

Penerapan Algoritma K-Means Dengan Optimasi Jumlah Cluster Untuk Pengelompokan Angkatan Kerja Propinsi Jatim

Endik Kuswantoro¹, Yoyon K Suprpto²

Jurusan Teknik Elektro
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia

endik14@mhs.ee.its.ac.id¹, yoyonsuprpto@gmail.com²

Abstract—This Jumlah angkatan kerja di Indonesia terus meningkat seiring dengan pertambahan jumlah penduduk. Semakin besar jumlah penduduk maka angkatan kerja jadi semakin besar. Hal itu dapat menjadi beban tersendiri bagi perekonomian. Karena jika meningkatnya angkatan kerja yang tidak diimbangi dengan bertambahnya lapangan kerja akan menyebabkan masalah pengangguran. Kondisi tersebut dapat menyebabkan kesejahteraannya menurun. Oleh karena itu permasalahan pengangguran juga tidak terlepas dari bagian jumlah angkatan kerja, Pada Propinsi Jawa Timur juga mengalami permasalahan tersebut. Dalam penelitian ini akan bertujuan untuk mendapatkan Pengelompokan Angkatan kerja pada wilayah propinsi Jawa timur dengan menggunakan algoritma K-Means, dengan pemodelan tersebut akan menghasilkan tingkat penganggurannya dari hasil masing – masing cluster yang dihasilkan, dan persebaran kelompok – kelompok tenaga kerja di pedesaan dan perkotaan, sehingga bisa memberikan informasi kebutuhan tenaga kerja apa saja yang ada di propinsi Jawa Timur. setelah dilakukan pengklasteran maka hasil yang didapat akan di visualisasi ke dalam grafik chart. Dari proses pengelompokan dengan algoritma K-Means dari jumlah sample sebanyak 17.576 sample rumah tangga didapat 2 (Dua) kluster yang optimal yang mampu mewakili analisa data yaitu kluster 1 memiliki jumlah anggota kluster paling banyak dan mempunyai karakteristik pengangguran sebanyak 7.936 rumah tangga sample terdiri dari 2.982 sample data tergolong pengangguran setengah dan pekerja paruh waktu sisanya sebesar 4.954 tergolong pekerja aktif diprosentasekan cluster 1 sebanyak 92,59 %. Kluster 2 (dua) kluaster yang mempunyai karakteristik pengangguran Terbuka sebanyak 730 rumah tangga atau sebanyak 8,42 % dari jumlah rumah tangga sample, faktor paling banyak penyebab pengangguran terbuka ialah merasa putus asa dalam memperoleh pekerjaan bisa dikarenakan cacat fisik atau lainnya.

Keywords—*Pengelompokan, Angkatan Kerja, K-means, Propinsi Jawa Timur.*

I. PENDAHULUAN

Masalah ketenagakerjaan memang masalah yang sering terjadi pada negara – negara berkembang seperti Indonesia,

semakin bertambahnya jumlah penduduk yang tidak diiringi dengan jumlah lapangan pekerjaan yang memadai menyebabkan jumlah pengangguran yang tinggi, Dengan kata lain lapangan kerja yang ada tidak mampu menampung (mempekerjakan) tenaga kerja yang ada, lebih-lebih tenaga kerja yang tidak terampil atau berpendidikan. Masalah ini akan menyebabkan semakin meningkatnya tingkat pengangguran sehingga jumlah penduduk miskin juga semakin besar dan memiliki efek-efek negatif yang lain pula. Masalah ketenagakerjaan di Indonesia sekarang ini sudah mencapai kondisi yang cukup memprihatinkan ditandai dengan jumlah penganggur dan setengah penganggur yang besar, pendapatan yang relatif rendah dan kurang merata. Sebaliknya pengangguran dan setengah pengangguran yang tinggi merupakan pemborosan pemborosan sumber daya dan potensi yang ada, menjadi beban keluarga dan masyarakat, sumber utama kemiskinan, dapat mendorong peningkatan keresahan sosial dan kriminal, dan dapat menghambat pembangunan dalam jangka panjang. Permasalahan dalam ketenagakerjaan juga disebabkan karena kekurangan kebutuhan kompetensi dan keahlian yang dibutuhkan oleh pengguna tenaga kerja, hal itu disebabkan oleh tidak meratanya kebutuhan kompetensi tenaga kerja di masing – masing daerah, karena tidak ada pemetaan oleh pemerintah untuk memetakan kebutuhan kompetensi keahlian tenaga kerja yang dibutuhkan oleh pengguna tenaga kerja. Dalam penelitian ini bertujuan untuk memetakan tingkat pengangguran dan kompetensi tenaga kerja yang bertempat di Propinsi Jawa Timur.

Dalam penelitian sebelumnya yang berkaitan dengan permasalahan ketenagakerjaan dan pengangguran yaitu menentukan tingkat pengangguran terbuka dan faktor penyebabnya dengan menggunakan metode klasifikasi dengan hasil akurasi 78,9 % sehingga diperlukan penelitian menggunakan metode yang berbeda [2].

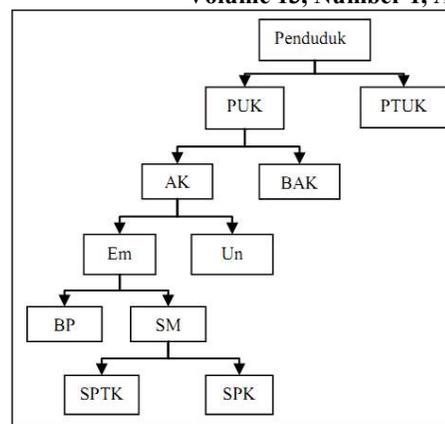
Dalam penelitian ini akan dilakukan pemodelan tingkat angkatan kerja di propinsi Jawa Timur dengan menggunakan algoritma K-means, K-Means merupakan salah satu metode data clustering non hirarki yang berusaha mempartisi data yang ada

ke dalam bentuk satu atau lebih cluster/kelompok. Metode ini mempartisi data ke dalam cluster/kelompok sehingga data yang memiliki karakteristik samadikelompokkan ke dalam satu cluster yang sama [3]

II. DASAR TEORI

A. Ketenagakerjaan dan Angkatan Kerja

Penduduk adalah semua orang yang mendiami suatu wilayah tertentu pada waktu tertentu dan terlepas dari warga Negara atau bukan warga negara [5]. Penduduk adalah mereka yang bertempat tinggal atau berdomisili didalam suatu wilayah Negara. Jumlah penduduk yang besar disuatu negara tidak otomatis akan menjadi modal pembangunan, bahkan dapat pula justru menjadi beban dan tanggungan penduduk lainnya. Penduduk dapat diklasifikasikan kedalam 5 level. Penduduk pada penelitian ini terbagi menjadi dua kategori dan merupakan level 2 yaitu penduduk usia kerja (PUK) dan penduduk tidak usia kerja (PTUK). Penduduk usia kerja sendiri terbagi menjadi dua kelompok, yaitu angkatan kerja (AK) dan bukan angkatan kerja (BAK). Level 4 dari pembagian penduduk adalah penduduk yang tergolong dalam angkatan kerja terbagi menjadi dua kategori, yaitu pengangguran (Unemploy) dan bekerja (Employ). Bekerja penuh (BP) dan setengah menganggur (SM) merupakan kategori dari angkatan kerja yang tergolong bekerja. Setengah menganggur dibagikembali menjadi dua kategori, yaitu setengah pengangguran tidak ketara (SPTK) dan setengah pengangguran ketara (SPK) [6]. Klasifikasi penduduk tersebut memperlihatkan bahwa penduduk yang merupakan angkatan kerja dan tergolong pengangguran dapat menjadi beban penduduk yang lainnya. Pengangguran terbuka merupakan bagian dari angkatan kerja yang tidak bekerja atau sedang mencari pekerjaan (baik bagi mereka yang belum pernah bekerja sama sekali maupun yang sudah pernah bekerja), atau sedang mempersiapkan suatu usaha, mereka yang tidak mencari pekerjaan karena merasa tidak mungkin untuk mendapatkan pekerjaan dan mereka yang sudah memiliki pekerjaan namun belum memulai pekerjaan [7]. Pengangguran adalah jumlah tenaga kerja dalam perekonomian yang secara aktif mencari pekerjaan tetapi belum memperolehnya [4]. Tingkat pengangguran dapat diketahui dari hasil Survei Angkatan Kerja Nasional (SAKERNAS) yang dilakukan secara triwulan setiap tahunnya mulai tahun 2011, yaitu Februari, Mei, Agustus, dan Nopember. Usia, pendapatan rumah tangga/keluarga, tingkat pendidikan, ketrampilan dan pelatihan kerja mempengaruhi seseorang untuk bekerja [6]. Digambarkan dalam Gambar 1 Penduduk dan tenaga kerja



Gambar 1 Penduduk dan Tenaga Kerja

Keterangan Gambar 1:

- PUK : Penduduk Usia Kerja
- PTUK : Penduduk Tidak Usia Kerja
- AK : Angkatan Kerja
- BAK : Bukan Angkatan Kerja
- Em : Employ (Pekerja)
- Un : Unemploy (Bukan Pekerja)
- BP : Bekerja Penuh
- SM : Setengah Menganggur
- SPTK : Setengah Menganggur Terpaksa
- SPK : Setengah Penganggur Kentara/Sukarela (Pekerja Paruh waktu)

B. Metoda K-Means

Algoritma *K-Means Clustering* merupakan algoritma pengelompokan iteratif yang melakukan partisi set data ke dalam sejumlah K cluster yang sudah di tetapkan di awal. Algoritma *K-Means Clustering* sederhana untuk diimplementasikan dan dijalankan, relatif cepat, mudah beradaptasi serta umum penggunaannya dalam paraktek [1]. Algoritma ini disusun atas dasar ide yang sederhana. Pada awalnya ditentukan berapa cluster yang akan dibentuk. Sebarang obyek atau elemen pertama dalam cluster dapat dipilih untuk dijadikan sebagai titik tengah (centroid point) cluster. Algoritma metode K-Means selanjutnya akan melakukan pengulangan langkah-langkah berikut sampai terjadi kestabilan (tidak ada obyek yang dapat dipindahkan) [4]:

Langkah-langkah dalam algoritma k-means adalah:

1. Ditentukan initial partition dengan k klaster berisi sample yang dipilih secaraacak, kemudian dihitung pusat klaster dari tiap-tiap klaster.
2. Membangkitkan partisi baru dengan penugasan setiap sample terhadap pusat

klaster terdekat denganpersamaan dari pengukuran jarak menggunakan Euclidean ditunjukkan ke dalam persamaan (2.1) berikut:

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2} \quad (1)$$

dimana:

$$d(x_i, x_j) = \text{jarak dari } x_i \text{ dan } x_j$$

- Menghitung pusat-pusat kluster baru. Dengan cara menghitung rata – rata pada masing – masing partisi, persamaan (2.2) merupakan persamaan perhitungan centroid baru

$$\bar{x}C_m = \sum_{i=1}^N \frac{x_i c_m}{N} \quad (2)$$

Dimana:

$\bar{x}C_m$ = centroid baru yang merupakan rata – rata dari data yang berada dalam cluster

$X_i C_m$ = data X ke i pada centroid m

N = Jumlah data

- Mengulangi langkah 2 dan 3 sampai nilai optimum dari fungsi kriteria dipenuhi (atau sampai anggota kluster telah stabil

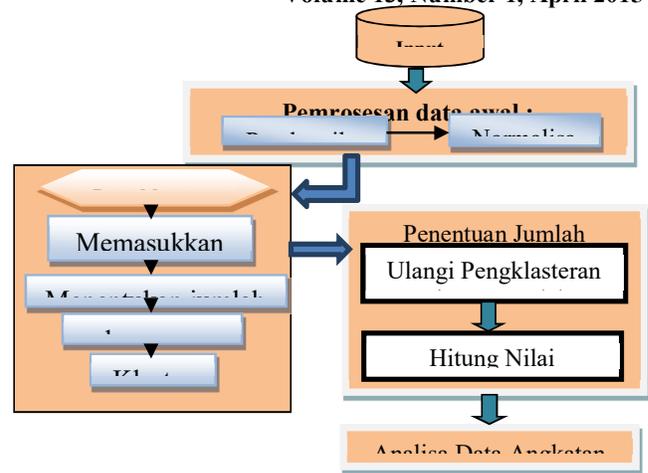
III. METODE PENELITIAN

Berdasarkan tujuan penelitian, berikut ini akan diuraikan metode dan tahapan-tahapan penelitian yang akan dilakukan. Proses dasar pengembangan secara umum pada penelitian ini terdiri dari beberapa tahapan sesuai Gambar 2 Metodologi penelitian.

Pengambilan Data Masukan; Data masukan yang digunakan dalam penelitian adalah data survey Angkatan Kerja atau SAKERNAS triwulan I bulan Februari Propinsi Jawa Timur yang didalamnya terdapat informasi data tentang Angkatan Kerja dan Pengangguran, data awal pengambilan data sebanyak 13.266 sample data **Pembersihan Data;** Pembersihan data dilakukan dengan memperhatikan konsep dan tujuan dari penelitian yaitu pemodelan angkatan kerja, data yang kurang relevan dibersihkan agar tidak melenceng dari tujuan **Normalisasi Data;** Normalisasi data dilakukan untuk memperbaiki data apabila ada data yang hilang terhapus atau ada beberapa data yang salah ketik seharusnya nilai nominal tapi di masukan nilai karakter.

Proses Pengklasteran; Data yang sudah melalui konversi data kemudian di lakukan pengklasteran ,pada tahap ini dilakukan pengklasteran dengan menggunakan algoritma K-means

Proses Penentuan Kluster Optimal; Proses penentuan kluster optimal dengan menentukan jumlah cluster yang ideal dengan cara melakukan uji coba pengklasteran untuk menghasilkan 2 sampai 10 cluster, setelah diklasterkan hasil dari pengelompokan di validasi dengan Shiloutte index dimana nilai yang paling mendekati 1 adalah jumlah kluster yang paling ideal. **Analisa Data Angkatan Kerja;** Pada proses ini adalah hasil analisa dari pengklasteran dari jumlah kluster yang optimal.



Gambar 2 Metodologi Penelitian

IV. PENGUJIAN DAN PEMBAHASAN

Pengklasteran; Tahapan ini dilakukan pengujian *klustering* menggunakan algoritma K-Means dengan aplikasi Orange Canvas. Data uji diambil dari data SAKERNAS (Survey Angkatan Kerja) Tahun 2014 sebanyak 13.866 record atau sebanyak 13.866 sample rumah tangga, data tersebut sudah melalui proses normalisasi data dan pembersihan data sampai tersisa 8.665 sample rumah tangga yang dijadikan data set. Data tersebut terdiri dari 6 fitur atau attribut data yang masing – masing fitur tersebut berhubungan dengan karakteristik pengangguran yaitu bekerja Seminggu yang lalu, masih_cari_kerja, jam_kerja_dalam_1_minggu, sementara_tidak_kerja, cari_usaha, alasan_tidak_kerja. Kemudian data tersebut dapat dikelompokkan menggunakan Algoritma K-means, untuk mengelompokkan data – data tersebut menjadi beberapa cluster dilakukan langkah – langkah sebagai berikut yaitu:

- Tentukan jumlah cluster yang diinginkan, dalam penelitian ini ditentukan menjadi 2 cluster
- Tentukan titik pusat awal cluster dari setiap cluster, dalam penelitian ini titik pusat cluster ditentukan secara random, dapat dilihat pada table 1 centroid awal sebagai berikut:

Tabel 1 Centroid Awal cluster

Titik Pusat Cluster	Bekerja Seminggu lalu	Sementara_tidak_kerja	Cari_kerja	Jam_kerja_seminggu
Cluster 1	1	1	1	10
Cluster 2	2	2	2	40

- Tempatkan setiap data pada cluster. Dalam penelitian ini digunakan metode k-means untuk mengalokasikan setiap data ke dalam suatu cluster, sehingga data akan dimasukan dalam suatu cluster yang memiliki jarak paling dekat dengan titik pusat dari setiap cluster. Untuk mengetahui cluster mana yang paling dekat dengan data, maka perlu dihitung jarak setiap data dengan titik pusat setiap cluster, perhitungan ini menggunakan perhitungan jarak euclidean. yang dirumuskan pada rumus 1 berikut [8]

$$D(i, j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2} \dots (1)$$

Dimana:

- D(i,j) : Jarak data ke i ke pusat j
- X_{ki} : Data ke i pada Atribut data ke k
- X_{kj} : Titik pusat ke j pada Atribut data ke k

Sebagai contoh, akan dihitung jarak dari data pertama ke pusat cluster pertama,

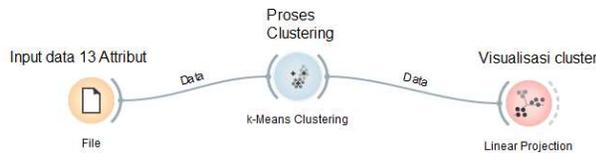
$$D(1,1): \sqrt{(2-1)^2 + (2-1)^2 + (1-1)^2 + (35-10)^2}$$

$$D(1,2): \sqrt{(1-2)^2 + (1-2)^2 + (2-2)^2 + (20-40)^2}$$

Dst...

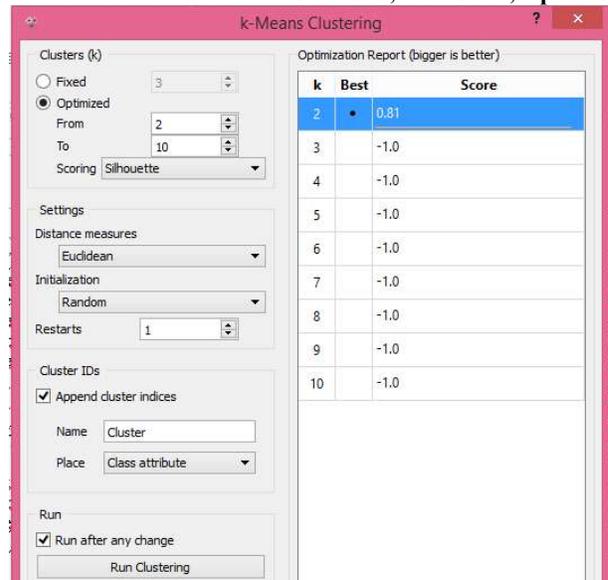
4. Setelah semua data ditempat ke dalam cluster yang terdekat, kemudian hitung kembali pusat cluster yang baru berdasarkan rata-rata anggota yang ada pada cluster tersebut
5. Setelah didapatkan titik pusat yang baru dari setiap cluster, lakukan kembali dari langkah ketiga hingga titik pusat dari setiap cluster tidak berubah lagi dan tidak ada lagi data yang berpindah dari satu cluster ke cluster yang lain.

Dalam penelitian ini proses pengelompokan data uji dilakukan dengan software Orange Canvas, berikut tampilan dashboard Orange Canvas pada gambar 3



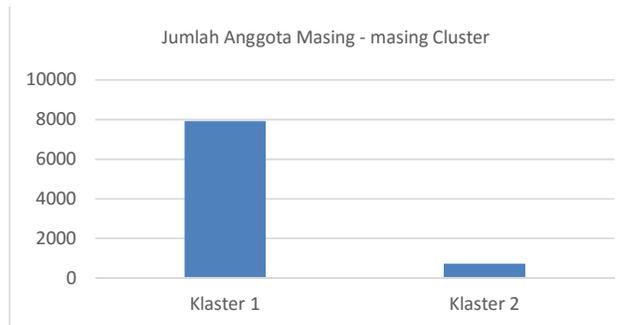
Gambar 3 Dashboard Orange Canvas untuk Alur Clustering

Kemudian setelah data tersebut dilakukan pengklasteran dengan menggunakan Algoritma K-Means pada software Orange Canvas yang menghasilkan 2 buah cluster optimal, kemudian 2 kluster tersebut mempunyai karakteristik masing – masing dimana cluster 1 memiliki jumlah anggota cluster paling banyak dan mempunyai karakteristik pengangguran sebanyak 7.936 rumah tangga sample terdiri dari 2.982 sample data tergolong pengangguran setengah dan pekerja paruh waktu sisanya sebesar 4.954 tergolong pekerja aktif dipersentasekan cluster 1 sebanyak 92,59 %. Kluster 2 (dua) kluster yang mempunyai karakteristik pengangguran Terbuka sebanyak 730 rumah tangga atau sebanyak 8,42 % dari jumlah rumah tangga sample, faktor paling banyak penyebab pengangguran terbuka ialah merasa putus asa dalam memperoleh pekerjaan bisa dikarenakan cacat fisik atau lainnya. Pada Gambar 4 Perhitungan jumlah cluster yang optimal dengan menggunakan Software Orange Canvas sebagai berikut:



Gambar 2 Penentuan Jumlah Cluster Optimal

Pada Gambar 3 Grafik Jumlah Cluster didapat jumlah anggota pada masing-masing kluster pada pengelompokan tingkat Angkatan kerja propinsi Jawa Timur, dimana cluster 1 mempunyai karakteristik Pekerja Aktif dan Pekerja Paruh Waktu, cluster 2 merupakan karakteristik Pengangguran Terbuka



Gambar 3 Jumlah anggota masing – masing Cluster

V. KESIMPULAN

Kesimpulan dari hasil penelitian ini adalah sebagai berikut:

- Metode clustering dengan algoritma *K-Means Clustering* untuk melakukan pengelompokan guna mengetahui tingkat pengangguran dan tingkat kompetensi tenaga kerja dalam satu propinsi dengan menggunakan metode *k-means clustering*.
- Data yang digunakan dari Data Survey Angkatan Kerja Nasional (SAKERNAS) Tahun 2014 Bulan Februari pada propinsi Jawa Timur, dengan jumlah data sample yang terpilih sebanyak 17.576 rumah tangga sample
- Data yang digunakan untuk pengelompokan berjumlah 13 atribut yaitu bekerja seminggu lalu, masih cari kerja, jam kerja dalam 1 minggu, sementara tidak kerja, cari usaha, alasan tidak kerja

- Dari pengelompokan dengan menggunakan metode K-means menghasilkan 2 (dua) kluster yang mampu mewakili analisa data yaitu kluster 1 memiliki jumlah anggota kluster paling banyak dan mempunyai karakteristik pengangguran sebanyak 7.936 rumah tangga sample terdiri dari 2.982 sample data tergolong pengangguran setengah dan pekerja paruh waktu sisanya sebesar 4.954 tergolong pekerja aktif diprosentasekan cluster 1 sebanyak 92,59 %. Kluster 2 (dua) kluaster yang mempunyai karakteristik pengangguran Terbuka sebanyak 730 rumah tangga atau sebanyak 8,42 % dari jumlah rumah tangga sample, faktor paling banyak penyebab pengangguran terbuka ialah merasa putus asa dalam memperoleh pekerjaan bisa dikarenakan cacat fisik atau lainnya.
 - Faktor yang mempengaruhi tingkat pengangguran dari hasil pengelompokan adalah tingkat pendidikan yang rendah, kurang ketrampilan karena mereka rata – rata tidak mengikuti pelatihan atau kursus
- [7] Prasetyo, Eko, "Data Mining Mengolah Data menjadi Informasi dengan Matlab," Andi-Yogyakarta, 2009.
 - [8] Febti Eka Pratiwi, Ismaini Zain, "klasifikasi pengangguran terbuka menggunakan CART (Classification and regression tree) di propinsi Sulawesi Utara," JURNAL SAINS DAN SENI POMITS Vol. 3, No.1, 2014.
 - [9] Andayani, Sri., 2007, Pembentukan Cluster dalam Knowledge Discovery dengan Algoritma K-Means, Semnas Matematika dan Pendidikan Matematika.
 - [10] Oyelade, Oladipupo, Obagbuwa, 2010, Application of K-Means Clustering Algorithm for Prediction of Students' Academic Performance, International Journal of Computer Science and Information Security, Volume 7
 - [11] Suryana. 2000. Ekonomi Pembangunan Problematika dan Pendekatan, Bandung: Salemba Empat
 - [12] Widodo, S T. 1990, Indikator Ekonomi, Yogyakarta: Kanisius.
 - [13] Sistem Informasi Rujukan Statistik. 2011. Survei Angkatan Kerja Nasional (SAKERNAS) [Online]. Available : diakses tanggal 17 Januari 2016.
 - [14] Oscar, Johan Ong, "Implementasi Algoritma K-Means Clustering untuk menentukan Strategi Marketing President University", Jurnal Ilmiah Teknik Industri, Vol.12 No.1, 2013.