

Analisis Clustering Tingkat Buta Aksara di Indonesia Berdasarkan Provinsi Tahun 2024 Menggunakan Algoritma K-Means

Fuja Fania Patmawati^{*1}, Muhamad Januaraka Naufal Putra², Siti Lutfiani Afifah³

^{1,2,3}Jurusan Administrasi Publik, UIN Sunan Gunung Djati Bandung, Indonesia

^{*}Korespondensi: fujafania02@gmail.com

KATA KUNCI	ABSTRAK
buta aksara, clustering, algoritma K-means	<i>Angka buta aksara di Indonesia masih menjadi isu penting, meskipun terjadi penurunan dalam beberapa tahun terakhir. Penelitian ini bertujuan untuk menganalisis tingkat buta aksara di setiap provinsi Indonesia pada tahun 2024, dengan menggunakan algoritma K-Means untuk mengelompokkan data berdasarkan tingkat buta aksara pada kelompok usia 15+, 15-45, dan 45+. Data yang diperoleh dari Badan Pusat Statistik (BPS) diolah menggunakan teknik data mining, dengan Elbow Method dan Silhouette Analysis untuk mengoptimalkan jumlah klaster. Hasil analisis menunjukkan bahwa provinsi-provinsi terbagi dalam tiga klaster: klaster rendah (25 provinsi), klaster sedang (11 provinsi), dan klaster tinggi (2 provinsi). Provinsi dengan tingkat buta aksara tinggi didominasi oleh wilayah terpencil dengan akses pendidikan yang terbatas. Berdasarkan teori kesenjangan sosial Pierre Bourdieu, penelitian ini memberikan rekomendasi strategis, termasuk peningkatan infrastruktur pendidikan, pelatihan guru, serta program literasi berbasis komunitas. Dengan hasil ini, penelitian diharapkan dapat mendukung perumusan kebijakan untuk mempercepat penurunan angka buta aksara di Indonesia.</i>
KEYWORDS	ABSTRACT
illiteracy, clustering, K-means algorithm	<i>Illiteracy in Indonesia is still an important issue, although there has been a decline in recent years. This study aims to analyze the illiteracy rate in each province in Indonesia in 2024, using the K-Means algorithm to group data based on the illiteracy rate in the 15+, 15-45, and 45+ age groups. Data obtained from the Central Statistics Agency (BPS) were processed using data mining techniques, with the Elbow Method and Silhouette Analysis to optimize the number of clusters. The results of the analysis show that the provinces are divided into three clusters: low cluster (25 provinces), medium cluster (11 provinces), and high cluster (2 provinces). Provinces with high illiteracy rates are dominated by remote areas with limited access to education. Based on Pierre Bourdieu's theory of social inequality, this study provides strategic recommendations, including improving educational infrastructure, teacher training, and community-based literacy programs. With these results, the study is expected to support policy formulation to accelerate the reduction of illiteracy in Indonesia.</i>

Latar Belakang

Angka buta aksara di Indonesia masih menjadi isu penting dalam bidang pendidikan dan pemberdayaan masyarakat, meskipun menurut Badan Pusat Statistik (BPS), angka buta

huruf dari tahun ke tahun mengalami penurunan. Berdasarkan Badan Pusat Statistik (BPS) Laporan Survei Indikator Kinerja Utama Pendidikan (SIKP) 2019 mencatat angka buta aksara penduduk usia 15-59 tahun sebesar 2,34% dan penduduk usia 60 tahun ke atas mencapai 8,63%. Selain itu, publikasi Indeks Pembangunan Manusia (IPM) pada tahun 2021 (Badan Pusat Statistik (BPS), 2021) juga mencatat angka buta aksara penduduk usia 15 tahun ke atas sebesar 2,06%, dan Laporan Survei Angkatan Kerja Nasional (Sakernas) Badan Pusat Statistik (BPS), 2018 menunjukkan angka buta aksara pada kelompok usia yang sama adalah 2,73%. Kajian Profil Pendidikan Dasar dan Menengah 2017 mencatat angka buta aksara pada penduduk usia 15-59 tahun sebesar 2,07% dan meningkat menjadi 9,87% pada penduduk usia 60 tahun ke atas (Kementerian Pendidikan dan Kebudayaan, 2017). Meskipun terdapat penurunan angka buta huruf secara keseluruhan, masih terdapat jutaan orang, terutama di daerah terpencil, yang belum memiliki kemampuan membaca dan menulis dikarenakan kurangnya infrastruktur dan sumber daya pendidikan yang merata. Penelitian ini dilakukan untuk mengetahui sejauh mana tingkat kebutaan aksara ditahun 2024, dengan fokus pada perbedaan tingkat buta aksara di setiap provinsi. Data mengenai buta aksara di setiap provinsi pada tahun 2024 akan digunakan untuk mengidentifikasi daerah-daerah dengan tingkat buta aksara yang tinggi, memahami penyebabnya, dan mengukur seberapa berhasil pemerintah dan lembaga terkait dalam upaya mengurangi buta aksara.

Buta aksara merupakan permasalahan yang masih menjadi tantangan besar bagi Indonesia dalam upaya mencapai pembangunan berkelanjutan. Angka buta aksara yang masih berkiraan di angka 2-3% menunjukkan bahwa masih ada kesenjangan akses dan kualitas pendidikan di berbagai wilayah. Permasalahan ini juga memiliki daya tarik tersendiri untuk kita bahas lebih dalam, mengingat pemerintah telah berupaya keras untuk menekan angka buta aksara dengan berbagai program, namun hasil yang didapat belum mencapai diangka optimal.

Tujuan dari penelitian ini adalah untuk mengidentifikasi faktor-faktor yang dapat mempengaruhi tingginya angka buta aksara di Indonesia, baik dari sisi geografis, sosial dan ekonomi, maupun kebijakan pemerintah. Selain itu, penelitian ini juga akan mengkaji efektivitas program-program perbaikan buta aksara yang telah dilaksanakan sejauh ini. Dengan menganalisis permasalahan tersebut secara menyeluruh, diharapkan penelitian ini dapat memberikan rekomendasi strategis dalam upaya percepatan perbaikan buta aksara di

Indonesia di masa yang akan mendatang. Berdasarkan data dari situs resmi BPS dan beberapa penelitian yang telah terbukti, maka kemudian penelitian dilakukan dengan menggunakan teknik data mining yang mengubah data menjadi informasi. Penelitian ini menggunakan teknik algoritma K-means untuk mengolah jumlah penduduk buta aksara di setiap provinsi.

Penelitian Terdahulu

Studi oleh Siregar & Wahiddin, (2020) menunjukkan bahwa wilayah-wilayah di bagian Timur Indonesia, seperti Papua dan Nusa Tenggara Barat memiliki persentase buta aksara yang lebih tinggi dibandingkan wilayah lainnya. Sementara itu, penelitian oleh Ningsih et al., (2019) menemukan bahwa buta aksara berkaitan dengan pendidikan yang menjadi salah satu indikator kemajuan sebuah bangsa. Rendahnya angka partisipasi sekolah di beberapa provinsi berdampak langsung pada tingginya angka buta aksara.

Selain itu, program literasi yang tidak merata juga telah menunda pencapaian tujuan buta aksara di beberapa daerah. Studi literatur ini menunjukkan bahwa untuk mencapai tujuan bebas buta aksara, intervensi pendidikan harus menjadi prioritas, terutama di daerah dengan angka buta aksara yang tinggi. Berdasarkan teori pembangunan manusia dari Amartya Sen, akses terhadap pendidikan dasar adalah hak asasi dan faktor penting yang mendukung kemampuan seseorang untuk berpartisipasi dalam masyarakat secara produktif.

Tinjauan Teoritis

Teori Kesenjangan Sosial (Social Inequality Theory)

Penelitian ini menggunakan teori dari Pierre Bourdieu, seorang sosiolog asal Prancis yaitu Teori Kesenjangan Sosial (Social Inequality Theory), (Siregar & Wahiddin, 2020). Bourdieu mengembangkan teori yang sangat berpengaruh mengenai kesenjangan sosial, terutama dalam konteks pendidikan. Bourdieu mengemukakan bahwa kesenjangan sosial dalam pendidikan tidak hanya berkaitan dengan perbedaan material (seperti akses ke sumber daya), tetapi juga dengan perbedaan simbolik yang lebih halus, yaitu bagaimana nilai-nilai, norma, dan budaya dari kelas sosial tertentu lebih dihargai dalam sistem pendidikan (Saputra & Nataliani, 2021).

Terdapat tiga aspek dalam teori ini, yaitu : (1) Menjelaskan perbedaan akses pendidikan antar wilayah; (2) Menganalisis faktor-faktor yang memengaruhi kesenjangan Pendidikan; dan (3) Mengidentifikasi akar permasalahan buta aksara dari perspektif sosial

Data Mining

Data mining merujuk pada kegiatan yang menggunakan metode analisis lanjutan untuk mencari dan mengambil data penting dari set data yang besar dan rumit. Proses untuk mengumpulkan dan mengambil informasi ini dapat dilakukan dengan bantuan perangkat lunak yang didukung oleh analisis statistik, matematika, atau teknologi kecerdasan buatan (AI). Menurut Turban dalam bukunya yang berjudul "Decision Support Systems and Intelligent Systems", data mining merupakan suatu terminologi yang digunakan untuk menjelaskan proses pengungkapan informasi dalam suatu basis data. (Turban, E., & Aronson, 2001). Menurut Hoffer, J. A., Prescott, M. B., & McFadden, (2002) tujuan dari data mining yaitu :

1. Eksplanatory: digunakan untuk menguraikan suatu kondisi penelitian.
2. Confirmatory: digunakan untuk memastikan sebuah pernyataan atau mempertegas suatu hipotesis.
3. Eksploratory: digunakan untuk menganalisa data yang memiliki hubungan yang baru atau mencari pola baru yang sebelumnya tidak terdeteksi.

Tahapan dalam data mining diantaranya sebagai berikut :

- a. Data Selection, merupakan langkah untuk memilih data yang relevan dan mempunyai standar yang cukup memuaskan untuk dilakukan analisis lebih lanjut. Data yang sudah dipilih dari database operasional disimpan dalam file khusus yang terpisah dari data operasional asli.
- b. Pre-processing/Cleaning, merupakan proses pembersihan data dengan membuang data salinan atau duplikat, memeriksa ketidak konsistenan data, dan memperbaiki kesalahan pada sebuah data.
- c. Transformation, pada data yang telah dipilih untuk memastikan format dan struktur data itu optimal sehingga algoritma data mining dapat berfungsi dengan baik.
- d. Data mining adalah proses yang meliputi penerapan cara dan strategi khusus untuk menemukan pola dan informasi penting dalam data yang telah dipilih.
- e. Interpretation / Evaluation Pola informasi yang dihasilkan dari kegiatan penggalian data perlu disajikan dalam bentuk yang jelas dan mudah dimengerti oleh semua pihak yang terlibat. Di tahap ini, akan di evaluasi apakah pola dan informasi yang diperoleh sesuai dengan fakta dan dugaan yang sudah ada.

Clustering

Clustering adalah teknik yang diterapkan untuk mengorganisir data ke dalam kategori atau kelompok berdasarkan kesamaan fitur tertentu. Metode-metode yang terdapat dalam clustering adalah sebagai berikut :

1. Partitioning Method, yaitu data dikelompokkan menjadi beberapa cluster tanpa struktur hierarki antara satu dengan lainnya
2. Exclusive Clustering, Ini adalah jenis cluster yang objeknya eksklusif untuk satu kelompok saja dan tidak bisa dimiliki oleh kelompok lainnya.
3. Overlapping Clustering, Merupakan pengelompokan di mana benda atau objek bisa dimasukkan ke beberapa kelompok atau kluster. Dengan cara ini, suatu objek yang tidak eksklusif pada suatu cluster, dapat menjadi bagian dari beberapa cluster yang berbeda jika setiap cluster memiliki karakteristik atau atribut yang serupa.
4. Hierarchical Clustering, Sebuah metode klusterisasi yang mengatur data ke dalam hierarki seperti bagan, di mana cluster besar akan dibagi menjadi sub-cluster yang lebih kecil secara bertahap.
5. Density-based Clustering, Metode ini adalah metode pengelompokan data yang mengidentifikasi kelompok (cluster) berdasarkan kepadatan data di suatu wilayah tertentu. Cluster ini berkaitan dengan kepadatan atau kerapatan objek. Kelompok objek yang semakin padat dan rapat akan dipisahkan oleh kelompok objek yang semakin renggang. Metode ini efektif dalam mendeteksi cluster yang bentuknya tidak beraturan,
6. Model-based Methods, Merupakan pendekatan yang menggunakan model untuk mengelompokkan data. Yang berarti dalam metode ini, diasumsikan bahwa data dalam setiap cluster (kelompok) mengikuti suatu distribusi atau model statistik tertentu.

Algoritma K-Means

Algoritma K-Means adalah teknik pengelompokan yang mengorganisir data sesuai dengan pusat cluster (centroid) yang terdekat dengan data tersebut. K-means merupakan salah satu algoritma yang bersifat unsupervised learning. Digunakan untuk mengelompokkan data ke dalam sekumpulan cluster berdasarkan kesamaan atribut. Proses Kerja Algoritma K-Means:

- a. Inisialisasi : Tentukan jumlah klaster (kkk) dan pilih titik awal untuk centroid (bisa secara acak).

- b. Pengelompokan : Setiap data ditugaskan ke kluster dengan centroid terdekat berdasarkan jarak (biasanya menggunakan jarak Euclidean).
- c. Pembaruan Centroid : Hitung kembali lokasi centroid dengan menjadikannya sebagai rata-rata dari seluruh titik data yang ada dalam kluster itu.
- d. Iterasi : Ulangi langkah 2 dan 3 hingga centroid tidak berubah atau mencapai batas iterasi tertentu.

R Studio

R Studio merupakan salah satu software Integrated Development Environment (IDE) yang sifatnya terbuka dan gratis, biasanya digunakan dalam pemrograman berorientasi statistik. R Studio juga merupakan program yang umum digunakan untuk mengolah dan menganalisis data.

Metode Penelitian

Metode penelitian ini menggunakan pendekatan kuantitatif dengan pengumpulan data melalui website internet. Sumber data utama diperoleh dari website Badan Pusat Statistik (BPS) pada Angka Buta Aksara Menurut Provinsi dan Kelompok Umur (Persen), 2024. Waktu pengumpulan data dilakukan selama kurang lebih satu minggu.

Populasi penelitian ini adalah seluruh masyarakat yang tinggal di Indonesia khususnya di daerah terpencil, dengan rentang usia 15 hingga 45 tahun keatas. Sampel penelitian ditentukan menggunakan teknik web scraping, dengan mengumpulkan data melalui website. Data asli yang didapat oleh peneliti terdiri dari tahun 2021 sampai tahun 2024 dengan lima atribut yaitu atribut provinsi, umur 15+, umur 15-45, umur 45+ dan atribut tahun. Setelah dilakukan seleksi data maka hanya empat atribut yang dipakai yaitu atribut provinsi, Umur 15+, Umur 15-45, dan Umur 45+, Sehingga didapatkan dataset tahun 2024 terlihat pada gambar 1.

Hasil dan Pembahasan

Seleksi Data

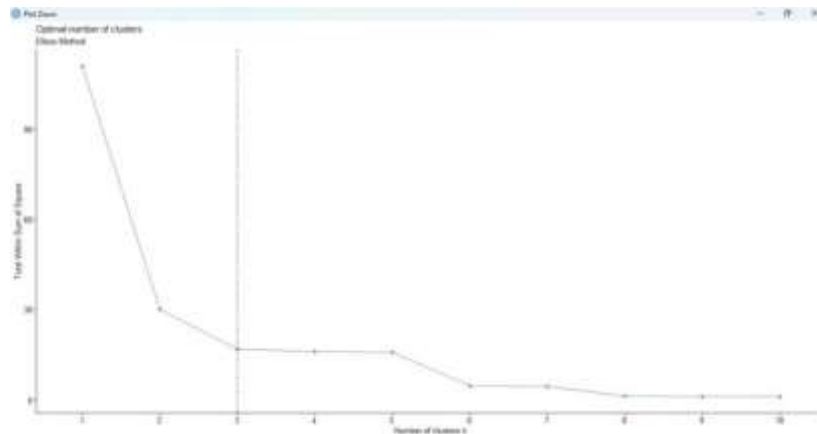
Seleksi data dilakukan dari sumber data utama dari BPS. Jumlah awal data adalah 38 provinsi rentan waktu dari 2024 yang tersedia. Selanjutnya, dilakukan penyaringan data dengan merujuk pada data per 2024, mengurangi jumlah sampel menjadi satu tahun data yang kemudian disimpan dalam sebuah file terpisah guna proses analisis lanjutan.

38 Provinsi	Angka Buta Aksara Menurut Provinsi dan Kelompok Umur (Persen)		
	15+	15-44	45+
	2024	2024	2024
ACEH	1,66	0,15	4,51
SUMATERA UTARA	0,91	0,12	2,27
SUMATERA BARAT	0,63	0,07	1,57
RIAU	0,8	0,04	2,3
JAMBI	1,98	0,11	4,89
SUMATERA SELATAN	1,32	0,13	3,35
BENGKULU	2,09	0,12	5,22
LAMPUNG	2,64	0,32	6,14
KEP. BANGKA BELITUNG	1,8	0,24	4,3
KEP. RIAU	0,8	0,09	2,39
DKI JAKARTA	0,35	0,09	0,77
JAWA BARAT	1,43	0,07	3,61
JAWA TENGAH	5,5	0,29	11,97
DI YOGYAKARTA	4,35	0,24	9,52
JAWA TIMUR	5,87	0,35	12,31
BANTEN	1,59	0,06	4,17
BALI	4,21	0,18	9,37
NUSA TENGGARA BARAT	30,17	1,08	29,09
NUSA TENGGARA TIMUR	4,85	1,23	11,23
KALIMANTAN BARAT	4,97	0,5	12,56
KALIMANTAN TENGAH	1,17	0,1	2,94
KALIMANTAN SELATAN	1,66	0,08	3,99
KALIMANTAN TIMUR	1,04	0,06	2,65
KALIMANTAN UTARA	2,63	0,32	6,61
SULAWESI UTARA	0,25	0,05	0,51
SULAWESI TENGAH	1,65	0,38	3,78
SULAWESI SELATAN	5,89	1,08	12,89
SULAWESI TENGGARA	3,71	0,44	9,78
GORONTALO	1,45	0,53	2,91
SULAWESI BARAT	5	1,19	11,7
MALUKU	0,47	0,15	1,05
MALUKU UTARA	1,2	0,08	3,26
PAPUA BARAT	2,11	1,01	4,85
PAPUA BARAT DAYA	2,16	0,48	5,7
PAPUA	1,63	0,49	3,29
PAPUA SELATAN	5,79	3,54	9,67
PAPUA TENGAH	15,31	13,01	20,93
PAPUA PEGUNUNGAN	29,53	22,74	47,45
INDONESIA	3,33	0,43	7,66

Gambar 1. Hasil Seleksi Data
Sumber: Data Diolah Peneliti (2024)

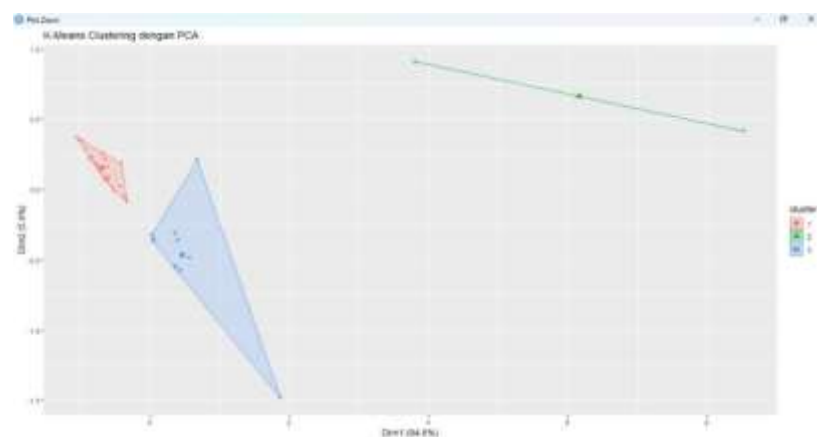
Elbow method

Hasil perhitungan menggunakan Elbow Method menunjukkan bahwa jumlah kluster yang optimal adalah 3. Titik optimal dalam Elbow Method terletak pada lekukan kedua, di mana menambah jumlah kluster berikutnya tidak menunjukkan pengurangan varians yang berarti. Elbow point pada titik ketiga menandakan jumlah kluster yang mencapai keseimbangan antara kompleksitas model (jumlah kluster) dan variasi yang dijelaskan oleh model. Nilai WSS (*Within Sum of Squares*) sebesar 16.88826 menunjukkan bahwa jarak total dalam cluster relatif kecil, yang berarti data dalam cluster cukup rapat dan terkelompok dengan baik.



Gambar 2. Hasil Elbow Method

K- Means Clustering dengan PCA



Gambar 3. Hasil klasterisasi menggunakan PCA

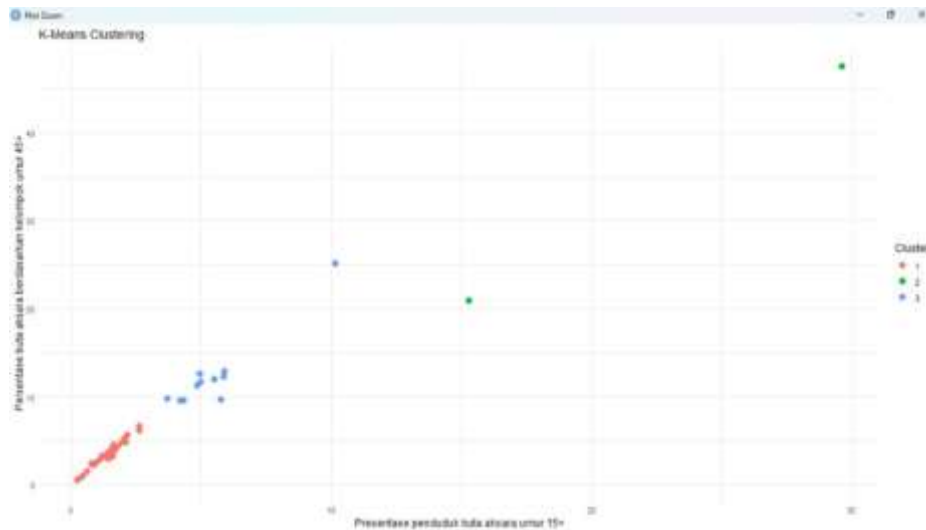
Hasil klasterisasi menggunakan metode optimasi Elbow, diperoleh jumlah anggota masing-masing klaster, yaitu Klaster 1 dengan 25 anggota, Klaster 2 dengan 2 anggota, dan Klaster 3 dengan 11 anggota. Selain itu, hasil perhitungan Sum of Squares (SS) untuk masing-masing klaster adalah 1.308087 untuk Klaster 1, 11.313904 untuk Klaster 2, dan 4.266264 untuk Klaster 3, dengan nilai ratio Between cluster SS/Total SS (mencapai 0.8478536 atau setara dengan 84.8 % yang berarti variasi dalam data pembagian cluster yang dihasilkan di indikasi bahwa cluster cukup baik dalam menggambarkan struktur data.. Selanjutnya, perhitungan klasterisasi menggunakan metode optimasi Silhouette juga dilakukan untuk mendapatkan hasil yang lebih akurat.

Provinsi	x1	x2	x3	cluster
ACEH	1,66	0,15	4,51	1
SUMATERA UTARA	0,91	0,12	2,27	1
SUMATERA BARAT	0,63	0,07	1,57	1
RIAU	0,8	0,04	2,3	1
JAMBI	1,98	0,11	4,89	1
SUMATERA SELATAN	1,32	0,13	3,35	1
BENGKULU	2,09	0,12	5,22	1
LAMPUNG	2,64	0,32	6,14	1
KEP. BANGKA BELITUNG	1,8	0,24	4,3	1
KEP. RIAU	0,8	0,09	2,39	1
DKI JAKARTA	0,35	0,09	0,77	1
JAWA BARAT	1,43	0,07	3,61	1
JAWA TENGAH	5,5	0,29	11,97	3
DI YOGYAKARTA	4,35	0,24	9,52	3
JAWA TIMUR	5,87	0,35	12,31	3
BANTEN	1,59	0,06	4,17	1
BALI	4,21	0,18	9,57	3
NUSA TENGGARA BARAT	10,17	1,08	25,09	3
NUSA TENGGARA TIMUR	4,85	1,23	11,23	3
KALIMANTAN BARAT	4,97	0,5	12,56	3
KALIMANTAN TENGAH	1,17	0,1	2,94	1
KALIMANTAN SELATAN	1,66	0,08	3,99	1
KALIMANTAN TIMUR	1,04	0,06	2,65	1
KALIMANTAN UTARA	2,63	0,32	6,61	1
SULAWESI UTARA	0,25	0,05	0,51	1
SULAWESI TENGAH	1,65	0,38	3,78	1
SULAWESI SELATAN	5,89	1,08	12,89	3
SULAWESI TENGGARA	3,71	0,44	9,78	3
GORONTALO	1,45	0,53	2,91	1
SULAWESI BARAT	5	1,19	11,7	3
MALUKU	0,47	0,15	1,05	1
MALUKU UTARA	1,2	0,08	3,26	1
PAPUA BARAT	2,11	1,01	4,85	1
PAPUA BARAT DAYA	2,16	0,48	5,7	1
PAPUA	1,63	0,49	3,29	1
PAPUA SELATAN	5,79	3,54	9,67	3
PAPUA TENGAH	15,31	13,01	20,93	2
PAPUA PEGUNUNGAN	29,63	22,74	47,45	2

Gambar 4. Tabel Hasil Clustered

K-means Clustering

Hasil clustering K-Means berdasarkan data dari 38 provinsi, dengan presentase penduduk buta aksara usia 15+ (sumbu X) dan 45+ (sumbu Y). Cluster 1 (merah) mencakup sebagian besar provinsi dengan nilai rendah pada kedua sumbu, menunjukkan wilayah dengan tingkat buta aksara rendah. Cluster 3 (biru) mencakup provinsi dengan nilai sedang, mencerminkan wilayah dengan tingkat buta aksara yang cukup signifikan, terutama pada kelompok usia 45+. Sementara itu, Cluster 2 (hijau) mencakup provinsi dengan nilai sangat tinggi pada kedua sumbu, menggambarkan wilayah dengan tingkat buta aksara yang sangat tinggi dan menjadi prioritas utama untuk intervensi. Secara keseluruhan, sebagian besar provinsi memiliki tingkat buta aksara rendah, tetapi ada beberapa provinsi yang membutuhkan perhatian lebih untuk meningkatkan literasi.



Gambar 5. Hasil Clustering Diagram Titik

Berdasarkan hasil evaluasi clustering yang diperoleh, kinerja clustering yang dilakukan menunjukkan hasil yang cukup baik. Nilai within-cluster sum of squares (WSS) sebesar 16.88826 mencerminkan pengelompokan data yang cukup rapat di dalam masing-masing cluster. Sementara itu, between cluster sum of squares (BSS) sebesar 94.11174 menunjukkan bahwa cluster-cluster yang terbentuk cukup terpisah dengan baik. Rasio BSS/TSS sebesar 0.8478536 mengindikasikan bahwa lebih dari setengah variasi data dapat dijelaskan oleh perbedaan antar-cluster, yang menandakan bahwa model clustering ini cukup solid dan efektif dalam membedakan kelompok-kelompok data.

Hasil Pengelompokkan dengan Perhitungan Manual

Cluster	Provinsi	Jumlah
Persentase tinggi	Papua tengah, Papua pegunungan	2
Persentase sedang	Jawa Tengah, DI Yogyakarta, Jawa Timur, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Sulawesi Selatan, Sulawesi Tenggara, Sulawesi Barat, Papua Selatan	11
Persentase rendah	Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, Jawa Barat, Banten, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Gorontalo, Maluku, Maluku Utara, Papua Barat, Papua Barat Daya, Papua	25

Tabel 1. Hasil Pengelompokkan Manual

Dari tabel 1 dapat dikatakan bahwa terdapat 2 provinsi yang termasuk kedalam cluster 2 dengan persentase tinggi yaitu Papua Tengah dan Papua Pegunungan. Kemudian terdapat 11

provinsi yang masuk kedalam cluster 3 dengan persentase sedang. Lalu 25 provinsi dengan persentase rendah yang berada pada cluster 1. Total keseluruhan provinsi yaitu 38.

Kesimpulan

Penelitian ini menghasilkan pengelompokan data pada pengangguran di tingkat Provinsi berdasarkan 3 variabel yang relevan. Hasil analisis menghasilkan tiga klaster dengan karakteristik yang berbeda-beda, yang masing-masing klaster menunjukkan tingkat kerentanan dan kebutuhan yang berbeda terkait jumlah presentase pengangguran nya. Variabel : x_1 = "Presentase angka buta aksara pada umur 15 tahun", x_2 = "Persentase angka buta aksara pada umur 15-44 tahun", x_3 = "Persentase angka buta aksara pada umur 45 ke atas".

Cluster 1 Tingkat buta aksara yang rendah (25 provinsi) Mencakup provinsi dengan fasilitas yang sudah memadai dari segi infrastrukturnya dan juga sumber daya pendidikan yang sudah merata, Cluster 2 Tingkat buta aksara yang tinggi (2 provinsi) Mencakup provinsi dengan fasilitas yang kurang memadai, seperti minimnya infrastruktur dan partisipasi sekolah juga kurangnya pendidikan yang merata, Cluster 3 Tingkat buta aksara dengan persentase sedang (11 provinsi) Mencakup provinsi dengan fasilitas yang cukup memadai di sebagian besar tingkat pendidikan, namun masih ada yang belum merata.

Rekomendasi Kebijakan

Rekomendasi kebijakan untuk menurunkan tingkat buta aksara di Indonesia berdasarkan penelitian ini antara lain:

Pertama, provinsi dengan Persentase Tinggi Buta Aksara, dengan karakteristik akses pendidikan terbatas, wilayah terpencil, dan dominasi kelas sosial bawah yang tidak memiliki modal ekonomi dan simbolik yang memadai. Rekomendasi kebijakan yakni:

1. Peningkatan Akses Material (Modal Ekonomi) melalui: bangun dan perbaiki infrastruktur pendidikan (sekolah, perpustakaan, dan akses internet) di daerah terpencil; Subsidi pendidikan gratis untuk anak-anak dari keluarga serba kekurangan; dan program bantuan operasional sekolah dan beasiswa berbasis keluarga tidak mampu.
2. Peningkatan Modal Budaya dan Simbolik, melalui: program pendidikan literasi berbasis komunitas dengan pendekatan budaya lokal agar lebih diterima;

pelibatan pemuka masyarakat atau tokoh adat untuk menanamkan pentingnya pendidikan; dan pelatihan guru yang memahami konteks budaya dan sosial daerah setempat.

Kedua, provinsi dengan Persentase Sedang, dengan karakteristik akses pendidikan mulai membaik, tetapi masih terdapat ketimpangan antara kelompok sosial bawah dan atas. Rekomendasi Kebijakan yakni: Peningkatan Kualitas Pendidikan, melalui: perlu upaya intensif untuk menurunkan tingkat buta aksara, terutama di daerah dengan akses pendidikan menengah; program peningkatan kapasitas guru melalui pelatihan berkala; penyediaan akses ke materi pendidikan berkualitas, seperti buku, alat tulis, dan teknologi.

Ketiga, provinsi dengan Persentase Rendah dengan karakteristik: akses pendidikan sudah merata dan lebih stabil, tetapi tetap perlu intervensi untuk memastikan semua kelompok sosial mendapatkan manfaat yang sama. Rekomendasi Kebijakan yakni: kebijakan bisa fokus pada peningkatan literasi teknologi atau program pendidikan lanjutan; pastikan sekolah-sekolah di seluruh wilayah memiliki standar pengajaran yang sama; dan program peningkatan kualitas guru dan penyebaran merata tenaga pendidik ke daerah-daerah yang lebih tertinggal.

Daftar Pustaka

- Badan Pusat Statistik (BPS). (2018). *Laporan Survei Angkatan Kerja Nasional (Sakernas)*.
- Badan Pusat Statistik (BPS). (2019). *Laporan Survei Indikator Kinerja Utama Pendidikan (SIKP)*.
- Badan Pusat Statistik (BPS). (2021). *Indeks Pembangunan Manusia (IPM)*.
- Hoffer, J. A., Prescott, M. B., & McFadden, F. R. (2002). *Modern Database Management*. New York: pearsson education.
- Kementerian Pendidikan dan Kebudayaan. (2017). *Kajian Profil Pendidikan Dasar dan Menengah*.
- Ningsih, S. R., Damanik, I. S., Windarto, A. P., Tambunan, H. S., Jalaluddin, J., & Wanto, A. (2019). Analisis K-Medoids Dalam Pengelompokan Penduduk Buta Huruf Menurut Provinsi. *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 1(September), 721. <https://doi.org/10.30645/senaris.v1i0.78>
- Saputra, E. A., & Nataliani, Y. (2021). Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means. *Journal of*

Information Systems and Informatics, 3(3), 424–439.
<https://doi.org/10.51519/journalisi.v3i3.164>

Siregar, A. M., & Wahiddin, D. (2020). Penerapan Algoritma K-Means Dalam Mengurangi Tingkat Buta Aksara Di Indonesia Sebagai Penunjang Keputusan. *Scientific Student Journal for*, 1, 55–60.

Turban, E., & Aronson, J. (2001). *Decision Support Systems and Intelligent Systems*. New Jersey: Prentice Hall.