



Penerapan Algoritma K-Medoids Untuk Pengelompokan Pengangguran Umur 25 tahun Keatas Di Sumatera Utara

Dedy Sutris Martua Simanjuntak^{1*}, Indra Gunawan², Sumarno³, Poningsih⁴, Ika Purnama Sari⁵

^{1*,2,3,4,5}Teknik Informatika, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

^{1*}dedysutris12@gmail.com, ²indra@amiktunasbangsa.ac.id, ³Sumarno@amiktunasbangsa.ac.id,

⁴Poningsih@amiktunasbangsa.ac.id, ⁵ikapurnamasari@amiktunasbangsa.ac.id

Received on 28/12/2022	Revised on 15/01/2023	Accepted on 27/01/2023
---------------------------	--------------------------	---------------------------

Abstract

Unemployment is a value that shows the number of working age population who are looking for work, or are preparing a business, or feel it is impossible to get a job, or already have a job but have not started working. The number of unemployed in Indonesia continues to increase along with the increase in the number of graduates from schools and universities. The greater the number of graduates from schools and universities, the higher the unemployment rate. This can be a burden on the economy. Because if the increase in the labor force is not matched by an increase in employment, it will cause unemployment problems. This condition can cause their welfare to decline. Therefore the problem of unemployment is also inseparable from the number of the workforce, this study will aim to get the unemployment rate grouping in the province of North Sumatra by using the K-Medoids algorithm, the grouping process with the K-Medoids algorithm uses a number of samples based on 34 districts, namely aged 25 years and over in North Sumatra and 3 clusters will be conducted. The cluster that has the most open unemployment characteristics is the Regency with the highest unemployment rate. For this grouping, it will later function to cluster which Regency/City is a concern for the Manpower Office to provide job opportunities. The grouping results were tested with rapidminer in grouping the data. The level of performance is indicated by the value of accuracy. The accuracy value is obtained by testing cluster results on training data and testing data. Comparison of the accuracy values between the algorithms used can be known the best algorithm in making the unemployment rate grouping. Based on the results of the grouping, there are three clusters, namely 14 districts in cluster zero, 2 districts in cluster one, and 17 districts in cluster two.

Keywords: Modeling, Open Unemployment Rate, K-medoids

Abstrak

Pengangguran adalah suatu nilai yang menunjukkan jumlah penduduk usia kerja yang sedang mencari pekerjaan, atau sedang mempersiapkan usaha, atau merasa tidak mungkin mendapatkan pekerjaan, atau sudah punya pekerjaan tetapi belum memulai bekerja. Jumlah pengangguran di Indonesia terus meningkat seiring dengan pertambahan jumlah lulusan dari sekolah maupun perguruan tinggi. Semakin besar jumlah lulusan dari sekolah maupun perguruan tinggi maka tingkat pengangguran jadi semakin besar. Hal itu dapat menjadi beban tersendiri bagi perekonomian. Karena jika meningkatnya angkatan kerja yang tidak diimbangi dengan bertambahnya lapangan kerja akan menyebabkan masalah pengangguran. Kondisi tersebut dapat menyebabkan kesejahteraannya menurun. Oleh karena itu permasalahan pengangguran juga tidak terlepas dari bagian jumlah angkatan kerja, Penelitian ini akan bertujuan untuk mendapatkan Pengelompokan tingkat pengangguran pada wilayah propinsi Sumatera Utara dengan menggunakan algoritma K-Medoids, Proses pengelompokan dengan algoritma K-Medoids ini menggunakan jumlah sample berdasarkan 34 Kabupaten yaitu umur 25 tahun keatas di Sumatera Utara dan akan di lakukan kluster sebanyak 3 kluster. Kluster yang mempunyai karakteristik pengangguran Terbuka terebanyak adalah Kabupaten dengan tingkat pengangguran yang tertinggi. Guna pengelompokan ini nantinya berfungsi untuk meng-cluster Kabupaten/Kota mana yang menjadi perhatian bagi dinas tenaga kerja untuk memberi lapangan pekerjaan. Hasil pengelompokan diuji dengan rapidminer dalam mengelompokkan data. Tingkat performance ditunjukkan dengan nilai akurasi. Nilai akurasi tersebut diperoleh dengan pengujian hasil klaster terhadap data training dan data testing. Perbandingan nilai akurasi antar algoritma yang digunakan dapat diketahui algoritma terbaik dalam membuat pengelompokan tingkat pengangguran. Berdasarkan Hasil pengelompokan terdapat tiga cluster, yaitu 14 kabupaten di cluster nol, 2 Kabupaten di cluster satu, dan 17 Kabupaten di Cluster dua.

Kata Kunci: Pemodelan, Tingkat Pengangguran terbuka , K-medoids

PENDAHULUAN

Jumlah pengangguran di Indonesia khususnya di Sumatera Utara mengalami peningkatan. Ada beberapa faktor yang mempengaruhi perkembangan pengangguran dari tahun ke tahun yang semakin meningkat. Pengangguran yang tinggi termasuk ke dalam masalah ekonomi dan sosial, Orang-orang yang menganggur suatu saat bisa kehilangan kepercayaan dirinya sehingga dapat menimbulkan tindakan kriminal, perselisihan dengan masyarakat dan sebagainya. Pengangguran yang sangat tinggi termasuk ke dalam pengalaman dan bidang pekerjaan tersebut, orang-orang pada suatu wilayah yang memiliki latar belakang pendidikan tinggi serta kesediaan lapangan pekerjaan yang tidak sesuai dengan bidang keilmuannya. Pengangguran yang sangat rendah termasuk ke dalam waktu

pekerjaan tersebut, kondisi ketika seseorang meninggalkan pekerjaan lama namun belum menemukan pekerjaan baru atau ketika lulusan baru yang sedang mencari pekerjaan. Pengangguran yang rendah termasuk ke keadaan dimana seseorang yang tergolong angkatan kerja namun tidak memiliki pekerjaan serta menurunkan tingkat kemiskinan, kondisi ketika seseorang mempunyai profesi namun termasuk ke dalam jenis pengangguran ini adalah penjual es, penjual jas hujan, karyawan di tempat wisata, dan buruh konstruksi karena pekerjaan tersebut termasuk pekerjaan musiman.

Masalah yang sering terjadi di beberapa wilayah di Indonesia termasuk di Provinsi Sumatera Utara, yang sangat berdampak pada masyarakat lainnya antara lain meningkatnya tindak kriminal sampai perselisihan

antara beberapa kelompok. Penyebab yang menimbulkan hal tersebut adalah tingkat pengangguran. Maka diperlukan pembangunan nasional guna mengurangi tingkat pengangguran tersebut. Upaya melaksanakan kegiatan pembangunan dengan tujuan mensejahterakan masyarakat perlu dilakukannya pengelompokan antara persentasi pengangguran berdasarkan pengelompokan umur dari 25 tahun keatas, untuk mempermudah dalam pembangunan ketenaga kerjaan. Hasil dari pengelompokan akan dilakukan perbandingan antara persentasi pengangguran berdasarkan umur. Pengelompokan persentasi pengangguran dilakukan menggunakan datamining dengan metode K-Medoids. Tujuan pengelompokan tersebut adalah untuk mengetahui Kabupaten/Kota mana yang tingkat penganggurannya paling tinggi, dilakukan teknik cluster menggunakan metode K-Medoids. Penerapan Algoritma K-Medoids dapat membantu dalam mengelompokkan data pengangguran berdasarkan tingkatan umur. Dengan demikian pihak instansi departemen ketenagakerjaan dan BUMN dapat memberi perhatian lebih terhadap tingkatan pengangguran yang ada di Indonesia terutama di Sumatera Utara, Agar kedepannya dapat mengurangi tingkat pengangguran.

Berdasarkan uraian diatas banyak cabang kecerdasan buatan dalam ilmu komputer yang dapat menyelesaikan permasalahan tersebut secara kompleks diantaranya sistem pendukung keputusan, sistem pakar, data mining dan lain sebagainya. Beberapa penelitian tentang data mining seperti:[1] Metode datamining dengan metode K-Means dapat diterapkan dalam perbandingan pengelompokan data tingkat pengangguran dan tenaga kerja. Dari hasil yang diperoleh dapat dilihat bahwa tingkat pengangguran lebih tinggi dibanding dengan tenaga kerja untuk Provinsi yang ada di Indonesia. Kemudian penelitain[2] Pada algoritma K-Means clustering ini dilakukan pengelompokan 19 Kabupaten/Kota di Provinsi Sumatera Barat berdasarkan indicator angkatan kerja pada tahun 2018 dan dibentuk cluster dengan 3 cluster. Cluster 0 terdiri dari 1 Kota, cluster 1 (satu) terdiri dari 11 Kabupaten/Kota, dan cluster 2 (dua) terdiri dari 7 Kabupaten/Kota. Nilai clustering diukur berdasarkan nilai DBI dan SI dengan nilai terbaik diperoleh menggunakan DBI, sehingga analisis data difokuskan berdasarkan validitas cluster terbaik dan dengan jumlah cluster 3. Kemudian Penelitian[3]Berdasarkan hasil analisis, diketahui bahwa cluster yang terbentuk menggunakan hierarchical clustering analysis ada sebanyak 3, yaitu cluster dengan TPK tinggi, menengah dan rendah berdasarkan IPM dan PDRB Kabupaten/Kota yang ada di Sulawesi selatan.

TINJAUAN PUSTAKA

2.1. Data Mining

Data Mining merupakan proses kegiatan untuk mengumpulkan data dalam jumlah yang besar untuk menemukan pengetahuan sehingga dapat menjadi informasi yang dapat digunakan. “Data mining adalah proses menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengetahui informasi yang bermanfaat yang terkait dari database yang besar. “Tujuan utama Data mining adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki”[4]

“Data Mining sering juga disebut Knowledge Discovery in Database (KDD) merupakan meliputi kegiatan pengumpulan, pemakaian data historis menemukan pola atau hubungan dalam set data berukuran besar” [5] Data mining merupakan salah satu bidang paling penting dalam penelitian yang bertujuan untuk memperoleh informasi dari data set. [6]

Data mining memiliki kelebihan sebagai alat analisis seperti berikut :

1. Data mining mampu menangani data dalam jumlah besar dan kompleks.
2. Data mining dapat menangani data dengan berbagai macam tipe atribut.
3. Data mining mampu mencari dan mengolah data secara otomatis. Disebut semi otomatis karena dalam beberapa teknik data mining, diperlukan parameter yang harus di input oleh user secara manual.
4. Data mining dapat menggunakan pengalaman ataupun kesalahan terdahulu untuk meningkatkan kualitas dan hasil analisa sehingga mendapat hasil yang terbaik.

2.2. Clustering

Clustering merupakan proses partisi sekumpulan objek data dari satu set menjadi beberapa kelas. Hal ini dapat dilakukan dengan menerapkan berbagai persamaan dan langkah-langkah mengenai jarak algoritma yaitu dengan Euclidean Distance. Analisis cluster ialah metode yang dipakai untuk membagi rangkaian data menjadi beberapa grup berdasarkan kesamaan-kesamaan yang telah ditentukan sebelumnya. Dalam menentukan cluster berdasarkan data yang telah tersedia dibutuhkan sebuah flowchart untuk memudahkan dalam menentukan alur perhitungan sebagai alur untuk menemukan hasil dari penerapan cluster terhadap data yang akan diproses.[7]

Menurut [8]” Strategi dasar dari algoritma k-medoids adalah untuk menemukan cluster k pada objek n dengan terlebih dahulu menemukan objek awal (medoid) secara acak sebagai perwakilan untuk setiap cluster. Setiap objek yang tersisa dikelompokkan dengan medoid yang paling mirip. Metode k-medoids menggunakan objek representatif sebagai titik yang menjadi acuan dan bukan rata – rata objek dari setiap cluster adalah kunci dari metode ini

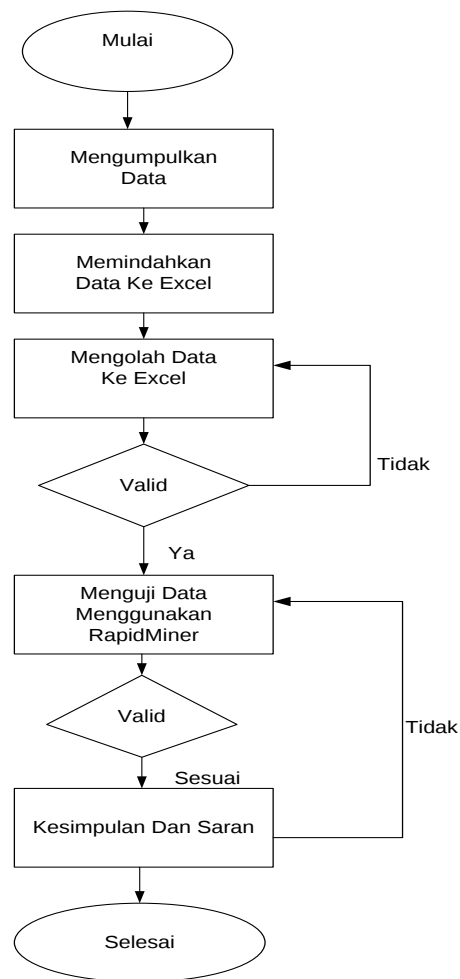
2.3. RapidMiner

Rapid Miner merupakan perangkat lunak yang dibuat oleh Dr. Markus Hofmann dari Institute of Teknologi Blanchardstown dan Ralf Klinkenberg dari rapid-i.com dengan tampilan GUI (Graphical User Interface) sehingga memudahkan pengguna dalam menggunakan perangkat lunak ini. Perangkat lunak ini bersifat open source dan dibuat dengan menggunakan program Java di bawah lisensi GNU Public Licence dan Rapid Miner dapat dijalankan di sistem operasi manapun.[9]

METODE PENELITIAN

3.1. Rancangan Penelitian

Rancangan penelitian ini pertama kali penulis melakukan pengamatan selanjutnya mengumpulkan data setelah itu data tersebut dipindahkan ke excel (xls) setelah itu data akan diolah melalui proses perhitungan dan mengikuti langkah-langkah proses perhitungan metode *K-Medoids* Selanjutnya hasil perhitungan tersebut akan diaplikasikan ke *RapidMiner* untuk melihat keakuratan hasil yang diperoleh. Flowchart Rancangan Penelitian dapat dilihat pada gambar 1



Gambar 1. Flowchart Penelitian

3.2. Prosedur Pengumpulan Data

Prosedur pengumpulan data yang penulis lakukan dalam penelitian tersebut diantaranya sebagai berikut:

1. Penelitian Kepustakaan (*Library Research*) yaitu memanfaatkan perpustakaan, buku, prosiding atau jurnal sebagai media untuk bahan referensi dalam menentukan parameter yang digunakan dalam penelitian.

2. Sumber data penelitian dari <https://sumut.bps.go.id/> yang digunakan dalam penelitian ini adalah data Pengangguran Berdasarkan Kabupaten/Kota di Provinsi Sumatera Utara dari tahun 2016 sampai 2020.

3.3. Anaisa Data

Proses analisis data dapat dilakukan setelah adanya pengumpulan data dan setelah data dikumpulkan dan diolah ke excel. Setelah hasil yang telah dapat di excel akan diaplikasikan ke *RapidMiner* untuk menyesuaikan hasil yang didapatkan. Penulis akan melakukan analisa data statistik deskriptif. Data deskriptif adalah metode-metode yang berkaitan dengan pengumpulan data penyajian suatu gugus data sehingga memberikan informasi yang berguna informasi yang dapat diperoleh dari statistika deskriptif ini. Jenis data yang digunakan dalam penelitian ini merupakan data sekunder, Data sekunder adalah data yang diperoleh tidak dari sumbernya langsung melainkan sudah dikumpulkan oleh pihak lain dan sudah di olah serta memiliki keterkaitan dengan permasalahan yang di teliti. Data penelitian dapat dilihat pada tabel 1

Tabel 1. Data Penelitian

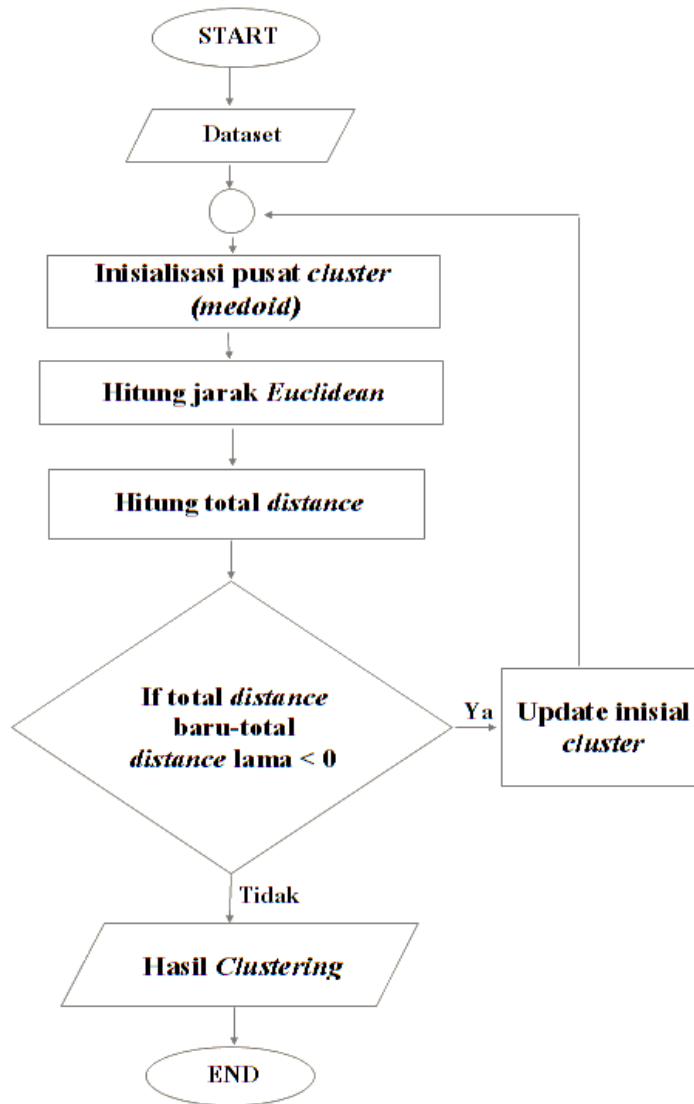
Tingkat Pengangguran Terbuka (TPT) Penduduk Umur 25 Tahun Keatas Menurut Kab/Kota (Persen)					
Kabupaten/Kota	2016	2017	2018	2019	2020
Nias	0.92	1.19	1.62	1.09	3.49
Mandailing Natal	5.78	5.75	4.43	6.37	6.50
Tapanuli Selatan	5.91	5.80	5.28	4.17	4.42
Tapanuli Tengah	4.98	7.39	6.38	7.26	7.54
Tapanuli Utara	2.56	1.89	1.42	1.33	2.94
Toba Samosir	3.47	2.18	2.15	1.26	2.50
Labuhan Batu	11.39	7.09	6.98	5.70	6.05
Asahan	5.82	5.95	5.26	6.86	7.24
Simalungun	5.75	5.62	5.10	4.39	4.58
Dairi	1.26	1.42	1.69	1.58	1.75
Karo	2.23	1.34	1.50	1.09	1.83
Deli Serdang	6.38	6.16	7.06	5.74	9.50
Langkat	8.02	3.57	4.67	5.30	7.02
Nias Selatan	0.40	1.28	3.77	2.25	4.15
Humbang Hasundutan	1.22	0.31	0.34	0.33	0.84
Pakpak Barat	2.88	0.49	0.43	0.19	1.93
Samosir	1.28	1.28	1.35	1.25	1.20
Serdang Bedagai	7.18	5.98	5.10	4.37	5.54
Batu Bara	6.32	5.00	5.39	6.69	6.48
Padang Lawas Utara	5.01	3.21	3.15	3.21	3.11
Padang Lawas	5.95	4.24	4.10	4.24	4.11
Labuhanbatu Selatan	4.15	5.68	4.79	4.80	4.90
Labuhanbatu Utara	8.75	6.35	5.67	5.84	6.82
Nias Utara	4.02	2.67	2.40	3.07	4.54
Nias Barat	2.96	1.23	1.23	1.63	1.71

Sibolga	10.25	9.29	8.61	7.40	8.00
Tanjungbalai	10.06	5.50	5.58	6.82	6.97
Pematangsiantar	9.47	8.80	12.14	11.09	11.50
Tebing Tinggi	10.46	9.73	7.23	8.60	9.98
Medan	11.00	9.46	8.25	8.53	10.74
Binjai	10.00	5.95	7.40	6.14	8.67
Padangsidempuan	6.96	3.78	5.18	4.34	7.45
Gunungsitoli	10.00	6.00	5.92	5.59	5.94

<https://sumut.bps.bps.go.id/>.

3.4. Pemodelan Metode

Setiap metode atau algoritma memiliki pemodelannya masing-masing sesuai pengerjaan, pada penelitian ini penulis menggunakan metode *K-Medoids* untuk mengelompokkan Tingkat Pengangguran Berdasarkan Kabupaten/Kota di Provinsi Sumatera Utara. Untuk pemodelan metode *K-Medoids* dapat dilihat pada Gambar 2



Gambar 2. Flowcaht Metode *K-Medoids*

Pada Gambar 2 menjelaskan Tentang langkah-langkah metode *K-Medoids clustering* berdasarkan bagan alir (*flowchart*) dalam mengelompokkan data tingkat pengangguran berdasarkan Kabupaten/Kota di provinsi Sumatera Utara yang terdiri dari:

1. Menentukan jumlah data yang akan di *cluster*, dimana sampel data jumlah pengangguran umur 25 tahun keatas yang digunakan dalam proses *cluster* tahun 2016-2020 yang terdiri dari 33 Kabupaten/Kota.
2. Menetapkan nilai *k* jumlah *cluster* data sebanyak 3 *cluster* ($k = 3$)
3. Menentukan nilai *centroid* (pusat *cluster*) awal yang ditentukan secara random berdasarkan nilai variabel data yang di *cluster* sebanyak *k* yang ditentukan sebelumnya.
4. Pilih secara acak objek pada masing-masing *cluster* sebagai kandidat *medoid* baru.
5. Hitung jarak setiap objek yang berada pada masing-masing *cluster* dengan kandidat *medoid* baru.

6. Hitung total simpangan (S) dengan menghitung nilai total *distance* baru–total *distance* lama. Jika $S < 0$, maka tukar objek dengan data *cluster* untuk membentuk sekumpulan k objek baru sebagai *medoid*.
7. Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan *medoid*, sehingga didapatkan *cluster* beserta anggota *cluster* masing-masing.

HASIL DAN PEMBAHASAN

Pada tahapan ini dilakukan analisis data Tingkat Pengangguran di Provinsi Sumatera Utara dengan tool *Rapidminer*. *Rapidminer* adalah sebuah *software* pilihan untuk melakukan ekstraksi data dengan metode-metode *data mining*. Berikut adalah data Pengangguran di Provinsi Sumatera Utara umur 25 tahun keatas setelah di hitung berdasarkan Hasil Persentase.

4.1. Menentukan Data Yang Akan Diolah

Berikut adalah data Jumlah Pengangguran berdasarkan Kabupaten/Kota yang diambil penulis melalui BPS Sumatera Utara. Kita lihat pada tabel 2

Tab 2. Data Pengangguran

Kabupaten	2016	2017	2018	2019	2020
Nias	211	156	130	241	867
Mandailing Natal	5083	6803	916	828	9616
Tapanuli Selatan	6733	7693	3669	4054	18552
Tapanuli Tengah	7383	6223	4521	1192	2739
Tapanuli Utara	12709	17484	14780	18030	33395
Toba Samosir	17846	14095	12782	11189	21969
Labuhan Batu	3430	942	2463	3305	2773
Asahan	17964	9047	7710	10625	13461
Simalungun	383796	284956	324434	381685	382309
Dairi	131877	129613	121647	259033	217003
Karo	486293	424210	413346	553208	52187
Deli Serdang	72119	72307	88009	74324	10775
Langkat	130618	161352	125193	147368	156583
Nias Selatan	1453	821	1390	1210	6314
Humbang Hasundutan	1787	1298	1362	2525	6070
Pakpak Bharat	9428	4074	6992	9070	7418
Samosir	6083	7885	4947	7008	7511
Serdang Bedagai	21040	13161	9395	15785	16386
Batu Bara	2996	539	483	1432	6732
Padang Lawas Utara	1679	1498	405	644	727
Padang Lawas	2958	1751	1497	1260	1286

Labuhanbatu Selatan	1274	1194	912	572	1388
Labuhanbatu Utara	3575	2588	2254	305	1166
Nias Utara	494	161	192	185	253
Nias Barat	251	87	91	73	73
Sibolga	112	126	164	62	52
Tanjungbalai	11224	6890	3823	5653	4164
Pematangsiantar	149	126	142	130	142
Tebing Tinggi	892	733	681	1415	1612
Medan	4015	3697	4038	5842	5412
Binjai	10224	6790	3714	5872	4164
Padangsidempuan	72519	711307	88882	74165	10652
Gunungsitoli	10211	6245	3164	5852	4773

Langkah- Langkah perhitungan metode *K-Medoids Clustering*

1. Menentukan Jumlah *cluster* dan *Centorid*

Data tersebut akan masuk ke tahapan *clustering* dengan menerapkan algoritma *K-Medoids* menggunakan *rapidminer* untuk mengcluster data menjadi dua *cluster*. Data yang sudah dipilih akan dimasukkan kedalam *tool rapidminer*. Kemudian Dalam penerapan algoritma *K-Medoids* dihasilkan nilai titik tengah atau *centroid* dari data yang didapat dengan ketentuan bahwa clusterisasi yang diinginkan adalah 3. Penentuan *cluster* dibagi atas 3 bagian yakni *cluster* tingkat menengah (C0), *cluster* tingkat tinggi (C1) dan *cluster* tingkat rendah (C2). Penentuan titik *cluster* ini dilakukan dengan mengambil nilai menengah (*sedang*) untuk tingkat menengah (C0), nilai terbesar (*maksimum*) untuk *cluster* tingkat tinggi (C1) dan nilai terkecil (*minimum*) untuk *cluster* tingkat rendah (C2). Nilai titik tersebut dapat diketahui pada tabel 3

Tabel 3 Tabel nilai Centroid

Nilai Tertinggi	486293	42410	413346	553208	382309
Nilai Menengah	6083	6223	3669	4054	6070
Nilai Terendah	112	87	91	62	52

2. Menghitung Jarak dari *Centroid*

Untuk menghitung jarak antara titik *Centroid* dengan titik tiap objek menggunakan *Euclidian Distance*. Rumus untuk menghitung jarak dari *Centroid* adalah :

$$De = \sqrt{(xi - si)^2 + (yi - ti)^2}$$

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

$$D_{A1,c1} = \sqrt{(211 - 486293)^2 + (156 - 42410)^2 + (130 - 413346)^2 + (241 - 553208)^2 + (867 - 382309)^2}$$

$$= 1018879,517$$

$$D_{A2,c2} = \sqrt{(5083 - 486293)^2 + (6803 - 42410)^2 + (916 - 413346)^2 + (828 - 553208)^2 + (9616 - 382309)^2}$$

$$= 1009907,872$$

Dan seterusnya sampai dengan $D_{A33,c1}$. Selanjutnya perhitungan untuk jarak dari *Centroid* ke-2 adalah sebagai berikut:

$$D_{A1,c1} = \sqrt{(211 - 6083)^2 + (156 - 6223)^2 + (130 - 3669)^2 + (241 - 4054)^2 + (867 - 6070)^2}$$

$$= 11199,26658$$

$$D_{A2,c2} = \sqrt{(5083 - 6083)^2 + (6803 - 6223)^2 + (916 - 3669)^2 + (828 - 4054)^2 + (9616 - 6070)^2}$$

$$= 5647,707588$$

Dan seterusnya sampai dengan $D_{A33,c2}$. Sehingga didapat tabel jarak dari *Centroid* dan mencari nilai minimal dari kedua *centroid*. Tabel jarak dari *centroid* ada pada tabel 4

Tabel 4. Tabel nilai Centroid

Kabupaten	C0	C1	C2	Jarak Terpendek
Nias	724129,31	846040,38	37777,31	37777,31
Mandailing Natal	716903,60	837894,12	29015,94	29015,94
Tapanuli Selatan	715261,21	831494,14	23000,13	23000,13
Tapanuli Tengah	716835,97	836911,70	29703,01	29703,01
Tapanuli Utara	702913,56	808191,82	15278,98	15278,98
Toba Samosir	706357,73	814312,21	9653,12	9653,12
Labuhan Batu	722430,23	840861,84	32742,89	32742,89
Asahan	711857,79	820985,56	14450,57	14450,57
Simalungun	752906,21	209882,90	752615,82	209882,90
Dairi	647831,86	497078,85	366353,57	366353,57
Karo	768281,82	209882,90	912167,23	209882,90
Deli Serdang	639000,47	698187,08	121302,56	121302,56
Langkat	577736,87	543628,94	285752,82	285752,82
Nias Selatan	723047,52	842659,49	33627,00	33627,00
Humbang Hasundutan	722419,38	841633,73	32649,64	32649,64
Pakpak Bharat	717687,94	830128,77	22239,30	22239,30
Samosir	714663,20	832259,09	23578,30	23578,30
Serdang Bedagai	706950,71	813356,64	8670,29	8670,29
Batu Bara	723259,85	842153,95	33182,23	33182,23
Padang Lawas Utara	722593,22	844418,02	36357,72	36357,72
Padang Lawas	722017,74	842695,80	34727,45	34727,45
Labuhanbatu Selatan	722868,37	844402,62	36175,32	36175,32

Labuanbatu Utara	721143,64	842259,45	34468,12	34468,12
Nias Utara	724103,00	846054,04	37955,69	37955,69
Nias Barat	724226,32	846361,90	38273,07	38273,07
Sibolga	724194,42	846396,65	38305,06	38305,06
Tanjungbalai	715467,64	832133,51	25256,28	25256,28
Pematangsiantar	724185,12	846326,58	38218,28	38218,28
Tebing Tinggi	723298,87	844369,78	36071,98	36071,98
Medan	719208,62	836646,64	28127,29	28127,29
Binjai	715644,45	832611,40	25445,52	25445,52
Padangsidempuan	200,00	730995,38	703499,03	200,00
Gunungsitoli	716244,09	832942,66	25557,57	25557,57

3. Menentukan *Cluster* atau Pengelompokan

Dalam menentukan *Cluster* dengan nilai *Cluster* berdasarkan nilai yang sama dari nilai *Cluster* dan diletakkan pada *Cluster* yang sesuai dengan nilai sama pada Iterasi 1. Selanjutnya dalam metode *K-Medoids*, perhitungan berhenti apabila *Cluster* pada iterasi yang dihasilkan sama pada iterasi sebelumnya. Maka selanjutnya mencari *Cluster* pada iterasi selanjutnya sampai nilai iterasinya sama. Untuk mencari nilai *Centroid* selanjutnya dengan menggunakan *Centroid* baru pada Iterasi ke-1 dengan menjumlahkan nilai sesuai yang tertera pada *Cluster* di tabel diatas. Adapun *Centroid* baru untuk mencari *Cluster* selanjutnya adalah dengan menjumlahkan nilai yang terpilih pada *Cluster* tersebut kemudian membagikannya sebanyak jumlah nilai sebagai berikut :

$$C_{1a} = (383796 + 486293)/2 = 435044,5$$

$$C_{1b} = (284956 + 42410)/2 = 709166$$

$$C_{1c} = (324434 + 413346)/2 = 368890$$

$$C_{1d} = (381865 + 553208)/2 = 934839$$

$$C_{1e} = (382309 + 52187)/2 = 217248$$

Maka, data *Centroid* baru Iterasi ke-2 adalah sebagai berikut : pada tabel 5

Tabel 5. Tabel nilai *Centroid* baru

C1	435044,5	354583	368890	467446,5	217248
C2	15175,8571	58671	12095,2143	12511,786	11443,7143
C3	20933,8235	22076	20412,6471	29063,765	24342,1176

Sehingga didapat tabel jarak dari *Centroid* dan mencari nilai minimal dari kedua *centroid*. Tabel jarak dari *centroid* adalah sebagai berikut : pada tabel 6

Tabel 6. Tabel nilai Iterasi Ke 2

Kabupaten	C0	C1	C2	Jarak Terpendek
Nias	11199,26658	1167916,701	822,8086047	822,8086047
Mandailing Natal	5647,707588	1158452,067	11716,71929	5647,707588
Tapanuli Selatan	12585,05955	1151857,183	20685,51324	12585,05955
Tapanuli Tengah	4658,595175	1158633,318	8086,155514	4658,595175
Tapanuli Utara	35158,96072	1128132,744	44165,25724	35158,96072
Toba Samosir	24229,52224	1135617,979	30980,19159	24229,52224
Labuhan Batu	6914,695655	1163202,835	4900,642509	4900,642509
Asahan	16225,52742	1143453,927	20698,92304	16225,52742
Simalungun	779365,3432	479172,5091	691421,6031	479172,5091
Dairi	393033,1436	814206,3139	381656,5238	381656,5238
Karo	936401,0434	437498,8255	811971,3965	437498,8255
Deli Serdang	144225,8433	1026193,288	136263,368	136263,368
Langkat	312306,1338	853021,5072	296376,7699	296376,7699
Nias Selatan	7997,513176	1164320,385	6522,488712	6522,488712
Humbang Hasundutan	7097,276309	1163356,013	6712,327838	6712,327838
Pakpak Bharat	7336,682834	1152966,497	14075,2389	7336,682834
Samosir	3898,477267	1153690,547	13688,47409	3898,477267
Serdang Bedagai	23423,72699	1135679,411	27743,33399	23423,72699
Batu Bara	7700,723927	1163932,057	6828,545233	6828,545233
Padang Lawas Utara	9620,30592	1166246,806	1665,190079	1665,190079
Padang Lawas	8073,108757	1164700,459	2746,245073	2746,245073
Labuhanbatu Selatan	9490,195941	1166238,854	1960,533091	1960,533091
Labuhanbatu Utara	7720,733838	1164159,626	3477,611393	3477,611393
Nias Utara	11352,57433	1168001,179	247,885054	247,885054
Nias Barat	11673,78148	1168294,953	61,56297589	61,56297589
Sibolga	11715,95762	1168315,307	98,84331035	98,84331035
Tanjungbalai	5752,227655	1154301,984	10376,12317	5752,227655
Pematangsiantar	11634,62307	1168246,052	117,6690274	117,6690274
Tebing Tinggi	9636,016293	1166287,985	2212,839127	2212,839127
Medan	3797,81108	1158662,119	9495,358392	3797,81108
Binjai	4940,582456	1154709,751	10392,29917	4940,582456
Padangsidempuan	716767,0414	802556,4976	720612,2316	716767,0414
Gunungsitoli	4712,844788	1155053,155	10122,5967	4712,844788

Adapun *Centroid* baru untuk mencari *Cluster* selanjutnya adalah dengan menjumlahkan nilai yang terpilih pada *Cluster* tersebut kemudian membagikannya, Maka, data *Centroid* baru Iterasi ke-3 adalah sebagai berikut : pada tabel 7

Tabel 7. Tabel nilai Centroid baru

C1	435044,5	354583	368890	467446,5	217248
C2	72519	711307	88652	74165	10775
C3	16527,27	16179,3	14248,87	19836,13	18779,2

Sehingga didapat tabel jarak dari *Centroid* dan mencari nilai minimal dari kedua *centroid*. Tabel jarak dari *centroid* adalah sebagai berikut : pada tabel 8

Tabel 8. Tabel nilai Iterasi Ke 3

Kabupaten	C0	C1	C2	Jarak Terpendek
Nias	11199,26658	1167916,701	822,8086047	822,8086047
Mandailing Natal	5647,707588	1158452,067	11716,71929	5647,707588
Tapanuli Selatan	12585,05955	1151857,183	20685,51324	12585,05955
Tapanuli Tengah	4658,595175	1158633,318	8086,155514	4658,595175
Tapanuli Utara	35158,96072	1128132,744	44165,25724	35158,96072
Toba Samosir	24229,52224	1135617,979	30980,19159	24229,52224
Labuhan Batu	6914,695655	1163202,835	4900,642509	4900,642509
Asahan	16225,52742	1143453,927	20698,92304	16225,52742
Simalungun	779365,3432	479172,5091	691421,6031	479172,5091
Dairi	393033,1436	814206,3139	381656,5238	381656,5238
Karo	936401,0434	437498,8255	811971,3965	437498,8255
Deli Serdang	144225,8433	1026193,288	136263,368	136263,368
Langkat	312306,1338	853021,5072	296376,7699	296376,7699
Nias Selatan	7997,513176	1164320,385	6522,488712	6522,488712
Humbang Hasundutan	7097,276309	1163356,013	6712,327838	6712,327838
Pakpak Bharat	7336,682834	1152966,497	14075,2389	7336,682834
Samosir	3898,477267	1153690,547	13688,47409	3898,477267
Serdang Bedagai	23423,72699	1135679,411	27743,33399	23423,72699
Batu Bara	7700,723927	1163932,057	6828,545233	6828,545233
Padang Lawas Utara	9620,30592	1166246,806	1665,190079	1665,190079
Padang Lawas	8073,108757	1164700,459	2746,245073	2746,245073
Labuhanbatu Selatan	9490,195941	1166238,854	1960,533091	1960,533091

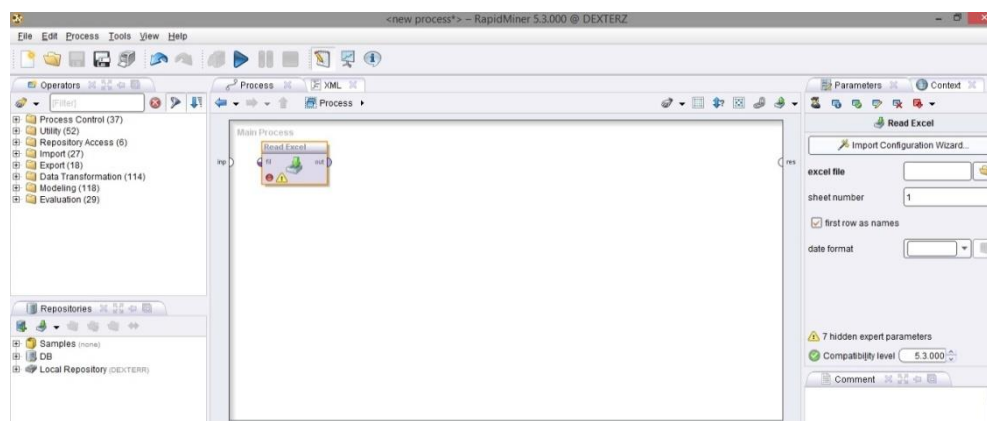
Labuhanbatu Utara	7720,733838	1164159,626	3477,611393	3477,611393
Nias Utara	11352,57433	1168001,179	247,885054	247,885054
Nias Barat	11673,78148	1168294,953	61,56297589	61,56297589
Sibolga	11715,95762	1168315,307	98,84331035	98,84331035
Tanjungbalai	5752,227655	1154301,984	10376,12317	5752,227655
Pematangsiantar	11634,62307	1168246,052	117,6690274	117,6690274
Tebing Tinggi	9636,016293	1166287,985	2212,839127	2212,839127
Medan	3797,81108	1158662,119	9495,358392	3797,81108
Binjai	4940,582456	1154709,751	10392,29917	4940,582456
Padangsidempuan	716767,0414	802556,4976	720612,2316	716767,0414
Gunungsitoli	4712,844788	1155053,155	10122,5967	4712,844788

Dari tabel *Cluster* Iterasi ke-2 dan tabel *Cluster* Iterasi ke-3 memiliki nilai *Cluster* yang sama atau tidak berubah pada *cluster* terakhir maka perhitungan dihentikan dan hasil yang diperoleh yaitu :

1. *Cluster* 0 (*C0*) mendapatkan 14 data Kabupaten/Kota dalam artian bahwa kelompok pertama termasuk kategori Kabupaten/Kota dengan tingkat pengangguran Menengah Pada Tahun 2016 - 2020.
2. *Cluster* 1 (*C1*) mendapatkan 2 data Kabupaten/Kota dalam artian bahwa kelompok pertama termasuk kategori Kabupaten/Kota dengan tingkat pengangguran Tertinggi Pada Tahun 2016 - 2020.
3. *Cluster* 2 (*C2*) mendapatkan 17 data Kabupaten/Kota dalam artian bahwa kelompok pertama termasuk kategori Kabupaten/Kota dengan tingkat pengangguran Terendah Pada Tahun 2016 - 2020.

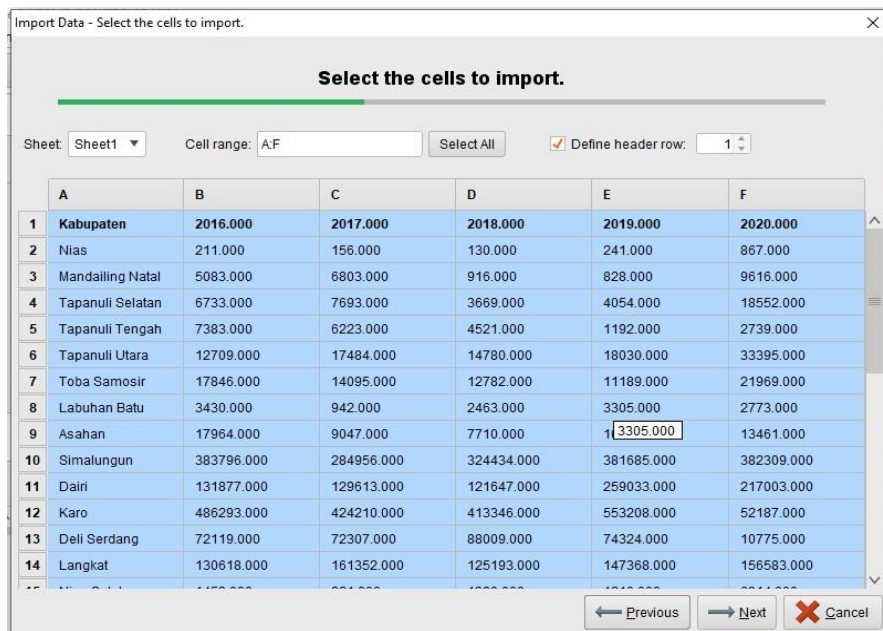
4.2. Pemodelan *Rapidminer*

Dalam menjalankan proses untuk membentuk pemodelan, tahap pertama yang dilakukan adalah *importing* data dari data yang sudah di transformasi terlebih dahulu ke dalam *microsoft excel* dengan format *.xls* agar dapat diakses menggunakan *software Rapidminer*. Atribut yang digunakan sebagai label adalah data pengangguran. Dalam melakukan *importing* data, dibutuhkan operator baru seperti gambar 3. *New Process* untuk *Import Data Excel* sebagai berikut :



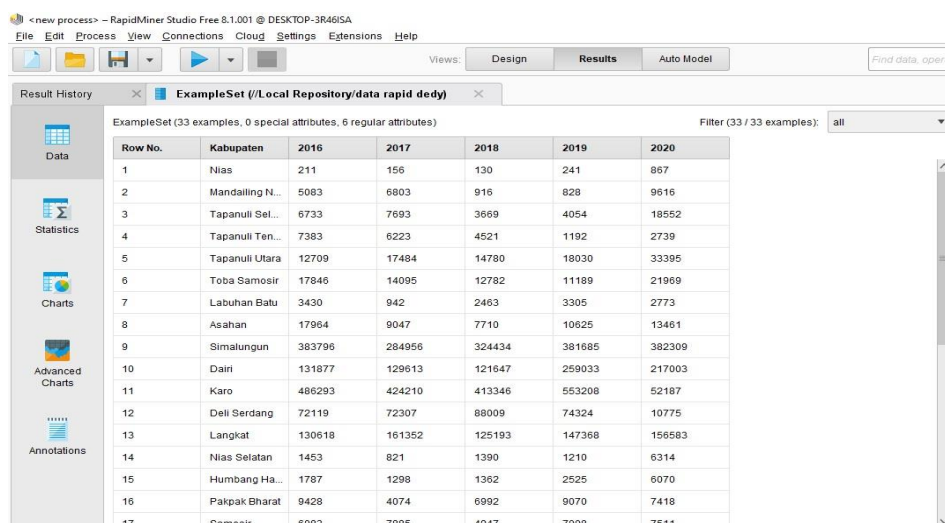
Gambar 3. New Process untuk Import Data Excel

Pilih data yang akan digunakan untuk di *import*. Lakukan *drag and drop* data yang dipilih kedalam *main process*. Berikut alur proses *import* data yang dilakukan. Pada *import* data wizard step 1, pilih data berformat .csv pada folder yang akan di *import*. Kita lita pada gambar 4



Gambar 4. Alur Proses Import Data Step 2

Pada data *import* wizard step 2, data pengangguran.csv dengan *import excel* dimasukan. File *excel* dapat berisi beberapa lembar, cukup menandai beberapa sel yang akan dimuat. Pada proses kali ini, peneliti memilih sel 1 untuk dilanjutkan prosesnya kemudian klik *next*. Lihat pada gambar 5



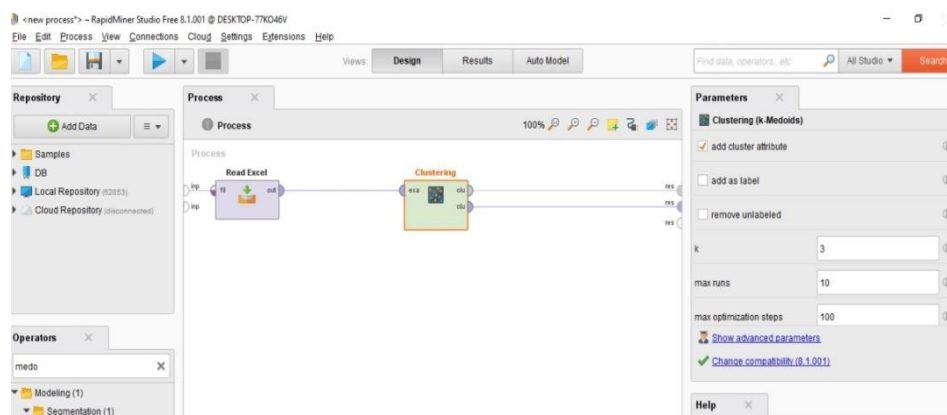
Gambar 5. Alur Proses *Import Data Step 3*

Pada *step* ke 3 ini tidak ada dilakukan apapun, setiap atribut dapat dianotasi. Yang paling penting pada *step* 3 ini adalahnya namanya mengandung atribut maka dari itu langsung ke data *import wizard step 4* dengan cara klik *Next* maka akan muncul form data *import wizard step 4*. Pada gambar 6

Kabupaten	2016	2017	2018	2019	2020
polynomial	integer	integer	integer	integer	integer
id	attribute	attribute	attribute	attribute	attribute
Nias	211	156	130	241	867
Mandailing Natal	5083	6803	916	828	9616
Tapanuli Selatan	6733	7693	3669	4054	18552
Tapanuli Tengah	7383	6223	4521	1192	2739
Tapanuli Utara	12709	17484	14780	18030	33395
Toba Samosir	17846	14095	12782	11189	21969
Labuhan Batu	3430	942	2463	3305	2773
Asahan	17964	9047	7710	10625	13461
Simalungun	383796	284956	324434	381685	382309
Dairi	131877	129613	121647	259033	217003

Gambar 6. Alur Proses *Import Data Step 4*

Langkah berikutnya *drag* dan *drop* operator *K-Medoids* lalu hubungkan pada *output*. Seperti gambar 7 berikut :



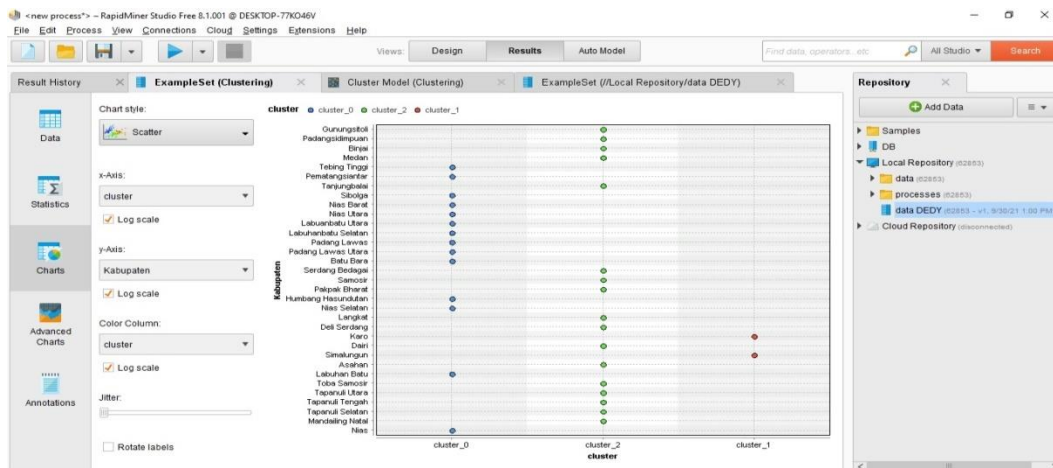
Gambar 7. *Drag* dan *Drop* Operator *K-Medoids*

Pada saat *tool run* diklik akan muncul pohon keputusan seperti gambar dibawah. Pada gambar 8 Data Hasil *Clustering*



Gambar 8. Data Hasil *Clustering*

Sehingga berdasarkan data hasil *clustering* pada gambar 4.6. didapatkan grafik hasil dari rapidminer seperti gambar 9



Gambar 9. Grafik *clustering*

ExampleSet (33 examples, 3 special attributes, 5 regular attributes)

Filter (33 / 33 examples)

Row ...	id	Kabupaten	cluster ↑	2016	2017	2018	2019	2020
1	1	Nias	cluster_0	211	156	130	241	867
7	7	Labuhan Batu	cluster_0	3430	942	2463	3305	2773
14	14	Nias Selatan	cluster_0	1453	821	1390	1210	6314
15	15	Humbang Hasundutan	cluster_0	1787	1298	1362	2525	6070
19	19	Batu Bara	cluster_0	2996	539	483	1432	6732
20	20	Padang Lawas Utara	cluster_0	1679	1498	405	644	727
21	21	Padang Lawas	cluster_0	2958	1751	1497	1260	1286
22	22	Labuhanbatu Selatan	cluster_0	1274	1194	912	572	1388
23	23	Labuanbatu Utara	cluster_0	3575	2588	2254	305	1166
24	24	Nias Utara	cluster_0	494	161	192	185	253
25	25	Nias Barat	cluster_0	251	87	91	73	73
26	26	Sibolga	cluster_0	112	126	164	62	52
28	28	Pematangsiantar	cluster_0	149	126	142	130	142
29	29	Tebing Tinggi	cluster_0	892	733	681	1415	1612

Gambar 10. Pengelompokan Berdasarkan Kabupaten

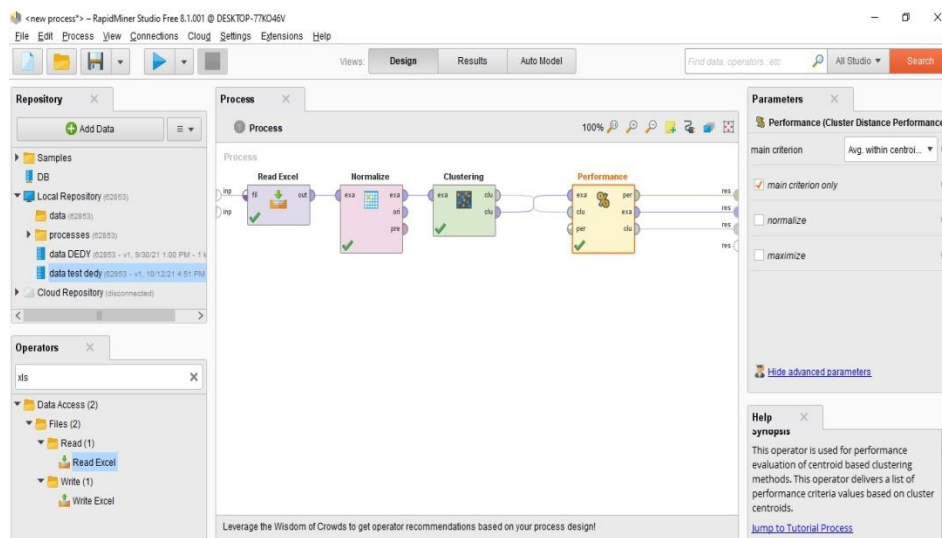
ExampleSet (33 examples, 3 special attributes, 5 regular attributes)

Filter (33 / 33 examples)

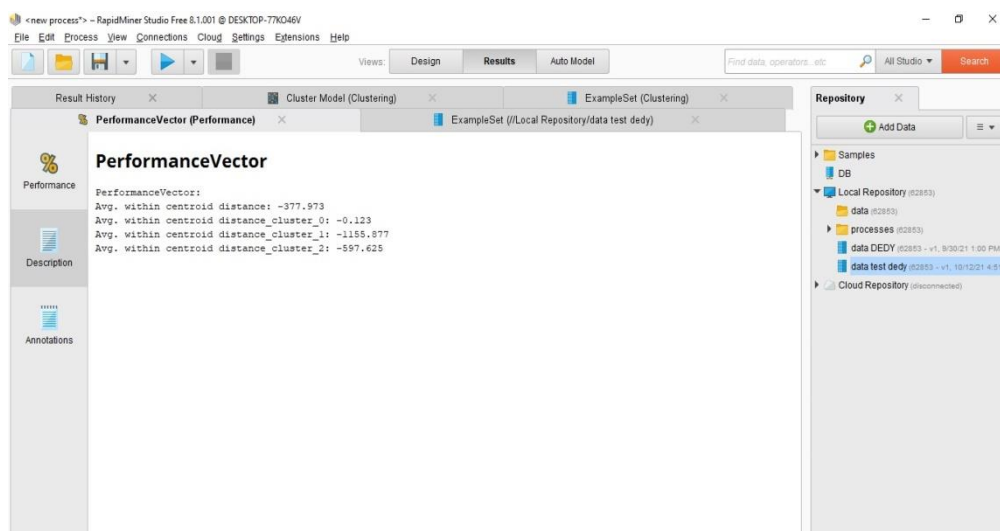
Row ...	id	Kabupaten	cluster ↑	2016	2017	2018	2019	2020
9	9	Simalungun	cluster_1	383796	284956	324434	381685	382309
11	11	Karo	cluster_1	486293	424210	413346	553208	52187
2	2	Mandailing Natal	cluster_2	5083	6803	916	828	9616
3	3	Tapanuli Selatan	cluster_2	6733	7693	3669	4054	18552
4	4	Tapanuli Tengah	cluster_2	7383	6223	4521	1192	2739
5	5	Tapanuli Utara	cluster_2	12709	17484	14780	18030	33395
6	6	Toba Samosir	cluster_2	17846	14095	12782	11189	21969
8	8	Asahan	cluster_2	17964	9047	7710	10625	13461
10	10	Dairi	cluster_2	131877	129613	121647	259033	217003
12	12	Deli Serdang	cluster_2	72119	72307	88009	74324	10775
13	13	Langkat	cluster_2	130618	161352	125193	147368	156583
16	16	Pakpak Bharat	cluster_2	9428	4074	6992	9070	7418
17	17	Samosir	cluster_2	6083	7885	4947	7008	7511
18	18	Serdang Bedagai	cluster_2	21040	13161	9395	15785	16386
27	27	Tanjungbalai	cluster_2	11224	6890	3823	5653	4164
30	30	Medan	cluster_2	4015	3697	4038	5842	5412
31	31	Binjai	cluster_2	10224	6790	3714	5872	4164

Gambar 11. Pengelompokan Berdasarkan Kabupaten

Berdasarkan gambar yang telah dijelaskan di atas, maka dapat diketahui bahwa pengujian data yang dilakukan dengan menggunakan *Performance Distance* memiliki nilai Akurasi seperti pada gambar 12 dan 13



Gambar 12. Proses Mencari nilai Performance



Gambar 13. Nilai Performance

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan pembahasan pada bab sebelumnya dapat disimpulkan bahwa:

1. Metode yang digunakan dalam perhitungan skripsi ini adalah penerapan *data mining* dengan metode *K-Medoids Clustering*.
2. Hasil dari penelitian menggunakan metode *K-Medoids* ini berupa 3 (tiga) buah *cluster* yaitu *cluster* tinggi, *cluster* sedang dan *cluster* rendah.
3. Penerapan Algoritma *K-Medoids* dapat diimplementasikan terhadap tingkat pengangguran tertinggi di Sumatera Utara.

5.2. Saran

Berdasarkan Penelitian tentang penerapan algoritma K-Medoids Clustering ini, ada beberapa saran untuk pengembangan penelitian ini kedepannya, diantaranya :

1. Kepada Pihak Dinas Ketenagakerjaan di Provinsi Sumatera Utara ada baiknya untuk menerapkan sistem seperti yang dirancang oleh penulis dalam penelitian ini. Hal ini diperlukan untuk memberikan solusi bagi Kabupaten/Kota yang memiliki angka pengangguran tertinggi.
2. Ada baiknya sistem dikembangkan berbasis android, sehingga dapat lebih banyak menyediakan informasi dan lebih mudah diakses oleh pihak Kemnaker

DAFTAR PUSTAKA

- [1] A. R. Damanik *et al.*, “Implementasi Metode K-Means Dalam Perbandingan Tingkat Pengangguran Dengan Tenaga Kerja Berdasarkan Provinsi,” *Pros. Semin. Nas. Ris. Dan Inf. Sci.*, vol. 2, pp. 24–31, 2020, [Online]. Available: <https://www.bps.go.id>.
- [2] D. Safira, E. D. Lestari, M. Iffa, and S. Annisa, “Berdasarkan Angkatan Kerja Menggunakan Algoritma K-Means,” vol. 6, no. 1, pp. 26–31, 2020.
- [3] A. Akramunnisa and F. Fajriani, “Hierarchical Clustering Analysis dalam Pengelompokan Tingkat Pengangguran di Sulawesi Selatan,” *Pros. Semant.*, vol. 2, no. 1, pp. 270–275, 2019, [Online]. Available: <https://journal.uncp.ac.id/index.php/semantik/article/view/1525/1336%0Ahttps://journal.uncp.ac.id/index.php/semantik/article/view/1525>.
- [4] N. Erlangga, S. Solikhun, and I. Irawan, “Penerapan Data Mining Dalam Mengelompokan Produksi Jagung Menurut Provinsi Menggunakan Algoritma K-Means,” *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 702–709, 2019, doi: 10.30865/komik.v3i1.1681.
- [5] L. Purba, S. Saifullah, and R. Dewi, “Pengelompokan Kasus Penyakit Aids Berdasarkan Provinsi Dengan Data Mining K-Medoids Clustering,” *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 687–694, 2019, doi: 10.30865/komik.v3i1.1679.
- [6] Z. Mustofa and Iman Saufik Suasana, “Algoritma Clustering K-Medoids Pada E-Government Bidang Information and Communication Technology Dalam Penentuan Status Edgi,” *J. Teknol. Inf. Dan Komun.*, vol. 9, no. 1, pp. 1–10, 2020, doi: 10.51903/jtikp.v9i1.162.
- [7] A. P. Windarto, “Penerapan Datamining Pada Ekspor Buah-Buahan Menurut Negara Tujuan

- Menggunakan K-Means Clustering Method,” *Techno.Com*, vol. 16, no. 4, pp. 348–357, 2017, doi: 10.33633/tc.v16i4.1447.
- [8] S. Defiyanti, M. Jajuli, and N. Rohmawati, “Optimalisasi K-MEDOID dalam Pengklasteran Mahasiswa Pelamar Beasiswa dengan CUBIC CLUSTERING CRITERION,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 3, no. 1, pp. 211–218, 2017, doi: 10.25077/teknosi.v3i1.2017.211-218.
- [9] S. Haryati, A. Sudarsono, and E. Suryana, “Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu),” *J. Media Infotama*, vol. 11, no. 2, pp. 130–138, 2015.