

PENGELOMPOKAN PERSENTASE BUTA HURUF UMUR 15-44 MENURUT PROVINSI MENGGUNAKAN ALGORITMA K-MEANS

Saifullah¹, Nani Hidayati²

¹ Program Studi Sistem Informasi, STIKOM Tunas Bangsa Pematangsiantar

² Program Studi Teknik Informatika, STIKOM Tunas Bangsa Pematangsiantar

Jl. Jend. Sudirman Blok A No. 1,2 & 3 Pematangsiantar

*E-mail: saifullah@amiktunasbangsa.ac.id

Abstract

Data Mining is a method that is often needed in large-scale data processing, so data mining has important access to the fields of life including industry, finance, weather, science and technology. In data mining techniques there are methods that can be used, namely classification, clustering, regression, variable selection, and market basket analysis. Illiteracy is one of the factors that hinder the quality of human resources. One of the basic things that must be fulfilled to improve the quality of human resources is the eradication of illiteracy among the community. The purpose of this study is to determine the clustering of illiterate communities based on provinces in Indonesia. The results of the study are illiterate data clustering according to the age proportion of 15-44 namely 1 high group node, low group has 27 nodes, and medium group 6 nodes. The results of this study become input for the government to determine illiteracy eradication policies in Indonesia based on provinces.

Kata Kunci: Illiterate, Data mining, K-Means Clustering

Abstrak

Data Mining termasuk metode yang sering dibutuhkan dalam pengolahan data berskala besar, maka data mining mempunyai akses penting pada bidang kehidupan diantaranya yaitu bidang industri, bidang keuangan, cuaca, ilmu dan teknologi. Pada teknik data mining terdapat metode-metode yang dapat digunakan yaitu klasifikasi, clustering, regresi, seleksi variabel, dan market basket analisis. Buta huruf merupakan salah satu faktor yang menghambat kualitas sumber daya manusia. Salah satu hal mendasar yang harus dipenuhi untuk meningkatkan kualitas sumber daya manusia adalah pemberantasan buta huruf di kalangan masyarakat Adapun tujuan penelitian ini adalah menentukan clustering masyarakat buta huruf berdasarkan propinsi di Indonesia. Hasil dari penelitian adalah data clustering buta huruf menurut propinsi umur 15-44 yaitu 1 node kelompok tinggi, kelompok rendah memiliki 27 node, dan kelompok sedang 6 node. Hasil penelitian ini menjadi bahan masukan kepada pemerintah untuk menentukan kebijakan pemberantasan buta huruf di Indonesia berdasarkan propinsi.

Kata Kunci: Buta Huruf, Data mining, K-Means Clustering

1. PENDAHULUAN

Data mining merupakan suatu langkah dalam knowledge discovery in databases (KDD) yang memiliki teknik menganalisa data untuk digali informasi tersembunyi dalam jumlah besar dan kompleks, sehingga menghasilkan output berupa karakteristik atau pola dari data tersebut. Salah satu teknik analisa Data Mining adalah analisis kelompok (cluster analysis) yang lebih dikenal dengan Clustering[1]

Analisis Pengelompokan/ Clustering merupakan proses membagi data dalam suatu himpunan ke dalam beberapa kelompok yang kesamaan datanya dalam suatu kelompok lebih besar dari pada kesamaan data tersebut dengan data dalam kelompok lain Potensi clustering adalah dapat digunakan untuk mengetahui struktur dalam data yang dapat dipakai lebih lanjut dalam berbagai aplikasi secara luas seperti klasifikasi, pengolahan gambar, dan pengenalan pola[2].

Buta huruf merupakan salah satu faktor yang menghambat kualitas sumber daya manusia. Salah satu hal mendasar yang harus dipenuhi untuk meningkatkan kualitas sumber daya manusia adalah pemberantasan buta huruf di kalangan masyarakat. Adapun tujuan penelitian ini adalah menentukan clustering masyarakat buta huruf. Hasil penelitian ini menjadi bahan masukan kepada pemerintah untuk menentukan kebijakan pemberantasan buta huruf di Indonesia berdasarakan propinsi.

2. METODE PENELITIAN

2.1. Data Mining

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari database yang besar. Tujuan utama data mining adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki[3].

2.2. Cluster

Algoritma Clustering mengelompokkan set dokumen ke dalam himpunan bagian atau klaster. Tujuan algoritma klaster adalah menciptakan klaster yang koheren secara internal, tetapi jelas berbeda satu sama lain. Dengan kata lain, dokumen dalam sebuah klaster harus semirip mungkin; dan dokumen dalam satu klaster harus sebeda mungkin dari dokumen dalam klaster lainnya[4]

2.3. K-Means

K-means adalah salah satu bentuk pengelompokan yang paling sederhana. Prosedurnya sederhana dan mudah untuk mengklasifikasikan data yang diberikan melalui sejumlah cluster. Penentuan centroid dilakukan dengan cara mengambil data pertama sebagai centroid pertama, data kedua sebagai centroid kedua, dan seterusnya hingga jumlah centroid yang diperlukan. Langkah berikutnya adalah

dengan menghitung jarak dari titik yang akan di cluster ke setiap centroid yang ada dan dikelompokkan sesuai dengan jarak terdekat kepada centroid-nya. Bila semua titik sudah masuk kedalam pengelompokan maka langkah pertama selesai. Kemudian langkah berikutnya, kita perlu menghitung kembali k-centroid baru sebagai barycenters dari kelompok yang dihasilkan. Setelah memiliki k-centroid yang baru, pengelompokan di uji kembali terhadap k-centroid. Penghitungan k-centroid yang baru dan pengelompokannya dilakukan berulang hingga k-centroid tidak bergerak lagi[5].

3. HASIL DAN PEMBAHASAN

3.1. Analisis Data

Analisis data merupakan cara yang digunakan untuk menjawab masalah penelitian melalui prosedur pengelolannya terutama masalah yang berkaitan dengan penelitian. Data yang diperoleh untuk penelitian ini adalah presentasi buta huruf 15-44 dari tahun 2011-2018.

Tabel 1. Data BPS/Penelitian (Persentase Penduduk Buta Huruf Umur 15-44)

Provinsi	2011	2012	2013	2014	2015	2016	2017	2018
Aceh	1.29	1.03	0.67	0.43	0.27	0.22	0.13	0.13
Sumatera Utara	1.64	1.1	0.9	0.66	0.51	0.37	0.38	0.26
Sumatera Barat	1.26	1.05	0.64	0.43	0.32	0.17	0.19	0.17
Riau	1.05	0.87	0.67	0.48	0.33	0.2	0.15	0.1
Jambi	1.43	1.33	0.89	0.57	0.49	0.35	0.29	0.29
Sumatera Selatan	1.17	0.91	0.72	0.52	0.48	0.35	0.25	0.22
Bengkulu	1.31	1.09	0.68	0.54	0.48	0.36	0.27	0.18
Lampung	1.12	1.01	0.7	0.42	0.34	0.21	0.16	0.14
Kep. Bangka Belitung	2.13	1.59	1.24	0.91	0.87	0.81	0.7	0.84
Kep. Riau	1.17	0.77	0.62	0.38	0.29	0.28	0.29	0.15
Dki Jakarta	0.42	0.2	0.14	0.08	0.06	0.08	0.06	0.07
Jawa Barat	0.99	0.93	0.63	0.41	0.29	0.23	0.26	0.26
Jawa Tengah	1.56	1.47	1.08	0.65	0.5	0.36	0.36	0.34
Di Yogyakarta	0.63	0.34	0.2	0.09	0.19	0.13	0.2	0.09
Jawa Timur	2.59	2.24	1.84	1.43	1.24	1.09	1.01	1.04
Banten	1.14	0.94	0.65	0.48	0.33	0.19	0.21	0.13
Bali	3.06	2.1	1.44	1.06	0.61	0.51	0.4	0.22
Nusa Tenggara Barat	5.83	5.73	4.86	3.54	3.31	3.26	3.2	2.72
Nusa Tenggara Timur	5.67	5.4	4.18	3.48	3.1	3.06	3.08	3.03
Kalimantan Barat	4.23	3.11	2.64	2.06	2	1.88	1.76	1.63
Kalimantan Tengah	1.31	1.05	0.66	0.32	0.3	0.19	0.23	0.12
Kalimantan Selatan	1.52	0.93	0.52	0.28	0.19	0.11	0.15	0.17

Provinsi	2011	2012	2013	2014	2015	2016	2017	2018
Kalimantan Timur	1.17	0.67	0.56	0.19	0.13	0.12	0.15	0.2
Kalimantan Utara	0	0	0	0	1.36	1.31	1.38	1.38
Sulawesi Utara	0.7	0.66	0.42	0.18	0.17	0.15	0.19	0.11
Sulawesi Tengah	3.04	2.48	1.74	1.38	0.91	0.82	0.84	0.81
Sulawesi Selatan	4.83	4.17	3.2	2.58	2.22	2.07	2.03	2.02
Sulawesi Tenggara	3.15	2.84	2.05	1.62	1.37	1.19	1.14	1.08
Gorontalo	3.27	2.88	1.44	1.1	0.61	0.49	0.5	0.41
Sulawesi Barat	6.45	5.86	4.73	3.93	3.33	3.06	3.16	3.29
Maluku	1.94	1.62	1.3	0.81	0.8	0.76	0.77	0.76
Maluku Utara	1.89	1.52	1	0.57	0.47	0.36	0.41	0.29
Papua Barat	5.37	4.53	2.93	2.27	2.09	1.97	1.94	1.83
Papua	34.55	33.4	31.44	28.5	28.47	28.21	24.66	21.2

- a. Menentukan nilai *Centroid* yang ditentukan secara manual atau acak yang diambil dari data.

Tabel 2. Nilai *Centroid Awal*

Centroid	2011	2012	2013	2014	2015	2016	2017	2018
Centroid 1	34.55	33.4	31.44	28.5	28.47	28.21	24.66	21.2
Centroid 2	4.83	4.17	3.2	2.58	2.22	2.07	2.03	2.02
Centroid 3	0	0	0	0	1.36	1.31	1.38	1.38

- b. Menghitung jarak dari *centroid*
 Untuk menghitung jarak antara titik centroid dengan titik tiap objek menggunakan Euclidian Distance.

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

Maka perhitungan untuk jarak dari *Centroid* ke-1 adalah sebagai berikut :

$$D_{x1,c1} = \sqrt{(1,29 - 34,55)^2 + (1,03 - 33,4)^2 + (0,67 - 31,44)^2 + (0,43 - 28,5)^2 + (0,27 - 28,47)^2 + (2,22 - 28,21)^2 + (0,13 - 24,66)^2 + (10,13 - 21,2)^2} = 80,70$$

$$D_{x2,c1} = \sqrt{(1,26 - 34,55)^2 + (1,05 - 33,4)^2 + (0,64 - 31,44)^2 + (0,43 - 28,5)^2 + (0,32 - 28,47)^2 + (0,17 - 28,21)^2 + (0,19 - 24,66)^2 + (0,17 - 21,2)^2} = 80,12$$

Dan seterusnya sampai dengan $D_{x34,c1}$. Selanjutnya perhitungan untuk jarak dari *Centroid* ke-2 adalah sebagai berikut :

$$D_{x1,c2} = \sqrt{(1,29-4,83)^2 + (1,03 - 4,17)^2 + (0,67 - 3,2)^2 + (0,43 - 2,58)^2 + (0,27 - 2,22)^2 + (2,22 - 2,07)^2 + (0,13 - 2,03)^2 + (10,13 - 2,02)^2} = 6,92$$

$$D_{x2,c2} = \sqrt{(1,26-4,83)^2 + (1,05 - 4,17)^2 + (0,64 - 3,2)^2 + (0,43 - 2,58)^2 + (0,32 - 2,22)^2 + (0,17 - 2,27)^2 + (0,19 - 2,03)^2 + (0,17 - 2,02)^2} = 6,34$$

Dan seterusnya sampai dengan $D_{x34,c2}$. Selanjutnya perhitungan untuk jarak dari *Centroid* ke-3 adalah sebagai berikut :

$$D_{x1,c3} = \sqrt{(1,29-0)^2 + (1,03 - 0)^2 + (0,67 - 0)^2 + (0,43 - 0)^2 + (0,27 - 1,36)^2 + (2,22 - 1,31)^2 + (0,13 - 1,38)^2 + (10,13 - 1,38)^2} = 2,98$$

$$D_{x2,c3} = \sqrt{(1,26-0)^2 + (1,05 - 0)^2 + (0,64 - 0)^2 + (0,43 - 0)^2 + (0,32 - 1,36)^2 + (0,17 - 1,31)^2 + (0,19 - 1,38)^2 + (0,17 - 1,38)^2} = 3,00$$

Dan seterusnya sampai dengan $D_{x34,c3}$. Sehingga didapat tabel jarak dari *Centroid* dan mencari nilai minimal dari ketiga *Centroid*. Tabel Jarak dari *Centroid* adalah sebagai berikut :

Tabel 4. Jarak *Centroid* Iterasi ke-1

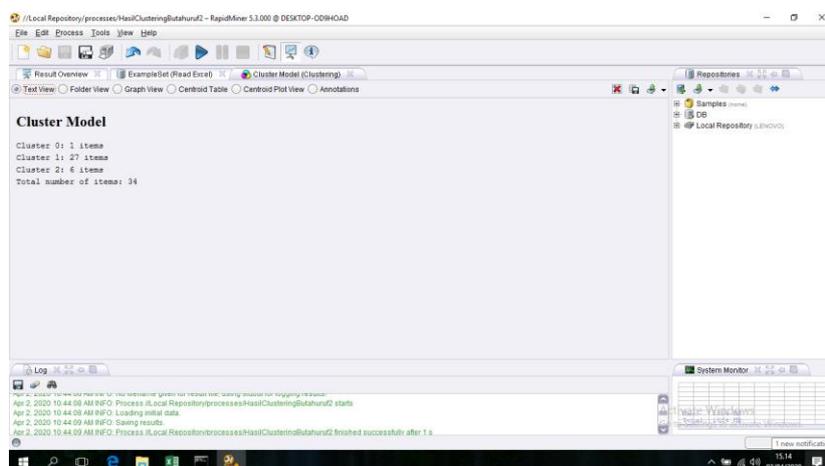
C1	C2	C3	Jarak Centroid
80.70	6.92	2.98	2.98
80.12	6.34	3.00	3.00
80.69	6.91	2.92	2.92
80.84	7.09	2.83	2.83
80.18	6.40	3.00	3.00
80.57	6.84	2.70	2.70
80.45	6.70	2.83	2.83
80.74	6.98	2.87	2.87
78.98	5.21	3.27	3.27
80.81	7.06	2.72	2.72
81.88	8.19	2.63	2.63
80.81	7.06	2.69	2.69
79.93	6.14	3.15	3.15
81.60	7.90	2.53	2.53
77.74	3.98	4.18	3.98
80.76	7.00	2.84	2.84
78.72	4.87	4.53	4.53
70.55	3.40	10.76	3.40

C1	C2	C3	Jarak Centroid
80.18	5.71	0.42	0.42
80.57	6.16	0.21	0.21
80.45	6.01	0.18	0.18
80.74	6.29	0.33	0.33
78.98	4.53	1.54	1.54
80.81	6.38	0.36	0.36
81.88	7.51	1.49	1.49
80.81	6.38	0.38	0.38
79.93	5.45	0.68	0.68
81.60	7.22	1.20	1.20
77.74	3.30	2.77	2.77
80.76	6.31	0.32	0.32
78.72	4.20	2.29	2.29
70.55	4.05	10.03	4.05
71.13	3.47	9.45	3.47
75.28	0.96	5.29	0.96
80.70	6.22	0.31	0.31
80.81	6.31	0.52	0.52
81.08	6.62	0.63	0.63
80.64	7.11	2.74	2.74
81.33	6.92	0.89	0.89
77.84	3.29	2.85	2.85
73.89	0.69	6.73	0.69
77.00	2.48	3.58	2.48
78.24	3.74	2.90	2.90
70.09	4.52	10.54	4.52
79.10	4.66	1.41	1.41
79.84	5.33	0.85	0.85
73.90	1.12	6.90	1.12
0.00	74.56	80.49	0.00
80.70	6.23	0.33	0.33
80.12	5.66	0.46	0.46
80.69	6.22	0.29	0.29

Dari tabel Jarak *Centroid* diatas, maka *Cluster* atau pengelompokkan Iterasi ke-2 adalah sebagai berikut :

Tabel 8 *Cluster* Iterasi ke-2

C1	C2	C3
		1
		1
		1

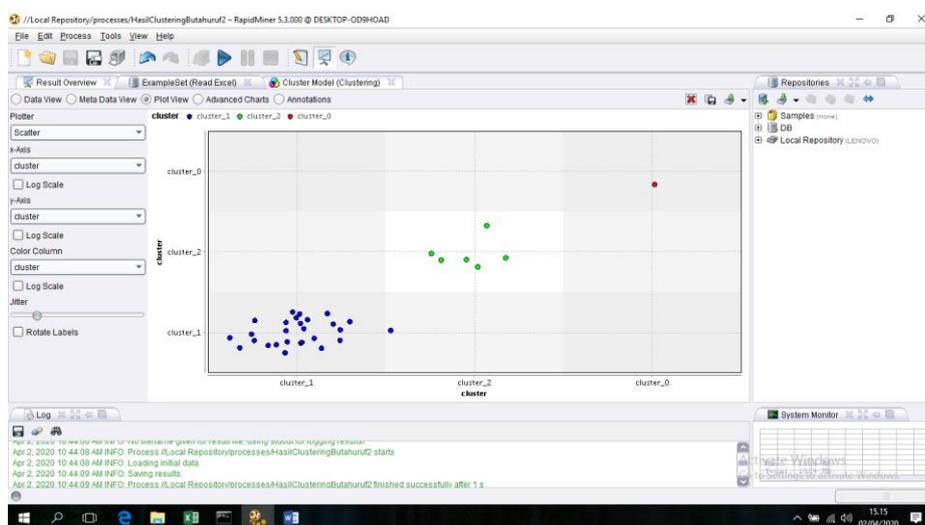


Gambar 1. Nilai *Cluster Model*

Keterangan :

1. Jumlah *Cluster 0* (Tinggi) berjumlah 1 *Items*
2. Jumlah *Cluster 1* (Rendah) berjumlah 27 *Items*
3. Jumlah *Cluster 2* (Sedang) berjumlah 6 *Items*
4. Jumlah keseluruhan *items* adalah 34

Sehingga dapat diketahui hasil pengelompokkan rapidminer dapat dilihat pada gambar sebagai berikut :



Gambar 2. Hasil Pengelompokkan

Berdasarkan pada gambar dapat dilihat bahwa kelompok tinggi memiliki 1 node, kelompok rendah memiliki 27 node, dan kelompok sedang 6 node.

4. SIMPULAN

Adapun kesimpulan yang dapat di ambil pada tehnik data mining dalam clustering persentase buta huruf umur 15-44 menurut provinsi dengan algoritma k-means ini adalah sebagai berikut :

- a. Algoritma k-means clustering ini dapat membantu peneliti mengelompokkan clustering persentase buta huruf umur 15-44 pada setiap provinsi di indonesia. Dari data clustering persentase buta huruf umur 15-44 menurut provinsi di indonesia dapat di ketahui terdapat 1 provinsi dengan *cluster* tingkat tinggi yaitu: Papua, dengan *cluster* tingkat sedang yaitu : NTB, NTT, Kalimantan Barat, Sulawesi Selatan, Sulawesi Barat dan Papua Barat dan dengan cluster rendah adalah 27 propinsi lainnya.
- b. Aplikasi rapidminer ini dapat membantu peneliti mengelompokkan produksi jagung pada setiap provinsi di indonesia. Perhitungan manual algoritma *k-means clustering* dalam mengelompokkan produksi jagung dan penerapan di *Rapidminer* menunjukkan hasil yang sama

DAFTAR PUSTAKA

- [1] Y. Darmi and A. Setiawan, "Penerapan metode clustering k-means dalam pengelompokan penjualan produk," *J. Media Infotama Univ. Muhammadiyah Bengkulu*, vol. 12, no. 2, pp. 148–157, 2016.
- [2] L. Maulida, "Penerapan Datamining Dalam Mengelompokkan Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Prov. Dki Jakarta Dengan K-Means," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 2, no. 3, p. 167, 2018.
- [3] N. Erlangga, S. Solikhun, and I. Irawan, "Penerapan Data Mining Dalam Mengelompokkan Produksi Jagung Menurut Provinsi Menggunakan Algoritma K-Means," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 702–709, 2019.
- [4] M. Robani and A. Widodo, "Algoritma K-Means Clustering Untuk Pengelompokan Ayat Al Quran Pada Terjemahan Bahasa Indonesia," *J. Sist. Inf. Bisnis*, vol. 6, no. 2, p. 164, 2016.
- [5] N. H. Kristanto, A. C. L. A, and H. B. S, "Implementasi K-Means Clustering untuk Pengelompokan Analisis Rasio Profitabilitas dalam Working Capital," *Juisi*, vol. 02, no. 01, pp. 9–15, 2016.