunan yang signifikan dan berbentuk siku. Untuk menghitung SSE menggunakan rumus berikut

$$SSE = \sum_{k=1}^K \sum_{i=1}^n \|x_i - c_k\|_2^2 \quad (2)$$

dimana K adalah jumlah cluster, x_i data ke - i , dan C_k centroid cluster. Sum of Square Error (SSE) merupakan rumus yang digunakan untuk mengukur perbedaan antara data yang diperoleh dengan model perkiraan yang telah dilakukan sebelumnya. SSE sering digunakan sebagai acuan penelitian terkait dalam menentukan optimal cluster.

Metode Silhouette

rumus yang digunakan adalah sebagai berikut (Izzadin, 2020); Hitung rata-rata jarak dari suatu dokumen misalkan i dengan semua dokumen lain yang berada dalam satu cluster.

$$a(i) = \frac{1}{|A|-1} \sum_{j \in A, j \neq i} d(i, j) \quad (3)$$

Lalu hitung rata-rata jarak dari dokumen i tersebut dengan semua dokumen di cluster lain, dan diambil nilai terkecilnya.

$$d(i, C) = \frac{1}{|A|} \sum_{j \in A, j \neq i} d(i, j) \quad (4)$$

dengan $d(i, C)$ adalah jarak rata-rata dokumen i dengan semua objek pada cluster lain C dimana $A \neq C$.

$$b(i) = \min_{C \neq A} d(i, C) \quad (5)$$

Nilai Silhouette Coefficient

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (6)$$

HASIL DAN PEMBAHASAN

Hasil dari analisis statistika deskriptif akan disajikan pada Tabel 2 dibawah ini.

Tabel 2. Statistik Deskriptif

| Variabel | Nama Variabel | Mean | Minimum | Maximum |
|----------------|-------------------------|---------------|-----------|-----------|
| X ₁ | Pencari Kerja Terdaftar | 3055,447 | 563 | 12.121 |
| X ₂ | TPT | 5,177 | 0 | 8,8 |
| X ₃ | Angkatan Kerja | 844.908,658 | 104,250 | 2.348.058 |
| X ₄ | Gaji Pegawai Formal | 2.323.516,921 | 1.441,196 | 3.861.713 |

Berdasarkan Tabel 2 di atas, dapat diketahui jika jumlah pencari kerja terkecil yaitu milik Mojokerto sebesar 563 orang. Sedangkan yang terbesar yaitu milik kota Surabaya, sebesar 121.121 orang. Hal ini menunjukkan bahwa jumlah lapangan pekerjaan yang ada di Surabaya tidak sebanding dengan jumlah angkatan kerjanya. Hal ini juga didukung dengan fakta bahwa TPT kota Surabaya tergolong tinggi (7,62). Angka ini tidak berbeda terlalu jauh jika dibandingkan dengan Kabupaten Sidoarjo yang memiliki TPT tertinggi sebesar 8,8.

Setelah mengetahui beberapa informasi awal dari data yang digunakan, kita dapat masuk ke dalam proses pengelompokan provinsi di Indonesia menggunakan algoritma *K-Means*. Namun sebelum itu, jumlah *cluster* optimal haruslah ditentukan terlebih dahulu. Penentuan jumlah *cluster* optimal ini akan dilakukan dengan menggunakan metode *Elbow* dan Metode *Silhouette*. Metode *Elbow* menggunakan nilai Total *Within Sum of Square* untuk menentukan jumlah *cluster* mana yang paling optimum. Dimana, apabila nilai total *within sum of square* suatu *cluster* mengalami penurunan yang paling banyak (membentuk *elbow*) dan pada *cluster* selanjutnya penurunan nilai total *sum of square* sudah stabil, maka *cluster* tersebut merupakan *cluster* paling optimum. Sedangkan pada metode *silhouette*, penentuan k optimal didasarkan pada rata-rata koefisien *silhouette*-nya. Nilai ini berkisar pada -1 sampai 1. Dimana semakin besar rata-ratanya (mendekati 1), maka semakin optimal pula *cluster* tersebut.

Setelah melakukan metode *elbow*, diketahui bahwa penurunan terbesar nilai total *within sum of square* terbesar terletak pada *cluster* 2. Lalu pada metode *silhouette* didapatkan rata-rata nilai *silhouette* terbesar terletak pada *cluster* yang berjumlah 2 yaitu sebesar 0,334. Karena kedua metode menunjukkan bahwa jumlah *cluster* optimum sebanyak 2, maka kabupaten / kota akan dikelompokkan menjadi 2 *cluster*.

Setelah proses pengelompokkan dilakukan pada 38 kabupaten / kota di Jawa Timur, didapatkan 2 *cluster* dimana terdapat 28 provinsi pada *cluster* 1, 10 provinsi pada *cluster* 2.

Namun sebelum itu, akan dilakukan feature scaling agar data memiliki rentang nilai (scale) yang sama. Hasil pengklasteran secara lengkap akan disajikan pada tabel di bawah ini.

Tabel 3. Center Cluster

| Variabel | Cluster | |
|----------|---------|-------|
| | 1 | 2 |
| μ_1 | -0,463 | 1,136 |
| μ_2 | -0,282 | 0,692 |
| μ_3 | -0,430 | 1,056 |
| μ_4 | -0,309 | 9,09 |

Tabel 3 di atas menunjukkan titik pusat 2 cluster yang telah didapat pada masing masing variabel.

Tabel 4. Hasil Pengelompokkan Wilayah Berdasarkan Kesehatan Ibu dan Anak

| Cluster | Kabupaten / Kota |
|---------|---|
| 1. | Pacitan, Ponorogo, Trenggalek, Tulungagung, Blitar, Lumajang, Banyuwangi, Bondowoso, Situbondo, Probolinggo, Nganjuk, Madiun, Magetan, Ngawi, Bojonegoro, Tuban, Bangkalan, Sampang, Pamekasan, Sumenep, Kediri, Blitar, Probolinggo, Pasuruan, Mojokerto, Madiun, Batu |
| 2. | Kediri, Kabupaten Malang, Jember, Pasuruan Sidoarjo, Mojokerto, Jombang, Lamongan, Gresik, Malang, Surabaya |

Berdasarkan hasil pengclustering di atas, *cluster 2* merupakan *cluster* yang berisi provinsi-provinsi dengan kualitas ketenagakerjaan yang kurang baik hal ini terbukti dari center cluster 2 yang menunjukkan jumlah pencari kerja terdaftar, TPT dan Angkatan Kerja yang lebih tinggi dibanding cluster 1. Meskipun kabupaten / kota yang ada pada cluster 1 memiliki rata-rata upah pekerja formal yang lebih tinggi, apabila jumlah pencari kerja, TPT, dan Angkatan Kerja tetap tinggi hal ini tetap akan mengancam perekonomian Jawa Timur di masa bonus demografi sebab masih banyak orang yang menganggur.

Sedangkan cluster 1 merupakan *cluster* yang berisi provinsi-provinsi dengan kualitas ketenagakerjaan yang baik. Meskipun kabupaten / kota yang ada pada cluster 2 memiliki rata-rata upah pekerja formal yang lebih rendah, apabila jumlah pencari kerja, TPT, dan Angkatan Kerja rendah, hal ini tidak akan terlalu mengancam perekonomian Jawa Timur sebab tidak banyak orang yang menganggur. Berdasarkan hasil diatas, diharapkan kepada pemerintah untuk bisa membuka peluang kerja yang lebih merata di provinsi Jawa Timur. Hal ini bertujuan untuk mengatasi jumlah pencari kerja dan pengangguran yang ada di kota-kota besar. Apabila ada banyak kesempatan kerja di seluruh kabupaten / kota, akan ada lebih sedikit orang yang memutuskan untuk merantau dan meningkatkan jumlah pengangguran di kota kota besar seperti Surabaya dan Sidoarjo.

KESIMPULAN

Penelitian kali ini menggunakan variabel pencari kerja terdaftar, TPT, Angkatan Kerja dan gaji pegawai formal untuk bisa mengetahui kualitas ketenagakerjaan yang ada di Jawa Timur. Agar bisa mendapatkan kualitas yang sesuai, digunakanlah metode K-Means. Sebelum dilakukan proses pengclusteran, jumlah cluster optimal ditentukan menggunakan metode elbow dan metode silhouette yang sama sama menunjukkan bahwa jumlah cluster yang optimum yaitu 2 sehingga, proses pengclusteran akan menghasilkan 2 cluster. Berdasarkan hasil pengclusteran, *cluster 2* merupakan *cluster* yang berisi provinsi-provinsi dengan kualitas ketenagakerjaan yang kurang baik. Sedangkan *cluster 1* merupakan *cluster* yang berisi provinsi-provinsi dengan kualitas ketenagakerjaan yang baik. Berdasarkan hasil tersebut, diharapkan kepada pemerintah untuk bisa membuka peluang kerja yang lebih merata di provinsi Jawa Timur agar permasalahan mengenai ketenagakerjaan ini dapat teratasi.

DAFTAR PUSTAKA

- BPS. (2022). *Tenaga Kerja Jawa Timur*. <https://research.tue.nl/en/publications/the-euclidean-distance-degree>
- Damanik, A. R., Defit, S., Hartama, D., Fikrul, P. P. P. A. N. W., Zer, I. R. H., Putra, U., Padang, I., Tunas, S., & Pematangsiantar, B. (2020). Implementasi Metode K-Means Dalam Perbandingan Tingkat Pengangguran Dengan Tenaga Kerja Berdasarkan Provinsi. *Prosiding Seminar Nasional Riset Dan Information Science (SENARIS)*, 2, 24–31. <https://www.bps.go.id>.
- Gamas, I. R. (2021). *Tenaga Kerja Sebagai Major Equipment Dalam Fase Bonus Demografi*. <https://www.ombudsman.go.id/artikel/r/artikel--tenaga-kerja-sebagai-major-equipment-dalam-fase-bonus-demografi>
- IRWANSYAH, E. (n.d.). *CLUSTERING*. <https://socs.binus.ac.id/2017/03/09/clustering/>
- Izzadin, F. M. (2020). Optimasi Jumlah Cluster K-Means Dengan Metode Elbow Dan Silhouette Pada Produktivitas Tanaman Pangan Di Provinsi Jawa Tengah Tahun 2018. *file:///C:/Users/VERA/Downloads/ASKEP_AGREGAT_ANAK_and_REMAJA_PRINT.docx*, 21(1), 1–9.
- Izzah, L. ', & Jananto, A. (2022). Penerapan Algoritma K-Means Clustering Untuk Perencanaan Kebutuhan Obat Di Klinik Citra Medika. *Progresif: Jurnal Ilmiah Komputer*, 18(1), 69. <https://doi.org/10.35889/progresif.v18i1.769>
- Kusnandar, V. B. (2020). *Era Bonus Demografi, 69% Penduduk Indonesia Masuk Kategori Usia Produktif pada Juni 2022*. <https://databoks.katadata.co.id/datapublish/2022/09/30/era-bonus-demografi-69-penduduk-indonesia-masuk-kategori-usia-produktif-pada-juni-2022>
- Listiani, L., Agustin, Y. H., & Ramdhani, M. Z. (2019). Implementasi algoritma k-means cluster untuk rekomendasi pekerjaan berdasarkan pengelompokkan data penduduk. *SENSITIF: Seminar Nasional Sistem Informasi dan Teknologi Informasi*, 761–769.
- Lkpd, P., Pemerintah, D. I., & Provinsi, K. D. I. (2021). *Analisis Faktor-Faktor Yang Mempengaruhi Tingkat. 1*(December 2020), 232–237. <https://doi.org/10.52062/jakd.v14i2.1457>
- Maliqi, R., Falgenti, K., Priani, S., Fithri, F., Suherman, M., & Nugraha, D. S. (2022). Penerapan Metode K-Means Clustering pada Data Tingkat Pengangguran Terbuka Tahun 2016-2018 dan 2019-2021. *Computer Science (CO-SCIENCE)*, 2(2), 109–116. <https://doi.org/10.31294/coscience.v2i2.1151>
- Meilia, A.T; Erlangga, G. (2022). Aktualisasi Program Kampus Mengajar Sebagai Ruang Kontribusi Mahasiswa Terhadap Pendidikan Dasar Di Indonesia. *Metodik Didaktik; Jurnal Pendidikan*

Ke-SD-an, 17(2), 120–128.

Rahman, A. T. (2017). *COAL TRADE DATA CLUSTERING USING K-MEANS (CASE STUDY PT. GLOBAL BANGKIT UTAMA)*.

Yunita, F. (2018). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru. *Sistemasi*, 7(3), 238.
<https://doi.org/10.32520/stmsi.v7i3.388>

Yusri, N. H. (2022). *Siapakah Generasi Indonesia 2045 Hadapi Bonus Demografi?*
<https://www.its.ac.id/news/2022/11/05/siapakah-generasi-indonesia-2045-hadapi-bonus-demografi/>