

**CLUSTERING DATA PENDUDUK MIGRASI
BERDASARKAN FAKTOR EKONOMI
MENGGUNAKAN METODE K-MEANS**

(Studi Kasus: Kecamatan Tomilito)

OLEH

RAFIL TILAHUNGA

T3117234

SKRIPSI

Untuk memenuhi salah satu syarat ujian

Guna memperoleh gelar sarjana



**PROGRAM SARJANA
TEKNIK INFORMATIKA
UNIVERSITAS IHSAN GORONTALO
GORONTALO
2022**

PERSETUJUAN SKRIPSI

CLUSTERING DATA PENDUDUK MIGRASI BERDASARKAN FAKTOR EKONOMI MENGGUNAKAN METODE *K-MEANS*

(Studi kasus: Kecamatan Tomilito)

Oleh

RAFIL TILAHUNGA

T3117234

SKRIPSI

untuk memenuhi salah satu syarat ujian
guna memperoleh gelar sarjana
program studi teknik informatika,
dan ini telah disetujui oleh pembimbing

Gorontalo 31 Mei 2022

Pembimbing Utama



Suhardi Rustam, S.Kom M.Kom
NIDN 0915088403

Pembimbing Pendamping



Sumarni, S.Kom M.Kom
NIDN 0926018604

PENGESAHAN SKRIPSI
CLUSTERING DATA PENDUDUK MIGRASI
BERDASARKAN FAKTOR EKONOMI
MENGGUNAKAN METODE K-MEANS

(Studi Kasus: Kecamatan Tomilito)

OLEH

RAFIL TILAHUNGA

T3117234

Diperiksa oleh panitia ujian Strata Satu (S1)
Universitas Ichsan Gorontalo

Gorontalo 31 Mei 2022

1. Ketua Pengaji

Irvan A. Salihi M.Kom



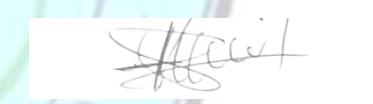
2. Anggota

Sudirman Melangi M.Kom



3. Anggota

Sarlis Mooduto M.Kom



4. Anggota

Suhardi Rustam S,Kom M.Kom



5. Anggota

Sumarni S,Kom M.Kom

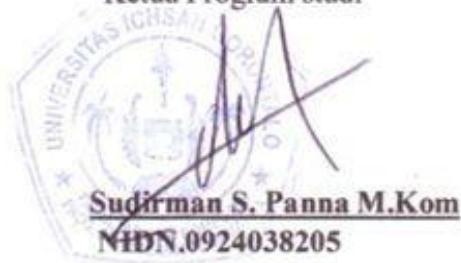


Mengetahui

Dekan Fakultas Ilmu Komputer



Ketua Program studi



PERNYATAAN SKRIPSI

Dengan ini menyatakan bahwa:

1. Tesis (tesis) ini asli dan belum pernah diajukan untuk memperoleh gelar akademik (gelar sarjana) di Universitas Ichsan Gorontalo atau pada perguruan tinggi lainnya.
2. Skripsi ini murni ide, rumusan, dan penelitian saya sendiri tanpa bantuan dari pihak lain, kecuali atas arahan tim pembimbing.
3. Dalam tesis (tesis) saya, tidak ada karya atau pendapat yang pernah diterbitkan oleh orang lain, kecuali dicantumkan secara tertulis sebagai referensi/sitasi dalam naskah yang juga dicantumkan dalam daftar pustaka.
4. Saya membuat pernyataan ini dengan sebenarnya dan apabila dikemudian hari terdapat kejanggalan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang diperoleh karena makalah ini, serta sanksi lainnya dalam sesuai dengan norma yang berlaku di Universitas Ichsan Gorontalo.

Gorontalo, 31 Mei 2022
Yang membuat pernyataan



Rafil Tilahunga

ABSTRACT

RAFIL TILAHUNGA. T3117234. MIGRATION POPULATION DATA CLUSTERING BASED ON ECONOMIC FACTORS USING K-MEANS METHOD

This study aims to obtain the results of the analysis of the K-Means clustering method on grouping population data on migration to obtain precise accuracy and determine the results of grouping migration population data. If seen from economic factors and the level of application, this study is applied research. If observed by the type of information processing, this study is quantitative. The method used in this study for the problem of grouping population data is descriptive. It aims to reveal the objective truth. In other words, it describes a situation or condition currently happening in society by using scientific procedures to answer a problem accurately. The method used in this research is the K-Means algorithm method. This method can group a lot of data based on predetermined characteristics. The use of this method is very effective. The data used in this study are up to 60 data. The data has been grouped into 3 clusters, namely low, middle, and high-income residents. By using the 3 clusters, the results of the K-Means method grouping will be obtained with a good level of accuracy. Based on the data analysis done, the results of the data grouping are 31 data for 'low-income' residents and 23 for 'middle-income' residents, and 6 for 'high-income' residents.

Keywords: Clustering, K-Means, Migration population, Assistance

ABSTRAK

RAFIL TILAHUNGA. T3117234. CLUSTERING DATA PENDUDUK MIGRASI BERDASARKAN FAKTOR EKONOMI MENGGUNAKAN METODE K-MEANS

Penelitian ini bertujuan untuk memperoleh hasil analisis metode k-means clustering pada pengelompokan data penduduk pada migrasi untuk mendapatkan akurasi yang tepat dan mengetahui hasil pengelompokan data penduduk migrasi. Dilihat dari faktor ekonomi dan dari tingkat penerapannya, penelitian ini merupakan penelitian terapan. juga dilihat dari jenis informasi yang diolah, penelitian ini merupakan penelitian kuantitatif. Metode yang digunakan dalam penelitian ini untuk masalah pengelompokan data populasi adalah metode deskriptif, dimana metode ini bertujuan untuk mengungkap kebenaran yang objektif. Dengan kata lain, menggambarkan situasi atau kondisi yang sedang terjadi di masyarakat dengan menggunakan prosedur ilmiah untuk menjawab suatu masalah secara akurat. Metode yang digunakan dalam penelitian ini adalah metode algoritma k-means, metode ini mampu mengelompokkan banyak data berdasarkan karakteristik yang telah ditentukan sebelumnya. Penggunaan metode ini sangat efektif dimana data yang digunakan dalam penelitian ini berjumlah 60 data. Data tersebut telah dikelompokkan menjadi 3 cluster yaitu penduduk berpenghasilan rendah, sedang dan tinggi. Dengan penelitian menggunakan 3 cluster tersebut akan didapatkan hasil pengelompokan metode k-means dengan tingkat akurasi yang baik. Dan berdasarkan penelitian data analisis yang telah dilakukan didapatkan hasil dari pengelompokan data tersebut yaitu penduduk berpenghasilan rendah sebanyak 31 data dan penduduk berpenghasilan sedang sebanyak 23 dan yang berpenghasilan tinggi sebanyak 6 data.



Kata kunci: Clustering, K-Means, Penduduk migrasi, Bantuan

KATA PENGANTAR

Assalamualaikum wr.wb

Alhamdulillah, Hirobbil Aalamiin, puji syukur penulis panjatkan kehadiran Allah SWT yang telah memberikan hidayah dan umur panjang sehingga penulis dapat menyelesaikan skripsi yang berjudul **CLUSTERING DATA PENDUDUK MIGRASI BERDASARKAN FAKTOR EKONOMI MENGGUNAKAN METODE K-MEANS** (Studi Kasus Kecamatan Tomilito) tepat waktu.

Penyusunan tesis ini merupakan salah satu syarat kelulusan bagi setiap mahasiswa ilmu komputer untuk menyelesaikan studinya dan memperoleh gelar sarjana. Penulis menyadari bahwa dalam penulisan atau penyusunan proposal ini, masih belum sempurna dan masih banyak kekurangan. Oleh karena itu, saran dan kritik akan diterima dengan senang hati untuk membangun demi penulisan laporan yang lebih baik.

Pada kesempatan ini penulis ingin mengucapkan terima kasih kepada semua pihak yang telah membantu dalam penyusunan proposal ini dalam hal memberikan bimbingan, dorongan, dan motivasi, serta bantuan dengan doa. Oleh karena itu, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Ibu Dr Hj. Juriko Abdusamad M.si selaku ketua Yayasan Pengembangan Ilmu Pengetahuan dan Teknologi (YPIPT) Ichsan Gorontalo
2. Bapak Dr. Abdul Gafar Latjokke, M.Si, selaku Rektor Universitas Ichsan Gorontalo
3. Bapak Jorri Karim M.Kom, selaku Dekan Fakultas Ilmu Komputer Universitas Ichsan Gorontalo
4. Bapak Sudirman S. Panna, M.Kom, selaku Pembantu Dekan 1 Bidang Akademik Fakultas Ilmu Komputer Universitas Ichsan Gorontalo
5. Ibu Irma Surya Kumala Idris, M.Kom selaku Pembantu Dekan II Bidang Administrasi Umum Fakultas Ilmu Komputer Universitas Ichsan Gorontalo.

6. Bapak Sudirman Melangi, M.Kom selaku Wakil Dekan I Bidang Kemahasiswaan Fakultas Ilmu Komputer Universitas Ichsan Gorontalo
7. Bapak sudirman panna, M.Kom selaku ketua program studi Teknik Informatika Fakultas Ilmu Komputer.
8. Bapak Suhardi Rustam, M,Kom selaku Penanggung Jawab Fakultas Ilmu Komputer Kampus 3 Ichsan Gorontalo Utara sekaligus Pembimbing Utama dalam penyusunan Proposal ini.
9. Ibu Sumarni, S.Kom selaku Sekretaris Program Studi Teknik Informatika Fakultas Ilmu Komputer Kampus 3 Ichsan Gorontalo Utara, serta Pembimbing
10. Kepada kedua orang tua tercinta, terima kasih telah memberikan nasehat, semangat, dan doa yang tiada hentinya untuk menjadikan pribadi penulis yang lebih baik lagi.
11. Kepada sahabat pena dakwah, kharisma, teman-teman kampus seperjuangan yang telah menjadi penyemangat untuk memberikan dukungan dan motivasi kepada penulis dalam menyelesaikan skripsi ini.

Terima kasih kepada semua pihak yang telah membantu penulis dalam penyusunan skripsi ini.

Semoga Allah SWT selalu membalas perbuatan baik yang telah diberikan kepada penulis, yang juga dibalas dengan kebaikan. Penulis hanya bisa berharap semoga penulisan penelitian ini dapat bermanfaat bagi kita semua, Aamiin.

Gorontalo,31 mei 2022

DAFTAR ISI

HALAMAN JUDUL	I
HALAMAN PERSETUJUAN	II
PENGESAHAN SKRIPSI	III
PERNYATAAN SKRIPSI.....	IV
ABSTRAK.....	V
KATA PENGANTAR.....	VII
DAFTAR ISI.....	IX
DAFTAR TABEL	XI
DAFTAR GAMBAR.....	XII
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Identifikasi Masalah.....	4
1.3 Rumusan Masalah.....	4
1.4 Tujuan Masalah	5
1.5 Manfatat Penelitian	5
BAB II LANDASAN TEORI.....	6
2.1 Tinjauan Studi	6
2.2 Tinjauan Pustaka.....	8
2.2.1 Migrasi	8
2.2.2 Data Mining	8
2.2.3 Clustering	9
2.2.4 Al goritma K-means.....	9
2.2.5 Contoh Penerapan Metode K-means Untuk Clustering	10
2.3 Perangkat Lunak Pendukung	14
2.4 Kerangka Pikir	15
BAB III METODE PENELITIAN	16
3.1 Objek Penelitian	16
3.2 Metode Penelitian.....	16

3.3 Sumber Data.....	.17
3.3.1 Variabel Perpindahan Penduduk	17
3.4 Pra Pengolahan Data.....	18
3.5 Hasil Clustering	18
3.6 Evaluasi.....	18
BAB IV PEMBAHASAN.....	19
4.1 Hasil pengumpulan data.....	21
4.2 Hasil pemodelan.....	21
4.2.1 Pra pengolahan data.....	21
4.2.2 Normalisasi data.....	21
4.2.3 Hasil hitungan k-means.....	22
4.2.4 Hasil clustering.....	28
BAB V PEMBAHASAN.....	27
5.1 Pembahasan model.....	29
5.2 Pembahasan tools.....	34
BAB VI KESIMPULAN DAN SARAN.....	37
6.1 Kesimpulan.....	37
DAFTAR PUSTAKA.....	38
LAMPIRAN.....	40

DAFTAR TABEL

Tabel 1.1	Data Set	2
Tabel 2.1	Penelitian Terdahulu	6
Tabel 2.2	Contoh Data Yang Akan Di Olah.....	10
Tabel 2.3	Penentuan Awal <i>cluster</i>	11
Tabel 2.4	Perhitungan <i>centroid</i> Terdekat	11
Tabel 2.5	Hasil Perhitungan Jarak Terdekat.....	12
Tabel 2.6	Hasil Pengelompokan <i>cluster</i>	13
Tabel 2.7	Hasil Perhitungan Pusat <i>cluster</i> Baru	14
Tabel 4.1	Data set penduduk	19
Tabel 4.2	Data sampel penduduk migrasi	22
Tabel 4.3	Penentuan awal cluster.....	22
Tabel 4.4	Menentukan jarak literasi 1	23
Tabel 4.5	Menentukan awal pusat cluster 2.....	24
Tabel 4.6	Awal pusat cluster literasi 2	25
Tabel 4.7	menghitung jarak literasi 2.....	25
Tabel 4.8	literasi 3 awal centroid	26
Tabel 4.6	mencari jarak terdekat literasi 3	26
Tabel 4.7	awal pusat centroid literasi 4	26
Tabel 4.8	Menentukan jarak literasi 4	27
Tabel 4.9	Awal centroid literasi 5	27
Tabel 4.10	Menentukan jarak literasi 5	27
Tabel 5.1	Data hasil cluster	35

DAFTAR GAMBAR

Gambar 2.7	Kerangka Pikir	15
Gambar 3.1	Model Usulan.....	16
Gambar 5.1	Model eksperimen	27
Gambar 5.2	Tampilan utama.....	30
Gambar 5.3	Tampilan impor data	30
Gambar 5.4	Penyimpanan data	31
Gambar 5.5	Data yang digunakan	31
Gambar 5.6	Tampilan desaign	32
Gambar 5.7	Tampilan data result	32
Gambar 5.8	Tampilan folder data	33
Gambar 5.9	Tampilan data pada folder	33
Gambar 5.10	Tampilan cluster model	34

BAB I

PENDAHULUAN

1.1 Latar Belakang

Migrasi adalah perpindahan orang dari satu tempat ke tempat lain dengan tujuan menetap, melampaui batas-batas politik/negara atau batas-batas administratif dalam suatu negara. Migrasi yang terjadi di Indonesia dapat berupa migrasi internal, migrasi internasional dan juga urbanisasi. Migrasi memiliki dampak ekonomi, baik bagi tempat yang ditinggalkan maupun bagi daerah tujuan. Migrasi tentunya membawa dampak, baik dampak maupun dampak sosial bagi daerah yang ditinggalkan, akan berdampak buruk sehingga berdampak pada pertumbuhan ekonomi yang dibutuhkan daerah yang dituju untuk mendapatkan tambahan tenaga kerja agar pertumbuhan ekonomi meningkat,[1] hal ini tidak hanya dapat terjadi di kota tetapi juga dapat terjadi di desa, misalnya tiga desa di kecamatan Tomilito, yaitu desa Milango, desa Bubode dan desa Leyao.

Ketiga desa ini merupakan desa yang berada di Kabupaten Gorontalo Utara tepatnya di Kecamatan Tomilito, dan merupakan Desa Mekar dari Desa Dambalo. Sejak pemekaran dari desa Dambalo pada tahun 2011, ketiga desa ini mulai melakukan pembangunan dalam hal penataan dan peningkatan infrastruktur dan sejak saat itu sudah ada warga yang merantau di ketiga desa ini untuk menetap dan mencari pekerjaan untuk meningkatkan taraf hidupnya. pendapatan sebelumnya. Dengan masuknya penduduk yang bermigrasi, ketiga desa ini mengalami kemajuan dalam hal peningkatan jumlah tenaga kerja produktif, yang berarti dapat mendorong pertumbuhan ekonomi yang lebih tinggi. Namun, tidak semua penduduk migrasi di ketiga desa tersebut berpenghasilan lebih atau memenuhi kebutuhan mereka, adapun penduduk migrasi memiliki pendapatan yang lebih kecil dan juga pendapatan sedang.

Dilihat dari cara pengelompokan data masyarakat yang dilakukan oleh kantor desa di ketiga desa tersebut hanya

dengan menggunakan microsoft office tentunya cara ini sedikit memakan waktu karena prosesnya yang lama ketika data sewaktu-waktu dibutuhkan, dan belum adanya sistem dengan akurasi yang tepat untuk mengatur pengelompokan data masyarakat berdasarkan faktor ekonomi, dalam hal ini penelitian penulis akan mengelompokkan data migrasi dalam suatu sistem berdasarkan faktor ekonomi. Pengelompokan ini akan menggunakan algoritma k-means dan akurasinya untuk mengetahui penduduk migrasi mana yang mendekati pendapatan rendah, sedang atau tinggi. Hasil klasterisasi yang memenuhi syarat berpenghasilan rendah akan menjadi bahan pertimbangan pemerintah desa untuk mendapatkan bantuan.

Tabel 1.1 Data Set

Data Migrasi Berdasarkan Ekonomi

No	Nama Dan Desa	Jumlah pinjaman	Jumlah Tanggungan	Jumlah Penghasilan
1	Risman Podomi, Milango	100	2	1.000.000
2	Maya Rifka Abas, Milango	500	1	750.000
3	Husain Rajak, Bubode	300	3	500.000
4	Aril Pratama Arfan, Leyao	400	1	250.000
5	Usman Yunus, Bubode	100	2	500.000
6	Ripon Yahya, Bubode	200	2	300.000
7	Hawaria Y. Mohamad, Milango	800	1	350.000
8	Suleman Dude, Leyao	250	4	400.000
9	Pian Harun, Leyao	400	2	500.000
10	Yunarti Laudji, Milango	370	2	300.000

Sumber Data: Desa Milango, Leyao, Bubode(2019,2020,2021)

Clustering adalah suatu metode untuk mencari dan mengelompokkan data yang memiliki karakteristik yang mirip (similarity) satu sama lain. Clustering adalah metode penambangan data yang tidak diawasi. Dalam data mining ada dua jenis metode clustering yang digunakan dalam pengelompokan data, yaitu hierarchical clustering dan non-hierarchical clustering. Hierarchical clustering adalah metode pengelompokan data yang dimulai dengan mengelompokkan dua atau lebih objek yang memiliki kemiripan terdekat.[2]

k-means adalah metode pengelompokan data non-hierarki yang mencoba mempartisi data menjadi satu atau lebih cluster/grup. Metode ini mempartisi data ke dalam cluster/grup sehingga data dengan karakteristik yang sama dikelompokkan ke dalam cluster yang sama.[3]

Pada penelitian ini data penduduk akan dikelompokkan menggunakan metode algoritma k-means untuk mencari cluster-cluster yang mendekati kemiripan atau memiliki karakteristik yang sama, metode ini dipilih dalam pengelompokan karena dapat mengelompokkan data yang ada menjadi beberapa kelompok, dimana data dalam satu kelompok memiliki karakteristik yang berbeda satu sama lain sehingga nantinya kantor desa akan lebih mudah dalam mengambil keputusan atau kebijakan dalam menentukan penduduk keimigrasian yang berpenghasilan rendah, sedang, dan tinggi. Variabel yang digunakan penulis dalam penelitian ini adalah : jumlah pinjaman, Jumlah tanggungan, jumlah pendapatan

Pada penelitian Yunita Ratna Sari, Arbi Sudewa, Dian Ayu Lestari, Tri Ika Jaya dengan judul “Penerapan Algoritma K-means Untuk Clustering Data Kemiskinan di Provinsi Banten Menggunakan Rapidminer” pada penelitian ini data mining digunakan untuk mengklasifikasikan data untuk miskin untuk menentukan hasil yang akurat. Penelitian ini menghasilkan 3 klaster yaitu klaster 0, klaster 1, klaster 2, masing-masing klaster memiliki hasil yang berbeda-beda dan hasil tersebut akan menjadi penentu keputusan pemerintah untuk langkah selanjutnya. Data mining telah banyak digunakan dan sangat berguna dalam

berbagai bidang, misalnya dalam bidang pemerintahan untuk mengelompokkan data masyarakat dan menentukan hasil sebagai pengambil keputusan.

Demikian juga dengan penelitian ini, penulis bertujuan untuk mengelompokkan data komunitas untuk mengetahui data komunitas yang memiliki persamaan atau karakteristik yang sama dapat dikelompokkan menjadi satu cluster dan data yang berbeda akan dikelompokkan ke dalam kelompok lain dan untuk mengetahui keakuratan model k-means. . sehingga dalam hal ini dapat mempermudah kantor desa dalam pendataan penduduk migrasi berdasarkan faktor ekonomi. Dalam penelitian ini penulis menggunakan tools rapidminer sebagai tools perangkat lunak pendukung dalam analisis data perpindahan penduduk

Berdasarkan uraian latar belakang di atas, maka perlu dilakukan penelitian yang berjudul **“CLUSTERING DATA PENDUDUK MIGRASI BERDASARKAN FAKTOR EKONOMI MENGGUNAKAN METODE K-MEANS”**, dalam penelitian ini diharapkan dapat mempermudah pihak kantor desa Milango, Leyao dan Bubode untuk menentukan penduduk migrasi berpenghasilan rendah, sedang dan tinggi

1.2 IDENTIFIKASI PERMASALAHAN

Identifikasi masalahnya adalah sebagai berikut:

1. Penduduk desa masih kesulitan dalam pendataan penduduk migrasi berdasarkan faktor ekonomi
2. Belum adanya sistem pengelompokan data dengan tingkat akurasi yang tepat.

1.3 RUMUSAN MASALAH

Rumusan masalahnya adalah sebagai berikut:

1. Bagaimana cara mendapatkan hasil pengelompokan penduduk migrasi menggunakan model K-Means Algoritma dan mendapatkan akurasi model K-Means

1.4 TUJUAN

Tujuan dari masalah tersebut adalah sebagai berikut:

1. Mendapatkan hasil clustering untuk model K-Means serta akurasinya dari penduduk migrasi berdasarkan faktor ekonomi yang nantinya akan direkomendasikan untuk mendapatkan bantuan.

1.5 MANFAAT PENELITIAN

Manfaat penelitian adalah sebagai berikut:

1. Manfaat Teoretis

Menambah pengetahuan tentang sistem yang dibuat dan sebagai modal dasar untuk karya tulis kedepannya.

2. Manfaat bagi praktisi

Peneliti dan pengguna dapat memahami kegunaan sistem sekaligus menjadi masukan bagi semua pihak yang terlibat dalam pembuatan aplikasi metode k-means.

3. Peneliti

Sebagai saran atau masukan bagi peneliti lain yang akan melakukan penelitian lebih lanjut terkait dengan metode k-means

BAB II

LANDASAN TEORI

2.1 TINJAUAN STUDI

Ada beberapa penelitian yang terkait dengan penelitian clustering data perpindahan penduduk berdasarkan tahun menggunakan metode k-means dan juga menjadi acuan bagi penulis diantaranya.

Tabel 2.1 Penelitian Terdahulu

NO	PENELITI	JUDUL	TAHUN	METODE	HASIL
1	Fredy Azizy	Kepindahan penduduk desa Tugurejo.	2017	Algoritma k-means	Terbukti mampu mengelola dan mengelompokan data warga secara otomatis
2	Dina Sunia, Kurniabudi, Pareja Alam Jusia	Pererapan data mining untuk clustering penduduk miskin.	2016	Clustering dan algoritma k-means	Jumlah data yang digunakan 512 mempunyai 16 atribut dan direkomendasikan lima cluster
3	Bayu Resi Indrawan	Penerapan algoritma k-means clustering untuk menentukan strategi promosi	2018	Algoritma k-means	Dalam penelitian ini dikelompokan tiga cluster

NO	PENELITI	JUDUL	TAHUN	METODE	HASIL
4	Sarah Ashari, Salsabila Khansa, Celvin Habib Maulana Surudin, Ika Nurlaili Isnainiyah	Clustering jumlah penduduk kota bandung berdasarkan jenis kelamin per kecamatan pada tahun 2012 dengan metode k-means	2018	Metode algoritma k-means	Penelitian k-means menghasilkan dua cluster terdapat tiga kali literasi hingga centroid tidak berubah dan menghasilkan clustering final
5	M. Rosyid Ridlo, Sofi Defiyanti, Aji Primajaya	Implementasi algoritma k-means untuk pemetaan produktivitas panen padi di kabupaten karawang	2017	Algoritma k-means	Menghasilkan aplikasi berbasis web, pengelompokan data dan mempresentasikan hasil pengelompokan, dengan jumlah tiga cluster dan

2.2 TINJAUAN PUSTAKA

2.2.1 Migrasi

Jalannya pembangunan kependudukan dinilai secara umum disebabkan oleh relokasi metropolitan provinsi, yang bergantung pada pembangunan reguler yang lebih rendah di ruang yang akan menyalurkan limpahan pekerjaan di daerah-daerah saat ini di daerah perkotaan dengan batas asimilasi yang lebih tinggi.[4]

Proses perpindahan penduduk diperkirakan sebagian besar disebabkan oleh migrasi desa-kota, yang didasarkan pada pertumbuhan alami yang lebih rendah di daerah yang akan menyalurkan surplus tenaga kerja di daerah modern di kota-kota dengan daya serap yang lebih tinggi.[4]

2.2.2 Data Mining

Penambangan informasi adalah interaksi yang menggunakan strategi faktual, aritmatika kekuatan otak buatan manusia, dan untuk menghapus dan mengenali data berharga dan informasi terkait dari kumpulan data yang sangat besar. Penambangan informasi merupakan perkembangan siklis untuk menyelidiki nilai tambah suatu indeks informasi sebagai informasi yang belum diketahui secara fisik.[5]

Menurut Fayyad dalam bukunya (Kusrini 2009) istilah data mining dan knowledge discovery in database (KDD) sering digunakan secara bergantian untuk menggambarkan proses penggalian informasi yang tersembunyi dalam database yang besar. Kedua istilah tersebut memiliki konsep yang berbeda, namun saling berkaitan. Dan salah satu tahapan dalam keseluruhan proses KDD adalah data mining.[6]

2.2.3 Clustering

Menurut Widodo (2013: 9) clustering atau klasifikasi adalah suatu metode yang digunakan untuk membagi suatu deret data menjadi beberapa kelompok berdasarkan kesamaan yang telah ditentukan sebelumnya. Cluster adalah sekelompok atau sekumpulan objek data yang mirip satu sama lain dalam cluster yang sama dan mirip dengan objek dari cluster yang berbeda. Objek akan dikelompokkan menjadi satu atau lebih cluster sehingga objek dalam satu cluster akan memiliki kemiripan yang tinggi antara satu dengan yang lainnya.[7]

Menurut Garcia (2002), clustering adalah mengelompokkan item data ke dalam sejumlah kecil kelompok sehingga setiap kelompok memiliki persamaan yang esensial.[8]

2.2.4 Algoritma K-Means

K-Means adalah metode clustering data non-hierarchical yang mencoba untuk mempartisi data yang ada menjadi satu atau lebih cluster, sehingga data dengan karakteristik yang sama dikelompokkan ke dalam cluster yang sama. Data dengan karakteristik yang berbeda dikelompokkan ke dalam kelompok lain. Metode tersebut termasuk dalam algoritma distance-based clustering yang membagi data menjadi beberapa cluster dan algoritma ini hanya bekerja pada atribut numerik.[9]

Tujuan dari pengelompokan data ini adalah untuk meminimalkan fungsi tujuan yang ditetapkan dalam suatu kelompok dan memaksimalkan variasi antar kelompok (Eko Prasetyo, 2012:178).

Algoritma dasar k-means dijelaskan sebagai berikut:

1. Tentukan nilai k sebagai jumlah cluster yang ingin Anda bentuk
2. Inisialisasi k sebagai centroid yang dapat dibangkitkan secara acak
3. Hitung jarak setiap data ke setiap centroid menggunakan persamaan Euclidean Distance, yaitu sebagai berikut $d(p,q)=$

4. Kelompokkan setiap data berdasarkan jarak terpendek antara data dan centroidnya.
5. tentukan posisi centriod baru (k).
6. Kembali ke langkah 3 jika posisi centroid baru dengan centroid lama tidak sama.[10]

2.2.5 Contoh penerapan metode k-means untuk *clustering*

Berikut ini contoh perhitungan algoritma k-means yang diambil dari penelitian sebelumnya yaitu oleh Nayuni Dwitri, Jose A Tampupolo, Sandi Prayoga, P.P.P.A.NW Fikrul Ilmi R.H Zer, Dedy Hartama dengan judul “Aplikasi K-means Algoritma Penentuan Penyebaran Pandemi COVID-19 Di Indonesia” Tabel 1.1 merupakan data yang diuji dan digunakan sebagai data pendukung.

Table 2.2 Contoh Data Yang Akan Diolah
Data Penyebaran Virus *Covid-19*

NO	Provinsi	Kasus Positif	Kasus Meninggal
1	Aceh	17	1
2	Bali	306	4
3	Banten	523	56
4	Bangka Belitung	28	1
5	Bengkulu	37	1
6	DI Yogyakarta	146	7
7	DKI Jakarta	5056	427
8	Jambi	62	0
9	Jawa Barat	1437	95
10	Jawa Tengah	959	66
11	Jawa Timur	1419	141
12	Kalimantan Barat	118	3
13	Kalimantan Timur	214	2
14	Kalimantan Tengah	189	7
15	Kalimantan Selatan	253	9
16	Kalimantan Utara	131	1
17	Kepulauan Riau	101	10
18	Nusa Tenggara Barat	330	6
19	Sumatera Selatan	278	6
20	Sumatera Barat	286	17
21	Sulawesi Utara	53	4
22	Sumatera Utara	179	21
23	Sulawesi Tenggara	71	2
24	Sulawesi Selatan	710	46

25	Sulawesi Tengah	75	3
26	Lampung	66	5
27	Riau	71	2
28	Maluku Utara	54	0
29	Maluku	32	2
30	Papua Barat	53	1
31	Papua	277	6
32	Sulawesi Barat	62	2
33	Nusa Tenggara Timur	12	0
34	Gorontalo	19	1

(Sumber: Dedy Hartama,2020)

a. Penentuan Pusat Awal *Cluster*

Tabel 2.3 Penentuan Awal *Cluster*

Klaster 1	Jawa Timur	1419,141
Klaster 2	Nusa Tenggara Barat	330,6
Klaster 3	Maluku	32,3

(Sumber: Dedy Hartama,2020)

b. Perhitungan Jarak *Cluster*

Tabel 2.4 Perhitungan *Centroid* Terdekat

Aceh C1	$\sqrt{(17 - 1419)^2 + (1 - 141)^2}$	14,972675
Aceh C2	$\sqrt{(17 - 330)^2 + (1 - 6)^2}$	313,0399336
Aceh C3	$\sqrt{(17 - 32)^2 + (1 - 2)^2}$	15,03329638

(Sumber: Dedy Hartama, 2020)

c. Pengelompokan Data

Pada tahap ini, jarak terdekat antara data dengan cluster akan menentukan data mana yang termasuk dalam cluster. Berikut adalah hasil perhitungan jarak

Tabel 2.5 Hasil perhitungan jarak terdekat

Provinsi	Jarak ke centroid		
	Cluster 1	Cluster 2	Cluster 3
ACEH	1408,972675	313,0399336	15,03329638
BALI	1121,400018	24,08318916	274,0072992
BANTEN	900,0227775	199,3715125	493,9605247
BANGKA BELITUNG	1398,027539	302,0413879	4,123105626
BENGKULU	1389,073072	293,042659	5,099019514
DI YOGYAKARTA	1280,033203	184,0027174	114,1095964
DKI JAKARTA	3648,227652	4744,714638	5041,944169
JAMBI	1364,305684	268,0671558	30,06659276
JAWA BARAT	49,39635614	1110,571925	1408,074572
JAWA TENGAH	466,0740285	631,8552049	929,2066509
JAWA TIMUR	0	1097,335865	1393,947632
KALBAR	1308,298513	212,0212254	86,00581376
KALTIM	1212,990519	166,068945	182
KALTENG	1237,277657	141,0035461	157,0795977
KALSEL	1173,447911	77,0584194	221,1108319
KALUT	1295,586354	199,0628042	99,00505038
KEP. RIAU	1324,494243	229,0349318	69,46221995
NTB	1097,335865	0	298,0268444
SUMSEL	1148,958659	52	246,0325182
SUMBAR	1139,765327	45,35416188	254,4425279
SULUT	1372,852869	277,0072201	21,09502311
SUMUT	1245,79292	151,7432041	148,2228053
SULTRA	1355,147593	259,0308862	39
SULSEL	715,3362846	382,0994635	679,4262285
SULTENG	1351,066246	255,0176464	43,011626634
LAMPUNG	1359,818003	264,0018939	34,13209633
RIAU	1345,743149	259	39,20459157
MALUKU UTARA	1372,263094	276,0652097	22,09072203
MALUKU	1393,947632	298,0286444	0
PAPUA BARAT	1373,15549	277,0451227	21,02379604
PAPUA	1149,951738	53	245,0326509
SULBAR	1364,100436	268,0298491	30
NTT	1414,047383	318,0565987	20,09975124
GORONTALO	1406,982587	311,0401903	13,03840481

(Sumber: Dedy Hartama, 2020)

Berdasarkan hasil perhitungan pada tabel di atas maka di dapatkan pengelompokan sebagai berikut:

Tabel 2.6 Hasil pengelompokan *cluster*

Provinsi	Jarak Terdekat
ACEH	C3
BALI	C2
BANTEN	C2
BANGKA BELITUNG	C3
BENGKULU	C3
DI YOGYAKARTA	C3
DKI JAKARTA	C1
JAMBI	C3
JAWA BARAT	C1
JAWA TENGAH	C1
JAWA TIMUR	C1
KALBAR	C3
KALTIM	C2
KALTENG	C2
KALSEL	C2
KALUT	C3
KEP. RIAU	C3
NTB	C2
SUMSEL	C2
SUMBAR	C2
SULUT	C3
SUMUT	C3
SULTRA	C3
SULSEL	C2
SULTENG	C3
LAMPUNG	C3
RIAU	C3
MALUKU UTARA	C3
MALUKU	C3
PAPUA BARAT	C3
PAPUA	C2
SULBAR	C3
NTT	C3
GORONTALO	C3

(Sumber: Dedy Hartama 2020)

d. Penetapan pusat cluster baru

Tentukan centroid baru untuk iterasi berikutnya dengan mencari nilai rata-rata dari setiap cluster. Berikut adalah nilai centroid baru:

Tabel 2.7 Hasil Perhitungan Pusat Cluster Baru

Centroid	C1	2217,8	182,25
	C2	336,6	15,9
	C3	69,35	3,55

(Sumber: Dedy Hartama, 2020)

Jika tahap iterasi telah mencapai hasil yang sama tanpa perpindahan lebih lanjut, perhitungan dihentikan.[11]

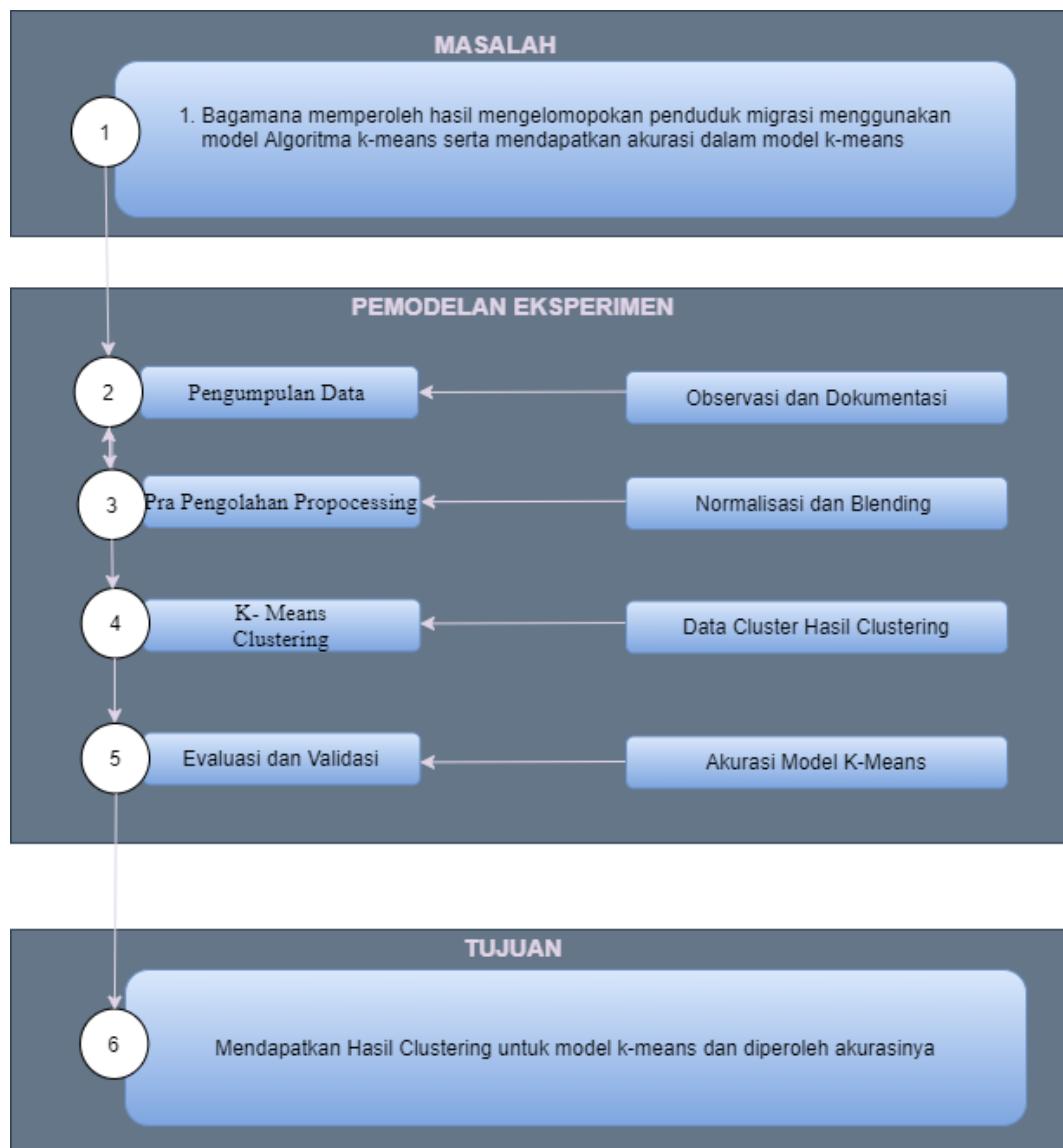
2.3 Perangkat Lunak Pendukung

Perangkat lunak pendukung yang digunakan penulis dalam mengelompokan data yaitu RapidMiner

RapidMiner adalah sebuah perangkat lunak untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi. Rapidminer menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. Aplikasi atau software ini adalah perangkat lunak berfungsi sebagai alat pembelajaran pada data mining rapidminer adalah salah satu software untuk pengolahan data mining. Pekerjaan yang dilakukan oleh rapidminer teks mining adalah berkisar dengan analisis teks mengekstrak pola-pola dari data set besar dan mengombinasikannya dengan metode statistika, kecerdasan buatan, dan database.

2.4 Kerangka Pikir

Adapun kerangka pikir pada penelitian ini sebagai berikut:



Gambar 2.1 kerangka pikir

BAB III

METODE PENELITIAN

3.1 Objek Penelitian

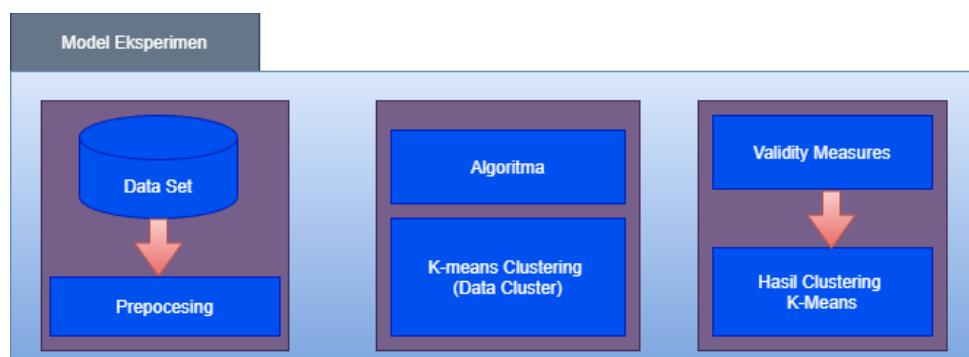
Berdasarkan uraian latar belakang dan kerangka pikir di atas, maka yang menjadi objek penelitian adalah “**Clustering Data Penduduk Migrasi Berdasarkan Factor Ekonomi Menggunakan Metode K-Means Studi Kasus Kecamatan Tomilito**”

3.2 Metode Penelitian

Metode yang digunakan dalam penelitian ini adalah metode k-means yaitu mengelompokkan data yang ada ke dalam kelompok-kelompok dan akan menghasilkan akurasi, dimana data dalam satu kelompok memiliki karakteristik yang sama, data populasi dikelompokkan menjadi tiga cluster yaitu yang memiliki tinggi, sedang dan rendah pendapatan. Kemudian setiap cluster dikelompokkan berdasarkan cluster mana yang diprioritaskan dan memenuhi persyaratan. Cluster dengan nilai terbesar pada centroid akhir adalah cluster yang direkomendasikan untuk populasi dengan pendapatan tinggi, sedang atau rendah. cluster dengan nilai tengah atau sedang adalah nilai cluster yang akan dipertimbangkan, sedangkan cluster dengan nilai terkecil pada centroid akhir adalah cluster berpenghasilan rendah.

3.3 Model Eksperimen

Adapun model eksperimen pada penelitian ini sebagai berikut:



3.4 Sumber Data

Data primer dalam penelitian ini sesuai dengan observasi lapangan dan wawancara langsung dengan kantor desa milango leyao dan bubode. Data primer yang diperoleh adalah data penduduk yang belum terdaftar di desa tetap penduduk.

Sebagai data sekunder dalam penelitian ini adalah mengumpulkan data atau informasi dengan membaca berbagai referensi seperti hasil penelitian sebelumnya, buku teks, jurnal terkait dari internet terkait penerapan metode k-means untuk pengelompokan data perpindahan penduduk. Pada penelitian ini digunakan beberapa cara untuk mengumpulkan data diantaranya Pengamatan

Observasi merupakan salah satu teknik pengumpulan fakta atau data yang cukup efektif untuk mempelajari dan mengamati secara langsung bagaimana mengelompokkan beberapa data perpindahan penduduk.

3. 5 Variabel Perpindahan Penduduk

Variabel terdiri dari:

1. Jumlah pinjaman

Jumlah tanggungan adalah jumlah yang dibayarkan oleh komunitas migrasi Data ini digunakan sebagai variabel input.

2. Jumlah tanggungan

Jumlah dari migrasi yang telah tinggal dan menetap di desa baru data ini juga digunakan sebagai data input.

3. Jumlah pendapatan

Pendapatan sebelumnya yaitu pendapatan sebelum penduduk pindah ke tempat atau daerah baru, data pendapatan sebelumnya ini dapat menjadi faktor atau alasan penduduk berpindah dan data ini dijadikan sebagai data masukan.

3. 6 Pra Pengolahan Data

Sebelum data diolah, data dikelompokkan menjadi beberapa variabel, hal ini dilakukan agar lebih mudah untuk memasukkan data ke dalam perhitungan Algoritma K-Means.

3. 7 Hasil Clustering

Hasil dari klasterisasi merupakan keluaran, pada data yang diperoleh dari proses klasterisasi menggunakan metode K-Means berdasarkan data yang diperoleh dari data migrasi.

3. 8 Evaluasi

Evaluasi bertujuan untuk mengetahui hasil performansi dan metode yang digunakan, evaluasi dilakukan terhadap output data yang dihasilkan akan dimasukkan ke dalam cluster hasil untuk menghitung nilai akurasi.

BAB IV

HASIL PENELITIAN

4.1 Hasil Pengumpulan Data

Pada penelitian ini penulis menggunakan metode Algoritma K-Means dalam menentukan penduduk migrasi yang berpenghasilan tinggi, sedang dan rendah yang ada pada tiga desa yang berada dikecamatan tomilito yaitu desa milango, bubode dan leyao hasil yang di dapatkan dari pengelompokan data ini akan menjadi acuan atau bahan pertimbangan bagi pemerintah sehingga pemerintah desa bisa menetapkan penduduk yang layak dibantu dan akan digolongkan pada jenis-jenis bantuan yang ada didesa.

Berikut ini adalah hasil pengumpulan data yang diperoleh dari lokasi penelitian.

Tabel 4.1 Data Set Penduduk Migrasi

No	Nama	Jumlah pinjaman	jumlah tanggungan	jumlah pendapatan
1	Husain Rajak	3	2	5
2	Ripon Yahya	6	2	3
3	Usman Yunus	5	2	3
4	Lian Yunus	3	1	4
5	Dian Simin	6	3	2
6	Dirman Hasim	2	4	5
7	Sunarti S. Ntuna	4	2	3
8	Bayur	5	2	10
9	Arianto Pola	6	2	2
10	Rosita A. Mohamad	3	5	6
11	Suleman Dude	3	0	2
12	Hadija Ibrahim	4	3	5
13	Aditia Djafar	2	5	4
14	Moh. Sandri Djafar	3	2	6
15	Riski Rajak	5	3	2
16	Wiwin Mawati Mahmud	4	2	3
17	Merlina Ibrahim	2	1	5
18	Ujang Pakaya	3	2	6

19	Pian Harun	4	1	3
20	Salma Yusuf	5	3	7
21	Abdulatif Eka Putra Harun	6	2	8
22	Fauzan Dulanimo	7	1	5
23	Nirmawati Tuu	5	2	6
24	Merlin Mahmud	3	0	5
25	Aril Pratama Arfan	4	1	10
26	Halma Taib	6	2	8
27	Fauzan Dulanimo	7	0	5
28	Deden Pakaya	5	1	9
29	Wirda	4	2	3
30	Ucin Bilali	3	1	5
31	Moh. Rafli Musa	5	2	7
32	Selvi Malanuwa	3	1	4
33	Hawariya Mohamad	6	1	8
34	Raflin Talib	5	3	9
35	Selvi Noe	6	4	5
36	Rahmat Aditia Yusuf	4	1	3
37	Rahmatia Muhamad	3	0	8
38	Yunarti Laudji	4	1	5
39	Karim Tahidina	6	1	7
40	Saldi Wombohuke	5	2	6
41	Ulyan Ayuba	6	3	8
42	Maya Rifka Abas	6	1	8
43	Djamin Midu	5	2	10
44	Wisma Puluhulawa	5	0	6
45	Rosna Madelu	7	2	7
46	Risman Podomi	1	2	10
47	Maryam Hipi	4	1	7
48	Ruslini Hasan Jafar	6	1	5
49	Erwin	10	3	7
50	Nur Intan	5	1	9
51	Rian	8	2	5
52	Melisa Sukamdi	4	2	10
53	Ayu Anggraini Ismail	4	2	7
54	Cicin Seu	4	1	10
55	Novita Laubunu	5	2	7
56	Abd Rahman Sedi	4	1	10
57	Romi Tintia	5	0	10
58	Al Fahri Tahidina	6	7	3

59	Denis Rajak	4	0	7
60	Yulan Djafar	5	2	8

Data set diatas merupakan data penduduk migrasi berdasarkan factor ekonomi yang akan di rekomendasikan mendapatkan bantuan.

4.2 Hasil Pemodelan

4.2.1 Pra Pengolahan Data

Data yang digunakan dalam penelitian ini adalah data yang berasal dari lokasi penelitian yaitu di kecamatan tomilito, pra pengolahan data itu sendiri adalah proses untuk mengidentifikasi data yang telah terkumpul, data yang sebelumnya tidak diketahui harus diproses untuk dapat dikenali. Hal ini bertujuan untuk dapat mengetahui jumlah data yang akan dikelompokkan menjadi cluster-cluster.

4.2.2 Normalisasi Data

Dalam penelitian ini terdapat 4 variabel yang digunakan untuk menentukan cluster yang diinginkan. Variabel-varibel tersebut adalah (nama, jumlah pinjaman, jumlah tanggungan, jumlah pendapatan). Dari variabel tersebut terdapat 3 variabel yang dapat dihitung yaitu jumlah pinjaman, jumlah tanggungan, jumlah pendapatan yang telah dikonversi.

4.2.3 Hasil Hitungan K-Means

Di ketahui data penduduk migrasi berdasarkan factor ekonomi :

Jumlah data : 60

Jumlah cluster : 3

Jumlah atribut : 5

Tabel 4.2 data sampel penduduk migrasi

No	nama	Jumlah pinjaman	jumlah tanggungan	jumlah pendapatan
1	Husain Rajak	3	2	5
2	Ripon Yahya	6	2	3
3	Usman Yunus	5	2	3
4	Lian Yunus	3	1	4
5	Dian Simin	6	3	2
6	Dirman Hasim	2	4	5
7	Sunarti S. Ntuna	4	2	3
8	Bayur	5	2	10
9	Arianto Pola	6	2	2
10	Rosita A. Mohamad	3	5	6
11	Suleman Dude	3	0	2
12	Hadija Ibrahim	4	3	5

Ket : C1= tinggi

C2= sedang

C3= rendah

1. Menentukan jumlah cluster = 3
2. menentukan nilai centroid awal cluster

dalam menentukan nilai centroid data yang diambil secara acak.

Tabel 4.3 penentuan awal cluster literasi 1

cluster	J1	J2	J3
	Jumlah pinjaman	Jumlah tanggungan	Jumlah pendapatan
c1	5	2	3
c2	3	5	6
c3	3	0	2

3. perhitungan jarak pada cluster
 - a. dibawah ini adalah rumus yang akan digunakan dalam menentukan persamaan *euclidiean*.

$$(x,y) = \sqrt{\sum (x_i - y_i)^2}$$

$$= \sum \frac{ai}{n}$$

b menghitung jarak terdekat

- a. $C1 = \sqrt{(x1 - y1)^2 + (x2 - y2)^2 + (x3 - y3)^2}$
- b. $= \sqrt{(3 - 0)^2 + (0 - 0)^2 + (5 - 2)^2}$
- c. $= \sqrt{(3)^2 + (0)^2 + (3)^2}$
- d. $\sqrt{9 + 0 + 9}$
- e. $\sqrt{3}$

Dari hasil perhitungan jarak terdekat dengan pusat cluster akan diperoleh data yang terdekat dengan cluster terdekat dari cluster. hasil perhitungan jarak terdekat dengan pusat cluster akan mendapatkan data yang terdekat dengan cluster terdekat dari cluster.

Tabel 4.4 menentukan jarak literasi 1

j1	j2	j3	c1	c2	c3	jarak terdekat	cluster
3	2	5	2,828427125	3,16227766	3,605551275	2,828427125	1
6	2	3	1	5,196152423	3,741657387	1	1
5	2	3	0	4,69041576	3	0	1
3	1	4	2,449489743	4,472135955	2,236067977	2,236067977	3
6	3	2	1,732050808	5,385164807	4,242640687	1,732050808	1
2	4	5	4,123105626	1,732050808	5,099019514	1,732050808	2
4	2	3	1	4,358898944	2,449489743	1	1
5	2	10	7	5,385164807	8,485281374	5,385164807	2
6	2	2	1,414213562	5,830951895	3,605551275	1,414213562	1
3	5	6	4,69041576	0	6,403124237	0	2
3	0	2	3	6,403124237	0	0	3
4	2	5	2,236067977	3,31662479	3,741657387	2,236067977	1

Pada klaster yang telah melalui proses perhitungan akan diketahui bahwa satu klaster akan dikelompokkan dengan klaster yang sama. Dan dari data di atas

kita dapat menghitung ulang centroid untuk menentukan awal dari pusat cluster baru.

Berikut cara menghitung centroid baru

J1= jumlah pinjaman

$$c_1 = \frac{3+4+3+3}{4} = 3,2$$

J2= jumlah tanggungan

$$C_1 = \frac{0+2+1+1}{4} = 1$$

J3= Jumlah pendapatan

$$C_1 = \frac{5+3+5+4}{4} = 4,25$$

Ulangi perhitungan pada literasi ke 2 hingga setiap cluster tidak mengalami perubahan

Tabel 4.5 menentukan awal pusat cluster literasi 2

No	Nama	j1	j2	j3	Cluster	Cluster baru		
						C1	C2	C3
1	Husain Rajak	3	2	5	1	4,8571429	3,333333333	3
2	Ripon Yahya	6	2	3	1	2,1428571	3,666666667	0,5
3	Usman Yunus	5	2	3	1	3,2857143	7	3
4	Lian Yunus	3	1	4	3			
5	Dian Simin	6	3	2	1			
6	Dirman Hasim	2	4	5	2			
7	Sunarti S. Ntuna	4	2	3	1			
8	Bayur	5	2	10	2			
9	Arianto Pola	6	2	2	1			
10	Rosita A. Mohamad	3	5	6	2			
11	Suleiman Dude	3	0	2	3			
12	Hadija Ibrahim	4	2	5	1			

Tabel 4.6 awal pusat cluster literasi 2

Cluster	j1	j2	j3
c1	4,857142857	2,142857143	3,285714286
c2	3,333333333	3,666666667	7
c3	3	0,5	3

Tabel 4.7 menghitung jarak literasi 2

Nama	J1	J2	J3	c1	c2	c3	jarak terdekat	Cluster
Husain Rajak	3	2	5	2,531435	2,6246693	2,5	2,5	3
Ripon Yahya	6	2	3	1,1866606	5,0881125	3,35410197	1,186660552	1
Usman Yunus	5	2	3	0,3499271	4,6427961	2,5	0,349927106	1
Lian Yunus	3	1	4	2,2946255	4,027682	1,11803399	1,118033989	3
Dian Simin	6	3	2	1,9219463	5,7057476	4,03112887	1,921946292	1
Dirman Hasim	2	4	5	3,81458	2,4267033	4,15331193	2,426703296	2
Sunarti S. Ntuna	4	2	3	0,914732	4,3843155	1,80277564	0,914732034	1
Bayur	5	2	10	6,7173245	3,8151744	7,43303437	3,815174381	2
Arianto Pola	6	2	2	1,7261494	5,9066817	3,5	1,726149425	1
Rosita A. Mohamad	3	5	6	4,3565573	1,6996732	5,40832691	1,699673171	2
Suleman Dude	3	0	2	3,1134992	6,209312	1,11803399	1,118033989	3
Hadija Ibrahim	4	2	5	1,9219463	2,6874192	2,6925824	1,921946292	1

Pada perhitungan iterasi ke-2 terdapat perbedaan hasil cluster dengan literasi pertama. Karena itu perlu dilakukan perhitungan kembali literasi 3 hingga setiap cluster sudah tidak ada perubahan.

Tabel 4.8 literasi 3 awal centroid

Cluster	j1	j2	j3
c1	5,166666667	2,166666667	3
c2	3,333333333	3,666666667	7
c3	3	1	3,666666667

Tabel 4.9 mencari jarak terdekat literasi 3

Jumlah pinjaman	jumlah tanggungan	jumlah pendapatan	C1	C2	C3	jarak terdekat	cluster
3	2	5	2,95334086	2,62466929	1,66666667	1,666666667	3
6	2	3	0,84983659	5,08811251	3,23178657	0,849836586	1
5	2	3	0,23570226	4,64279609	2,33333333	0,23570226	1
3	1	4	2,65622958	4,02768199	0,33333333	0,333333333	3
6	3	2	1,54560308	5,70574759	3,9721251	1,545603083	1
2	4	5	4,16999867	2,4267033	3,43187671	2,426703296	2
4	2	3	1,1785113	4,38431548	1,56347192	1,178511302	1
5	2	10	7,00396713	3,81517438	6,71648056	3,815174381	2
6	2	2	1,31233465	5,90668172	3,57460176	1,312334646	1
3	5	6	4,66071048	1,69967317	4,63081466	1,699673171	2
3	0	2	3,22317993	6,209312	1,94365063	1,943650632	3
4	2	5	2,32139805	2,68741925	1,94365063	1,943650632	3

Tabel 4.10 awal pusat cluster lliterasi 4

cluster	j1	j2	j3
c1	5,4	2,2	2,6
c2	3,333333333	3,666666667	7
c3	3,25	1,25	4

Tabel 4.11 menentukan jarak literasi 4

j1	j2	j3	c1	c2	c3	jarak terdekat	cluster
3	2	5	3,4	2,62466929	1,27475488	1,274754878	3
6	2	3	0,74833148	5,08811251	3,02076149	0,748331477	1
5	2	3	0,6	4,64279609	2,15058132	0,6	1
3	1	4	3,02654919	4,02768199	0,35355339	0,353553391	3
6	3	2	1,16619038	5,70574759	3,82426464	1,166190379	1
2	4	5	4,53431362	2,4267033	3,18198052	2,426703296	2
4	2	3	1,46969385	4,38431548	1,45773797	1,457737974	3
5	2	10	7,4135012	3,81517438	6,29483916	3,815174381	2
6	2	2	0,87177979	5,90668172	3,48209707	0,871779789	1
3	5	6	5,01597448	1,69967317	4,25734659	1,699673171	2
3	0	2	3,31058907	6,209312	2,37170825	2,371708245	3
4	2	5	2,78567766	2,68741925	1,45773797	1,457737974	3

Tabel 4.12 awal centroid literasi 5

cluster	j1	j2	j3
c1	5,75	2,25	2,5
c2	3,333333333	3,666666667	7
c3	3,4	1,4	3,8

Tabel 4.13 menentukan jarak literasi 5

j1	j2	j3	C1	C2	C3	Jarak terdekat	cluster
3	2	5	3,724916106	2,624669291	1,4	1,4	3
6	2	3	0,612372436	5,088112507	2,785677655	0,612372436	1
5	2	3	0,935414347	4,642796092	1,886796226	0,935414347	1
3	1	4	3,372684391	4,027681991	0,6	0,6	3
6	3	2	0,935414347	5,70574759	3,544009029	0,935414347	1
2	4	5	4,834769901	2,426703296	3,18747549	2,426703296	2
4	2	3	1,837117307	4,384315479	1,166190379	1,166190379	3
5	2	10	7,541551564	3,815174381	6,431174076	3,815174381	2
6	2	2	0,612372436	5,906681716	3,218695388	0,612372436	1
3	5	6	5,232112384	1,699673171	4,23792402	1,699673171	2
3	0	2	3,588175024	6,209312003	2,315167381	2,315167381	3
4	2	5	3,061862178	2,687419249	1,469693846	1,469693846	3

Dari hasil perhitungan diatas telah ditemukan setiap cluster yang sudah tidak berpindah-pindah maka dari itu perhitungan diberhentikan pada literasi ke 4 karena hasil yang di dapatkan dari literasi ke 5 sudah tidak berpindah seperti yang ada pada literasi ke 4

4.2.4 hasil clustering

Setelah melalui proses perhitungan maka data tersebut diolah pada proses perhitungan, sehingga didapatkan hasil cluster yaitu pada C1 jumlah pinjaman ada 4 data dan pada C2 jumlah tanggungan ada 3 data, terakhir C3 memiliki 5 data. Dengan hasil tersebut, data yang dikelompokkan menggunakan algoritma k-means memiliki akurasi clustering yang tinggi. Hal ini akan sangat membantu dalam penelitian ini.

Dalam analisis data ini ada perbedaan cara kerja antara perhitungan manual dengan menggunakan tool, ketika melakukan perhitungan atau mengelompokan data yang sangat banyak dengan cara manual atau hanya menggunakan Microsoft exel ini hanya akan menguras waktu yang sangat lama dikarenakan banyaknya data dan proses mencari data yang sudah tidak ada perubahan itu memakan waktu yang lama sehingga memerlukan perangkat lunak pendukung untuk memudahkan analisis dalam pengelompokan data. Pada penelitian ini penulis menggunakan tool rapidminer untuk membantu, memudahkan dan mempercepat dalam proses pengelompokan data sehingga dapat menghasilkan data dengan akurasi yang tepat dengan cara cepat.

BAB V

PEMBAHASAN

5.1 Pembahasan Tools

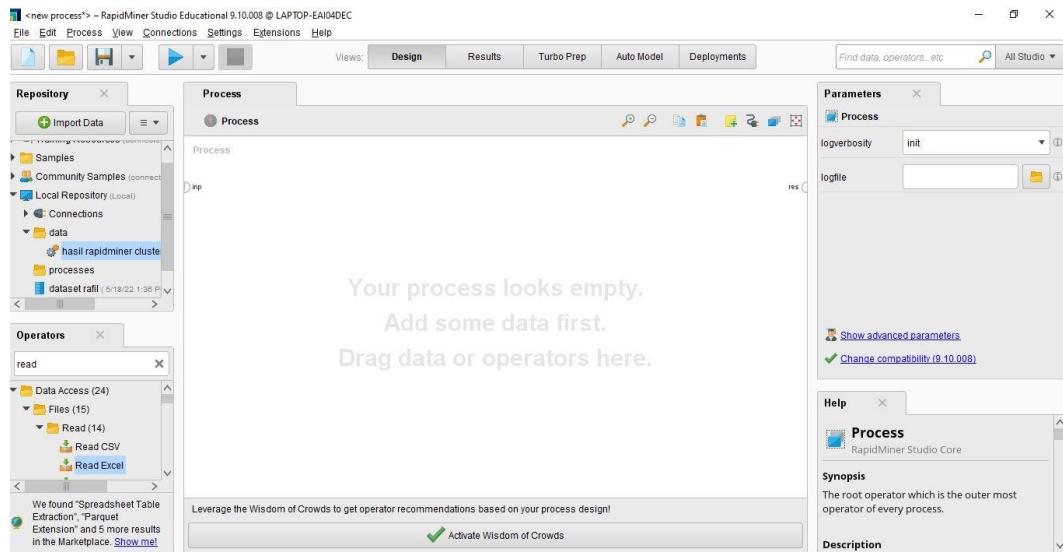
Dalam penelitian ini penulis menggunakan tools Rapidminer sebagai perangkat lunak pendukung untuk membantu dalam mengelompokan data yang banyak secara cepat dengan akurasi ketepatan yang tidak diragukan dalam mengelompokan data.

RapidMiner adalah platform perangkat lunak data ilmu pengetahuan yang dikembangkan oleh perusahaan dengan nama yang sama yang menyediakan lingkungan terpadu untuk pembelajaran mesin(*meachine learning*) pembelajaran mendalam (*deep learning*) penambangan teks (*teks mining*) dan analisis prediktif (*predictive analytics*). Aplikasi ini digunakan untuk aplikasi bisnis dan komersial serta untuk penelitian, pendidikan, pelatihan, pembuatan prototype dengan cepat, dan pengembangan aplikasi serta mendukung semua proses pembelajaran mesin termasuk persiapan data, visualisasi hasil, validasi pengoptimalan. Rapidminer dikembangkan dengan model open core. Terdapat dua versi edisi dari rapidminer, yang pertama adalah rapidminer basic edition (gratis) yang dibatasi kemampuannya pada 1 prosesor logis dan maksimal 10.000 baris data, tersedia dengan lisensi AGPL, yang kedua adalah rapidminer versi komersial dengan dimulai dari \$ 2.500.

Dikalangan akademik rapidminer digunakan oleh mahasiswa dosen atau peneliti dengan latar belakang system informasi untuk pemodelan system berbasis kecerdasan buatan (optimasi, pengenalan pola gambar/teks/grafik, peramalan).

Adapun tampilan saat membuka aplikasi RapidMiner sebagai berikut:

1. tampilan utama rapidminer



Gambar 5.1 tampilan utama

Gambar diatas merupakan gambar tampilan utama setelah terbuka rapidminer.

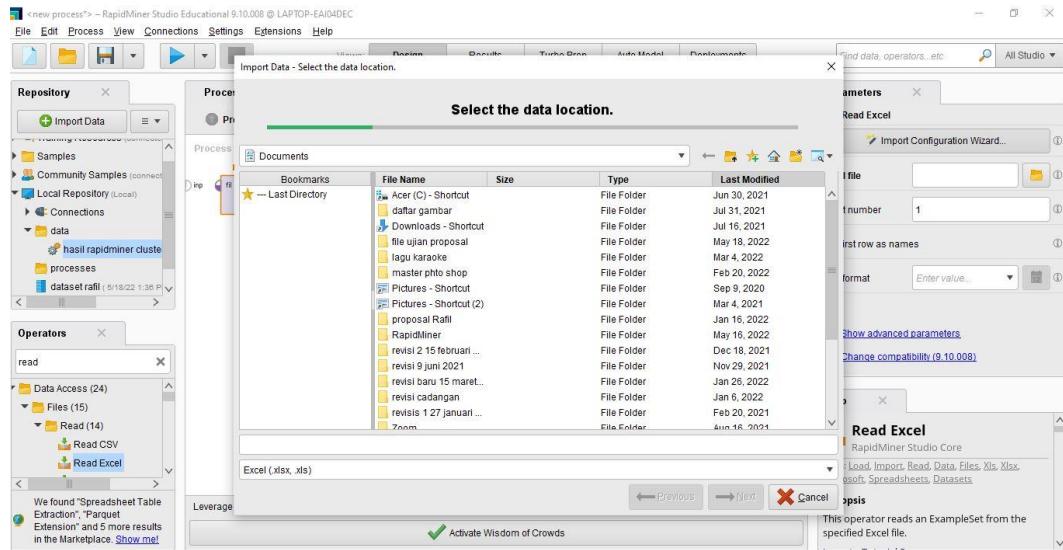
2. tampilan impor data



Gambar 5.2 tampilan impor data

Tampilan gambar diatas merupakan gambar impor data atau memasukan data analisis ke penyimpanan rapidminer

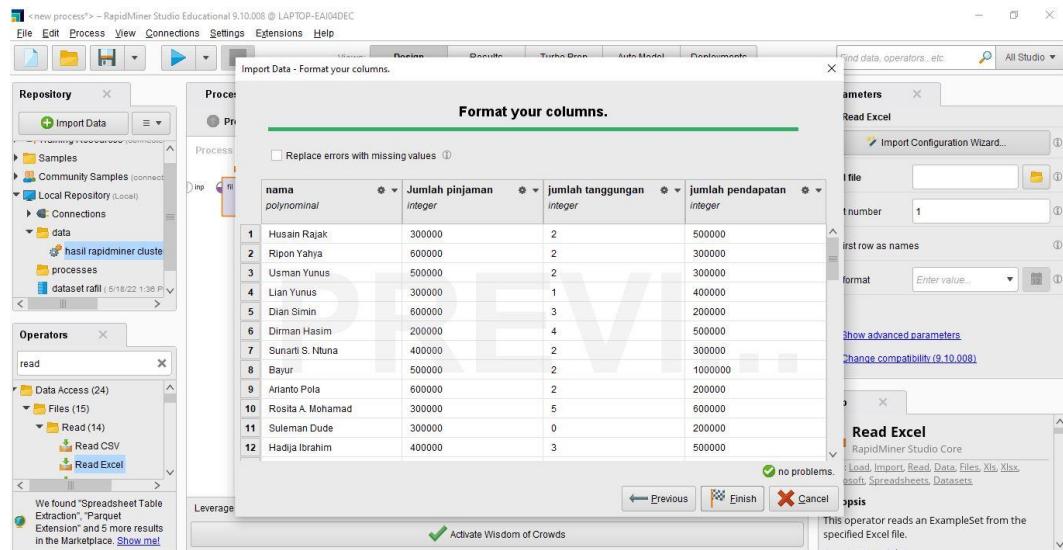
3. tampilan penyimpanan data



Gambar 5.3 penyimpanan data

Tampilan diatas merupakan tampilan lokasi data dan mencari data yang akan di impor ke rapidminer.

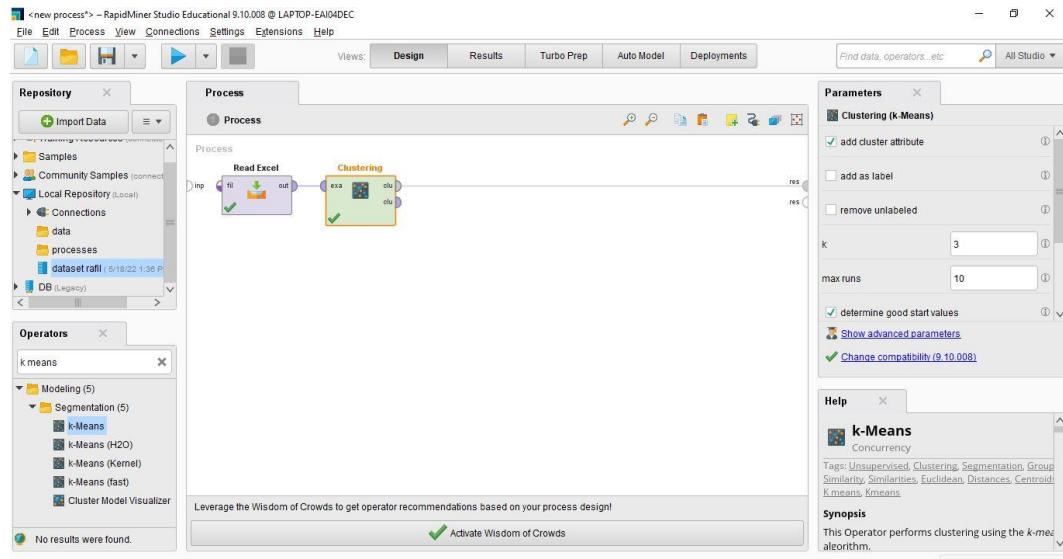
4. tampilan data yang digunakan



Gambar 5.4 data yang digunakan

Gambar ini merupakan tampilan data yang telah diimporkan dan akan digunakan dalam proses clustering pada rapidminer.

5. tampilan desain



Gambar 5.5 tampilan desain

Gambar diatas merupakan tampilan dari desain yang digunakan, tampilan terdiri dari yang sudah di import sebelumnya dan sistem operasi k-means yang digunakan. Setelah keduanya terhubung, lalu klik ikon segitiga biru (run) di kiri atas.

6. tampilan data result

The screenshot shows the RapidMiner Studio interface. The main area displays a table titled 'Result History' for 'ExampleSet (Clustering)'. The table has columns: 'Row No.', 'nama', 'Jumlah pin...', 'jumlah tang...', and 'jumlah pend...'. The data is as follows:

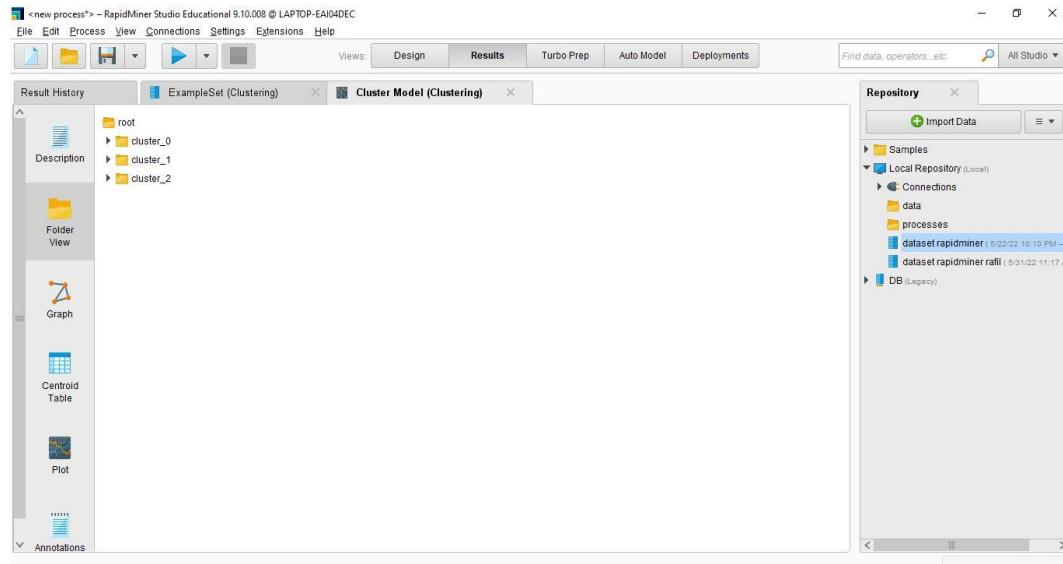
Row No.	nama	Jumlah pin...	jumlah tang...	jumlah pend...
1	Husain Rajak	300000	2	500000
2	Ripon Yahya	600000	2	300000
3	Usman Yunus	500000	2	300000
4	Lian Yunus	300000	1	400000
5	Dian Simrin	600000	3	200000
6	Dirman Hasim	200000	4	500000
7	Sunardi S. Nurdin	400000	2	300000
8	Bayur	500000	2	1000000
9	Arianto Pola	600000	2	200000
10	Rosita A. Moh...	300000	5	600000
11	Suleiman Dudo	300000	0	200000
12	Hadija Ibrahim	400000	3	500000
13	Adilla Djafar	200000	5	400000
14	Moh. Sandri...	300000	2	600000

The 'Annotations' panel on the left shows a list of annotations for each row. The 'Repository' panel on the right shows the local repository structure.

Gambar 5.6 tampilan data result

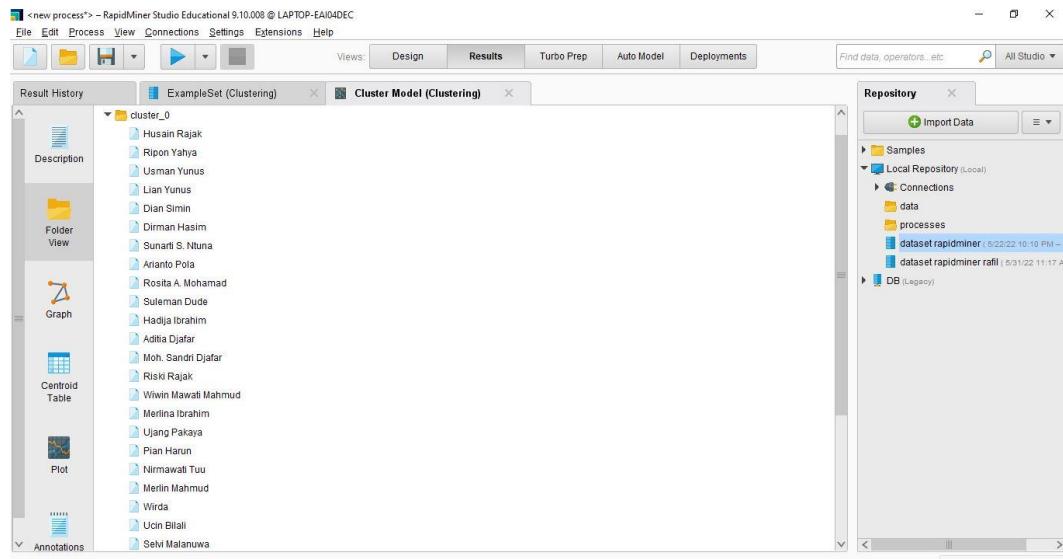
Data hasil adalah semua data yang sudah diolah, lengkap dengan data cluster yang sudah dibagi melalui proses cepat

7. tampilan folder cluster



Gambar 5.7 folder cluster

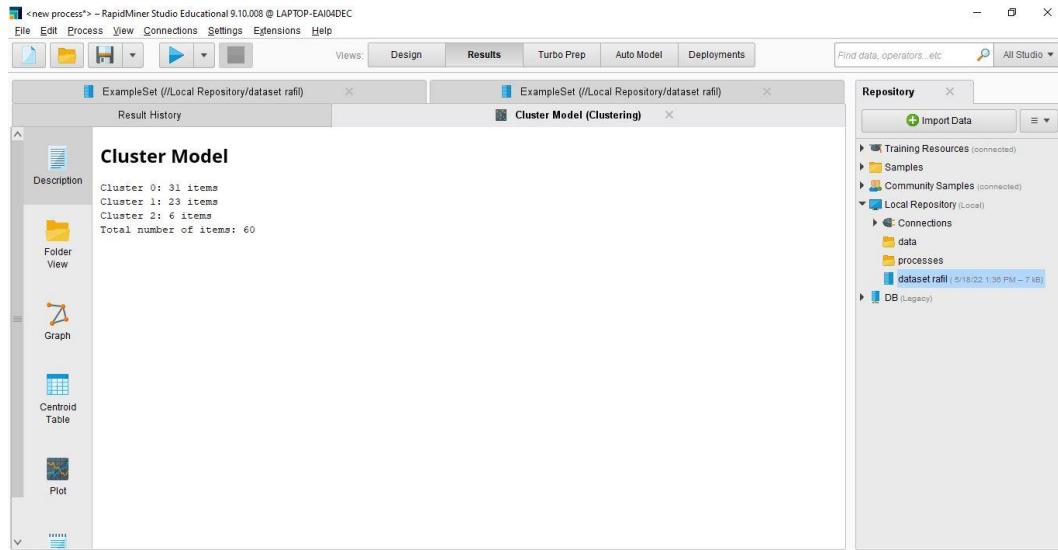
8. tampilan data pada folder



Gambar 5.8 tampilan data dalam folder

Tampilan ini adalah tampilan folder yang berisi data dari tiap-tiap cluster yang sudah melewati proses perhitungan di rapidminer

9. tampilan cluster model



Gambar 5.9 tampilan model cluster

Tampilan diatas merupakan tampilan dari hasil clustering semua data yang digunakan pada tampilan ini, dapat dilihat bahwa data telah dibagi menjadi 3 cluster dengan masing-masing data yang berbeda.

5.2 pembahasan hasil algoritma k-means

Dari hasil algoritma diketahui bahwa penggunaan algoritma k-means sangat membantu dan memudahkan penelitian ini, dengan keunggulan akurasinya yang presisi mampu mengatasi masalah penjawaban dan penyelesaian data pengelompokan. Data yang digunakan dalam penelitian ini berjumlah 60 data. Data ini dikelompokkan menjadi 3 cluster yaitu cluster pertama C1 jumlah pinjaman 31 data, cluster kedua C2 23 data dan terakhir cluster ketiga data C3 Total pendapatan ada 6 data.

Di bawah ini adalah tabel hasil dari seluruh proses perhitungan cluster.

Tabel 5.1 data hasil cluster

No	Nama	Jumlah pinjaman	jumlah tanggungan	jumlah pendapatan	Cluster
1	Husain Rajak	3	2	5	cluster_0
2	Ripon Yahya	6	2	3	cluster_0
3	Usman Yunus	5	2	3	cluster_0
4	Lian Yunus	3	1	4	cluster_0
5	Dian Simin	6	3	2	cluster_0
6	Dirman Hasim	2	4	5	cluster_0
7	Sunarti S. Ntuna	4	2	3	cluster_0
8	Bayur	5	2	10	cluster_1
9	Arianto Pola	6	2	2	cluster_0
10	Rosita A. Mohamad	3	5	6	cluster_0
11	Suleman Dude	3	0	2	cluster_0
12	Hadija Ibrahim	4	3	5	cluster_0
13	Aditia Djafar	2	5	4	cluster_0
14	Moh. Sandri Djafar	3	2	6	cluster_0
15	Riski Rajak	5	3	2	cluster_0
16	Wiwin Mawati Mahmud	4	2	3	cluster_0
17	Merlina Ibrahim	2	1	5	cluster_0
18	Ujang Pakaya	3	2	6	cluster_0
19	Pian Harun	4	1	3	cluster_0
20	Salma Yusuf	5	3	7	cluster_1
21	Abdulatif Eka Putra Harun	6	2	8	cluster_1
22	Fauzan Dulanimo	7	1	5	cluster_2
23	Nirmawati Tuu	5	2	6	cluster_0
24	Merlin Mahmud	3	0	5	cluster_0
25	Aril Pratama Arfan	4	1	10	cluster_1
26	Halma Taib	6	2	8	cluster_1
27	Fauzan Dulanimo	7	0	5	cluster_2
28	Deden Pakaya	5	1	9	cluster_1
29	Wirda	4	2	3	cluster_0
30	Ucin Bilali	3	1	5	cluster_0
31	Moh. Rafli Musa	5	2	7	cluster_1
32	Selvi Malanuwa	3	1	4	cluster_0
33	Hawariya Mohamad	6	1	8	cluster_1
34	Raflin Talib	5	3	9	cluster_1
35	Selvi Noe	6	4	5	cluster_0
36	Rahmat Aditia Yusuf	4	1	3	cluster_0

37	Rahmatia Muhamad	3	0	8	cluster_1
38	Yunarti Laudji	4	1	5	cluster_0
39	Karim Tahidina	6	1	7	cluster_1
40	Saldi Wombohuke	5	2	6	cluster_0
41	Ulyan Ayuba	6	3	8	cluster_1
42	Maya Rifka Abas	6	1	8	cluster_1
43	Djamin Midu	5	2	10	cluster_1
44	Wisma Puluhulawa	5	0	6	cluster_0
45	Rosna Madelu	7	2	7	cluster_2
46	Risman Podomi	1	2	10	cluster_2
47	Maryam Hipi	4	1	7	cluster_1
48	Ruslini Hasan Jafar	6	1	5	cluster_0
49	Erwin	10	3	7	cluster_2
50	Nur Intan	5	1	9	cluster_1
51	Rian	8	2	5	cluster_2
52	Melisa Sukamdi	4	2	10	cluster_1
53	Ayu Anggraini Ismail	4	2	7	cluster_1
54	Cicin Seu	4	1	10	cluster_1
55	Novita Laubunu	5	2	7	cluster_1
56	Abd Rahman Sedi	4	1	10	cluster_1
57	Romi Tintia	5	0	10	cluster_1
58	Al Fahri Tahidina	6	7	3	cluster_0
59	Denis Rajak	4	0	7	cluster_1
60	Yulan Djafar	5	2	8	cluster_0

BAB VI

KESIMPULAN DAN SARAN

6.1 Kesimpulan

Berdasarkan penelitian yang dilakukan oleh peneliti maka bisa disimpulkan sebagai berikut

1. dalam penelitian ini peneliti menggunakan metode algoritma k-means yang sangat membantu dalam mengolah data selain itu dapat mengelompokan data yang banyak. Penggunaan metode ini sangat efektif dimana data yang digunakan dalam penelitian ini berjumlah 60 data. Data tersebut dikelompokkan menjadi data penduduk berpenghasilan rendah (C1) 31 data, penduduk berpenghasilan sedang (C2) 23 data, data penduduk berpenghasilan tinggi (C3) 6 data. Dengan hasil tersebut penelitian ini dapat diselesaikan dengan baik
2. clustering data penduduk migrasi berdasarkan faktor ekonomi menggunakan metode k-means telah mendapatkan hasil data cluster dari variabel yang ditentukan serta dari pengelompokan tersebut mendapatkan akurasi yang baik dan data tersebut akan direkomendasikan untuk mendapatkan bantuan dari pemerintah desa bagi penduduk yang berpenghasilan rendah.

6.2 Saran

Berdasarkan penelitian yang telah dilakukan karya ilmiah ini masih banyak terdapat kekurangan sehingga kritik dan saran yang membangun sangat diperlukan dari si pembaca guna menyempurnakan karya ilmiah ini dan kekurangan dari penelitian ini setidaknya bisa memacu kepada pembaca untuk menciptakan ide yang lebih kreativitas lagi .

DAFTAR PUSTAKA

- [1] S. Dewi, D. Listyowati, and B. E. Napitupulu, “Dampak Ekonomi Dari Migrasi : Kasus Di Indonesia,” *J. Ekon.*, vol. 10, no. 2, pp. 47–56, 2019.
- [2] S. Riama and Chafid, “Algorithma K-Means Clustering Strategi Pemasaran Penerimaan Mahasiswa Baru Universitas Satya Negara Indondesia [Algorithm K-Means Clustering Strategy Marketing Admission Universitas Satya Negara Indonesia],” *Semin. Nas. Cendekiawan ke 4*, no. 2, pp. 685–690, 2018, [Online]. Available: <https://journal.ubm.ac.id/index.php/alu/article/view/1367>.
- [3] S. Ashari, S. Khansa, C. H. M. Surudin, and I. N. Isnainiyah, “KLUSTERING JUMLAH PENDUDUK KOTA BANDUNG BERDASARKAN JENIS KELAMIN PER KECAMATAN PADA TAHUN 2012 DENGAN METODE K-MEANS Sarah Ashari , Salsabilah Khansa , Celvin Habib Maulana Surudin , Ika Nurlaili Isnainiyah,” pp. 22–28, 2018.
- [4] D. Purnomo, “Fenomena Migrasi Tenaga Kerja Dan Perannya Bagi Pembangunan Daerah Asal: Studi Empiris Di Kabupaten Wonogiri,” *J. Ekon. Pembang. Kaji. Masal. Ekon. dan Pembang.*, vol. 10, no. 1, p. 84, 2009, doi: 10.23917/jep.v10i1.810.
- [5] Y. Mardi, “Data Mining : Klasifikasi Menggunakan Algoritma C4.5,” *Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017, doi: 10.22202/ei.2016.v2i2.1465.
- [6] D. Astika, “Penerapan Data Mining Untuk Menganalisis Penjualan Barang Dengan Pada Supermarket Sejahtera,” *Astik. D., Stud. P., Inform. T., Malikussaleh, U. (n.d.). PENERAPAN DATA Min. UNTUK MENGANALISIS PENJUALAN BARANG DENGAN PADA Supermark. SEJAHTERA.*, vol. Vol. 6 No., 2017.
- [7] Y. D. Darmi and A. Setiawan, “Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk,” *J. Media Infotama*, vol. 12, no. 2, pp. 148–157, 2017, doi: 10.37676/jmi.v12i2.418.
- [8] R. A. Asroni, “Penerapan Metode K-Means Untuk Clustering Mahasiswa

- Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang,” *Ilm. Semesta Tek.*, vol. 18, no. 1, pp. 76–82, 2015.
- [9] M. W. Talakua, Z. A. Leleury, and A. W. Talluta, “Acluster Analysis By Using K-Means Method for Grouping of District/City in Maluku Province Industrial Based on Indicators of Maluku Development Index in 2014,” *Barekeng J. Ilmu Mat. dan Terap.*, vol. 11, no. 2, pp. 119–128, 2017.
- [10] A. Asroni, H. Fitri, and E. Prasetyo, “Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokkan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik),” *Semesta Tek.*, vol. 21, no. 1, pp. 60–64, 2018, doi: 10.18196/st.211211.
- [11] N. Dwitri, J. A. Tampubolon, S. Prayoga, F. Ilmi Zer, and D. Hartama, “Penerapan Algoritma K-Means Dalam Menentukan Tingkat Kepuasan Pembelajaran Online Pada Masa Pandemi Covid-19 di Indonesia,” *Jti (Jurnal Teknol. Informasi)*, vol. 4, no. 1, pp. 101–105, 2020.

RIWAYAT HIDUP MAHASISWA

NAMA : RAFIL TILAUNGGA
NIM : T3117234
TEMPAT TANGGAL LAHIR : DAMBALO 21 MEY 1998
AGAMA : ISLAM
EMAIL : rafiltelahunga21@gmail.com

**RIWAYAT PENDIDIKAN**

1. Tahun 2010 menyelesaikan pendidikan di Sekolah Dasar Negeri 2
Dambalo
2. Tahun 2013 Menyelesaikan pendidikan di Sekolah Menengah Pertama
Negeri 3 Kwandang
3. Tahun 2016 Menyelesaikan pendidikan di Sekolah Menengah Atas Negeri
1 Gorontalo Utara
4. Tahun 2017 Telah di terima menjadi mahasiswa Perguruan Tinggi di
Universitas Ichsan Gorontalo

