

# Analisis Perbandingan Algoritma K-Medoids Clustering dan Self Organizing Maps (SOM) dalam Cluster Data (Studi Kasus : Data Indikator Pasar Tenaga Kerja setiap Provinsi di Indonesia tahun 2017-2019)

Fadhiila Senjaliana<sup>1,\*</sup> , Ayundyah Kesumawati<sup>2</sup>

<sup>1,2</sup> Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Indonesia Jl. Kaliurang No.Km 14,5, Krawitan, Umbulmartani, Kec. Ngemplak, Kab. Sleman, Daerah Istimewa Yogyakarta.

\* Corresponding author: fadhiilassenjaliana@gmail.com

**Abstract:** Analisis clustering merupakan analisis yang bertujuan untuk penempatan sekumpulan obyek dalam dua tau lebih kelompok berdasarkan kesamaan objek karakteristiknya. Terdapat beberapa algoritma clustering yaitu Algoritma K-Medoids dan Self Organizing Maps (SOM). Studi kasus dalam penelitian ini yaitu terkait ketenagakerjaan, tenaga kerja merupakan salah satu bidang untuk memajukan perekonomian suatu negara dari segi kuantitas dan kualitas sebagai sasaran untuk mengembangkan pasar serta sumber daya proses produksi dan distribusi barang dan jasa. Perbaikan masalah ketenagakerjaan di Indonesia menggunakan pasar tenaga kerja memberikan jalan keluar untuk menciptakan kondisi sinergi antara penjual dan pemberi tenaga kerja sehingga perlu diberlakukannya kerjasama baik antara pihak yang terkait, yaitu penjual tenaga kerja, pembeli tenaga kerja, dan pemerintah. Oleh sebab itu, peneliti tertarik melakukan penelitian terkait ketenagakerjaan di Indonesia tahun 2017-2019 menggunakan Algoritma K-Medoids dan Self Organizing Maps (SOM). Berdasarkan hasil cluster menggunakan Algoritma K-Medoids didapatkan hasil pada rata-rata data tahun 2017 sampai 2019 yaitu cluster satu terdiri dari 29 provinsi dan cluster dua terdapat 5 provinsi. Sedangkan hasil menggunakan algoritma SOM didapatkan pada rata-rata data tahun 2017 sampai 2019 yaitu cluster satu terdiri dari 28 provinsi dan cluster dua terdapat 6 provinsi. Selanjutnya berdasarkan hasil perbandingan metode cluster terbaik menggunakan validasi internal (indeks connectivity, dunn , dan silhouette) didapatkan algoritma terbaik yaitu Algoritma K-Medoids.

**Kata Kunci :** Ekonomi, Indikator Pasar Tenaga Kerja, K-Medoids Clustering, Self Organizing Maps.

## Pendahuluan

Analisis *clustering* merupakan suatu metode *unsupervised learning* dan termasuk dalam *machine learning*. Analisis *clustering* bertujuan dalam mencari pola data yang memiliki kemiripan sehingga kemungkinan dalam mengelompokkan data-data yang mirip tersebut. Pada penentuan *cluster* yang baik ketika suatu anggota dalam *cluster* memiliki kemiripan namun antar anggota *cluster* memiliki perbedaan yang cukup signifikan. Penggunaan

analisis *clustering* banyak digunakan dalam berbagai bidang seperti profiling data, rekomendasi produk, dan lain-lainnya [1].

Terdapat beberapa algoritma *clustering* dalam analisis *cluster* yang dapat diterapkan seperti Algoritma *K-Medoids* dan Algoritma *Self Organizing Maps* (SOM). *K-Medoids Clustering* atau *Partitioning Around Medoids* (PAM) merupakan varian dari algoritma *K-Means* berdasarkan dari penggunaan *medoids* bukan dengan pengamatan *mean* yang dimiliki oleh setiap *cluster* bertujuan untuk mengurangi sensitivitas dari partisi jika terdapat nilai ekstrim yang ada dalam dataset [2]. *K-Medoids Clustering* juga bertujuan untuk mengatasi kelemahan dari *k-means clustering* yang sensitif akan data *outlier* karena suatu objek dengan suatu nilai yang besar mungkin secara substansial menyimpang dari distribusi data [3]. Sedangkan algoritma SOM merupakan algoritma yang memungkinkan digunakan untuk menggambarkan data multidimensi ke dimensi yang lebih kecil pada satu atau lebih dimensi. Pada SOM memiliki kemampuan yaitu sebuah jaringan akan belajar dengan dasar pengetahuan parameter jaringan tanpa terdapat pengetahuan awal terlebih dahulu terkait segmen dan karakteristiknya serta tanpa harus mengetahui berapa kelompok yang akan terbentuk, dimana dilakukannya pengorganisasian sendiri hubungan-hubungan interkoneksi dalam dirinya atas masukan yang ada sehingga target tidak dibutuhkan [4].

Studi kasus yang dibahas dalam penelitian ini yaitu terkait ketenagakerjaan, dimana tenaga kerja adalah penduduk yang disebut dalam usia kerja. Menurut UU No. 13 tahun 2003 Bab I pasal 1 ayat 2 menjelaskan bahwa tenaga kerja adalah orang-orang yang sekiranya mampu melakukan pekerjaan untuk menghasilkan barang atau jasa baik dalam memenuhi kebutuhan individu maupun untuk lingkungan sekitar. Tenaga kerja yang memadai dari segi kuantitas dan kualitas menjadi aspek penting dalam pembangunan ekonomi, yaitu sebagai sasaran untuk menciptakan dan mengembangkan pasar serta sumber daya untuk menjalankan proses produksi dan distribusi barang dan jasa [5].

Masalah ketenagakerjaan yang kerap terjadi di Indonesia diantaranya yaitu jumlah angkatan kerja tidak berbanding lurus dengan kesempatan kerja. Peningkatan jumlah dari angkatan kerja yang tidak diimbangi dengan lapangan kerja mengakibatkan beban tersendiri dari sektor perekonomian. Tidak tertampungnya angka kerja dalam lapangan kerja akan menyebabkan pengangguran. Kemudian pada persebaran tenaga kerja yang tidak merata di setiap daerah. Selanjutnya terjadi krisis ekonomi banyak mengakibatkan sektor industri mengalami gulung tikar. Akibat dari kejadian tersebut, tenaga kerja pada sektor industri kehilangan pekerjaan dan semakin sempitnya lapangan kerja yang ada. Di sisi lain jumlah angkatan kerja terus meningkat. Dengan demikian pengangguran akan semakin banyak [6].

Pada permasalahan ketenagakerjaan di Indonesia, dalam penyelesaiannya perlu dilakukan pengelolaan akan pasar tenaga kerja lebih baik. Pasar tenaga kerja tersebut merupakan suatu pasar yang mempertemukan antara pembeli dan penjual tenaga kerja. Penyebutan pembeli tenaga kerja yaitu lembaga atau orang-orang yang memerlukan tenaga kerja, sedangkan sebagai penjual tenaga kerja yaitu para pencari kerja atau pemilik tenaga kerja. Pasar tenaga kerja diselenggarakan dengan tujuan untuk mengkoordinasi

pertemuan antara para pencari kerja dan lembaga-lembaga atau orang-orang yang membutuhkan tenaga kerja [7].

Data ketenagakerjaan merupakan hal penting dalam menggambarkan apa saja indikator pasar tenaga kerja di Indonesia. Salah satu sumber yang dijadikan patokan dalam penyajian indikator ini yaitu Survei Angkatan Kerja Nasional (Sakernas) yang dilakukan oleh Badan Pusat Statistika (BPS) bertujuan untuk memonitor dinamika ketenagakerjaan, agar nantinya para pengambil keputusan lebih cepat dalam pengambilan kebijakan untuk memperbaiki dan mengatasi permasalahan ketenagakerjaan khususnya di Indonesia. Sejak Triwulan I Februari 2011, BPS melakukan penyusunan publikasi yang merujuk pada publikasi Organisasi Buruh Internasional (*International Labour Organization-ILO*), yaitu Indikator Pasar Tenaga Kerja (*Key Indicators of the Labour Market-KILM*). ILO telah mengeluarkan KILM pada tahun 1999 dalam melengkapi program pengumpulan data secara rutin dan untuk meningkatkan penyebaran data pada unsur utama dari pasar tenaga kerja dunia [8].

Dalam perbaikan masalah ketenagakerjaan di Indonesia, pasar tenaga kerja ini dirasakan bisa memberikan jalan keluar untuk memenuhinya agar menciptakan kondisi yang sinergi antara penjual dan pemberi tenaga kerja sehingga perlu diberlakukannya kerjasama yang baik antara semua pihak yang terkait, yaitu penjual tenaga kerja, pembeli tenaga kerja, dan pemerintah. Berdasarkan uraian yang telah dipaparkan, peneliti tertarik untuk melakukan penelitian terkait masalah ketenagakerjaan di Indonesia dalam kurung waktu 3 tahun yaitu tahun 2017 sampai 2019 berdasarkan indikator pasar tenaga kerja menggunakan algoritma *K-Medoids Clustering* dan *Self Organizing Maps* (SOM). Dari hasil penelitian diharapkan bisa menghasilkan pengelompokan yang mampu memberikan informasi yang tepat mengenai ketenagakerjaan setiap provinsi di Indonesia dari tahun 2017 sampai 2019. *Software* yang digunakan dalam penelitian ini adalah *software R*, *Tableau for Student*, dan *Ms. Excel*. Oleh karena itu, dalam penyusunan Tugas Akhir ini penulis mengambil judul “Analisis Perbandingan Algoritma *K-Medoids Clustering* dan *Self Organizing Maps* (SOM) dalam *Cluster Data* (Studi Kasus : Data Indikator Pasar Tenaga Kerja setiap Provinsi di Indonesia tahun 2017-2019)”

## **Tinjauan Pustaka**

Penelitian merupakan hal yang dilakukan untuk membangun suatu kerangka berfikir, dalam membangun hal tersebut peneliti perlu menggunakan acuan dari beberapa penelitian terdahulu yang relevan dengan topik pembahasan. Selain itu, riset terdahulu juga dapat digunakan dengan tujuan agar mengetahui hubungan antara penelitian yang akan dilakukan dengan penelitian yang sudah terdahulu, sehingga nantinya setelah menggunakan acuan tersebut dapat menghindari akan adanya sebuah duplikasi dengan penelitian yang akan dilakukan. Berikut ini terdapat beberapa penelitian terdahulu yang akan dijadikan acuan dalam penelitian ini.

Penelitian yang dilakukan oleh Mety Agustini pada tahun 2017 tentang “*Model-Based Clustering* dengan Distribusi T Multivariate Menggunakan Kriteria *Integrated Completed Likelihood* dan *Minimum Message Length* Studi Kasus Pengelompokan Provinsi di Indonesia

Menurut Indikator Pasar Tenaga Kerja Tahun 2012-2015”. Analisis *cluster* salah satu alat statistik yang banyak digunakan dalam penentuan kelompok dalam sekumpulan data. Metode *clustering* yang kerap digunakan yaitu *clustering* berdasarkan ukuran jarak. Namun *clustering* menggunakan jarak akan sangat sulit dilakukan jika kondisi objek yang ada saling tumpang tindih. *Model-based clustering* (MBC) disarankan dalam penelitian ini dengan pendekatan yang didasarkan pada model *finite mixture*. Metode *clustering* mempunyai asumsi jika data yang dihasilkan berasal dari beberapa distribusi probabilitas dan kemudian *cluster* yang terbentuk dimulai oleh masing-masing distribusi probabilitas tersebut. Pada *model-based clustering* distribusi *t* multivariat digunakan untuk mengakomodasi keberadaan outlier. Distribusi *t* multivariat dianggap dapat mengatasi outlier dibandingkan distribusi normal multivariat. Pemilihan model terbaik dari beberapa model yang tersedia dilakukan dengan *Integrated Completed Likelihood* (ICL) dan *Minimum Message Length* (MML). Kelompok optimal pada MBC-ICL digunakan dalam analisis pasar tenaga kerja Indonesia berdasarkan indikator Bekerja Menurut Lapangan Usaha. Sedangkan kelompok optimal RMBC-MML digunakan dalam analisis pasar tenaga kerja Indonesia berdasarkan indikator Rasio Penduduk Bekerja Terhadap Jumlah Penduduk (EPR), Pekerja Rentan, dan Pekerja Sektor Informal [9].

Penelitian yang dilakukan oleh Desy Rizki Ramadhanty, Muthia Dishanur Izzati, Matiin Laugiwa Prawira, Novendri Isra Asriny, Annisa Shofura, Chairani Uli Albi, dan Edy Widodo tentang “Perbandingan Algoritma *K-Medoids* dan *Self Organizing Maps* dalam Pengelompokan Kejadian Bencana Alam Hidrometeorologi di Pulau Jawa Tahun 2017-2019” pada tahun 2020. Pada peneliti akan melakukan pengelompokan kejadian bencana alam hidrometeorologi di Indonesia untuk menentukan daerah mana saja yang menjadi fokus pemerintah terkhusus di Pulau Jawa dalam penanggulangan bencana dengan menggunakan metode *K-Medoids Clustering* dan *Self Organizing Maps* (SOM). Dari hasil penelitian yang telah didapatkan, penggunaan metode terbaik adalah *Metode Self Organizing Maps* dengan hasil *cluster* satu sebanyak 108 anggota yang merupakan kelompok dengan kerentanan bencana yang relatif rendah, lalu *cluster* dua sebanyak 10 anggota merupakan kelompok kota dan kabupaten yang memiliki kerentanan terjadinya bencana yang relatif tinggi [10].

Perbandingan dengan penelitian sebelumnya didapatkan kesimpulan keterbaruan dari penelitian yaitu pada penelitian ini menggunakan data ketenagakerjaan dengan indikator pasar tenaga kerja terbaru yaitu data tahun 2017 sampai 2019 di Indonesia dengan menggunakan metode yang telah digunakan pada studi kasus lain yaitu analisis *clustering* dengan Algoritma *K-Medoids Clustering* dan *Self Organizing Maps* dengan validasi *cluster* (indeks *connectivity*, indeks *dunn*, dan indeks *silhouette*).

## Metodologi Penelitian

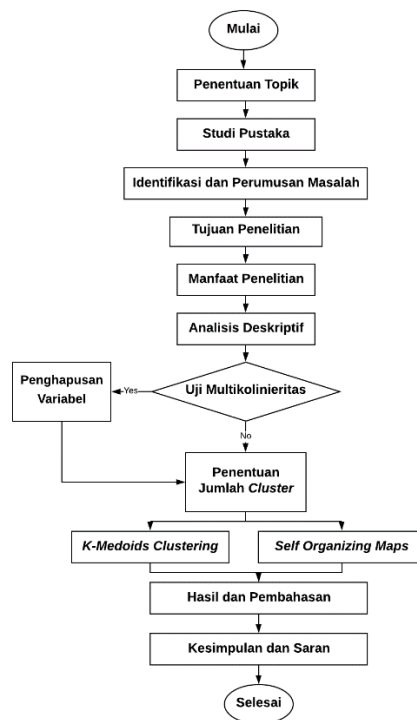
Populasi yang digunakan dalam penelitian ini yaitu seluruh provinsi di Indonesia berdasarkan data indikator pasar tenaga kerja dan sampel yang digunakan dalam penelitian ini yaitu seluruh provinsi di Indonesia berdasarkan data indikator pasar tenaga kerja periode 2017 sampai 2019. Sumber data yang digunakan dalam penelitian berasal

dari *website* resmi Badan Pusat Statistika (BPS) Pusat yaitu <https://www.bps.go.id/>. Sedangkan jenis data yang digunakan dalam penelitian adalah data sekunder berasal dari Badan Pusat Statistika (BPS) Pusat bersumber dari hasil Survei Angkatan Kerja Nasional (Sakernas) yaitu data Indikator Pasar Tenaga Kerja tahun 2017 sampai 2019. Penelitian ini menggunakan 9 indikator KILM yang terdiri dari 18 variabel indikator seperti rincian tabel 4.1 berikut :

**Tabel 1. Variabel Penelitian**

No	Indikator Pasar Tenaga Kerja (KILM)	Variabel Indikator (Persen)
1	Tingkat Partisipasi Angkatan Kerja-TPAK (KILM 1)	Persentase dari Tingkat Partisipasi Angkatan Kerja-TPAK ( $X_1$ )
2	Rasio Penduduk Bekerja Terhadap Jumlah Penduduk Usia Kerja (KILM 2)	Rasio Penduduk Bekerja Terhadap Jumlah Penduduk Usia Kerja ( $X_2$ )
3	Penduduk Bekerja Menurut Status Pekerjaan (KILM 3)	<ul style="list-style-type: none"> <li>- Persentase dari Status Pekerjaan Penduduk Bekerja dengan Upah/Gaji (<math>X_3</math>)</li> <li>- Persentase dari Status Pekerjaan Pengusaha (<math>X_4</math>)</li> <li>- Persentase dari Status Pekerjaan Berusaha Sendiri dan Berusaha Dibantu Buruh Tidak Tetap/Tidak Dibayar (<math>X_5</math>)</li> <li>- Persentase dari Status Pekerjaan Pekerja Bebas (<math>X_6</math>)</li> <li>- Persentase dari Status Pekerjaan Pekerja Keluarga (<math>X_7</math>)</li> </ul>
4	Penduduk Bekerja Menurut Lapangan Pekerjaan (KILM 4)	-
5	Penduduk Bekerja Menurut Jenis Pekerjaan Utama (KILM 5)	-
6	Pekerja Paruh Waktu (KILM 6)	Persentase dari Pekerja Paruh Waktu ( $X_8$ )
7	Penduduk Bekerja Menurut Jumlah Jam Kerja (KILM 7)	<ul style="list-style-type: none"> <li>- Persentase dari Jam Kerja &lt;35 jam (<math>X_9</math>)</li> <li>- Persentase dari Jam Kerja 35-39 jam (<math>X_{10}</math>)</li> <li>- Persentase dari Jam Kerja 40-48 jam (<math>X_{11}</math>)</li> <li>- Persentase dari Jam Kerja 49-59 jam (<math>X_{12}</math>)</li> <li>- Persentase dari Jam Kerja &gt;60 jam (<math>X_{13}</math>)</li> </ul>
8	Penduduk Bekerja di Sektor Informal (KILM 8)	<ul style="list-style-type: none"> <li>- Persentase dari Bekerja Sektor Formal (<math>X_{14}</math>)</li> <li>- Persentase dari Bekerja Sektor Informal (<math>X_{15}</math>)</li> </ul>
9	Pengangguran (KILM 9)	<ul style="list-style-type: none"> <li>- Persentase dari Pengangguran Pendidikan Dasar ke Bawah (<math>X_{16}</math>)</li> <li>- Persentase dari Pengangguran Pendidikan Menengah (<math>X_{17}</math>)</li> <li>- Persentase dari Pengangguran Pendidikan Tinggi (<math>X_{18}</math>)</li> </ul>

*Software* yang digunakan dalam penelitian ini adalah *software R*, *Tableau for Student*, dan *Ms. Excel*. Terdapat beberapa metode analisis data yang digunakan dalam penelitian ini adalah analisis deskriptif yang digunakan untuk melihat gambaran dari masing-masing variabel, selanjutnya menggunakan menggunakan analisis *cluster* atau analisis kelompok yaitu algoritma *K-Medoids Clustering* dan *Self Organizing Maps* (SOM) merupakan suatu teknik analisis statistik yang bertujuan untuk penempatan sekumpulan obyek dalam dua tau lebih kelompok berdasarkan kesamaan obyek berdasarkan berbagai karakteristik. Berikut pada **Gambar 1** terdapat *flowchart* yang akan menggambarkan tahapan-tahapan dalam penelitian :

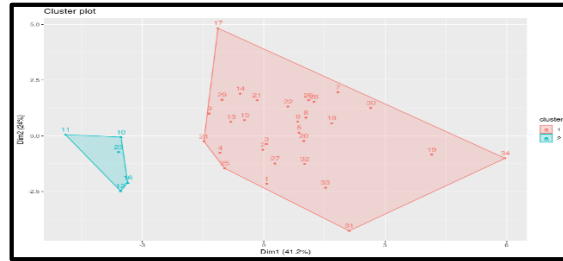


**Gambar 1.** *Flowchart* Penelitian

## Hasil dan Pembahasan

Pada penelitian ini, penulis terfokus pada bidang ekonomi menggunakan partisipasi dan indikator tenaga kerja serta indikator pengangguran. Pada partisipasi dan indikator tenaga kerja terdiri dari KILM 1 (Tingkat Partisipasi Angkatan Kerja-TPAK), KILM 2 (Rasio Penduduk Bekerja Terhadap Jumlah Penduduk Usia Kerja/*Employment to Population Ratio-EPR*), KILM 3 (Penduduk Bekerja Menurut Status Pekerjaan Utama), KILM 4 (Penduduk Bekerja Menurut Lapangan Pekerjaan Utama), KILM 5 (Penduduk Bekerja Menurut Jenis Pekerjaan Utama), KILM 6 (Pekerja Paruh Waktu), KILM 7 (Penduduk Bekerja Menurut Jumlah Jam Kerja), KILM 8 (Penduduk Bekerja di Sektor Informal). Sedangkan pada indikator pengangguran terdiri dari KILM 9 (Pengangguran). Gambaran umum kondisi pasar tenaga kerja di Indonesia tahun 2017-2019 yang bersumber dari hasil Survei Angkatan Kerja Nasional (Sakernas) yaitu pada tahun 2017-2019, tingkat partisipasi angkatan kerja di Indonesia mengalami peningkatan, penduduk bekerja berdasarkan status pekerjaan didominasi oleh pengusaha. Selanjutnya jenis pekerjaan dan lapangan pekerjaan didominasi oleh pekerjaan dan tenaga kerja pertanian, kehutanan, dan perikanan. Kemudian pada pengangguran di didominasi oleh pengangguran lulusan sekolah menengah.

Dari hasil analisis menggunakan *cluster* dengan algoritma *K-Medoids* didapatkan hasil berupa plot visualisasi yang terdiri dari dua *cluster* seperti berikut :



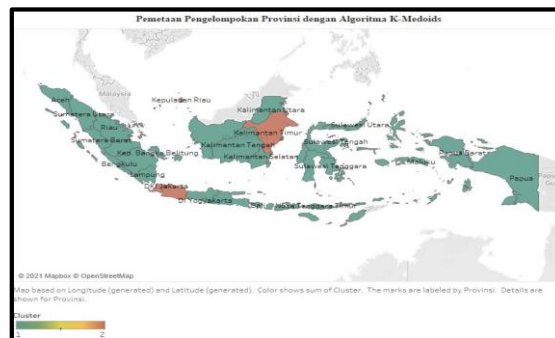
**Gambar 2.** Plot Visualisasi *K-Medoids Clustering*

Pada **Gambar 2** yaitu plot visualisasi *K-Medoids Clustering* yang terdiri dari dua *cluster* dibedakan dengan dua warna yang berbeda yaitu merah dan biru. Warna merah menjelaskan hasil *cluster* 1 sedangkan warna biru menjelaskan hasil *cluster* 2. Terlihat bahwa plot visualisasi warna merah lebih banyak memiliki anggota *cluster* dibandingkan plot visualisasi warna biru. Selanjutnya berikut ini tampilan melalui **Tabel 2** hasil *k-medoids clustering* :

**Tabel 2** Hasil *K-Medoids Clustering*

Cluster	Jumlah Provinsi	Provinsi
1	29	Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Jawa Tengah, DI Yogyakarta, Jawa Timur, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua.
2	5	Kepulauan Riau, DKI Jakarta, Jawa Barat, Banten, Kalimantan Timur.

Dari *cluster* menggunakan algoritma *k-medoids* didapatkan hasil bahwa pada rata-rata data tahun 2017 sampai 2019 *cluster* satu terdiri dari 29 provinsi yaitu Provinsi Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Jawa Tengah, DI Yogyakarta, Jawa Timur, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua. Sedangkan *cluster* dua terdapat 5 provinsi yaitu Provinsi Kepulauan Riau, DKI Jakarta, Jawa Barat, Banten, Kalimantan Timur. Selanjutnya membuat peta berdasarkan analisis algoritma *K-Medoids* dan didapatkan hasil seperti pada **Gambar 3** berikut :



**Gambar 3** Pemetaan Pengelompokan Provinsi

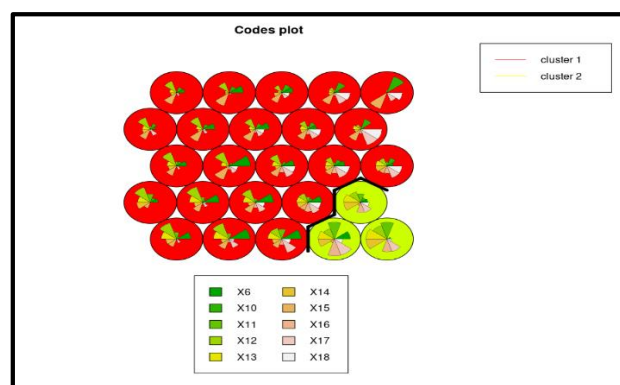
Pada **Gambar 3** merupakan hasil pemetaan pengelompokan provinsi di Indonesia dari data indikator pasar tenaga kerja berdasarkan hasil pada **Tabel 3**. Kemudian penulis melakukan segmentasi terhadap perhitungan rata-rata nilai setiap *cluster* dan di dapatkan hasil seperti pada **Tabel 3** berikut :

**Tabel 3** Rata-Rata Hasil *K-Medoids Clustering*

Cluster	X1	X4	X6	X10	X11	X12	X13	X14	X15	X16	X17	X18
1	70.92	3.50	6.73	10.33	25.23	14.59	12.30	37.81	62.19	2.51	6.67	6.19
2	68.06	3.96	6.24	8.09	37.23	15.39	17.47	58.19	41.81	5.51	9.34	5.55

Berdasarkan hasil pada **Tabel 3** merupakan perhitungan rata-rata nilai setiap *cluster* pada data indikator pasar tenaga kerja dimana warna biru tua menunjukkan hasil rata-rata tinggi dan warna biru muda menunjukkan hasil rata-rata rendah. Namun dari hasil rata-rata tersebut, hasil rata-rata *cluster* tinggi bukan mengartikan bahwa indikator tersebut memiliki kualitas baik begitupun sebaliknya. Dari hasil rata-rata *cluster* 1 lebih menonjol dan lebih baik pada variabel persentase tingkat partisipasi angkatan kerja ( $X_1$ ), persentase status pekerjaan pekerja bebas ( $X_6$ ), persentase jam kerja 35 sampai 39 jam ( $X_{10}$ ), persentase bekerja di sektor informal ( $X_{15}$ ), persentase pengangguran pendidikan dasar ke bawah ( $X_{16}$ ), dan persentase pengangguran pendidikan tingkat menengah ( $X_{17}$ ). Sedangkan hasil rata-rata *cluster* 2 lebih menonjol dan lebih baik pada variabel persentase status pekerja pengusaha ( $X_4$ ), persentase jam kerja 40 sampai 48 jam ( $X_{11}$ ), persentase jam kerja 49 sampai 59 jam ( $X_{12}$ ), persentase jam kerja diatas 60 jam ( $X_{13}$ ), persentase bekerja di sektor formal ( $X_{14}$ ), dan persentase pengangguran pendidikan tinggi ( $X_{18}$ ).

Selanjutnya didapatkan sebuah model berupa diagram kipas dari hasil analisis algoritma SOM menggunakan diagram kipas dengan tampilan *hexagonal* dengan *grid* berukuran 5 x 5. Diagram kipas merupakan ilustrasi dari distribusi variabel pemetaan, diagram kipas yang telah dibentuk terdapat 2 warna berbeda dan memiliki karakteristik tersendiri yang menunjukkan hasil dari *cluster* setiap provinsi seperti **Gambar 4** berikut :



**Gambar 4** Diagram Kipas

Pada **Gambar 4** merupakan gambar *codes plot*, vektor bobot kode atau *codes* terdiri dari nilai-nilai yang dinormalisasi berdasarkan variabel asli yang digunakan untuk menghasilkan SOM. Setiap vektor bobot node mewakili atau serupa dari sampel yang dipetakan ke node tersebut dengan memvisualisasikan vektor bobot di seluruh peta sehingga dapat melihat pola dalam distribusi sampel dan variabel. Visualisasi *default* dari



vektor bobot yaitu diagram kipas, dimana representasi kipas individu dari besarnya setiap variabel dalam vektor bobot ditampilkan untuk setiap node menggunakan tampilan *hexagonal* dengan *grid* 5x5 yang menunjukkan distribusi dari variabel pada pemetaan. Proses dalam memahami diagram pada algoritma SOM yaitu saat diagram telah memiliki suatu warna kemudian diberi batasan dengan vektor-vektor yang tervisualisasi dalam plot pemetaan.

Berdasarkan 4 diagram kipas dapat dilihat menggunakan tampilan *hexagonal grid* 5x5, diagram di atas terbentuk berdasarkan hasil olah data menggunakan algoritma SOM menggunakan 12 variabel, namun yang terlihat hanya 10 variabel karena variabel tersebut lebih dominan. Dapat diketahui pewarnaan dari masing-masing variabel yaitu gradasi warna hijau ke kuning dan warna orange menuju putih, untuk variabel persentase status pekerjaan pekerja bebas ( $X_6$ ) diberi warna hijau tua, persentase jam kerja 35 sampai 39 jam ( $X_{10}$ ) diberi warna hijau tua menuju muda, persentase jam kerja 40 sampai 48 jam ( $X_{11}$ ) diberi warna hijau muda menuju tua, persentase jam kerja 49 sampai 59 jam ( $X_{12}$ ) diberi warna hijau muda, persentase jam kerja diatas 60 jam ( $X_{13}$ ) diberi warna kuning, persentase bekerja di sektor formal ( $X_{14}$ ) diberi warna orange, persentase bekerja di sektor informal ( $X_{15}$ ) diberi warna orange menuju coklat, persentase pengangguran pendidikan dasar ke bawah ( $X_{16}$ ) diberi warna merah muda menuju tua, persentase pengangguran pendidikan menengah ( $X_{17}$ ) diberi warna merah muda, dan persentase pengangguran pendidikan sekolah tinggi ( $X_{18}$ ) diberi warna putih. Pola tersebut dapat dilihat dengan memeriksa warna yang dominan pada hasil diagram kipas.

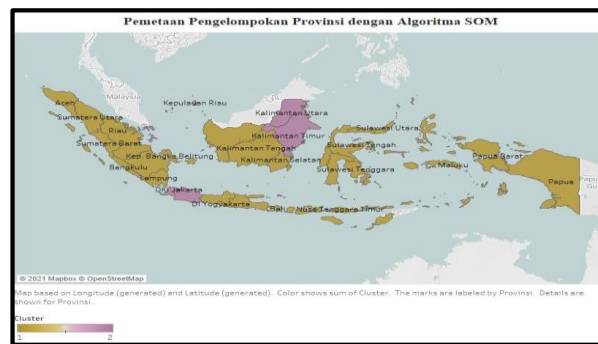
Selanjutnya dalam proses memahami diagram pada algoritma SOM ketika diagram tersebut telah memiliki warna dan diberikan batasan dengan vektor-vektor yang tervisualisasi dalam plot pemetaan. Model yang telah terbentuk kemudian dibuat menjadi 2 *cluster* tidak menunjukkan tingkatan namun memiliki karakteristik tersendiri, dapat dilihat terdiri dari dua warna yang berbeda menunjukkan kondisi dari setiap *cluster* yang terbentuk yaitu warna merah dan warna hijau mendekati kuning. *Cluster* 1 dengan lingkaran warna merah memiliki karakteristik yang lebih menonjol pada variabel persentase status pekerjaan pekerja bebas ( $X_6$ ), persentase jam kerja 35 sampai 39 jam ( $X_{10}$ ), persentase bekerja di sektor informal ( $X_{15}$ ), persentase pengangguran pendidikan dasar ke bawah ( $X_{16}$ ), dan persentase pengangguran pendidikan tingkat menengah ( $X_{17}$ ). Kemudian untuk *cluster* 2 memiliki karakteristik lebih menonjol pada variabel persentase jam kerja 40 sampai 48 jam ( $X_{11}$ ), persentase jam kerja 49 sampai 59 jam ( $X_{12}$ ), persentase jam kerja diatas 60 jam ( $X_{13}$ ), persentase bekerja di sektor formal ( $X_{14}$ ), dan persentase pengangguran pendidikan tinggi ( $X_{18}$ ). Berikut ini tampilan melalui **Tabel 4** hasil algoritma SOM :

**Tabel 4** Hasil Algoritma SOM

Cluster	Jumlah Provinsi	Provinsi
1	28	Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Jawa Tengah, DI Yogyakarta, Jawa Timur, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Utara, Sulawesi Tengah,

Cluster	Jumlah Provinsi	Provinsi
2	6	Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua.
		Kepulauan Riau, DKI Jakarta, Jawa Barat, Banten, Kalimantan Timur, Kalimantan Utara.

Dari hasil pada **Tabel 4** yaitu hasil *cluster* menggunakan algoritma SOM didapatkan hasil bahwa *cluster* satu terdiri dari 28 provinsi yaitu Provinsi Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Jawa Tengah, DI Yogyakarta, Jawa Timur, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua. Sedangkan *cluster* dua terdiri dari 6 provinsi yaitu Provinsi Kepulauan Riau, DKI Jakarta, Jawa Barat, Banten, Kalimantan Timur, Kalimantan Utara. Kemudian penulis membuat peta berdasarkan analisis algoritma SOM dan didapatkan hasil seperti pada **Gambar 5** berikut :



**Gambar 5** Pemetaan Pengelompokan Provinsi

Pada **Gambar 5** merupakan hasil pemetaan pengelompokan provinsi di Indonesia dari data indikator pasar tenaga kerja berdasarkan hasil pada **Tabel 5**. Setelah dilakukannya tahap *clustering* menggunakan *k-medoids* dan SOM, penulis akan melakukan pemilihan metode terbaik dari hasil validasi *cluster* menggunakan validasi internal dengan tiga metode yaitu indeks *connectivity*, *dunn*, dan *silhouette* untuk pembentukan jumlah *cluster*. Pengambilan jumlah *cluster* dengan melihat nilai indeks *connectivity* paling kecil, nilai indeks *dunn* mendekati satu, dan nilai indeks *silhouette* paling besar. Berikut hasil dari perbandingan validasi *cluster k-medoids* dan *Self Organizing Maps* (SOM) pada rata-rata data indikator pasar tenaga kerja tahun 2017 sampai 2019 dari **Tabel 6**

**Tabel 6** Validasi *Cluster*

Validasi Cluster	Nilai Validasi Cluster	Metode Cluster Terbaik	Cluster
<i>connectivity</i>	7,49	<i>k-medoids</i>	2
<i>dunn</i>	0,29	<i>k-medoids</i>	2
<i>silhouette</i>	0,32	<i>k-medoids</i>	2

Berdasarkan **Tabel 6** setelah dilakukannya perbandingan hasil validasi *cluster k-medoids* dan *Self Organizing Maps* (SOM) pada data, didapatkan metode algoritma terbaik

yaitu algoritma *k-medoids clustering* pada rata-rata data tahun 2017 sampai 2019. Terdapat beberapa provinsi yang perlu lebih diperhatikan diantaranya provinsi-provinsi yang memiliki tingkat partisipasi angkatan kerja yang rendah, tingginya jumlah jam kerja melebihi jam normal, kurangnya lapangan pekerjaan baik di sektor formal maupun informal, serta tingginya tingkat pengangguran terbuka dari setiap jenjang pendidikan. Pada tahun 2017 sampai 2019, provinsi-provinsi yang perlu diperhatikan yaitu pada *cluster* 2 diantaranya Provinsi Kepulauan Riau, DKI Jakarta, Jawa Barat, Banten, dan Kalimantan Timur tingkat partisipasi angkatan kerja yang rendah, persentase status pekerjaan pekerja bebas yang rendah, persentase jam kerja yang tinggi dari jam kerja normal, persentase bekerja di sektor informal yang rendah, persentase pengangguran pendidikan dasar ke bawah dan persentase pengangguran pendidikan tingkat menengah yang tinggi.

Berdasarkan hasil analisis menggunakan metode *clustering* didapatkan *k* optimal sebanyak 2 *cluster* dengan algoritma terbaik yaitu *K-Medoids* pada studi kasus data indikator pasar tenaga kerja di Indonesia tahun 2017 sampai 2019. Dalam pencatatan yang dilakukan oleh BPS melalui Sakernas tidak dilakukannya pengelompokan terkait indikator pasar tenaga kerja, namun tujuan dilakukannya penyusunan indikator pasar tenaga kerja dirancang sebagai penyedia indikator inti dari pasar tenaga kerja serta peningkatan akan ketersediaan indikator-indikator ketenagakerjaan dalam memantau perkembangan pasar tenaga kerja saat ini. Indikator-indikator ini nantinya menghasilkan kolaborasi yang melibatkan ILO dengan para ahli dari Organisasi Kerjasama Ekonomi dan Pembangunan (*Organisation for Economic Cooperation and Development-OECD*) serta beberapa perwakilan nasional dari Departemen Tenaga Kerja dan kantor statistik berbagai negara [11].

## Kesimpulan

Dari analisis yang telah dilakukan oleh penulis, didapatkan beberapa kesimpulan yaitu :

1. Gambaran umum kondisi pasar tenaga kerja di Indonesia tahun 2017-2019 yang bersumber dari hasil Survei Angkatan Kerja Nasional (Sakernas) yaitu pada tahun 2017-2019, tingkat partisipasi angkatan kerja di Indonesia mengalami peningkatan, penduduk bekerja berdasarkan status pekerjaan didominasi oleh pengusaha. Selanjutnya jenis pekerjaan dan lapangan pekerjaan didominasi oleh pekerjaan dan tenaga kerja pertanian, kehutanan, dan perikanan. Kemudian pada pengangguran di didominasi oleh pengangguran lulusan sekolah menengah.
2. Berdasarkan hasil *cluster* menggunakan algoritma *k-medoids* didapatkan hasil bahwa pada rata-rata data tahun 2017 sampai 2019 *cluster* satu terdiri dari 29 provinsi dan *cluster* dua terdapat 5 provinsi.
3. Berdasarkan hasil *cluster* menggunakan algoritma SOM didapatkan hasil bahwa pada rata-rata data tahun 2017 sampai 2019 yaitu *cluster* satu terdiri dari 28 provinsi dan *cluster* dua terdapat 6 provinsi.

4. Berdasarkan hasil perbandingan metode *cluster* terbaik didapatkan hasil yaitu menggunakan algoritma *k-medoids clustering* pada data indikator pasar tenaga kerja.
5. Dari hasil rata-rata *cluster* 1 lebih menonjol dan lebih baik pada variabel persentase tingkat partisipasi angkatan kerja ( $X_1$ ), persentase status pekerjaan pekerja bebas ( $X_6$ ), persentase jam kerja 35 sampai 39 jam ( $X_{10}$ ), persentase bekerja di sektor informal ( $X_{15}$ ), persentase pengangguran pendidikan dasar ke bawah ( $X_{16}$ ), dan persentase pengangguran pendidikan tingkat menengah ( $X_{17}$ ). Sedangkan hasil rata-rata *cluster* 2 lebih menonjol dan lebih baik pada variabel persentase status pekerja pengusaha ( $X_4$ ), persentase jam kerja 40 sampai 48 jam ( $X_{11}$ ), persentase jam kerja 49 sampai 59 jam ( $X_{12}$ ), persentase jam kerja diatas 60 jam ( $X_{13}$ ), persentase bekerja di sektor formal ( $X_{14}$ ), dan persentase pengangguran pendidikan tinggi ( $X_{18}$ ).

## References

- [1] Wolfram, "Fuzzy Clustering," 26 April 2016. [Online]. Available: <https://reference.wolfram.com/legacy/applications/fuzzylogic/Manual/12.html>.
- [2] C. Vercillis, Business Intelligence: Data Mining and Optimization for, Milan: WILEY, 2009.
- [3] J. K. M. Han, Data Mining: Concept and Techniques, Waltham: Morgan Kauffman Publisher, 2006.
- [4] Septianusa and Supriyaningsih, "Karakteristik dan Segmentasi Pertanian Padi Menggunakan Algoritma SOM Kohonen.," Diunduh tanggal 25 Me 2019, 2014.
- [5] R. N. Sari, "Analisis Pengaruh Inflasi, Indeks Pembangunan Manusia, Investasi dan Upah Minimum Provinsi terhadap Penyerapan Tenaga Kerja di Indonesia Tahun 2017," 19 Oktober 2019. [Online]. Available: <http://eprints.ums.ac.id/78146/1/NASKAH%20PUBLIKASI.pdf>.
- [6] H. Wijayanto, "Dinamika Permasalahan Ketenagakerjaan dan Pengangguran di Indonesia," 10 Agustus 2019. [Online]. Available: <http://jurnaladministratio.fisip.unila.ac.id/index.php/administratio/article/view/82/56>.
- [7] Bitar, "Pasar Tenaga Kerja : Pengertian, Fungsi, Jenis, Dan Ciri Beserta Kelebihan & Kekurangannya Lengkap," 2 November 2020. [Online]. Available: <https://www.gurupendidikan.co.id/pasar-tenaga-kerja-pengertian-fungsi-jenis-dan-ciri-beserta-kelebihan-kekurangannya-lengkap/>.
- [8] BPS, Indikator Pasar Tenaga Kerja di Indonesia, BPS RI, 2019.
- [9] M. Agustini, "Model-Based Clustering dengan Distribusi t Multivariat Menggunakan Kriteria Integrated Completed Likelihood dan Minimum Message Length (Pengelompokan Provinsi di Indonesia Menurut Indikator Pasar Tenaga Kerja tahun 2012-2015)," 2017.
- [10] D. R. Ramadhanty, M. D. Izzati and M. L. Prawira, "Perbandingan Algoritma K-Medoids dan Self Organizing Maps dalam Pengelompokan Kejadian Bencana Alam Hidrometeorologi di Pulau Jawa Tahun 2017-2019," ICITDA, Yogyakarta, 2020.
- [11] Sakernas, Indikator Pasar Tenaga Kerja di Indonesia, Jakarta: Badan Pusat Statistika, 2019.