PENERAPAN METODE K-MEANS UNTUK PENGELOMPOKKAN KABUPATEN/KOTA DI PROVINSI SUMATERA UTARA BERDASARKAN INDIKATOR INDEKS PEMBANGUNAN MANUSIA

Horainim Sibarani, Solikhun, Widodo Saputra, Indra Gunawan, Zulaini Masruro Nasution
Program Studi Teknik Informatika
STIKOM Tunas Bangsa Pematangsiantar, Jl. Jendral Sudirman Blok A, No.1,2&3 Pematangsiantar
sibaranihorainim@gmail.com

ABSTRAK

Pembangunan manusia adalah harapan utama bagi sebuah negara dalam mensejahterahkan masyarakat.Sebagian faktor penting untuk mengukur derajat manusia ialah bahwa suatu masyarakat maju secara kuantitas serta kualitas, yang dapat diamati pada tiga aspek kehidupan, yaitu kesempatan hidup, pendidikan dan kehidupan yang cocok.Penelitian ini membicarakan perihal penggunaan algoritma *k-means*pada pengelompokkan kabupaten/kota di Sumatera Utara berdasarkan karakteristik wilayah pada empat indeks pembangunan manusia (IPM). Empat kriteria tersebut adalah angka harapan hidup (AHH), angka melek huruf(AMH), rata - rata lama sekolah (RLS) dan pengeluaran per kapita (PPK). Berdasarkan data Indeks Pembangunan Manusia, Angka Harapan Hidup, Angka Melek Huruf, Rata-rata Lama Sekolah, dan Pengeluaran Per Kapita, ditemukan bahwa terdapat perbedaan yang cukup besar. Algoritma *k-means* berhasil diterapkan pada data indikator Indeks Pembangunan Manusia Kabupaten Dan Kota Di Provinsi Sumatera Utara untuk mengelompokkan jumlah IPM dari yang terendah, sedang dan terkecil. Adapun hasil yang didapatkan dalam mengelompokkan data indikator indeks pembangunan manusia di sumatera utara yaitu *cluster* 0 (terendah) berjumlah 13, *cluster* 1 (sedang) berjumlah 1, dan *cluster* 2 (tertinggi) berjumlah 19.

Kata kunci : *IPM, Pembangunan, K-Means.*

1. PENDAHULUAN

Indeks Pembangunan Manusia (IPM) adalah alat pengukur dalam melakukan penilaian keberhasilan pembangunan manusia, dilihat dari sisi akibatnya (kesehatan dan kesejahteraan) maupun yang bersifat non-fisik (intelektualitas). Manusia selalu menjadi hal utama dalam setiap program pendapatan pembangunan, Berdasarkan penilaian ini, daftar negara menurut IPM diberikan setiap tahun. Indeks Kemiskinan Manusia, yang berfokus pada kemiskinan masyarakat.

Badan Pusat Statistik (BPS) melakukan perhitungan dalam satu dekade ini pembangunan manusia di Indonesia terus mengalami peningkatan yang pesat. Khususnya provinsi Sumatera Utara, Pertumbuhan IPM Nias Barat (0,61 persen), lebih tinggi dari Kota Medan (0,01 persen) sehingga pada tahun 2020 berhasil memperkecil disparitas IPM Sumatera Utara.

Penurunan kesenjangan IPM cukup menggembirakan, Tetapi hal ini menyebabkan Upaya pembangunan pemerintah tidak merata, atau dengan kata lain upaya pembangunan pemerintah di daerah tidak tepat sasaran. dan Pemerintah tentu saja bidang prioritas harus diperhitungkan agar kesenjangan juga berkurang. Jika tidak dilakukan secara merata maka Kenaikan besar di suatu wilayah tertentu akan memperbesar disparitas yang tidak seimbang oleh pembangunan wilayah lainnya. Oleh karena itu, salah satu alternatifnya adalah dengan mengelompokkan kabupaten/kota berdasarkan indikator IPM sehingga pemerintah dapat mengambil atau memutuskan kebijakan dan strategi yang baik dan tepat untuk mencapai tujuan pembangunan.

Metode K-Means merupakan metode non-*hierarchicalclustering* yang berusaha membantu penyertaan variabel kelompok dalam kelas hasil perhitungan akhir. Pendekatan ini membagi data menjadi cluster/grup sedemikian rupa sehingga data dengan kualitas yang sama dikelompokkan bersama.

sistem kerja pada Algoritma ini yaitu dengan membagi data menjadi beberapa kelompokuntuk menganalisis faktor kesamaan dan ketidaksamaan yang melekat pada kumpulan data. Penelitian ini membahas penggunaan algoritma *k-means* untuk mengelompokkan kabupaten/kota di Provinsi Sumatera Utara berdasarkan kesamaan karakteristik kabupaten/kota ditinjau dari empat ukuran indikator kematian IPM regional. Ukuran indikator tersebut adalah Angka Harapan Hidup (AHH), Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), dan Pengeluaran Per Kapita (PPK). Secara umum, *output* penelitian berupa hasil klasterisasi dapat dimanfaatkan sebagai pedoman untuk mengkarakterisasi pengelompokan sebaran di Provinsi Sumatera Utara berdasarkan parameter profil IPM.

Penelitian sebelumnya tahun 2017 dari M. W. Talakua, Z. A. Leleury dan A. W. Talluta dengan judul "Analisis *Cluster* Dengan Menggunakan Metode *K-Means* Untuk Pengelompokkan Kabupaten/Kota Di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014" bertujuan untuk mengelompokan kabupaten/kota di

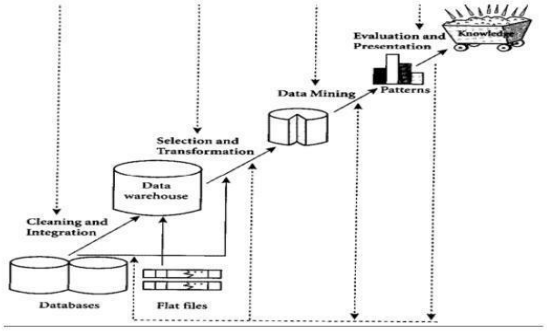
Provinsi Maluku berdasarkan kemiripan karakteristik daerah yang ditinjau dari empat ukuran indikator mortalitas IPM di daerah. Data yang dipakai adalah adalah Angka Harapan Hidup (AHH), Angka Melek Huruf (AMH), Rata-rata Lama Sekolah (RLS), Pengeluaran Per Kapita (PPK). Dari penelitian tersebut dapat disimpulkan *Cluster* 1 yaitu Kota Ambon yang mempunyai angka IPM dan AHH sangat maksimal (sangat banyak sedangkan jumlah AMH, angka RLS dan PPK berbanding balik dengan angka IPM yaitu terjadi pada tahun 2014. *Cluster* 2 yaitu MTB, Kepulaan Aru, SBB, SBT, MBD, dan Buru Selatan mempunyai angka IPM, AHH, AMH, RLS, dan angka PPK sangat rendah terjadi pada tahun 2014. *Cluster* 3 yaitu Maluku Tenggara, Maluku Tengah, Buru, Tual mempunyai angka IPM, AHH, AMH, RLS, dan angka PPK sangat banyak terjadi pada tahun 2014.

2. TINJAUAN PUSTAKA

2.1 Data *Mining*

Data mining adalah teknologi yang menggabungkan pendekatan statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk berinteraksi dengan database besar dan mengungkap informasi yang dapat digunakan. Widiastari, Solikhun, & Irawan, 2021 [18]

Data mining adalah *subjek* yang menyelidiki bagaimana mengekstrak pengetahuan atau mengungkap pola dari data sehingga temuan pengolahan data dapat digunakan untuk membuat keputusan di masa depan. Pradipta *et al.*, 2019 [1]



Gambar 1. Tahapan Data *Mining*
(Sumber: Handoko, 2018)

2.2 *Clustering*

Clustering adalah metode analisis data atau data mining yang melakukan proses pemodelan tanpa pengawasan (*unsupervised*) dan mengelompokkan data menggunakan skema partisi. *K-Means* merupakan salah satu metode clustering data non-*hierarchical* atau *PartitionalClustering*. Ada dua jenis pengelompokan data yang biasa digunakan dalam proses pengelompokan data: Hirarkis dan Non-Hierarki. Pendekatan *K-MeansClustering* berusaha untuk membagi data yang ada ke dalam berbagai kelompok, dengan data dalam satu kelompok memiliki ciri yang sama dan data pada kelompok lain memiliki karakteristik yang berbeda. Sy and Syam, 2016 [15]

2.3 *Algoritma K-Means*

K-means adalah algoritma partisi data yang membagi data ke dalam kelompok yang berbeda. *K-Means* mampu rata-rata jarak antara setiap titik data dan clusternya dengan membelah secara iteratif. Setiap titik data dalam algoritma *K-Means* harus menjadi bagian dari *cluster* tertentu pada satu titik dalam proses, tetapi dapat berpindah ke *cluster* lain pada titik berikutnya.

K-Means merupakan algoritma yang termasuk ke dalam *datamining* bidang *Clustering*atau pengklasteran data. *K-Means Clustering* adalah suatu metode penganalisaan data yang melakukan proses pemodelan tanpa supervisi dan merupakan metode yang termasuk mengelompokkan data secara sistem partisi Sembiring, Octaviana and Saepudin, 2020 [12]

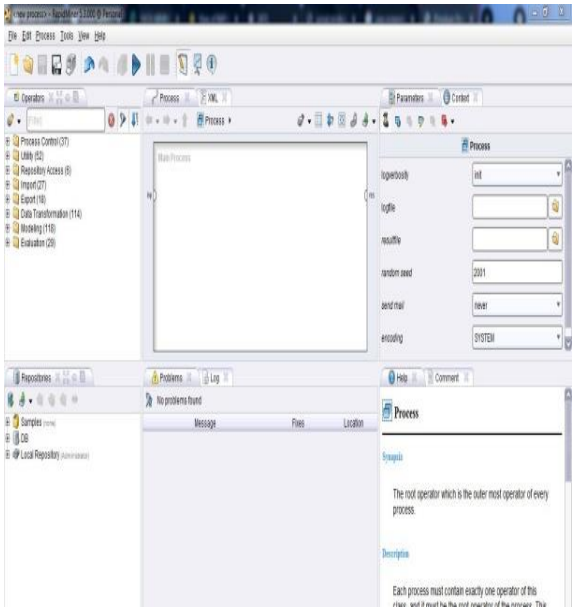
2.4 Indeks Pembangunan Manusia (IPM)

Indeks Pembangunan Manusia (IPM) adalah indikator komposit tunggal yang, meskipun tidak dapat mengukur semua aspek pembangunan manusia, mengukur tiga dimensi dasar manusia yang dianggap mencerminkan kemampuan dasar penduduk (kemampuan dasar). Umur panjang dan kesehatan, pengetahuan dan keterampilan, dan akses ke sumber daya yang dibutuhkan untuk mempertahankan tingkat hidup yang adil adalah tiga bakat dasar. Pembangunan manusia, menurut UNDP, adalah proses peningkatan pilihan masyarakat dalam hal pendapatan, kesehatan, pendidikan, lingkungan fisik, dan sebagainya.

2.5 *RapidMiner*

Rapidminer adalah *platform* alat ilmu data terkenal yang menyediakan lingkungan terintegrasi untuk produksi data pembelajaran mesin untuk pengembangan teks dan analisis prediktif. Perangkat lunak *Rapidminer* menawarkan sejumlah algoritma, termasuk klasifikasi, pengelompokan, dan asosiasi. Widiastari, Solikhun, & Irawan, 2021 [18]

RapidMiner menggunakan kombinasi statistik, kecerdasan buatan, dan teknologi *database* untuk mengekstrak pola dari kumpulan data yang sangat besar. *Rapid Miner* memudahkan pengguna untuk mempekerjakan operator untuk menghitung volume data yang sangat besar. Operator ini beroperasi dengan data untuk mengubahnya. Kita cukup menghubungkannya ke node hasil untuk memeriksa hasil data dengan node pada operator. Hasil *RapidMiner* juga dapat direpresentasikan menggunakan grafik. *RapidMiner* telah menjadi salah satu aplikasi paling populer untuk mengekstrak data menggunakan teknik *data*. *mining.* Rahmat C.T.I. *et al.*, 2017 [10]



Gambar 2. Tampilan *RapidMiner*
(Sumber: Penulis)

2.6 Perbedaan Penelitian Sebelumnya

Berdasarkan penelitian yang diakukan oleh (Talakua, Leleury and Talluta, 2017) berjudul "Analisis Cluster Dengan Menggunakan Metode *K-Means* Untuk Pengelompokkan Kabupaten/Kota Di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia 2014" dapat disimpulkan bahwa penelitian tersebut dengan penelitian ini memiliki perbedaan yaitu masalah yang diteliti berbeda, variabel penelitian berbeda, jumlah *cluster* yang digunakan dan studi kasus yang berbeda.

3. METODE PENELITIAN

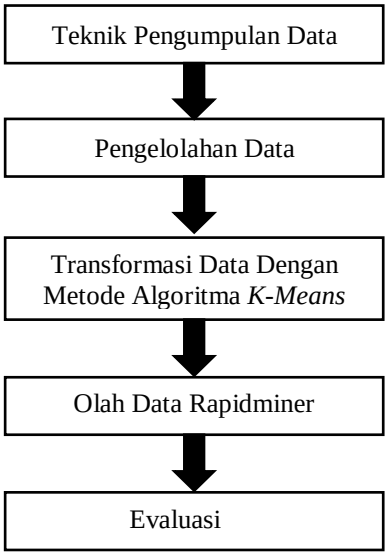
Metode penelitian yang dipakai adalah metode sistematis yang bertujuan untuk memperoleh jawaban yang sesuai dengan pertanyaan tentang objek penelitian dan mengetahui urutan yang sistematis.

Dalam penelitian ini, penulisakan membahas, Penerapan Metode *K-Means* Untuk Pengelompokkan Kabupaten/Kota Di Provinsi Sumatera Utara Berdasarkan Indikator Indeks Pembangunan Manusia (IPM).

3.1 Rancangan Penelitian

Berikut merupakan sebuah gambar rancangan penelitian sebagai berikut:

1. Identifikasi Masalah
   Upaya dalam mendapatkan sebuah gambaran dari bentuk permasalahan yang diteliti.
2. Teknik Pengumpulan Data
   Bagian ini dilakukan seorang peneliti dalam mengumpulkan sebuah data yang akan berguna sebagai bukti nyata penelitian.
3. Pengolahan Data
   Pengolahan data dilakukan dengan menggunakan metode *k-means* dan aplikasi *Rapidminer*
4. Transformasi Data Dengan Metode Algoritma *K-Means*
   Bagian ini untuk mengubah data mentah ke dalam bentuk angka/*numerikal* yang kemudian dilakukan pengelompokkan menjadi 3 *cluster* yaitu rendah, sedang, dan tinggi.
5. Olah Data di *RapidMiner*
   Untuk melakukan pengujian data peneliti menggunakan *software RapidMiner Versi 5.3*. Pengujian dilakukan setelah melakukan perhitungan manual di aplikasi *Excel 2007*.
6. Evaluasi
   Melakukan pembandingan supaya hasil kegiatan dapat dianalisis agar setiap rencana-rencana mencapai tujuan.



Gambar 3. Rancangan Penelitian
(Sumber: Penulis)

3.2 Prosedur Pengumpulan Data

Berikut ini adalah prosedur digunakan untuk penelitian terdiri dari:

1. Memanfaatkan berbagai referensi seperti jurnal, internet dalam melakukan faktor, parameter dan label yang digunakan dalam penelitian.
2. Sumber data dalam penelitian ini diproses dari website resmi pemerintah yang beralamatkan sumut.bps.go.id. Data yang diambil adalah data Penerapan Metode *K-Means* Untuk Pengelompokkan Kabupaten/Kota Di Provinsi Sumatera Utara Berdasarkan Indikator Indeks Pembangunan Manusia (IPM). Data ini akan menggunakan 3 *cluster* yang terendah, sedang dan tinggi dalam pengolahannya.

3.3 Algoritma *K-Means*

K-means adalah algoritma partisi data yang membagi data ke dalam kelompok yang berbeda. *K-Means* mampu rata-rata jarak antara setiap titik data dan clusternya dengan membelah secara iteratif. Setiap titik data dalam algoritma K-Means harus menjadi bagian dari *cluster* tertentu pada satu titik

dalam proses, tetapi dapat berpindah ke cluster lain pada titik berikutnya.

Secara umum algoritma *K-Means* memiliki langkah-langkah dalam pengelompokan, diantaranya Iin Parlina, Agus Perdana Windarto, Anjar Wanto, 2018 [6]:

1. Inisilisasi: menentukan nilai K centroid yang diinginkan dan metrik ketidakmiripan (jarak) yang diinginkan.
2. Memilih K data dari set X sebagai centroid. Untuk menentukan centroid dapat menggunakan persamaan (1).
   $\left[\frac{Jumlah\ Data}{Jumlah\ Class+1}\right]$.....(1)
3. Mengalokasikan semua data ke centroid terdekat dengan matrik jarak yang telah ditetapkan
4. Menghitung kembali centroid C berdasarkan data yang mengikuti cluster masing – masing.
5. Mengulangi langkah 3 dan 4 hingga kondisi konvergen tercapai.

   Berikut ini adalah rumus untuk menentukan jumlah cluster:

   $\left[k = \sqrt{\frac{n}{2}}\right]$.....*(2)*

   Keterangan:

   K = klaster

   N = jumlah data

   Menghitung jarak pada ruang jarak Euclidean menggunakan formula:

$$D(x_2,x_1) = \|x_2 - x_1\|_2 = \sqrt{\sum_{j=1}^{p} |x_{2j} - x_{1j}|^2} \quad (3)$$

3.4 *RapidMiner*

Rapidminer adalah *platform* alat ilmu data terkenal yang menyediakan lingkungan terintegrasi untuk produksi data pembelajaran mesin untuk pengembangan teks dan analisis prediktif. Perangkat lunak *Rapidminer* menawarkan sejumlah algoritma, termasuk klasifikasi, pengelompokan, dan asosiasi. Widiastari, Solikhun, & Irawan, 2021 [18]

RapidMiner menggunakan kombinasi statistik, kecerdasan buatan, dan teknologi database untuk mengekstrak pola dari kumpulan data yang sangat besar.*RapidMiner* memudahkan pengguna untuk mempekerjakan operator untuk menghitung volume data yang sangat besar.Operator ini beroperasi dengan data untuk mengubahnya.Kita cukup menghubungkannya ke node hasil untuk memeriksa hasil data dengan node pada operator.Hasil *RapidMiner* juga dapat direpresentasikan menggunakan grafik.*RapidMiner* telah menjadi salah satu program paling populer untuk mengambil data menggunakan teknik *datamining*. Rahmat C.T.I. *et al.*, 2017 [10]

4. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data IPM Sumatera Utara berdasarkan indikator. Untuk mendukung proses penelitian, penulis menggunakan *software Rapidminer 5.3* untuk memvalidasi hasil perhitungan manual menggunakan aplikasi untuk mengelompokkan IPM di Sumatera Utara. Data IPM Sumatera Utara diambil dari situs resmi yang beralamatkan www.bps.go.id. Data yang diperoleh akan diubah ke dalam data *Excel* 2007 kemudian dikelolah dengan algoritma *k-means* untuk menentukan jumlah klaster IPM

4.1 Perhitungan Manual Metode *K-Means*

Berikut ini penulis melakukan beberapa langkah-langkah dalam mengelompokkan data Indeks Pembangunan Manusia (IPM) menggunakan algoritma *k-means clustering:*

1. Langkah awal yang penulis lakukan adalah dengan mengurangi beberapa atribut yang tidak diperlukan.

Tabel 1 Data yang telah di *Tranformasi*

No	Kabupaten/Kota	Angka Harapan Hidup	Harapan Lama Sekolah	Rata-rata Lama Sekolah	Pengeluaran per kapita
1	Nias	69,75	12,57	5,36	6,898
2	Mandailing Natal	62,60	13,32	8,62	9,684
3	Tapanuli Selatan	64,91	13,24	9,28	11,236
4	Tapanuli Tengah	67,15	13,06	8,62	10,071
5	Tapanuli Utara	68,63	13,69	9,85	11,648
6	Toba Samosir	70,08	13,45	10,52	12,154
7	Labuhanbatu	69,93	12,73	9,24	11,150
8	Asahan	68,26	12,60	8,79	10,890
9	Simalungun	71,22	12,78	9,60	11,308
10	Dairi	69,00	13,10	9,58	10,350
11	Karo	71,40	12,76	9,79	12,349
12	Deli Serdang	71,73	13,35	10,09	12,225
13	Langkat	68,80	13,05	8,65	11,071
14	Nias Selatan	68,74	12,23	5,85	6,974
15	Humbang Hasudutan	69,27	13,28	9,54	7,850
16	Pakpak Barat	65,74	13,86	9,03	8,170
17	Samosir	71,27	13,47	9,43	8,422
18	Serdang Bedagai	68,68	12,60	8,54	10,950
19	Batu Bara	68,96	12,63	8,06	10,410
20	Padang Lawas Utara	67,17	12,87	9,37	9,987
21	Padang Lawas	67,09	13,03	9,01	8,807
22	Labuhanbatu Selatan	68,71	13,00	8,75	11,495
23	Labuhanbatu Utara	69,46	13,04	8,40	11,779
24	Nias Utara	69,43	13,03	6,58	6,064
25	Nias Barat	68,96	12,94	6,49	5,830
26	Sibolga	69,01	13,16	10,40	11,473
27	Tanjungbalai	63,27	12,50	9,44	11,321
28	Pematangsiantar	73,55	14,45	11,16	12,372
29	Tebing Tinggi	70,87	12,72	10,31	12,876
30	Medan	73,14	14,74	11,39	14,890
31	Binjai	72,38	13,62	10,93	10,997
32	Padangsidimpuan	69,41	14,54	11,00	10,856
33	Gunungsitoli	71,19	13,74	8,61	7,980

2. Menentukan nilai pusat (*centroid*) secara random. *Centroid* yang saya ambil adalah C1 (Batubara), C2 (Nias), C3 (Labuhanbatu) yang dimana nilai centroid tersebut di ambil dari nilai rata-rata yang sebelumnya dihitung. nilai rata-rata diambil akan menjadi nilai terendah, sedang, tertinggi.

Tabel 2 Nilai *Centroid*

Cluster	Nama	AHH	HLS	RLS	PPK
C1	Batu Bara	66.96	12.63	8.06	10.410
C2	Nias	6.750	12.57	5.36	6.898
C3	Labuhanbatu	69.93	12.73	9.24	11.150

3. Penentuan nilai *Centroid* dapat dilakukan dengan menghitung rata-rata ataupun mengambilnya secara acak. Setelah ditentukannya nilai *centroid*, setiap data akan dilakukan penghitungan terhadap pusat *cluster* sehingga masing-masing dapatmenemukan jarak terdekat dari setiap data dengan *centroid*. Berikut ini contoh perhitungan jarak terdekat setiap data pada *cluster* pertama (C1) :

D (1,1)

$$:\sqrt{(6.750-66.96)^2+(12.57-12.63)^2+(5.36-8.06)^2+(6.898-10.410)^2}$$

$$= 60{,}37277$$

Perhitungan akan diteruskan sampai dengan D (1,33) perhitungan jarak terdekat setiap data pada *cluster* pertama (C2) :

D (2,1)

$$:\sqrt{(6.750-6.750)^2+(12.57-12.57)^2+(5.36-5.36)^2+(6.898-6.898)^2}$$

$$= 0$$

Perhitungan akan diteruskan sampai dengan D (2,33) perhitungan jarak terdekat setiap data pada *cluster* pertama (C3) :

D (3,1)

$$:\sqrt{(6.750-69.93)^2+(12.57-12.73)^2+(5.36-9.24)^2+(6.898-11.150)^2}$$

$$= 63{,}44188$$

Perhitungan akan diteruskan sampai dengan D (3,33)

Berikut adalah hasil dari perhitungan jarak data dengan titik pusat pada iterasi 1 :

Tabel 3 Hasil Perhitungan Jarak Data Ke *Cluster*

No	Nama	C1	C2	C3
1	Nias		I	
2	Mandailing Natal	I		
3	Tapanuli Selatan	I		
4	Tapanuli Tengah	I		
5	Tapanuli Utara			I
6	Toba Samosir			I
7	Labuhanbatu			I
8	Asahan	I		
9	Simalungun			I
10	Dairi			I
11	Karo			I
12	Deli Serdang			I
13	Langkat			I
14	Nias Selatan	I		
15	Humbang Hasudutan			I
16	Pakpak Bharat	I		
17	Samosir			I
18	Serdang Bedagai			I
19	Batu Bara	I		
20	Padang Lawas Utara	I		
21	Padang Lawas	I		
22	Labuhanbatu Selatan			I
23	Labuhanbatu Utara			I
24	Nias Utara			I
25	Nias Barat	I		
26	Sibolga			I
27	Tanjungbalai	I		
28	Pematangsiantar			I
29	Tebing Tinggi			I
30	Medan			I
31	Binjai			I
32	Padangsidimpuan			I
33	Gunungsitoli			I

Pada tabel 3 hasil yang didapat dalam mengelompokkan objek satu data berdasarkan jarak ke *centroid* terdekat. Jika data terdapat nilai maka masuk kelompok *cluster* 1 (C1) begitu juga dengan nilai sedang maka masuk ke *cluster* 2 (C2) dan nilai tertinggi maka masuk ke *cluster* 3 (C3).

4. Jika pada iterasi 1 hasil cluster telah didapatkan langkah selanjutnya untuk memastikan ulang apakah nilai cluster di iterasi 1 benar atau sama, maka penulis akan mencoba iterasi 2.
5. Selanjutnya memperbarui nilai centroid dengan cara sebagai berikut:

C1 p
(AHH) :
(62.60+64.91+67.15+68.26+68.74+65.74+66.96 +67.17+67.09+68.96+63.27)/11 = 66,4409
(HLS) :
(13.32+13.24+13.06+12.60+12.23+13.86+12.63 +12.87+13.0+12.94+12.50)/11
= 12,9345
(RLS) :

(8.62+9.28+8.62+8.79+5.85+9.03+8.06+9.37+9 .01+6.49+9.44)/11
= 8.4145
(PPK):
(9.684+11.236+10.071+10.890+6.974+8.170+1 0.410+9.987+8.807+5.830+11.321)/11 = 9.0226

C2
(AHH) : $6.750/1 = 6.750$
(HLS) : $12.57/1 = 12.57$
(RLS) :$5.36/1 = 5.36$
(PPK) :$6.898/1 = 6.898$

C3(AHH):
(68.63+70.08+69.93+71.22+69.00+71.40+71.73 +68.80+69.27+71.27+68.68+68.71+69.46+69.4 3+69.01+73.55+70.87+73.14+72.38+69.41+71. 19)/21
= 70,3409
(HLS) :
(13.69+13.45+12.73+12.78+13.10+12.76+13.35 +13.05+13.28+13.47+12.60+13.00+13.04+13.0 3+13.16+14.45+12.72+14.74+13.62+14.54+13. 74)/21 = 13,3476
(RLS):
(9.85+10.52+9.24+9.60+9.58+9.79+10.09+8.65 +9.54+9.43+8.54+8.75+8.40+6.58+10.40+11.16 +10.31+11.39+10.93+11.00+8.61)/21 = 9.6361
(PPK) :

(11.648+12.154+11.150+10.890.+11.308+10.350+12.349+12.225+11.071+7.850+8.422+10.950+11.495+11.779+11.473+12.327+14.890+10.997+10.856+7.980) /21 = 10.9647

Tabel 4 Perhitungan *Centroid* baru pada iterasi 2

Cluster Baru	AHH	HLS	RLS	PPK
C1	66.44090909	12.93454545	8.414545455	9.398181818
C2	6.75	12.57	5.36	6.898
C3	70.34095238	13.34761905	9.636190476	10.96471429

6. Setelah mendapatkan nilai centroid yangbaru maka langkah selanjutnya menghitung kembali jarak setiap data terhadap pusat *cluster*.

Perhitungan jarak terdekat setiap data pada *cluster* pertama (C1) :

E (1,1)

$:\sqrt{(67.50-66.4409)^2+(12.57-12.93455)^2+(5.36-8.4145)^2+(6.898-9.3981)^2}$

= 59,82238

Perhitungan akan diteruskan sampai dengan E (1,33). Perhitungan jarak terdekat setiap data pada *cluster* pertama (C2) :

E (2,1)

$:\sqrt{(6.750-6.750)^2+(12.57-12.57)^2+(5.36-5.36)^2+(6.898-6.898)^2}$

= 0

Perhitungan akan diteruskan sampai dengan E (2,33). Perhitungan jarak terdekat setiap data pada *cluster* pertama (C3) :

E(3,1)

$:\sqrt{(6.750-70.256)^2+(12.57-13.347)^2+(5.36-9.636)^2+(6.898-10.964)^2}$

= 63.86891

Perhitungan akan diteruskan sampai dengan E (3,33)

Berikut table 5 hasil perhitungan jarak data dengan titik pusat pada iterasi 2 :

Tabel 5 Perhitungan Jarak Data iterasi 2

No	Nama	C1	C2	C3
1	Nias		I	
2	Mandailing Natal	I		
3	Tapanuli Selatan	I		
4	Tapanuli Tengah	I		
5	Tapanuli Utara			I
6	Toba Samosir			I
7	Labuhanbatu			I
8	Asahan			I
9	Simalungun			I
10	Dairi			I
11	Karo			I
12	Deli Serdang			I
13	Langkat			I
14	Nias Selatan	I		
15	Humbang Hasudutan			I
16	Pakpak Bharat	I		
17	Samosir			I
18	Serdang Bedagai			I
19	Batu Bara	I		
20	Padang Lawas Utara	I		
21	Padang Lawas	I		
22	Labuhanbatu Selatan			I
23	Labuhanbatu Utara			I
24	Nias Utara	I		
25	Nias Barat	I		
26	Sibolga			I
27	Tanjungbalai	I		
28	Pematangsiantar			I
29	Tebing Tinggi	I		
30	Medan			I
31	Binjai			I
32	Padangsidimpuan			I
33	Gunungsitoli			I

Tabel 6 Hasil *Cluster* iterasi 2

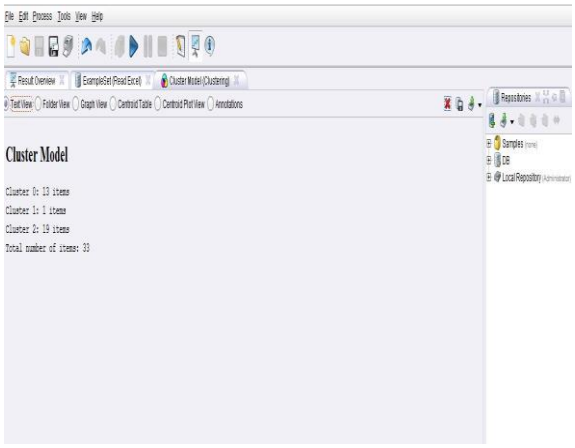
Cluster	Nilai
C1	12
C2	1
C3	20

Berdasarkan dari perhitungan manual diatas, mendapatkan hasil bahwa iterasi 2 bernilai berbeda, maka dari itu akan dilakukan lagi iterasi ke 3 sampai ke 4. dan ternyata mendapatkan hasil akhir iterasi 3 dan iterasi 4 bernilai sama yaitu C1 = 13, C2 = 1, dan C3 = 19. Posisi iterasi tidak berubah, maka proses berhenti atau selesai dan disimpulkan:

1. *Cluster* rendah (C1) dengan jumlah data 13 yaitu mandailing natal, tapanuli selatan, tapanuli tengah, asahan, nias selatan, humbang hasudutan, pakpak bharat, batu bara, padang lawas utara, padang lawas, nias utara, nias barat, tanjungbalai.
2. *Cluster* sedang (C2) dengan jumlah data 1 yaitu nias
3. *Cluster* Tinggi (C3) dengan jumlah data 19 yaitu tapanuli utara, toba samosir, labuhanbatu, simalungun, dairi, karo, deli serdang, langkat, samosir serdang bedagai, labuhanbatu selatan, labuhanbatu utara, sibolga, pematangsiantar, tebing tinggi, medan, binjai, padangsidimpuan, gunungsitoli

Setelah melakukan perhitungan manual dan didapatkan hasil dari perhitungan manual, maka selanjutnya mengimplementasikan hasil dari perhitungan manual menggunakan *software Rapidminer 5.3.*

Dari hasil perhitungan manual dan *Rapidminer* didapatkan hasil yang sama yang dimana *Cluster* rendah (C1) dengan jumlah data 13, *Cluster* sedang (C2) dengan jumlah data 1, dan *Cluster* Tinggi (C3) dengan jumlah data 19. Dapat dilihat di gambar 3 :



Gambar 3 Hasil *Clustering* di *RapidMiner*
(Sumber: Penulis)

5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Setelah melakukan perhitungan dan pembahasan algoritma *K-Means* dan implementasi *Rapidminer* pada bab sebelumnya, penulis dapat manarik kesimpulan yaitu algoritma *k-means* berhasil diterapkan pada data indikator Indeks Pembangunan Manusia Kabupaten Dan Kota Di Provinsi Sumatera Utara untuk mengelompokkan jumlah IPM dari yang terendah, sedang dan terkecil.Adapun hasil yang didapatkan dalam mengelompokkan data indikator indeks pembangunan manusia di sumatera utara yaitu *cluster* 0 (terendah) berjumlah 13 yaitu (Mandailing Natal, Tapanuli Selatan, Tapanuli Tengah, Asahan, Nias Selatan, Humbang Hasudutan, Pakpak Bharat, Batu Bara, Padang Lawas Utara, Padang Lawas, Nias Utara, Nias Barat, Tanjungbalai), *cluster* 1 (sedang) berjumlah 1 yaitu (Nias), *cluster* 2 berjumlah 19 yaitu (Tapanuli Utara, Toba Samosir, Labuhanbatu, Simalungun, Dairi, Karo, Deli Serdang, Langkat, Samosir Serdang Bedagai, Labuhanbatu Selatan, Labuhanbatu Utara, Sibolga, Pematangsiantar, Tebing Tinggi, Medan, Binjai, Padangsidimpuan, Gunungsitoli). Hasil tersebut sama dengan hasil yang telah diimplementasikan ke dalam *software Rapidminer* versi 5.3.

5.2 Saran

Untuk penulis berikutnya yang akan melakukan penelitian dengan algoritma yang sama, penulis akan memberikan beberapa sebuah saran sebagai berikut: Peneliti selanjutnya diharapkan Melakukan pengujian dengan metode dan algoritma yang berbeda supaya hasil yang didapatkan lebih berbeda dari sebelumnya, Untuk peneliti selanjuntnya diharapkan aplikasi yang digunakan sebaiknya berbeda agar terdapat keberagaman ilmu. Penelitian ini diharapkan dapat menjadi referensi dibidang informasi dan ilmu pengetahuan untuk penelitian selanjutnya

DAFTAR PUSTAKA

[1] A. Pradipta, D. Hartama, A. Wanto, S. Saifullah, and J. Jalaluddin, "The Application of Data Mining in Determining Timely Graduation Using the C45 Algorithm," *IJISTECH (International J. Inf. Syst. Technol.*, vol. 3, no. 1, p. 31, 2019, doi: 10.30645/ijistech.v3i1.30.

[2] Aprilla Dennis. Belajar Data Mining dengan RapidMiner. *Innov. Knowl. Manag. Bus. Glob. Theory Pract. Vols 1 2***5**, 1–5 (2013).

[3] Darsyah, M. Y. Klasifikasi Indeks Pembangunan Manusia (Ipm) Dengan Pendekatan K-Nearset Neighbor (K-Nn). *Semin. Nas. Pendidikan, Sains dan Teknol. Fak. Mat. dan Ilmu Pengetah. Alam Univ. Muhammadiyah Semarang* 29–35 (2014).

[4] G. N. W. Paramartha, D. E. Ratnawati, and A. W. Widodo, "Analisis Perbandingan Metode K-Means Dengan Improved Semi-Supervised Analisis Perbandingan Metode K-Means Dengan Improved Semi- Supervised K-Means Pada Data Indeks Pembangunan Manusia (IPM)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. Vol. 1, no. 9, pp. 813–824, 2017, [Online]. Available: http://j-ptiik.ub.ac.id.

[5] Handoko, K. (2018). Pengelompokkan Data Mining Pada Jumlah Penumpang Di Bandara Hang Nadim. *Computer Based Information System Journal*, *6*(2), 60. https://doi.org/10.33884/cbis.v6i2.708

[6] Iin Parlina, Agus Perdana Windarto, Anjar Wanto, M.Ridwan Lubis, 'Memanfaatkan Algoritma K-Means Dalam Menentukan Pegawai Yang Layak Mengikuti Asessment Center', *Memanfaatkan Algoritma K-Means Dalam Menentukan Pegawai Yang Layak Mengikuti Asessment Center Untuk Clustering Program Sdp*, 3.1 (2018), 87–93

[7] N. Putu, E. Merliana, P. Studi, M. Teknik, F. T. Industri, and U. A. Jaya, "ANALISA PENENTUAN JUMLAH CLUSTER TERBAIK PADA METODE K-MEANS," pp. 978–979, 2015.

[10] Rahmat C.T.I., Brilian, et al. "Implemetasi K-Means Clustering Pada Rapidminer Untuk Analisis Daerah Rawan Kecelakaan." *Seminar Nasional Riset Kuantitatif Terapan 2017*, no. April, 2017, pp. 58–60

[11] R. Dewi Kusumah, B. Warsito, and M. Abdul Mukid, "Perbandingan metode k – means dan self organizing map (Studi kasus: pengelompokan kabupaten/kota di jawa tengah berdasarkan indikator indeks pembangunan manusia 2015)," *J. Gaussian, Vol 6 No 3 tahun 2017*, vol. 6, pp. 429–437, 2017.

[12] Sembiring, Falentino, et al. "Implementasi Metode K-Means Dalam Pengklasteran Daerah Pungutan Liar Di Kabupaten Sukabumi (Studi Kasus : Dinas Kependudukan Dan Pencatatan Sipil)." *Jurnal Tekno Insentif*, vol. 14, no. 1,

2020, pp. 40–47, doi:10.36787/jti.v14i1.165.

[13] S. Nelson Butarbutar, Agus Perdana Windarto2, Dedi Hartama3, "1 , 2 , 3 , 4," vol. 1, no. 2012, 2016.

[14] Sutrisno, et al. "Penerapan Data Mining Pada Penjualan Menggunakan Metode Clustering Study Kasus Pt . Indomarco." *Penerapan Data Mining Pada Penjualan Menggunakan Metode Clustering*, vol. Vol.x No.x, no. Data Mining, 2013, pp. 1–11,http://eprints.binadarma.ac.id/78/1/PENERAPAN DATA MINING PADA PENJUALAN MENGGUNAKAN METODE CLUSTERING STUDY KASUS PT. INDOMARCO PALEMBANG.pdf.

[15] Sy, Hasyrif, and Asrul Syam. *Data Mining Menggunakan Algoritma K-Means Pengelompokan Penyebaran Diare Di Kota Makassar*. no. 1, 2016, pp. 73–82.

[16] Talakua, M. W., et al. *ANALISIS CLUSTER DENGAN MENGGUNAKAN METODE PROVINSI MALUKU BERDASARKAN INDIKATOR INDEKS PEMBANGUNAN MANUSIA TAHUN 2014 CLUSTER ANALYSIS BY USING K-MEANS METHOD FOR GROUPING OF DISTRICT / CITY IN MALUKU PROVINCE INDUSTRIAL BASED ON INDICATORS OF MALUKU DEV. 2017*, pp. 119–28.

[17] Wardhani, Anindya Khrisna. "K-Means Algorithm Implementation for Clustering of Patients Disease in Kajen Clinic of Pekalongan." *Jurnal Transformatika*, vol. 14, no. 1, 2016, p. 30,doi:10.26623/transformatika.v14i1.38.

[18] Widiastari, Afrialita. *Analisa Datamining Dengan Metode Klasifikasi C4 . 5 Sebagai Faktor Penyebab Tanah Longsor*. no. 3, 2021, pp. 247–55.