JNKTI
Jurnal Nasional Komputasi dan Teknologi Informasi
Fakultas Teknik Universitas Serambi Mekkah Jl. T. Imum Lueng Bata , Telp. (0651)26160 Batoh – Banda Aceh

SURAT KETERANGAN PUBLIKASI JURNAL

Nama Penulis : Bustami Yusuf[1], Rike Mahara[2], Hendri Ahmadian[3] , Sri Wahyuni[4], Khairan AR[5]
Afiliasi : [1,2,3,4,5]Universitas Islam Negeri Ar-Raniry Banda Aceh
Judul Artikel : Analisis Clustering Penduduk Miskin Di Provinsi Aceh Menggunakan Algoritma K-Means Dan X-Means
Email : bustamiyusoef@ar-raniry.ac.id

Dengan ini menerangkan bahwa artikel dengan judul tersebut di atas telah direview dan dinyatakan diterima serta memenuhi syarat untuk dipublikasikan pada Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI) Vol 5 No 1 Tahun 2022 yang akan terbit pada bulan Februari 2022.

Banda Aceh, 15 Februari 2022
Ketua Dewan Editor,

Zulfan, ST., MT
NIDN. 0115088402

Jurnal Nasional Komputasi dan Teknologi Informasi,
No. xx Vol. xx
P-ISSN. xxxx-xxxx;
E-ISSN. xxxx-xxxx,

ANALISIS *CLUSTERING* PENDUDUK MISKIN DI PROVINSI ACEH MENGGUNAKAN ALGORITMA *K-MEANS* DAN *X-MEANS*

Bustami Yusuf[1], Rike Mahara[2], Hendri Ahmadian[3] , Sri Wahyuni[4], Khairan AR[5]
[1] Program Studi Teknologi Informasi, [2] Program Sudi Pendidikan Teknologi Informasi
Universitas Islam Negeri Ar-Raniry Banda Aceh
Corresponding author's e-mail: bustamiyusoef@ar-raniry.ac.id

Abstrak-Analisis *Clustering* merupakan salah satu metode yang bertujuan untuk mengelompokan data/objek berdasarkan kemiripan dan ketidakmiripan karakteristiknya. Sehingga objek yang berada pada satu *cluster* memiliki kemiripan yang besar dan sangat kecil bila dibandingkan dengan *cluster* lain. Algoritma *K-means* merupakan salah satu metode pengelompokan non hierarki yang paling umum digunakan, namun pada algoritma ini pengguna harus mengetahui jumlah kelompok yang akan dikelompokan. Sedangkan algoritma *X-means* merupakan pengembangan dari algoritma *K-means*, dimana algoritma ini mampu mengelompokan datanya sendiri tanpa menginput jumlah k. Algoritma *X-means* akan berhenti melakukan pengelompokan hingga batas *cluster* maksimal tercapai. Algoritma *X-means* juga tau kapan dan dimana *centroid* baru harus muncul berdasarkan perhitungan nilai *Bayesian Information Criterion* (BIC). Penelitian ini bertujuan untuk membandingkan hasil analisis *clusteringK-means* dengan *X-means* menggunakan data penduduk miskin provinsi Aceh tahun 2018 dan data Iris. Hasil pengujian *clustering* dengan data Iris memiliki tingkat kemurnian dan hasil pengelompokan yang baik dari masing-masing algoritma. Yaitu nilai *purity* pada *K-means* sebanyak 0.89 dan pada *X-means* 0.88 dengan nilai *Davies Bouldin Index* (DBI) masing-masing 0.16. Dan hasil pengujian *clustering* dengan data penduduk miskin menunjukkan bahwa pada algoritma *K-means* didapatkan tingkatan kemiskinan dengan kategori Sangat Miskin berada pada *cluster* nol, kategori Miskin pada *cluster* dua, kateogori Rentan Miskin berada pada *cluster* tiga dan kategori Tidak Miskin berada pada *cluster* satu. Dan pada algoritma *X-means* diperoleh tingkatan kemiskinan dengan kategori Sangat Miskin berada pada *cluster* satu, kategori Miskin pada *cluster* tiga, kategori Rentan Miskin berada pada *cluster* dua dan kategori Tidak Miskin berada pada *cluster* nol. Berdasarkan Waktu yang dibutuhkan selama proses *clustering* algoritma *X-means* memiliki kecepatan yang lebih baik yaitu 0.06*second*. Sedangkan *K-means* membutuhkan waktu selama 0.22 *second*. Sehingga dapat disimpulkan bahwa algoritma *X-means* lebih cepat dalam melakukan pengelompokan dibandingkan algoritma *K-means*.

Keyword : Penduduk Miskin, *K-means*, *X-means*, *Purity*, *Davies Bouldin Index* (DBI)

1. Pendahuluan

Kemiskinan adalah suatu masalah serius yang sudah ada dari generasi ke generasi. Kemiskinan merupakan suatu keadaan yang berada dibawah standar kebutuhan minimum, baik dalam bidang makanan dan bukan makanan yang disebut sebagai garis kemiskinan[1]. Dimana khusus wilayah Sumatera pada tahun 2018 angka kemiskinan tertinggi berada di provinsi Aceh yaitu sebesar 15,68% dan angka terendah berada di Kepulauan Riau yaitu sebesar 4,77% [2].

Bila dilihat secara nasional Aceh merupakan provinsi termiskin ke-6. Pemerintah Aceh giat memberikan bantuan sosial berbasis keluarga seperti beasiswa untuk anak yatim[3], Jaminan Kesehatan Aceh (JKA)[4], serta membangun dan merehab rumah penduduk miskin[5]. Kendala utama yang dihadapi adalah persoalan penargetan, khususnya penargetan individu atau rumah tangga yang kurang mampu[6]. Oleh karenanya perlu dilakukan hal-hal yang lebih optimal serta menyeluruh dalam penanggulangan kemiskinan. Untuk itu diperlukan suatu cara pengelompokan data keluarga miskin yang lebih efisien agar pemerataan bantuan lebih tepat sasaran. Pengelompokan ini dapat diselesaikan dengan menggunakan salah satu metode dalam data mining, yaitu metode *clustering*. Dengan metode *clustering* pengelompokkan data dapat dilakukan dengan lebih akurat sesuai dengan tingkat kemiripan dari karakteristik datanya[7]. Ada beberapa algoritma *clustering* yang dapat digunakan untuk mengelompokan data, diantaranya algoritma *k-means* dan *x-means*[8].

2. Dasar Teori

2.1 Clustering

Clustering atau pengelompokan merupakan suatu teknik dalam data mining yang digunakan untuk menganalisis data yang bertujuan untuk memecahkan permasalahan dalam pengelompokan data atau lebih tepatnya memisahkan dari dataset ke dalam *subset*. Dalam pengelompokan suatu objek, teknik ini hanya berdasarkan pada tingkat kemiripan karakteristik dari suatu atribut pada data objek. Dengan demikian, data objek akan memiliki tingkat kemiripan yang sama di dalam *cluster* yang sama, serta memiliki kemiripan yang berbeda di antara cluster yang berbeda. Terdapat 2 jenis metode pada clustering yaitu *hierarchical clustering* dan *non-hierarchical clustering*[9]. Dan algoritma *K-means* merupakan salah satu metode dari *non-hierarchical clustering*.

2.2 K-means

K-means pertama kali diperkenalkan oleh MacQueen JB pada tahun 1976. Algoritma *K-means*merupakan salah satu metode *non-hirarchi*yang sangat umum digunakan. algoritma ini akan melakukan pengelompokan dengan teknik membagi atau memisahkan objek ke k lain[1].

Algoritma *K-means* sering digunakan karena memiliki kemampuan dalam mengelompokan jumlah data yang besar dengan waktu yang cepat dan efisien. Algoritma *K-means* menentukan pusat *cluster* (*centroid*) awal secara *random*, sehingga hasil pengelompokan yang didapat tidak menjanjikan hasil yang khas. Karena penentuan titik pusat (*centroid*) awal sangat bergantung pada efisiensi keaslian algoritma *K-means*.

Tahapan yang dilakukan pada algoritma K-means yaitu sebagai berikut :

1. Menentukan jumlah k, dimana k merupakan jumlah *cluster*
2. Menetapkan nilai awal titik pusat *cluster*secara random k-record yang menjadi lokasi pusat *cluster* awal.
3. Setiap data akan dicari titik pusat cluster terdekatnya. Artinya setiap titik pusat *cluster* "memiliki" s*ubset* dari data, sehingga dapat merepresentasikan sebuah partisi dari *dataset*. Dan akan didapatkan k *cluster* yaitu C1, C2, …, Ck. Untuk menentukan data masuk ke C1, C2 dan seterusnya maka perlu dihitung jarak terdekat setiap data dengan titik pusat cluster. Untuk menghitung jarak setiap data digunakan teori jarak *Euclidean* yang dirumuskan sebagai berikut :

$$\boldsymbol{D(ij) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + .. + .. + (X_{ki} - X_{kj})^2}} \qquad \textbf{(1)}$$

D(i,j) = Jarak data ke-*i* ke pusat *clusterj*
X_{ki} = Data ke-*i* pada atribut data ke *k*
X_{kj} = Titik pusat ke-*j* pada atribut ke *k*

Setiap k *cluster* yang dihasilkan akan dicari *centroid*nya dan akan memperbaharui lokasi setiap pusat *cluster* untuk menghasilkan *centroid* baru. Untuk menentukan *centroid* baru digunakan persamaan berikut:

$$\boldsymbol{v_{ij} = \frac{1}{N_i} \sum_{k=0}^{n} x_j} \qquad \textbf{(2)}$$

Keterangan :
v_{ij} = *centroid*/rata-rata *cluster* ke-i untuk variabel ke-j
N_i = jumlah data yang menjadi anggota *cluster* ke-i
i,k = indeks dari *cluster*
j = indeks dari variabel
x_{ij} = nilai data ke-*k* yang ada di dalam *cluster* tersebut untuk variabel ke-*j*

4. Ulangi langkah 3 sampai 5, sampai data tidak mengalami perpindahan.

Algoritma akan berhenti saat titik pusat *cluster* tidak lagi berubah. Artinya, algoritma berakhir ketika seluruh *cluster* C1, C2, …, Ck, semua *record* yang dimiliki oleh masing-masing pusat *cluster* tetap berada dalam *cluster* tersebut[10].

2.3 X-means

X-means merupakan algoritma penyempurnaan dari *K-means*. kesulitan dalam penentuan jumlah *cluster* yang digunakan serta waktu proses yang masih terhitung lama pada jumlah data yang besar dalam *clustering* sehingga mendorong penyempurnaan dari algoritma *k-means* ke algoritma *X-means*.

Proses kerja dari teknik ini adalah dengan mencari ruang diantara tempat *cluster* dan jumlah *cluster* untuk melakukan optimasi *Bayesian Information Criterion* (BIC) serta memberikan keputusan apakah *centroid* harus dibagi atau tidak. Proses ini hanya dapat digunakan untuk data Numeric[11]. Algoritma *X-means* terdiri dari dua operasi yang berulang hingga eksekusi berakhir, yaitu:

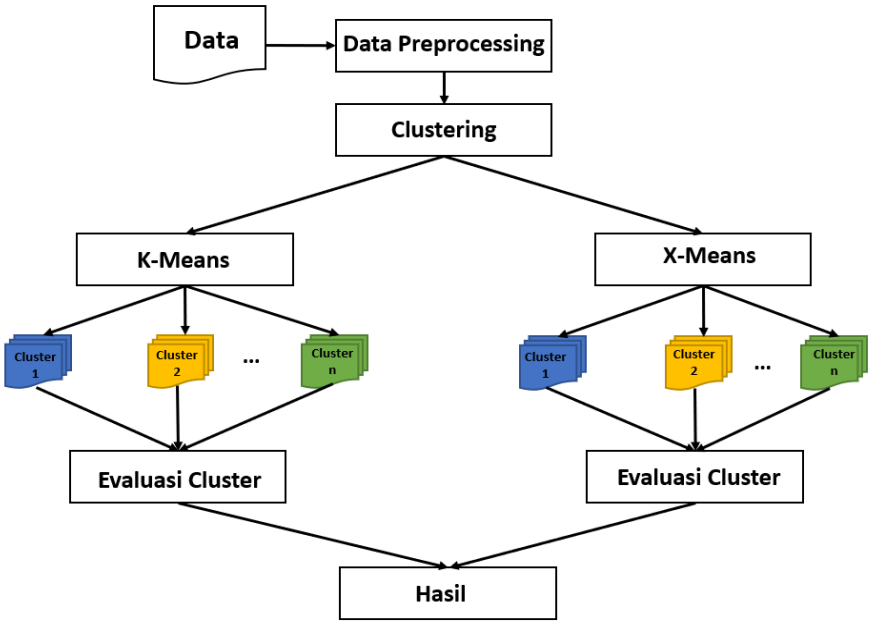
1. *Improve-Params*
Merupakan eksekusi algoritma *K-means* konvensional hingga mencapai konvergen atau tidak terjadi perpindahan point antar *cluster*.

2. *Improve-Structure*
Menemukan jika dan di mana titik pusat *cluster (centroid*) baru harus muncul. Hal ini dicapai dengan membiarkan beberapa *centroid* dibagi menjadi dua berdasarkan BIC[12].

3. Metode Penelitian

3.1 Tahapan Penelitian

Pada sub bahasan tahapan penelitian akan dibahas mengenai tahapan-tahapan yang akan dilakukan dalam penelitian. Untuk tahapannya dapat dilihat pada gambar 1.

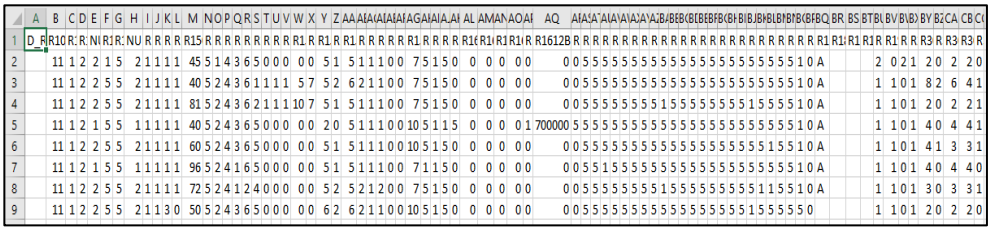


Gambar 1. Flowchart Metodologi Penelitian

1. Data

Pengumpulan data yang akan digunakan dalam penelitian ini yaitu data penduduk miskin di provinsi Aceh dan data Iris. Dimana data BPS digunakan untuk mengelompokan data dan data Iris digunakan untuk menguji keakuratan suatu *cluster* yang dihasilkan dari masing-masing algoritma.

Tabel 1. Data BPS



a. Data Badan Pusat Statistik (BPS)

Data penduduk miskin yang didapat dari Badan Pusat Statistik (BPS) Provinsi Aceh teridiri dari 88 variabel. Adapun jumlah data yang diterima adalah sebanyak 11.462 data. Tabel 1 merupakan contoh tampilan file dalam format *excel* yang diterima dari BPS dengan jumlah 88 variabel.

b. Data Iris

Data Iris merupakan data multivariat yang diperkenalkan oleh Sir Ronald Aylmer Fisher (1936) sebagai contoh analisis diskriminan. Data iris terdiri dari 4 variabel yang akan diukur yaitu*Sepal Length, Sepal Width, Petal Length dan Petal Width*. Dan memiliki 3 jenis bunga iris yaitu *Iris-Setosa, Iris Versicolor* dan *Iris Virginica* dengan jumlah data sebanyak 150 data[13]. Sampel data Iris ditampilkan dalam tabel 2.

Tabel 2. Data Iris

No	Sepal Length	Sepal Width	Petal Length	Petal Width
1	5.1	3.5	1.4	0.2
2	4.9	3	1.4	0.2
3	4.7	3.2	1.3	0.2
4	4.6	3.1	1.5	0.2
5	5	3.6	1.4	0.2
6	5.4	3.9	1.7	0.4
7	4.6	3.4	1.4	0.3
8	5	3.4	1.5	0.2
9	4.4	2.9	1.4	0.2
10	4.9	3.1	1.5	0.1

2. *Data Preprocessing*

Pada tahap ini dilakukan persiapan data untuk diproses di aplikasi Weka dengan memilih variabel yang berpengaruh serta membuang variabel yang tidak berpengaruh. Ada 2 tahap pemrosesan data yang dilakukan yaitu *data selection* dan *data transformation*.

a. *Data Selection*

Data selection atau seleksi data merupakan pemilihan data atau variabel yang akan digunakan dalam penelitian[14]. Adapun pada penelitian ini hanya melakukan proses seleksi pada variabel. Dimana jumlah variabel yang sebelumnya berjumlah 88 variabel, setelah dilakukan seleksimaka diperoleh 13 variabel. Variabel sebelum diseleksi dapat dilihat pada gambar 4 dan sesudah dilakukan seleksi dapat dilihat pada tabel dibawah :

Tabel 3. Variabel Setelah Diseleksi

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Kabupate	Kota/Desa	Status T.Tinggal	L.Lantai	Atap	Dinding	Lantai	MCK	P.Tinja	SAM	S.Penerangan	Jenis B.Bakar	K.Barang
2	1	2	1	45	4	3	6	5	0	5	1	7	5
3	1	2	1	40	4	3	6	1	1	5	1	7	5
4	1	2	1	81	4	3	6	2	1	5	1	7	5
5	1	2	1	40	4	3	6	5	0	2	1	10	5
6	1	2	1	60	4	3	6	5	0	5	1	10	5
7	1	2	1	96	4	1	6	5	0	5	1	7	5
8	1	2	1	72	4	1	2	4	0	5	1	7	5
9	1	2	3	50	4	3	6	5	0	6	1	10	5
10	1	2	1	60	4	1	6	1	1	2	1	7	5

b. Data Transformation

Setelah data di seleksi dan telah didapatkan atribut yang mengacu kepada indikator kemiskinan. Selanjutnya data diubah ke dalam format yang bisa diterima di perangkat *data mining* yang akan memprosesnya[14]. Perangkat lunak *data mining* menerima format *.csv*. berhubung data yang diperoleh dalam format *.xls* maka bisa langsung disimpan dalam format *.csv*. Setelah data diubah ke dalam format *.csv* maka selanjutnya diubah ke dalam format *.arff* pada weka. Adapun tampilan data dalam format *.csv* dan *.arff* dapat dilihat pada gambar 1 dan gambar 2 :

```
Kabupaten,Kota/Desa,Status
T.Tinggal,L.Lantai,Atap,Dinding,Lantai,MCK,P.Tinja,SAM,S.Penerangan,Jenis
B.Bakar,K.Barang
1,2,1,45,4,3,6,5,0,5,1,7,5
1,2,1,40,4,3,6,1,1,5,1,7,5
1,2,1,81,4,3,6,2,1,5,1,7,5
1,2,1,40,4,3,6,5,0,2,1,10,5
1,2,1,60,4,3,6,5,0,5,1,10,5
1,2,1,96,4,1,6,5,0,5,1,7,5
1,2,1,72,4,1,2,4,0,5,1,7,5
1,2,3,50,4,3,6,5,0,6,1,10,5
1,2,1,60,4,1,6,1,1,2,1,7,5
```

Gambar 2. Data BPS dalam format .csv

```
@relation 'file BPS'@attribute Kabupaten numeric@attribute Kota/Desa numeric
@attribute 'Status T.Tinggal' numeric@attribute L.Lantai numeric@attribute
Atap numeric@attribute Dinding numeric@attribute Lantai numeric@attribute MCK
numeric@attribute P.Tinja numeric@attribute SAM numeric@attribute S.Penerangan
numeric@attribute 'Jenis B.Bakar' numeric@attribute K.Barang numeric@data
1,2,1,45,4,3,6,5,0,5,1,7,5
1,2,1,40,4,3,6,1,1,5,1,7,5
1,2,1,81,4,3,6,2,1,5,1,7,5
1,2,1,40,4,3,6,5,0,2,1,10,5
1,2,1,60,4,3,6,5,0,5,1,10,5
1,2,1,96,4,1,6,5,0,5,1,7,5
1,2,1,72,4,1,2,4,0,5,1,7,5
1,2,3,50,4,3,6,5,0,6,1,10,5
1,2,1,60,4,1,6,1,1,2,1,7,5
1,2,1,46,7,3,6,5,0,5,1,10,5
1,2,1,32,4,3,6,1,1,2,1,1,5
1,2,1,90,4,3,2,1,3,8,1,1,5
```

Gambar 3. Data BPS dalam format .arrf

4. Hasil dan Pembahasan

4.1 Pengujian Algoritma *K-means* dan *X-means*

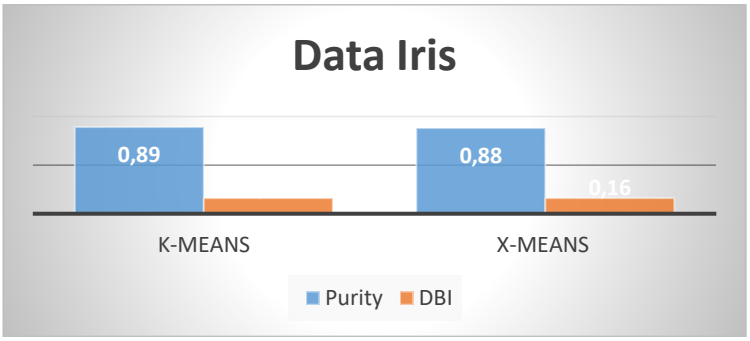
1. Data Iris

Berdasarkan hasil pengujian dengan data Iris maka didapatkan hasilnya seperti yang ditampilkan dalam tabel 4.

Tabel 4. Nilai *Purity* dan DBI Data Iris

Algoritma/ Metode	K-Means				X-Means			
	C1	C2	C3	Total	C1	C2	C3	Total
Purity	0.94	1	0.72	**0.89**	0.84	0.80	1	**0.88**
DBI				**0.16**				**0.16**

Dari tabel diatas dapat dilihat nilai *purity* pada masing-masing algoritma tidak berbeda jauh. Nilai *purity* tertinggi pada cluster menggunakan algoritma K-means adalah pada C2 dan menggunakan *X-means* nilai *purity* tertinggi yaitu pada C3. Dari tabel 4 diperoleh grafik sebagai berikut :



Gambar 5. Nilai *purity* dan DBI data Iris

Berdasarkan grafik diatas dapat dilihat bahwanilai *purity* pada *K-means* yaitu 0.89 dan pada *X-means*0.88 dan untuk nilai DBI tidak terdapat perbedaan yaitu sama-sama mendapatkan nilai 0.16.

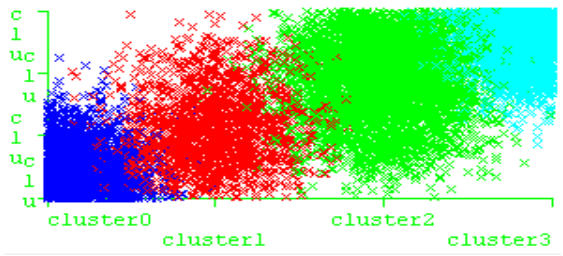
2. Data Badan Pusat Statistik (BPS)

Pada tahap ini akan dilakukan pengelompokan data BPS dengan menggunakan aplikasi Weka. Data akan dikelompokan menjadi 4 *cluster*. Adapun hasil yang diperoleh dapat dilihat pada tabel dibawah :

Tabel 5. Hasil Cluster menggunakan K-means dan X-means

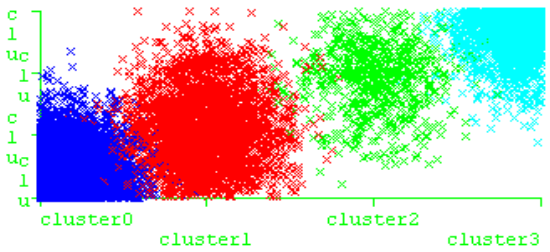
K-means		X-means	
Cluster	Jumlah Data	Cluster	Jumlah Data
0	2507	0	5672
1	1686	1	3262
2	4732	2	703
3	2537	3	1825

Dari hasil *cluster* diatas dapat divisualisasikan pada gambar plot dibawah :



Gambar 6. Visualisasi *cluster* dengan K-means

Dari gambar 6 visualisasi plot diatas dan berdasarkan tabel 5 dapat dilihat bahwa hasil pengelompokan menggunakan algoritma *K-means*diketahui bahwa jumlah data pada cluster 0 (warna biru) adalah sebanyak 2507, pada *cluster* 1 (warna merah) berjumlah 1686, pada cluster 2 (warna hijau) berjumlah 4732 data dan pada cluster 3 (warna biru muda) berjumlah 2537 data.



Gambar 7. Visualisasi Cluster dengan X-means

Dari gambar 7 visualisasi plot dapat dilihat bahwa hasil pengelompokan menggunakan algoritma *X-means* diketahui bahwa jumlah data pada *cluster* 0 (warna biru) adalah sebanyak 5672, pada *cluster* 1 (warna merah) berjumlah 3262, pada *cluster* 2 (warna hijau) berjumlah 703 data dan pada *cluster* 3 (warna biru muda) berjumlah 1825 data.

Waktu yang dibutuhkan algoritma *K-means* dalam melakukan pengelompokan yaitu selama 0.22 *second* dengan jumlah iterasi sebanyak 6 iterasi, sedangkan pada algoritma *X-means* yaitu selama 0.06 *second* dengan jumlah iterasi sebanyak 1 iterasi.

4.2 Analisis Hasil *Clustering*

1. Kategori

Kategori merupakan istilah, tanda, simbol yang dimaksudkan untuk mengenali suatu data atau kelompok 1 dengan yang lainnya[15]. Pada penelitian ini dilakukan proses kategori untuk setiap kelompok yang telah selesai di *clustering* dengan menentukan suatu kelompok masuk kategori Sangat Miskin (SM), Miskin (M), Rentan Miskin (RM) atau Tidak Miskin (TM). Untuk menentukan kategori setiap kelompok maka perlu beberapa variabel yang akan digunakan sebagai variabel yang dapat mengukur ke empat kategori diatas. Ada 8 variabel yang digunakan antara lain Status Tempat Tinggal, Atap Terluas, Dinding Terluas, Lantai Terluas, Sumber Penerangan Bahan Bakar, Fasilitas Tempat Buang Air Besar (BAB) dan Sumber Air Minum[16]. Masing-masing variabel memiliki kriteria tertentu untuk menentukan suatu cluster masuk kategori SM, M, RM dan TM. Dari setiap variabel memiliki kriteria masing-masing yaitu sebagai berikut :

Tabel 4.1. Kriteria Variabel Untuk Pengkategorian Cluster

Variabel / Kriteria	SM	M	RM	TM	Penjelasan
Status Tempat Tinggal	Dinas	Bebas Sewa	Kontrak	Milik Sendiri	-
Atap	Jerami	Bambu	Asbes	Beton	SM (Jerami dan lainnya), M (Bambu dan kayu), RM (Asbes dan Seng), TM (Beton dan Genteng),
Dinding	Bambu 3	Bambu 2	Bambu 1	Tembok	SM (Bambu dan lainnya), M (Bambu dan Batang Kayu, RM (Plasteran anyaman bambu dan Kayu), TM (Tembok)
Lantai	Bambu	Kayu	Ubin	Keramik	SM (Bambu, Tanah dan lainnya), M (Kayu dan Semen/Bata merah), RM (Parket dan Ubin), TM (Marmer dan Keramik),
Penerangan	Bukan Listrik	Non PLN	PLN	PLN 1	PLN 1 (PLN dengan Meteran)
Bahan Bakar	Minyak Tanah	Kayu Bakar	Gas	Listrik /Elpiji	SM (Kayu Bakar, lainnya), M (Minyak tanah, Briket, Arang), RM (Gas Kota, Biogas), TM (Listrik, Elpiji (5.5 kg,12 kg, 3 kg),
Fasilitas BAB	Tidak Ada	Umum 2	Umum 1	Milik Sendiri	SM (Tidak Menggunakan dan Tidak ada fasilitas), M (Ada, Umum) RM (Ada, digunakan dengan ART tertentu), TM (Ada, Milik Sendiri),
Sumber Air Minum	Sungai	Sumur Tak Terlindung	Leding	Isi Ulang	SM (Mata Air Tidak Terlindung, Sungai, Air Hujan dll), M (Sumur Tidak Terlindung, Mata Air Terlindung, Mata Air Tak Terlindung), RM (Leding, Sumur Bor/Pompa, Sumur Terlindung), TM (Air Kemasan Bermerk, Air Isi Ulang)

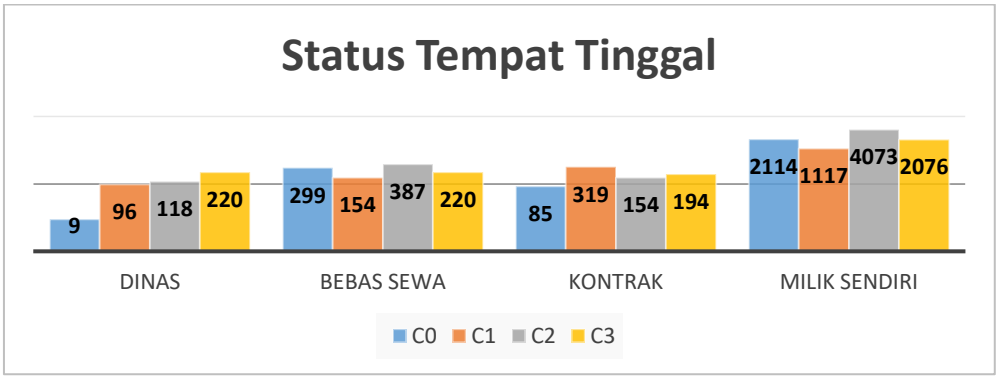
Untuk mencari kategori setiap kelompok, pada penelitian ini dilakukan secara manual yaitu menggunakan aplikasi Excel. Perhitungannya dilakukan dengan melihat setiap *cluster* yang memiliki jumlah data paling banyak/dominan dari setiap kriteria. Bila *cluster* 1 dan 2 sudah diketahui data mana yang paling dominan maka selanjutnya hanya membandingkan *cluster* 3 dan 4. Hasil dari perhitungan di Excel dapat dilihat sebagai berikut :

a. Status Tempat Tinggal

Tabel 4.2. Variabel Status Tempat Tinggal menggunakan K-means

Cluster	Status Tempat Tinggal			
	Dinas	Bebas Sewa	Kontrak	Milik Sendiri
C0	9	299	85	2114
C1	96	154	319	1117
C2	118	387	154	4073
C3	220	220	194	2076

Dari tabel diatas diperoleh grafik sebagai berikut :



Grafik 4.1. Variabel Status Tempat Tinggal dengan K-means

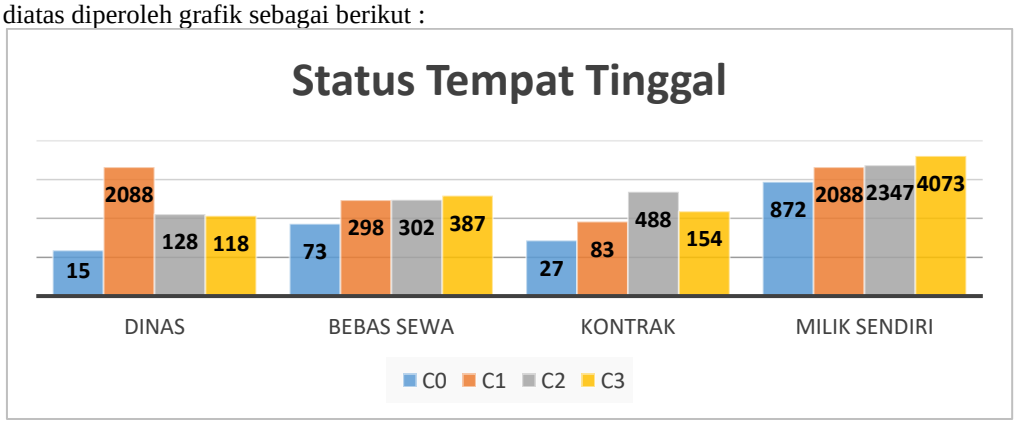
Dari grafik diatas dapat dilihat bahwa data yang paling banyak padarumah dengan status Dinas adalah cluster 3,status bebas sewa pada cluster 2, status Kontrak pada cluster 1 dan status Milik Sendiri pada cluster 0.

Dapat disimpulkan bahwa berdasarkan tabel status tempat tinggal menggunakan algoritma K-means didapatkan kategori Sangat Miskin berada di C3, Miskin berada di C2, Rentan Miskin berada di C1 dan Tidak Miskin C0.

Tabel 4.3. Variabel Status Tempat Tinggal menggunakan X-means

Cluster	Status Tempat Tinggal			
	Dinas	Bebas Sewa	Kontrak	Milik Sendiri
C0	15	73	27	872
C1	2088	298	83	2088
C2	128	302	488	2347
C3	118	387	154	4073

Dari tabel diatas diperoleh grafik sebagai berikut :



Grafik 4.2. Variabel Status Tempat Tinggal menggunakan X-means

Dari grafik diatas dapat dilihat bahwa data yang paling banyak pada rumah dengan status Dinas adalah *cluster* 1, status bebas sewa pada *cluster* 3, status Kontrak pada *cluster* 2 dan status Milik Sendiri pada *cluster* 0.

Dapat disimpulkan bahwa berdasarkan tabel status tempat tinggal menggunakan algoritma *X-means* didapatkan kategori Sangat Miskin berada di C1, Miskin berada di C3, Rentan Miskin berada di C2 dan Tidak Miskin C0.

Hasil kategori 8 variabel menggunakan algoritma *K-means* dan *X-means* dapat dilihat pada tabel dibawah :

Tabel 4.4. Perbandingan Hasil Kategori Setiap Variabel

Algoritma / Variabel	K-means				X-means			
	C0	C1	C2	C3	C0	C1	C2	C3
S.T. Tinggal	TM	RM	M	SM	TM	SM	RM	M
Atap	SM	TM	M	RM	RM	SM	TM	M
Dinding	M	TM	SM	RM	RM	M	TM	SM
Lantai	SM	RM	M	TM	TM	SM	RM	M
S.Penerangan	SM	TM	M	RM	TM	SM	RM	M
B. Bakar	M	RM	SM	TM	TM	M	RM	SM
F. BAB	SM	TM	RM	M	TM	SM	M	RM
S.A. Minum	SM	TM	M	RM	TM	SM	RM	M
Kategori Dominan	SM	TM	M	RM	TM	SM	RM	M

Berdasarkan tabel diatas dapat dilihat bahwa pada algoritma *K-means* jumlah kategori paling banyak pada C0 yaitu kategori SM (warna biru)dengan jumlah sebanyak 5, pada C1 kategori paling banyak yaitu TM (warna merah) dengan jumlah sebanyak 5, pada C2 kategori paling banyak yaitu M (warna hijau) dengan jumlah sebanyak 5 dan pada C3 jumlah kategori paling banyak yaitu RM (warna biru muda) dengan jumlah 4.Untuk hasil kategori kedua algoritma dapat dilihat pada tabel 4.22.

Tabel 4.5. Hasil Kategori

Algoritma / Cluster	K-Means	X-Means
Cluster 0	Sangat Miskin	Tidak Miskin
Cluster 1	Tidak Miskin	Sangat Miskin
Cluster 2	Miskin	Rentan Miskin
Cluster 3	Rentan Miskin	Miskin

Dari tabel diatas dapat disimpulkan bahwa tingkatan kemiskinan pada algoritma *K-means* C0 merupakan kategori Sangat Miskin, C1 Tidak Miskin, C2 Miskin dan C3 Rentan Miskin. Sedangkan tingkatan kemiskinan pada algoritma *X-means* C0 merupakan kategori Tidak Miskin, C1 Sangat Miskin, C2 Rentan Miskin dan C3 Miskin.

5. Kesimpulan dan Saran

A. Kesimpulan

1. Berdasarkan pengujian menggunakan data Iris dengan algoritma *K-means* dan *X-means*, digunakan 2 metode untuk menilai keakuratan suatu *cluster* yaitu dengan *purity* dan DBI. Nilai *purity* yang diperoleh *K-means* yaitu 0.89 dan *X-means* 0.88 serta nilai DBI yang didapat dari kedua algoritma memiliki nilai yang sama yaitu 0.16. Waktu yang diperlukan untuk pengelompokan sangat cepat yaitu 0.02 pada K-means dan 0.00000001 pada *X-means*. Dari pengujian tersebut algoritma *K-means* dan *X-means* cocok digunakan untuk melakukan pengelompokan karena memiliki tingkat akurasi yang baik dengan nilai *purity* yang mendekati 1 dan nilai DBI yang mendekati 0 serta tidak membutuhkan waktu yang lama dalam proses pengelompokan.
2. Pada pengelompokan data BPS waktu yang diperoleh pada algoritma *X-means* lebih cepat daripada *K-means* yaitu 0.22 pada *K-means* dan 0.06 pada *X-means*. *X-means* memiliki kelebihan pada jumlah iterasi yang tidak banyak sehingga membuat proses pengelompokan menjadi lebih cepat.
3. Data iris kurang tepat jika dikelompokan menggunakan *X-means* karena jumlah data yang sedikit mempengaruhi jumlah kelompok yang dihasilkan.
4. Berdasarkan pengelompokan data BPS menggunakan algoritma *K-means* dan *X-means* didapatkan tingkatan kemiskinan menggunakan algoritma *K-means*yaitu Sangat Miskin berada pada *cluster* 0, Miskin berada pada *cluster* 2, Rentan Miskin berada pada *cluster* 3 dan Tidak Miskin berada pada *cluster* 1. Sedangkan menggunakan algoritma *X-means* didapatkan tingkatan kemiskinan yaitu Sangat Miskin berada pada *cluster* 1, Miskin berada pada *cluster* 3, Rentan Miskin *cluster* 2 dan Tidak Miskin berada pada *cluster* 0.

B. Saran

Saran yang dapat diberikan untuk pengembangan penelitian selanjutnya adalah :

1. Dalam penelitian ini hanya menggunakan aplikasi Weka dalam melakukan pengelompokan. Diharapkan pada penelitian selanjutnya dapat menggunakan aplikasi tambahan untuk mengukur serta memperkuat perbandingan tingkat akurasi dari masing-masing algoritma. Disarankan membandingkan algortima *X-means* dengan aplikasi berbeda Seperti *Rapid Miner*, MAT LAB, RStudio dll.
2. Untuk mengelompokan data yang sedikit serta sudah diketahui jumlah kelompoknya disarankan untuk menggunakan *K-means*.
3. Untuk penelitian selanjutnya dapat menggunakan data yang lebih banyak lagi, untuk mengetahui jumlah iterasi dan tingkat kecepatan dalam melakukan pengelompokan.

DAFTAR PUSTAKA

[1] D. R. Utami, "Aplikasi Monitoring Keluarga Miskin Menggunakan Metode K-Means Clustering Berbasis Mobile Gis (Studi Kasus : Pkh Kec Kedungkandang Kota Malang) Skripsi Oleh : Dwi Rahayu Utami." pp. 1–110, 2018.

[2] News.detik.com, "Peringkat Kemiskinan di Sumatera: Aceh Tertinggi, Kepri Terendah." 2019.

[3] https://baitulmal.acehprov.go.id, "Pendidikan." 2019.

[4] L. K. Damayanti, "Aplikasi Algoritma Cart Untuk Mengaklasifikasikan Data Nasabah Asuransi Jiwa Bersama Bumiputera 1912 Surakarta," no. July. p. 92, 2011.

[5] P. Subarkah, "Perbandingan Kinerja Algoritma Cart dan Naive Bayesian untuk Mendiagnosa Penyakit Diabetes Melitus," p. 17, 2016.

[6] W. Isdijoso and A. Suryahadi, "Penetapan Kriteria dan Variabel Pendataan Penduduk Miskin yang Komprehensif dalam Rangka Perlindungan Penduduk Miskin di Kabupaten / Kota." pp. 1–25, 2016.

[7] N. F. Kahar, L. Hadjaratie, S. Suhada, and I. R. Padiku, "I Mplementasi D Ata M Ining D Alam P Enentuan T Ingkat K Emiskinan M Enggunakan F Uzzy C-M Eans," pp. 1–10.

[8] R. Awaliah, "Analisis Clustering Untuk Mengelompokkan Tingkat Kesejahteraan Kabupaten/Kota Berdasarkan Sosial Ekonomi Rumah Tangga Di Wilayah Provinsi Sulawesi Selatan." pp. 1–86, 2018.

[9] NUR ARISKA, "Analisis Cluster Dengan Metode Ensemble Rock Untuk Data Berskala Campuran Kategorik Dan Numerik." pp. 1–113, 2017.

[10] R. Fatah, "Perancangan Model Prediksi Kelulusan Mahasiswa Tepat Waktu pada UIN Raden Fatah," vol. 4, pp. 49–62, 2018.

[11] T. Mardiana, R. D. Nyoto, P. Studi, and T. Informatika, "Kluster Bag-of-Word Menggunakan Weka," vol. 1, no. 1, pp. 1–5, 2015.

[12] A. Wijayanto and R. Adhitama, "Penggunaan X-Means Clustering Method untuk Mengelompokkan Potensi Sekolah Menengah Unggul di Kabupaten Banyumas," vol. 2, no. 1, p. 9, 2019.

[13] B. Yusuf, "Analisis dan perbandingan kualitas pengelompokan dokumen (document clustering) dengan menggunakan metode k-means dan k-medians bustami yusuf," vol. 1, no. 2, pp. 1–10, 2015.

[14] E. P. A. Pratiwi, Lhorend Mutiara, Diana, "Penerapan K-Means Clustering Untuk Memprediksi Minat Nasabah Pada Pt . Asuransi Jiwa Bersama 1912 Bumiputera Prabumulih Lhorend Mutiara Pratiwi[1] , Diana[2] , Eka Puji Agustini[3] Dosen Universitas Bina Darma[2] · [3] , Mahasiswa Universitas Bina Darma[1] Jalan Jend," p. 16.

[15] T. Suharseno, R. Hidayat, D. Ayu, and L. Dewi, "Kategori Produk Terhadap Keputusan Perpindahan Merek Dengan Kebutuhan Mencari Variasi Sebagai," vol. 18, no. 2, pp. 176–182, 2013.

[16] A. Sitepu, "Karakteristik Keluarga Menurut Peringkat Kemiskinan: Studi Pendahuluan Untuk Perumusan Kriteria Fakir Miskin (Family characteristics ratings by poverty: preliminary studies for the formulation of criteria for poor people)," vol. 17, no. 01.