

BAB II TINJAUAN PUSTAKA

(Khasanah, 2016) melakukan studi komparasi terhadap performa dua algoritma, *K-means* dan SOM dengan menggunakan tiga dataset dari UCI *Machine Learning Repository* yaitu Iris, Sonar, dan Wine. Iris memiliki 4 variabel dengan 150 baris data, untuk Sonar memiliki 60 variabel dengan 208 baris data sedang Wine memiliki 13 variabel dengan 178 baris data. Kesimpulannya, bahwa analisis *percent misclassified* menunjukkan *K-means* lebih unggul dari SOM, seperti ditunjukkan oleh grafik visualisasi keluaran (secara terpisah dan bersamaan dengan PCA).

Studi berikutnya diambil oleh penulis adalah, (Elvi, 2021). Peneliti melakukan penelitian tentang pengelompokan perkara perceraian berdasar Kelurahan di Kota Jambi. Metode yang digunakan adalah kuantitatif dengan analisis *cluster* memakai metode *K-means* yang diuji dengan tools RapidMiner. Data diperoleh dengan melakukan wawancara dan observasi di Pengadilan Agama Jambi. Masalah yang dianalisa berdasar 4 kriteria faktor penyebab perceraian. Hasil pengclusteran perkara perceraian berdasarkan kelurahan di Kota Jambi dengan data sebanyak 62 kelurahan, diperoleh hasil dengan *cluster* tinggi (C2) sebanyak 12 kelurahan, *cluster* sedang (C1) sebanyak 20 kelurahan dan *cluster* rendah (C0) sebanyak 30 kelurahan.

Penelitian ketiga, karya (Asriny et al., 2021). Peneliti melakukan penelitian tentang kejadian bencana hidrometeorologi yang ada di Indonesia khususnya di Pulau Jawa kemudian dikelompokkan sesuai kasus kejadiannya menggunakan algoritma K-Medoids dan SOM *clustering*. Data diperoleh dari Badan Nasional Penanggulangan Bencana selama tiga tahun terakhir dengan 6868 kejadian. Berdasarkan hasil penelitian diketahui bahwa metode yang cocok digunakan adalah Metode SOM karena dilihat berdasarkan nilai standar deviasi yang lebih kecil. *Cluster* satu berisi 108 anggota yang merupakan kelompok dengan kerentanan bencana relatif rendah, sedangkan *cluster* dua berisi 10 anggota

merupakan kelompok kota dan kabupaten yang memiliki kerentanan bencana relatif tinggi.

Penelitian keempat, oleh (Melo Riveros et al., 2019), melakukan penelitian tentang studi komparatif penerapan *K-means* dan *SOM clustering* untuk mendiagnosis pasien dengan masalah kolom tulang belakang. Data diambil dari *UCI Machine Learning Repository* mengenai *vertebral column dataset*. Jumlah data 310 sampel, dengan masing-masing 6 karakteristik biomekanik yang berasal dari bentuk orientasi panggul dan kolom lumbal. Hasil menunjukkan model yang dilatih dengan SOM mengungguli model yang dilatih dengan *K-means*, dengan peningkatan deteksi pasien masalah tulang belakang.

(Wang & Ferrus, 2021) melakukan penelitian tentang pengkombinasian dua algoritma tersebut untuk mengelompokkan data OAM yang dihasilkan oleh jaringan radio seluler. Data ini mencakup pengukuran kinerja yang beragam dan beragam serta indikator yang mencirikan perilaku radio seluler. Peneliti menjelaskan metodologi untuk mengelompokkan radio seluler berdasarkan kombinasi metode SOM dan *K-means*. Metodologi ini diterapkan untuk mengekstraksi pola sel dari karakterisasi jangka panjang dan jangka pendek perilaku. Secara keseluruhan, bahwa metode SOM-K sekarang dapat digunakan untuk menyesuaikan kinerja jaringan, dan algoritma tersebut berhasil menggunakan data besar untuk mengelompokkan sel dengan pola karakteristik yang hampir sama.

Penelitian keenam yang ditulis oleh (Galvan et al., 2021), mengangkat tema sebaran COVID-19 di Brazil. Para peneliti mencoba membuat analisa dan evaluasi dari hasil penambangan data sebagai modal informasi untuk negara melakukan upaya menahan laju sebaran COVID-19 di Brazil. Dalam studi tersebut, jaringan saraf tipe *Self-Organizing Maps* tanpa pengawasan digunakan untuk menilai penyebaran COVID-19 spasial dan temporal. Hasil disajikan dalam bentuk peta dengan warna sebagai penanda berikut keterangannya. Dalam pengungkapan hasil, pendekatan yang dilakukan menunjukkan bahwa penyebaran penyakit di Brasil tidak memiliki perilaku standar, berubah menurut wilayah, negara bagian, atau kota. Analisis menunjukkan bahwa kota dan negara bagian di

wilayah utara dan timur laut negara itu adalah yang paling terpengaruh oleh penyakit ini, dengan jumlah kasus dan kematian tertinggi tercatat per 100.000 penduduk. Dari analisa tersebut maka peneliti mengambil simpulan, bahwa pengelompokan SOM mampu mengelompokkan kota, negara bagian, dan wilayah secara spasial menurut kasus virus corona mereka, dengan perilaku serupa. Dengan demikian, penggunaan strategi serupa untuk menangani penyebaran virus di kota, negara bagian, dan wilayah ini dapat dimanfaatkan.

Penelitian ketujuh oleh (Kartikasari, 2021). Peneliti mengambil tema pengelompokan wilayah Indonesia berdasar konsumsi pangan. Penelitian dilakukan dengan tujuan untuk mengelompokkan wilayah Indonesia berdasarkan konsumsi pangan berdasarkan data konsumsi kalori per kapita sehari dari berbagai komoditas pangan. Pengelompokan wilayah dilakukan dengan metode SOM dengan terlebih dahulu ditentukan jumlah *cluster* optimum menggunakan nilai *Davies-Bouldin Index* (DBI) terkecil. Dalam pengujiannya dengan SOM, menghasilkan diagram kipas yang menunjukkan kesamaan *intra cluster* terbesar, yang mana tingkat kesamaan diketahui berdasarkan jarak. Hasil penelitian menunjukkan bahwa hasil kelompok/*cluster* optimum yang terbentuk sejumlah 4 kelompok dengan jumlah anggota untuk kelompok 1 sebanyak 22 propinsi, kelompok 2 sebanyak 10 propinsi, kelompok 3 sebanyak 1 propinsi, dan kelompok 4 sebanyak 1 propinsi. Dasar-dasar penelitian sebelumnya yang menjadi tinjauan pustaka pada penelitian ini dirangkum dalam Tabel 2.1 di halaman selanjutnya.

Tabel 2.1 Tinjauan Pustaka

Nama, Tahun	Permasalahan	Akibat	Data	Metode	Hasil
(Khasanah, 2016)	Banyak peneliti mencoba menggunakan dan membandingkan kinerja algoritma <i>K-means</i> dan SOM, namun saat ini belum ada solusi pasti karena setiap peneliti memberikan kesimpulan yang berbeda	Tidak ada petunjuk pasti algoritma mana yang lebih unggul dari keduanya, semuanya tergantung banyak data, atribut dan alat yang digunakan	Dataset dari UCI <i>Machine Learning Repository</i> berupa data Iris 150 baris, Sonar 208 baris dan Wine 178 baris	<i>K-means</i> dan SOM <i>clustering</i> dengan penilaian <i>percent misclassified</i>	Bahwa analisis <i>percent misclassified</i> menunjukkan <i>K-means</i> lebih unggul dari SOM, seperti ditunjukkan oleh grafik visualisasi keluaran (secara terpisah dan bersamaan dengan PCA)
(Elvi, 2021)	Angka peningkatan perceraian antar daerah tidak sama dan belum tersaji secara memadai dalam hal sebaran dalam wilayah Jambi	Daerah dengan peningkatan kasus perceraian signifikan tidak diketahui	Daftar wilayah, faktor penyebab, jumlah kasus	<i>K-means clustering</i>	<i>Clustering</i> perkara perceraian berdasarkan kelurahan di Kota Jambi dengan data sebanyak 62 kelurahan, maka diperoleh hasil pengelompokan dengan <i>cluster</i> tinggi (C2) sebanyak 12 kelurahan, <i>cluster</i> sedang (C1) sebanyak 20 kelurahan dan <i>cluster</i> rendah (C0) sebanyak 30 kelurahan
(Satria et al. 2023)	Angka peningkatan perceraian antar daerah tidak sama dan belum tersaji secara memadai dalam hal pengelompokan berdasar wilayah di Pekanbaru	Tidak diketahui kelompok wilayah dalam kasus tertentu	Daftar wilayah, selisih umur, pendidikan P, pendidikan T, Alasan, Nomor Perkara	<i>K-Medoids</i> , <i>David Bouldin Index</i>	Pengelompokan kasus perceraian terkait atribut bagian/kota, perbedaan umur, pendidikan penggugat, pendidikan tergugat dan alasan perceraian menunjukkan bahwa jumlah <i>cluster</i> yang

Nama, Tahun	Permasalahan	Akibat	Data	Metode	Hasil
					terbentuk adalah 3 <i>cluster</i> , jumlah data di setiap <i>cluster</i> , yaitu. <i>cluster</i> 1 terdiri dari 565 data, <i>cluster</i> 2 terdiri dari 395 data dan <i>cluster</i> 3 terdiri dari 288 data
(Wang & Ferrus, 2021)	Data OAM yang dihasilkan oleh jaringan radio seluler sangat besar termasuk keragaman atribut. Meskipun ada metode pengelompokan yang canggih seperti <i>Self-Organizing Maps</i> (SOM), namun tidak ada literatur yang menunjukkan penerapan metode kumpulan data OAM ini dengan sejumlah besar atribut (>20) yang dikumpulkan dari penerapan jaringan langsung	Ketika menggunakan SOM tidak ada informasi yang diperoleh secara eksplisit tentang <i>cluster</i> setelah pelatihan SOM pada data yang mendasarinya, sehingga teknik <i>K-means</i> untuk mengelompokkan unit SOM harus dilakukan	Data yang digunakan mencakup data 63 sel LTE yang dioperasikan di 11 base station (BS) sel makro dan 6 BS sel mikro yang mencakup area seluas sekitar 3,2 km ² . Dalam semua kasus, digunakan bandwidth saluran 20 dan 10 MHz	Kombinasi SOM dan <i>K-means</i>	SOM-K telah memberikan hasil penggabungan metode konvensional dan keahlian yang diverifikasi berdasarkan perilaku pola sel jangka panjang dan pendek. Hasil SOM-K dapat dipahami secara rutin, sehingga meningkatkan kepercayaan terhadap analisis baru dan penerapannya di bidang jaringan LTE. Untuk memberikan pola sel yang efektif, pengelompokan data berdimensi tinggi perlu menggunakan SOM-K, karena <i>K-means</i> saja tidak cukup

Nama, Tahun	Permasalahan	Akibat	Data	Metode	Hasil
(Asriny et al., 2021)	Indonesia merupakan negara yang rawan terhadap bencana. Berdasar data BNPB paling tidak terjadi 6868 bencana dalam 3 tahun terakhir. Perlu untuk melakukan pengelompokan kasus bencana alam hidrometeorologi khususnya di Pulau Jawa	Mitigasi menjadi tidak tepat jika informasi tidak tersaji secara akurat	Dataset kasus bencana dengan 7 variabel	<i>K-medoids</i> , SOM, <i>internal validation</i>	Berdasar SOM, <i>cluster 1</i> terdiri 108 anggota dan <i>cluster 2</i> terdiri 10 anggota. Berdasar <i>K-Medoids</i> , <i>cluster 1</i> terdiri dari 86 anggota dan <i>cluster 2</i> terdiri dari 32 anggota. Metode <i>clustering</i> terbaik pada penelitian ini adalah metode <i>Self-Organizing Maps</i> dengan nilai standar deviasi rasio sebesar 0,517399. Dimana nilai ini lebih kecil dibandingkan dengan metode <i>K-Medoids</i>
(Galvan et al., 2021)	Penyebaran virus Covid-19 di Brasil yang terjadi secara spasial dan temporal belum dipetakan secara komprehensif	Tidak diketahui perilaku penyebaran virus	Data kota, negara bagian, pasien terpapar, dan pasien meninggal	SOM, geo-spasial	Pengelompokan SOM mampu mengelompokkan kota, negara bagian, dan wilayah secara spasial menurut kasus virus corona, dengan perilaku serupa
(Syarli & Qashlim, 1970)	Informasi pemetaan pendidikan di Dinas Pendidikan Kabupaten Polewali Mandar masih kurang memadai	Tidak diketahui tingkat kondisi pendidikan suatu wilayah di Kabupaten tersebut	APK, APM, PAJ, R-S/G, AM	SIG, SOM	Dihasilkan sebuah sistem informasi pemetaan pendidikan yang menghasilkan sebuah laporan berupa klasifikasi tingkat pendidikan berdasarkan wilayah kecamatan dan penyajian informasinya dalam bentuk peta

Nama, Tahun	Permasalahan	Akibat	Data	Metode	Hasil
(Anis & Isnanto, 2014)	Memvisualisasi hasil pengelompokan data dalam bentuk web-gis	Tidak diterangkan	Peta rupa bumi Kota Semarang dan DPT pada KPU Kota Semarang	SIG, SOM	Aplikasi yang telah dibuat dapat digunakan untuk mempermudah melihat data DPT secara visual sehingga dapat digunakan untuk membantu dalam pengambilan keputusan
(Organizing et al., 2018)	Banyak calon siswa SMK merasa tidak mempunyai bakat dan tidak menguasai dalam beberapa materi pelajaran	Kesulitan menentukan pilihan jurusan di SMK	Kuisisioner calon peserta didik, dan literatur	SOM	pengelompokan menggunakan <i>Self Organizing Map</i> (SOM) dapat membantu calon siswa mengambil keputusan berdasarkan perhitungan skill, bakat, serta minat yang dimiliki siswa. Pengelompokan perlu dilakukan agar dapat membantu menentukan jurusan pada SMK yang diminati
(Febianto & Palasara, 2019)	Indikator-indikator yang menentukan kriteria kemiskinan suatu wilayah tidak ditentukan secara baik	Data kemiskinan disetiap daerah berbeda yang berpengaruh pada pengambilan keputusan	Data informasi kemiskinan kabupaten/ kota tahun 2018	KDD, <i>K-means</i>	Hasil penelitian ini didapat 5 <i>cluster</i> dengan pemetaan karakteristik dari setiap kelompok yang terbentuk berdasarkan nilai rata-rata tertinggi dan terendah dari setiap indikator kemiskinan provinsi Jawa Barat tahun 2018

Nama, Tahun	Permasalahan	Akibat	Data	Metode	Hasil
(Kartikasari, 2021)	Belum terdapat visualisasi sebaran pangan yang dapat membantu memantau angka kecukupan pangan di Indonesia berdasarkan data rata-rata konsumsi kalori per kapita sehari	Informasi tidak tersaji secara akurat yang mempengaruhi pengambilan keputusan	Data rata-rata konsumsi kalori per kapita per hari Bulan Maret 2020 seluruh Indonesia menurut kelompok komoditas (kkal)	<i>Davies-Bouldin Index</i> , SOM	Penggunaan algoritma <i>self-organizing map</i> menghasilkan pengelompokan wilayah yang menunjukkan angka kecukupan kalori masih dibawah standar
(Annas et al., 2022)	Pertumbuhan kepadatan penduduk, pembangunan pusat industri dan peningkatan volume kendaraan di Makassar memicu peningkatan polusi udara. Sementara itu, belum ada pengetahuan awal tentang kondisi pencemaran udara di wilayah tersebut.	Makassar menjadi menderit polusi udara parah dan kondisi tersebut akan berpengaruh buruk bagi kesehatan manusia maupun lingkungan. Sementara itu, penerapan regulasi lingkungan menjadi kurang efektif imbas dari informasi sebaran titik pencemaran yang belum ada atau tidak akurat	Data hasil pengukuran nilai parameter kualitas udara. Pengukuran dilakukan siang dan malam hari dengan variabel yang digunakan adalah Total Partikel Tersuspensi (TSP), Sulfur Dioksida (SO ₂), Nitrogen Dioksida (NO ₂), Karbon Dioksida (CO), dan Lapisan Ozon (O ₃)	<i>K-means</i> , SOM, SIG	Diperoleh jumlah <i>cluster</i> optimal adalah 7 <i>cluster</i> regional, sementara pada malam hari, terkonfirmasi sebanyak 4 <i>cluster</i> regional. Hal ini menunjukkan adanya perbedaan distribusi karakteristik pencemaran pada setiap kelompok wilayah yang terbentuk antara siang dan malam. Integrasi <i>K-means</i> dan SOM telah dapat mengelompokkan karakteristik wilayah di Kota Makassar, sedang GIS mampu memvisualisasinya secara memadai

Nama, Tahun	Permasalahan	Akibat	Data	Metode	Hasil
(Melo Riveros et al., 2019)	Kasus nyeri punggung menimpa setidaknya 90% populasi manusia selama hidup mereka. Mendiagnosis orang dengan masalah punggung sulit karena berbagai kondisi medis yang mungkin ada. Sejumlah besar pemeriksaan medis diperlukan untuk menentukan masalah yang mempengaruhi pasien	Deteksi yang lemah terhadap banyak pasien, mempengaruhi lama waktu dan jumlah pemeriksaan.	Database terdiri dari kumpulan data biomedis dengan 310 sampel, masing-masing dengan 6 karakteristik biomekanik berasal dari bentuk orientasi panggul dan kolom lumbar	Algoritma <i>K-means</i> dan SOM	Model yang dilatih dengan SOM mengungguli model yang dilatih dengan <i>K-means</i> , dengan peningkatan deteksi pasien masalah tulang belakang. Selain itu, ditemukan bahwa model SOM dan <i>K-means</i> menghasilkan presisi yang sama dibandingkan dengan model yang diperoleh dengan algoritma berbeda yang dilaporkan di tempat lain. Nilai yang dihasilkan adalah kesepakatan dengan orang-orang ahli dokter ortopedi
(Priyo, 2024)	Data perkara perceraian di PA Yogyakarta yang telah di anonimisasi sudah dipublikasi di <i>website</i> , akan tetapi belum bisa menggambarkan informasi karakteristik perceraian di kelurahan-kelurahan lingkup Kota Yogyakarta	Data yang disajikan belum mampu memberikan dukungan bagi pengambil keputusan dalam menentukan langkah sosialisasi kesehatan keluarga dan upaya pencegahan perceraian	Data terdiri dari 589 register cerai pada tahun 2021 yang dirangkum dalam 45 baris data sesuai kelurahan di wilayah Kota Yogyakarta	Algoritma SOM dan <i>K-means clustering</i> , <i>internal validation Silhouette Coefficient</i>	Diperoleh 2 <i>cluster</i> optimal berdasar pengukuran <i>Silhouette</i> . Dalam perbandingan rasio standar deviasi S_w terhadap S_b , model yang dilatih dengan SOM mengungguli model yang dilatih dengan <i>K-means</i> dalam kasus pengelompokan ini. <i>Clustering</i> dengan SOM menunjukkan, <i>cluster</i> 1 berisikan 11 kelurahan dengan kasus perceraian cukup tinggi,

Nama, Tahun	Permasalahan	Akibat	Data	Metode	Hasil
					<p>sedang <i>cluster 2</i> berisikan 34 kelurahan dengan kasus perceraian rendah hingga sedang. Variabel rentang “usia perkawinan 0-7 tahun” dan alasan berupa “perselisihan dan pertengkaran terus menerus” menjadi rata-rata kejadian tertinggi di kedua <i>cluster</i>.</p>