



DATA MINING

Penerapan Algoritma K-Means Clustering dan K-Medoids Clustering

Mochamad Wahyudi ■ Masitha ■ Risna Saragih ■ Solikhun

Data Mining :
Penerapan Algoritma K-Means
Clustering dan K-Medoids
Clustering

UU 28 tahun 2014 tentang Hak Cipta

Fungsi dan sifat hak cipta Pasal 4

Hak Cipta sebagaimana dimaksud dalam Pasal 3 huruf a merupakan hak eksklusif yang terdiri atas hak moral dan hak ekonomi.

Pembatasan Perlindungan Pasal 26

Ketentuan sebagaimana dimaksud dalam Pasal 23, Pasal 24, dan Pasal 25 tidak berlaku terhadap:

- a. penggunaan kutipan singkat Ciptaan dan/atau produk Hak Terkait untuk pelaporan peristiwa aktual yang ditujukan hanya untuk keperluan penyediaan informasi aktual;
- b. Penggandaan Ciptaan dan/atau produk Hak Terkait hanya untuk kepentingan penelitian ilmu pengetahuan;
- c. Penggandaan Ciptaan dan/atau produk Hak Terkait hanya untuk keperluan pengajaran, kecuali pertunjukan dan Fonogram yang telah dilakukan Pengumuman sebagai bahan ajar; dan
- d. penggunaan untuk kepentingan pendidikan dan pengembangan ilmu pengetahuan yang memungkinkan suatu Ciptaan dan/atau produk Hak Terkait dapat digunakan tanpa izin Pelaku Pertunjukan, Produser Fonogram, atau Lembaga Penyiaran.

Sanksi Pelanggaran Pasal 113

1. Setiap Orang yang dengan tanpa hak dan/atau tanpa izin Pencipta atau pemegang Hak Cipta melakukan pelanggaran hak ekonomi Pencipta sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf c, huruf d, huruf f, dan/atau huruf h untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 3 (tiga) tahun dan/atau pidana denda paling banyak Rp500.000.000,00 (lima ratus juta rupiah).
2. Setiap Orang yang dengan tanpa hak dan/atau tanpa izin Pencipta atau pemegang Hak Cipta melakukan pelanggaran hak ekonomi Pencipta sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf a, huruf b, huruf e, dan/atau huruf g untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 4 (empat) tahun dan/atau pidana denda paling banyak Rp1.000.000.000,00 (satu miliar rupiah).

Data Mining : Penerapan Algoritma K-Means Clustering dan K-Medoids Clustering

Penulis:

Mochamad Wahyudi
Masitha
Risna Saragih
Solikhun

Penerbit Yayasan Kita Menulis

Data Mining: Penerapan Algoritma K-Means Clustering dan K-Medoids Clustering

Copyright © Yayasan Kita Menulis, 2020

Penulis:

Mochamad Wahyudi

Masitha

Risna Saragih

Solikhun

Editor: Janner Simarmata

Desain Cover: Tim Kreatif Kita Menulis

Penerbit

Yayasan Kita Menulis

Web: kitamenulis.id

e-mail: press@kitamenulis.id

Kontak WA: +62 821-6453-7176

Mochamad Wahyudi, dkk.

Data Mining: Penerapan Algoritma K-Means Clustering dan K-Medoids Clustering

Yayasan Kita Menulis, 2020

xii; 64 hlm; 16 x 23 cm

ISBN: 978-623-7645-81-8 (print)

E-ISBN: 978-623-7645-82-5 (online)

Cetakan 1, Mei 2020

- I. Data Mining: Penerapan Algoritma K-Means Clustering dan K-Medoids Clustering
- II. Yayasan Kita Menulis

Katalog Dalam Terbitan

Hak cipta dilindungi undang-undang

Dilarang memperbanyak maupun mengedarkan buku tanpa
ijin tertulis dari penerbit maupun penulis

Kata Pengantar

Puji syukur kehadiran Allah, SWT atas limpahan rahmat dan karunia-Nya sehingga Buku Data Mining Penerapan Algoritma K-Means Clustering dan K-Medoids Clustering telah dapat kami selesaikan. Buku ini membahas tentang penerapan Algoritma K-Means dan Algoritma K-Medoids dalam mengelompokkan data.

Semoga buku ini dapat membantu para mahasiswa yang tertarik melakukan penelitian tentang Data Mining dalam rangka membuat tugas akhir baik berupa skripsi maupun tesis untuk menyelesaikan studinya.

Buku ini membahas :

Bab 1 : Data Mining

Bab 2 : Konsep Algoritma K-Means Clustering

Bab 3 : Konsep Algoritma K-Medoids Clustering

Bab 4 : Rapidminer dan Pemasangan Perangkat Lunak

Bab 5 : Penerapan Algoritma K-Means Clustering

Bab 6 : Penerapan Algoritma K-Medoids Clustering

Terimakasih disampaikan kepada semua pihak yang telah memberikan kontribusi dalam penyelesaian buku ini. Pembuatan buku ini merupakan kolaborasi antara dosen Universitas Bina Sarana Informatika di Jakarta dan dosen serta mahasiswa dari AMIK & STIKOM Tunas Bangsa di Pematang Siantar, Sumatera Utara. Semoga kedepan kerjasama ini lebih baik lagi sehingga dapat menghasilkan buku-buku yang lebih berkualitas.

Kami menyadari masih banyak terdapat kekurangan dalam buku ini, untuk itu kritik dan saran yang membangun terhadap penyempurnaan buku ini sangatlah diharapkan. Semoga buku ini dapat memberi manfaat bagi semua pihak yang membutuhkan.

Jakarta, 2 Mei 2020

Penulis

Daftar Isi

Kata Pengantar	v
Daftar Isi	vii
Daftar Gambar	ix
Daftar Tabel	xi

Bab 1 Pengantar Data Mining

1.1 Data Mining	1
1.2 Tahapan Data Mining	3

Bab 2 Algoritma K-Means Clustering

2.1 Pendahuluan	5
2.2 K-Means Clustering	6

Bab 3 Algoritma K-Medoid Clustering

3.1 K-Means Clustering	9
3.2 Langkah-langkah K-Medoids Clustering	10

Bab 4 Rapidminer dan Pemasangan Perangkat Lunak

4.1 Rapidminer	13
4.2 Pemasangan Perangkat Lunak RapidMiner	14

Bab 5 Penerapan Algoritma K-Means Clustering

5.1 Pengolahan Data	19
5.2 Pengujian dengan RapidMiner	36
5.2.1 Tampilan Menu Utama RapidMiner	36
5.2.2 Tampilan Menu Excel Rapidminer	37
5.2.3 Tampilan Read Excel RapidMiner	38
5.2.4 Pemrosesan RapidMiner	40
5.2.5 Tampilan Keluaran Sistem	41
5.2.6 Validasi data	42

Bab 6 Penerapan Algoritma K-Medoids Clustering

6.1 Analisis Data Manual K-Medoids	45
6.2 Pengujian Dengan RapidMiner	53

6.3 Prosedur Kerja RapidMiner.....	53
6.4 Validasi Data.....	59
Daftar Pustaka.....	61
Biodata Penulis	63

Daftar Gambar

Gambar 4.1: Tampilan cara mengecek kapasitas Perangkat Keras	14
Gambar 4.2: Tampilan End User License Agreement.....	15
Gambar 4.3: Tampilan pembuatan account.....	15
Gambar 4.4: Tampilan pengisian account	16
Gambar 4.5: Tampilan confirmation account	16
Gambar 4.6: Tampilan akhir pemasangan perangkat lunak RapidMiner	17
Gambar 4.7: Tampilan Jendela Kerja Rapidminer.....	17
Gambar 5.1: Tampilan Awal RapidMiner 5.3.....	37
Gambar 5.2: “New Process” untuk Import Data Excel RapidMiner.....	38
Gambar 5.3: Input Data Read Excel RapidMiner	38
Gambar 5.4: Tampilan Read Excel RapidMiner.....	39
Gambar 5.5: Tampilan Data Import awal RapidMiner.....	39
Gambar 5.6: Tampilan Data Import kedua RapidMiner	40
Gambar 5.7: Pemrosesan Data	40
Gambar 5.8: Nilai Kluster Model RapidMiner	41
Gambar 5.9: Hasil Akhir Pengelompokkan RapidMiner	41
Gambar 6.1: Tampilan Awal Rapidminer 5.3	53
Gambar 6.2: Tampilan New Process.....	54
Gambar 6.3: Input Data Excel	54
Gambar 6.4: Pemrosesan Data Tahap 1.....	55
Gambar 6.5: Alur proses import step 1	55
Gambar 6.6: Data import data tahap 2	55
Gambar 6.7: Data import tahap 3	56
Gambar 6.8: Data import Tahap 4.....	56
Gambar 6.9: Pemrosesan Data Tahap 1.....	57
Gambar 4.10: Pemrosesan Data Tahap 1.....	57
Gambar 6.11: Nilai kluster Model RapidMiner	58
Gambar 6.12: Folder view RapidMiner	58

Daftar Tabel

Tabel 5.1: Persentase wanita yang berumur 15 s/d 49 tahun dan berstatus kawin yang sedang menggunakan alat kontrasepsi untuk keluarga berencana (Persen)	20
Tabel 5.2: Nilai Rata-Rata Persentase wanita yang berumur 15 s/d 49 tahun dan berstatus kawin yang sedang menggunakan alat kontrasepsi untuk keluarga berencana (Persen)	21
Tabel 5.3: Nilai Centroid Awal (Iterasi 1)	22
Tabel 5.4: Hasil Perhitungan Centroid pada Setiap Kluster (Iterasi 1).....	23
Tabel 5.5: Hasil Perhitungan Jarak Data pada Iterasi 1	24
Tabel 5.6: Hasil Kluster pada Iterasi 1	25
Tabel 5.7: Nilai Centroid Baru (Iterasi 2)	26
Tabel 5.8: Hasil Perhitungan Centroid pada Setiap Kluster (Iterasi 2).....	27
Tabel 5.9: Hasil Perhitungan jarak Data pada Iterasi 2.....	28
Tabel 5.10: Hasil Kluster pada Iterasi 2	29
Tabel 5.11: Nilai Centroid Baru (Iterasi 3)	29
Tabel 5.12: Hasil Perhitungan Centroid pada Setiap Kluster (Iterasi 3).....	30
Tabel 5.13: Hasil Perhitungan jarak Data pada Iterasi 3.....	31
Tabel 5.14: Hasil Kluster pada Iterasi 3	32
Tabel 5.15: Nilai Centroid Baru (Iterasi 4)	33
Tabel 5.16: Hasil Perhitungan Centroid pada Setiap Kluster (Iterasi 4).....	33
Tabel 5.17: Hasil Perhitungan jarak Data pada Iterasi 4.....	34
Tabel 5.18: Hasil Kluster pada Iterasi 4	35
Tabel 5.19: Hasil Perhitungan K-Means dengan Aplikasi RapidMiner.....	42
Tabel 6.1: Data Awal.....	45
Tabel 6.2: Medoids Awal.....	46
Tabel 6.3: Hasil Perhitungan Algoritma K-Medoids Iterasi ke-1	46
Tabel 6.4: Hasil Perhitungan Algoritma K-Medoids Iterasi ke-2	48
Tabel 6.5: Medoid Baru (Non-Medoid 2) Iterasi ke-3.....	49
Tabel 6.6: Data Perhitungan Algoritma K-Medoids Hasil Iterasi ke-3	50
Tabel 6.7: Hasil Pengklasteran dengan K-Medoids Clustering	52
Tabel 6.8: Hasil Perhitungan K-Means dengan Perangkat Lunak RapidMiner	59

Bab 1

Pengantar Data Mining

1.1 Data Mining

Data mining merupakan proses iteratif dan interaktif untuk menemukan pola-pola atau model baru yang shahih (sempurna), bermanfaat dan dapat dimengerti dalam suatu database yang sangat besar (massive databases). Data mining berisi pencarian trend atau pola yang diinginkan dalam basisdata besar untuk membantu pengambilan keputusan diwaktu yang akan datang (Syahdan dan Sindar, 2018).

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai basisdata besar. Istilah data mining memiliki hakikat sebagai disiplin ilmu yang tujuan utamanya adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki. Data mining, sering juga disebut sebagai *Knowledge Discovery in Database* (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar (Mustafa, Ramadhan dan Thenata, 2018).

Kemampuan data mining untuk mencari informasi bisnis yang berharga dari basisdata yang sangat besar, dapat dianalogikan dengan penambangan logam mulia dari lahan sumbernya, teknologi ini dipakai untuk:

1. Prediksi trend dan sifat-sifat bisnis, di mana data mining mengotomatisasi proses pencarian informasi memprediksi di dalam basis data yang besar.

2. Penemuan pola-pola yang tidak diketahui sebelumnya, di mana data mining “menyapu” basis data, kemudian mengidentifikasi pola-pola yang sebelumnya tersembunyi dalam satu sapuan (Angga Ginanjar Maburur, 2012).

Data mining bukanlah suatu bidang yang sama sekali baru. Salah satu kesulitan untuk mendefinisikan data mining adalah kenyataan bahwa data mining mewarisi banyak aspek dan teknik dari bidang-bidang ilmu yang sudah mapan terlebih dahulu. Berawal dari beberapa disiplin ilmu, data mining bertujuan untuk memperbaiki teknik tradisional sehingga bisa menangani:

- a Jumlah data yang sangat besar
- b Dimensi data yang tinggi
- c Data yang heterogen dan berbeda sifat Pengelompokan

Data mining dibagi menjadi beberapa kelompok, yaitu :

1. Deskripsi

Deskripsi merupakan cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data yang dimiliki.

2. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model yang dibangun menggunakan record lengkap yang menyediakan nilai variabel target sebagai nilai prediksi.

3. Prediksi

Penerapan data mining untuk analisis pola pembelian konsumen menerka sebuah nilai yang belum diketahui dan juga memperkirakan nilai untuk masa mendatang.

4. Klasifikasi

Dalam klasifikasi terdapat target variabel kategori, misalnya penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu : tinggi, sedang dan rendah.

5. Pengklasteran

Merupakan pengelompokan record, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan.

6. Asosiasi

Asosiasi bertugas menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja (Maulana dan Fajrin, 2018).

1.2 Tahapan Data Mining

Tahapan data mining dibagi menjadi tujuh bagian, yaitu :

1. Pembersihan data (data cleaning) sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus KDD. Proses cleaning mencakup antara lain membuang empat duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Juga dilakukan proses *enrichment*, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal
2. Integrasi data (data integration), merupakan penggabungan data dari berbagai basisdata ke dalam satu basisdata baru. Tidak jarang data yang diperlukan untuk data mining tidak hanya berasal dari satu basisdata tetapi juga berasal dari beberapa basisdata atau file teks. Integrasi data dilakukan pada atribut-aribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggandan lainnya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk ternyata menggabungkan produk dari kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada.

3. Seleksi Data (Data Selection), data yang ada pada basisdata sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari basisdata. Sebagai contoh, sebuah kasus yang meneliti faktor kecenderungan orang membeli dalam kasus market basket analisis, tidak perlu mengambil nama pelanggan, cukup dengan id pelanggan saja.
4. Transformasi data (Data Transformation), data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining. Beberapa metode data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan clustering hanya bisa menerima masukan data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut transformasi data.
5. Proses mining, adalah sebuah proses yang paling utama pada saat metode diterapkan untuk mencari pengetahuan tersembunyi dan berharga dari data.
6. Evaluasi pola (pattern evaluation), untuk mengidentifikasi pola-pola menarik ke dalam *knowledge based* yang ditemukan. Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai.
7. Presentasi pengetahuan (Knowledge Presentation), merupakan penyajian dan visualisasi pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau lima aksi dari hasil analisis yang didapat. Ada kalanya hal ini harus melibatkan orang yang tidak memahami data mining. Karenanya presentasi hasil data mining dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses data mining. Dalam presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil data mining (Eska, 2016).

Bab 2

Algoritma K-Means Clustering

2.1 Pendahuluan

Analisa klaster merupakan kegiatan yang menganalisa kumpulan objek untuk menemukan kesamaan dan perbedaan sehingga membentuk suatu klaster yang sama maupun berbeda dengan objek tersebut (Hermawati, 2013). Pengklasteran bertujuan untuk mengelompokkan dan memahami struktur data. Klasterisasi hanya tahap awal untuk kemudian dilanjutkan dengan pengolahan inti dan pelabelan kelas pada tiap kelompok. Hal ini nantinya nantinya dapat digunakan sebagai data latih.

Menurut Jhonson dan Wichern (2005), metode K-Means digunakan sebagai alternatif metode klaster untuk data dengan ukuran yang lebih besar. Hal ini dikarenakan metode ini memiliki kecepatan yang lebih tinggi jika dibandingkan dengan metode hirarki. Metode K-Means dapat digunakan untuk menjelaskan algoritma dalam penentuan suatu objek ke dalam klaster tertentu berdasarkan rata-rata terdekat (Ramdhani dan Hoyyi, 2015). K-Means merupakan salah satu metode data clustering non hierarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster atau kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lainnya.

K-Means adalah metode clustering berbasis jarak yang membagi data ke dalam sejumlah cluster dan algoritma ini hanya bekerja pada atribut numerik. Algoritma K-Means termasuk *partitioning clustering* yang memisahkan data ke k daerah bagian yang terpisah. Algoritma K-Means sangat terkenal karena kemudahan dan kemampuannya untuk mengkluster data yang besar dan data outlier dengan sangat cepat. Dalam Algoritma K-Means, setiap data harus termasuk ke kluster

tertentu dan bisa dimungkinkan bagi setiap data yang termasuk kluster tertentu pada suatu tahapan proses, pada tahapan berikutnya berpindah kekluster lainnya.

Pada dasarnya penggunaan algoritma dalam melakukan proses clustering tergantung dari data yang ada dan kongklusi yang ingin dicapai. Untuk itu digunakan Algoritma K-Means yang di dalamnya membuat aturan sebagai berikut :

1. Jumlah kluster perlu dimasukan
2. Hanya memiliki atribut bertipe numerik.

2.2 K-Means Clustering

Algoritma K-Means merupakan metode non-hierarki yang pada awalnya mengambil sebagian banyaknya komponen populasi untuk dijadikan pusat kluster awal. Pada tahap ini pusat kluster dipilih secara acak dari sekumpulan populasi data. Berikutnya K-Means menguji masing-masing komponen di dalam populasi data dan menandai komponen tersebut ke salah satu pusat kluster yang telah didefinisikan tergantung dari jarak minimum antar komponen dengan tiap-tiap kluster. Posisi pusat kluster akan dihitung kembali sampai semua komponen data digolongkan ke dalam tiap-tiap pusat kluster dan terakhir akan terbentuk posisi pusat kluster yang baru (Darmi dan Setiawan, 2016).

Algoritme klastering K-Means dapat membagi data berdasarkan jarak antar data pada kelompok yang telah ditetapkan. Algoritma ini bergantung pada fungsi untuk mengukur data yang mempunyai ciri khas sama. Jarak itu sendiri dihitung menggunakan fungsi euclidean. Kemudian data dimasukkan dalam kelompok yang mempunyai jarak terdekat.

Langkah-langkah pengelompokan data adalah :

1. Pilih jumlah klaster.
2. Inisialisasi awal dan pusat klaster dilakukan secara random.
3. Setiap data ditempatkan ke pusat klaster terdekat berdasarkan jarak antar obyek. Pada tahap ini jarak dihitung dengan menentukan kemiripan atau ketidakmiripan data dengan Metode Jarak Euclidean

(Euclidean Distance) dengan rumus seperti di bawah ini (Witanto, Ratnawati dan Anam, 2019) :

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

dimana :

$d(x, y)$ = ukuran ketidakmiripan

$x_i = (x_1, x_2, \dots, x_i)$ yaitu variabel data

$y_i = (y_1, y_2, \dots, y_j)$ yaitu variabel pada titik pusat.

4. Hitung pusat klaster yang baru dengan keanggotaan yang baru dengan cara menghitung rata-rata obyek pada klaster. Penghitungan bisa juga dengan menggunakan median.
5. Hitung kembali jarak tiap objek dengan pusat klaster yang baru, hingga klaster tidak berubah, maka proses pengklasteran selesai.

Bab 3

Algoritma K-Medoid Clustering

3.1 K-Means Clustering

Metode k-medoid dikembangkan oleh Leonard Kaufman dan Peter J. Rousseeuw pada tahun 1987. Algoritma K-Medoid sering disebut juga Algoritma Partitoning Around Medoid (PAM). Metode K-Medoid memiliki kesamaan dengan Metode K-Means yaitu sama-sama termasuk Metode Partitioning. Metode Partitioning merupakan metode pengelompokan data ke dalam sejumlah cluster tanpa adanya struktur hirarki antara satu dengan yang lainnya. Metode K-Medoid memiliki keunggulan dibandingkan dengan Metode K-Means. K-Medoid memiliki kinerja yang lebih optimal jika jumlah data yang digunakan berjumlah sedikit. Algoritma ini menggunakan objek pada kumpulan objek untuk mewakili sebuah cluster. Objek yang terpilih untuk mewakili sebuah cluster disebut medoid (Rofiqi, 2017).

K-Medoid adalah teknik berbasis objek yang representatif. Dalam metode ini kita memilih objek aktual untuk mewakili cluster alih-alih mengambil nilai rata-rata objek dalam cluster sebagai titik referensi. Partitioning Around Method (PAM) adalah salah satu Algoritma K-Medoid pertama. Strategi dasar dari algoritma ini adalah :

1. Temukan objek yang representatif untuk setiap cluster.
2. Kemudian setiap objek yang tersisa dikelompokkan dengan objek representatif yang paling mirip.
3. Kemudian secara iteratif ganti salah satu medoid dengan non-medoid selama “kualitas” pengelompokan diberlakukan (Tiwari dan Singh, 2012).

Metode K-Medoids adalah bagian dari partitioning clustering. Metode K-Medoids cukup efisien dalam dataset yang kecil. Langkah awal K-Medoids adalah mencari titik yang paling representatif (medoids) dalam dataset dengan menghitung jarak dari kelompok dalam semua kemungkinan kombinasi dari medoids sehingga jarak antar titik dalam suatu kluster kecil sedangkan jarak titik antar kluster besar (Indriani et al., 2019).

Perbedaan dari kedua metode ini yaitu Metode K-Medoids menggunakan objek sebagai perwakilan (medoid) sebagai pusat kluster untuk setiap kluster, sedangkan Metode K-Means menggunakan nilai rata-rata (mean) sebagai pusat kluster. Metode K-Medoids memiliki kelebihan untuk mengatasi kelemahan pada Metode K-Means yang sensitif terhadap noise dan outlier, di mana objek dengan nilai yang besar yang memungkinkan menyimpang pada distribusi data. Kelebihan lainnya yaitu hasil proses clustering tidak bergantung pada urutan masuk dataset.

3.2 Langkah-langkah K-Medoids Clustering

Langkah-langkah Metode K-Medoids clustering di antaranya:

1. Inialisasi pusat cluster sebanyak k (jumlah cluster)
2. Alokasikan setiap data (obek) ke kluster terdekat menggunakan persamaan ukuran jarak Euclidian Distance dengan persamaan :

$$d_{ij} = \sqrt{\sum_{a=1}^p (x_{ia} - x_{ja})^2} = \sqrt{(x_i - x_j)'(x_i - x_j)} \dots \dots \dots (3.1)$$

di mana $I = 1, \dots, n$; $j = 1, \dots, n$ dan p adalah banyak variabel, serta V adalah matrik varian kovarian.

3. Pilih secara acak objek pada masing-masing kluster sebagai kandidat medoid baru
4. Hitung jarak setiap objek yang berada pada masing-masing kluster dengan kandidat medoid baru.
5. 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. Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan medoid, sehingga didapatkan kluster beserta anggota kluster masing-masing (Sundari et al., 2019).

Bab 4

Rapidminer dan Pemasangan Perangkat Lunak

4.1 Rapidminer

RapidMiner merupakan perangkat lunak yang dibuat oleh Dr. Markus Hofmann dari Institute of Technology Blanchardstown dan Ralf Klinkenberg dari rapid-i.com dan merupakan platform perangkat lunak ilmu data yang dikembangkan oleh perusahaan bernama sama dengan menyediakan lingkungan terintegrasi untuk persiapan data. Dengan tampilan Graphical User Interface (GUI) sehingga memudahkan pengguna dalam menggunakan perangkat lunak ini (Kasus et al., 2015)

Ada beberapa tahap dalam menggunakan RapidMiner:

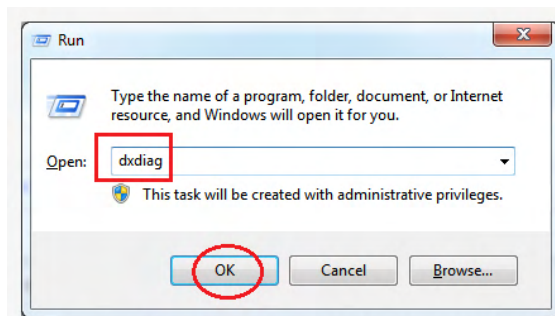
1. Untuk menganalisa, dibutuhkan data training. Data training yang akan dimasukkan ke dalam perangkat lunak Rapid Miner bisa dalam format .csv, .xls, .mdb dan lain-lain. Data yang digunakan penulis adalah data dalam format .csv.
2. Buka perangkat lunak Rapid Miner, kemudian akan ditampilkan tampilan awal. Untuk memasukkan data training yang telah dibuat sebelumnya, pilih menu File - Import Data - Import CSV file.
3. Tampilan jendela “Data Wizard” dengan total lima langkah. Pada langkah pertama ini tentukan nama file yang berisi data training dalam direktori kemudian pilih “Next”
4. Data training yang sebelumnya disimpan, akan tersimpan otomatis ke dalam **“Repositories”**. Pilih tab **“Repositories”** -

- “NewLocalRepository”** - **“data_training”**. Geser **“data_training”** ke area **“Main Process”**. Untuk menambahkan model, pilih tab **“Operators”** - **“Modelling”** - **“Classification”** kemudian geser **“Clustering”** ke area **“Main Process”** dan hubungkan.
5. Untuk melihat hasilnya, pilih **“Process”** - **“Run”** maka akan tampil hasil (Kasus et al., 2015).

4.2 Pemasangan Perangkat Lunak RapidMiner

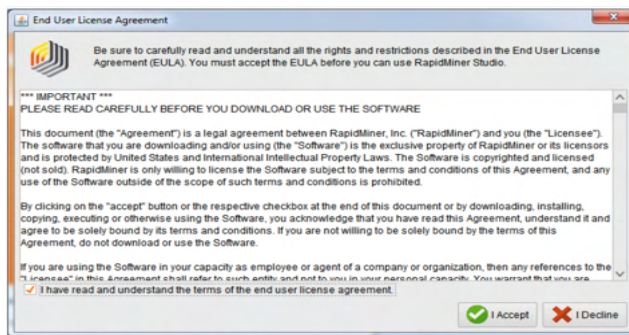
Langkah-langkah pemasangan perangkat lunak RapidMiner adalah :

1. RapidMiner dapat diunduh melalui beberapa tautan, tetapi terlebih dahulu sesuaikan dengan kapasitas perangkat keras yang akan digunakan
2. Terdapat dua versi dalam sistem operasi Microsoft Windows, yakni 32 Bit dan 64 Bit. Ketika akan menjalankan RapidMiner sebaiknya terlebih dulu mengetahui jenis perangkat komputer yang akan digunakan dengan cara menekan kedua tombol pada papan keyboard komputer secara bersamaan, yaitu “Logo Windows” dan huruf “R”, ketikkan “dxdiag” kemudian klik “OK”.



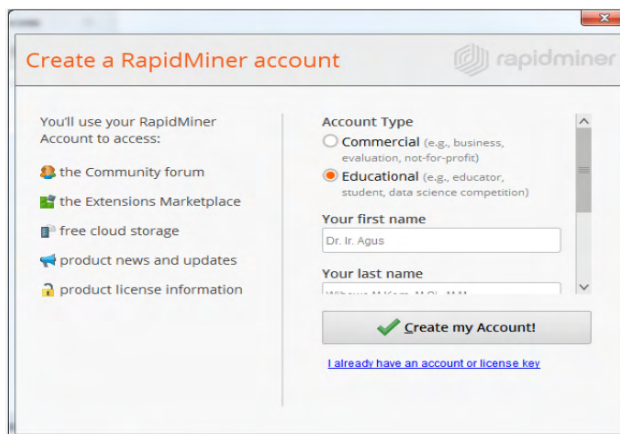
Gambar 4.1: Tampilan cara mengecek kapasitas Perangkat Keras

3. Setelah terlihat jenis sistem operasi yang akan digunakan, selanjutnya yaitu mengunduh aplikasi yang sesuai dengan versi perangkat keras yang akan dipakai.
4. Lakukan pemasangan seperti pada umumnya, kemudian jalankan, dan ketika keluar jendela dialog “End User License Agreement”, beri tanda centang pada opsi “I have read and understand the terms of end user license agreement” lalu klik tombol “I Accept”



Gambar 4.2: Tampilan End User License Agreement

5. Berikutnya pada kotak dialog “Create a RapidMiner account” pilih “Educational”

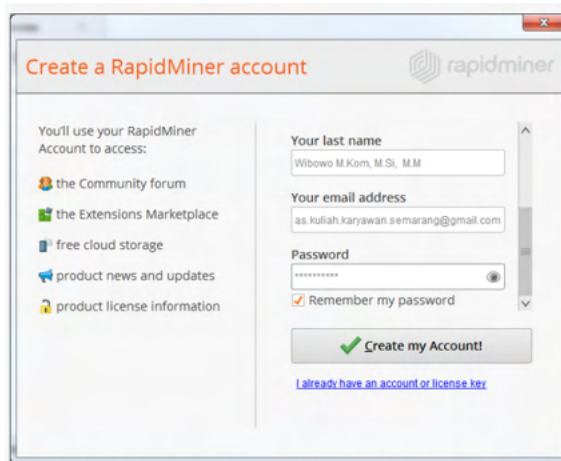


Gambar 4.3: Tampilan pembuatan account

6. Kemudian isikan :

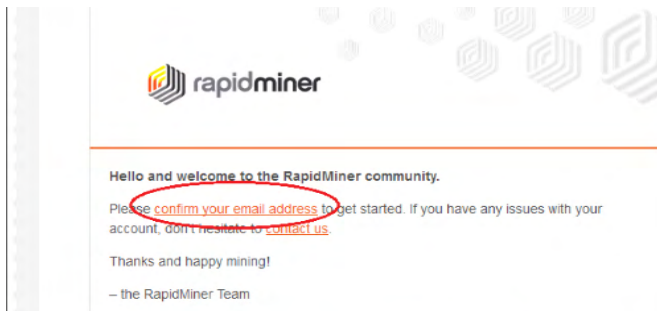
- Your First Name
- Your Last Name
- Your Email
- Password

7. Dan klik “Create my Account” (tetapi harus diingat dalam instalasi harus terkoneksi jaringan Internet)

A screenshot of a web browser window showing the 'Create a RapidMiner account' page. The page has a light blue header with the RapidMiner logo. On the left, there's a list of benefits: 'the Community forum', 'the Extensions Marketplace', 'free cloud storage', 'product news and updates', and 'product license information'. On the right, there are input fields for 'Your last name' (filled with 'Wibowo M Kom, M Si, M M'), 'Your email address' (filled with 'as.kuliah.karyawan.semarang@gmail.com'), and 'Password' (masked with dots). There is a 'Remember my password' checkbox which is checked. At the bottom right is a green 'Create my Account!' button and a link that says 'I already have an account or license key'.

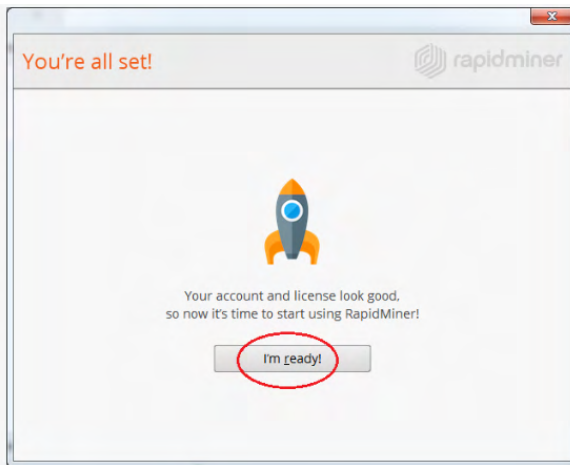
Gambar 4.4: Tampilan pengisian account

8. Selanjutnya buka alamat e-mail anda, lakukan konfirmasi dengan cara menekan tautan “confirm your email address”



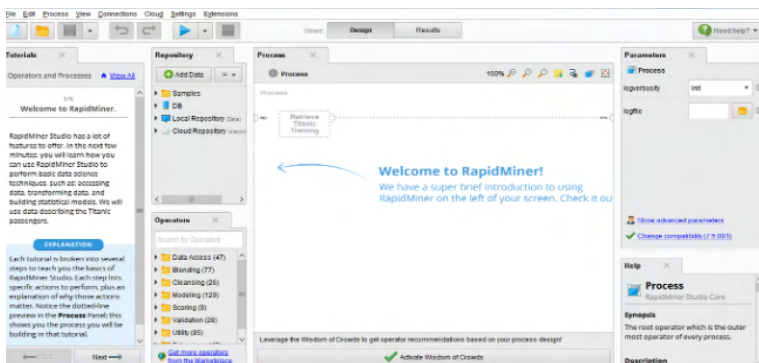
Gambar 4.5: Tampilan confirmation account

9. Ketika sistem akan membuka aplikasi browser Internet secara otomatis, lalu isikan “e-mail” dan “password” pada saat registrasi pertama kali.
10. Selanjutnya kembali pada perangkat lunak RapidMiner, dan tekan tombol “I’m Ready”



Gambar 4.6: Tampilan akhir pemasangan perangkat lunak RapidMiner

11. Selanjutnya akan keluar tampilan lingkungan terpadu dari RapidMiner, maka proses pemasangan perangkat lunak RapidMiner-pun telah selesai.



Gambar 4.7: Tampilan Jendela Kerja Rapidminer

Bab 5

Penerapan Algoritma K-Means Clustering

5.1 Pengolahan Data

Data persentasi wanita yang berumur 15 s/d 49 tahun dan berstatus kawin yang sedang menggunakan alat kontrasepsi untuk keluarga berencana (Persen) berdasarkan provinsi yang diambil dari Badan Pusat Statistik (BPS) tingkat nasional yang dihitung dengan menggunakan 2 nilai kluster, yaitu : tinggi dan rendah. Data berikut selanjutnya akan diolah menggunakan perangkat lunak RapidMiner.

Berikut uraian perhitungan manual proses clustering wanita yang berumur 15 s/d 49 tahun dan berstatus kawin yang sedang menggunakan alat kontrasepsi untuk keluarga berencana (Persen) dengan menggunakan sebuah Algoritma K-Means. Proses clustering dimulai dengan menentukan sebuah data yang akan dikluster. Dalam hal ini variabel data yang akan digunakan adalah data dari tahun 2005 s/d 2015, dan nilai rata-rata dari data setiap provinsi mulai dari tahun 2005 s/d 2015 sebagai data yang akan ditentukan nilai klusternya. Berikut adalah langkah-langkah penyelesaian dalam mengelompokkan wanita yang berumur 15 s/d 49 tahun dan berstatus kawin yang sedang menggunakan alat kontrasepsi untuk keluarga berencana (Persen) berdasarkan provinsi dengan menggunakan Algoritma K-Means:

Tabel 5.1: Persentase wanita yang berumur 15 s/d 49 tahun dan berstatus kawin yang sedang menggunakan alat kontrasepsi untuk keluarga berencana (Persen)

Provinsi	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Aceh	0	43,04	42,8	42,4	49,08	49,55	49,87	52,53	52,69	52,09	46,92
Sumatera Utara	42,51	45,08	45,53	41,91	49,71	48,67	50,25	52,86	51,79	51,87	49,06
Sumatera Barat	47,59	49,06	48,37	47,32	50,57	53,13	53,47	51,96	51,71	53,2	48,53
Riau	49,8	53,69	54,17	52,41	56,53	56,29	56,74	57,39	58,43	56,29	54,42
Jambi	62,94	61,63	64,66	62,16	66,72	65,8	67,32	68,05	68,5	67,5	64,16
Sumatera Selatan	59,42	62,44	61,97	62,92	64,63	65,78	66,37	67,23	67,98	66,47	68,06
Bengkulu	66,39	70,08	67,3	67,62	68,46	68,98	70,47	70,34	71,42	70,61	67,83
Lampung	65,97	64,49	64,03	64,58	67,81	69,28	66,41	67,74	69,55	69,36	67,35
Kep. Bangka Belitung	63,72	63,44	63,57	64,3	66,16	68,17	65,05	67,21	69,05	67,06	64,99
Kep. Riau	49,51	55,41	51,2	53,07	55,54	51,9	50,91	52,22	50,21	47,19	47,05
Dki Jakarta	54,13	55,25	54,69	52,68	56,62	57,42	55,19	57	57,55	55,14	54,75
Jawa Barat	62,88	62,84	62,28	60,51	63,67	64,57	64,09	65,53	65,12	65,36	64,67
Jawa Tengah	61,32	62,1	60,65	59,19	63,67	63,85	63,7	64,5	64,54	63,88	62,15
Di Yogyakarta	62,15	61,13	56,11	57,42	62,21	61,93	60,85	59,89	63,04	61,41	59,33
Jawa Timur	59,72	59,52	59,65	59,54	63,72	64,16	64,53	65,38	66,11	65,33	63,79
Banten	58,61	60,33	56,64	58	60,51	62,18	62,38	62,9	62,11	62,71	61,16
Bali	68,2	67,43	67,22	65,06	67,85	65,17	63,85	64,33	62,8	64,64	60,03
Nusa Tenggara Barat	55,71	54,82	52,44	53,07	57,88	57,75	59,68	58,67	60,34	58,79	59,07
Nusa Tenggara Timur	33,8	32,63	34,35	35,91	40,77	39,89	41,46	40,75	43,7	44,92	42,08
Kalimantan Barat	61,29	59,49	61,26	60,73	64,05	65,59	66,73	68,56	67,1	69,07	65,76
Kalimantan Tengah	67,08	66,64	67,46	68,4	70,34	68,16	70,85	72,49	72,88	72,07	68,5
Kalimantan Selatan	64,85	66,7	63,27	64,25	67,76	68,03	70,05	70,02	69,91	70,8	70,13
Kalimantan Timur	54,52	54,67	55,8	55,29	58,32	61,01	60,52	61,82	62,88	60,74	59,86
Sulawesi Utara	70,01	69,75	67,07	65,19	67,54	68,38	68,48	68,14	65,24	68,29	66,67
Sulawesi Tengah	54,97	54,68	56,83	55,91	61,5	61,08	58,25	60,8	59,7	60,38	57,55
Sulawesi Selatan	41,88	42,59	43,67	43,18	48,65	50,01	50,18	52,07	51,91	53,04	48,38
Sulawesi Tenggara	47,4	46,8	46,61	46,34	50,72	52,6	53,3	53	54,26	54,1	48,66
Gorontalo	59,91	61,24	64,22	59,54	62,83	64,22	61,6	65,08	65,13	66,83	64,78
Sulawesi Barat	0	38,82	38,47	45,23	49,78	48,83	47,84	50,92	47,93	49	47,69
Maluku	28,08	30,13	30,09	32,1	36,36	39,54	41,84	41	39,77	41,71	43,21
Maluku Utara	44,49	39,61	41,9	43,33	48,58	53,13	50,92	52,58	53,13	52,93	51,73
Papua Barat	0	31,73	28,29	26,69	36,47	38,48	37,84	41,25	42,91	42,12	43,96
Papua	32,8	31,22	31,92	27,71	33,71	26,97	23,91	24,77	23,87	27,88	23,37

1. Menentukan jumlah data yang akan dikluster, dimana sampel data Persentase Wanita Berumur 15-49 Tahun dan Berstatus Kawin yang Sedang Menggunakan/Memakai Alat KB (Persen) yang akan digunakan dalam proses clustering dengan jumlah data sebanyak 33 provinsi dan diambil dari tahun 2005 s/d 2015. Berikut yaitu cara mencari nilai rata-rata :

$$R1 = 0 + 43,04 + 42,8 + 42,2 + 49,08 + 49,55 + 49,87 + 52,53 + 52,69 + 52,09 + 46,92 / 11 = 43,72$$

$$R2 = 42,51 + 45,08 + 45,53 + 41,91 + 49,71 + 48,67 + 50,25 + 52,86 + 51,79 + 51,87 + 49,06 / 11 = 48,11$$

Lakukan pencarian rata-rata tersebut sampai R33, berikut nilai rata-rata yang telah didapat :

Tabel 5.2: Nilai Rata-Rata Persentase wanita yang berumur 15 s/d 49 tahun dan berstatus kawin yang sedang menggunakan alat kontrasepsi untuk keluarga berencana (Persen)

Provinsi	Rata-Rata
Aceh	43,72
Sumatera Utara	48,11
Sumatera Barat	50,45
Riau	55,11
Jambi	65,40
Sumatera Selatan	64,84
Bengkulu	69,05
Lampung	66,96
Kep, Bangka Belitung	65,70
Kep, Riau	51,29
Dki Jakarta	55,49
Jawa Barat	63,77
Jawa Tengah	62,69
Di Yogyakarta	60,50
Jawa Timur	62,86

Provinsi	Rata-Rata
Banten	60,68
Bali	65,14
Nusa Tenggara Barat	57,11
Nusa Tenggara Timur	39,11
Kalimantan Barat	64,51
Kalimantan Tengah	69,53
Kalimantan Selatan	67,80
Kalimantan Timur	58,68
Sulawesi Utara	67,71
Sulawesi Tengah	58,33
Sulawesi Selatan	47,78
Sulawesi Tenggara	50,34
Gorontalo	63,22
Sulawesi Barat	42,23
Maluku	36,71
Maluku Utara	48,39
Papua Barat	33,61
Papua	28,01

- Menetapkan nilai K jumlah kluster sebanyak 2 kluster, kluster yang dibentuk yaitu : kluster tertinggi dan kluster rendah.
- Menentukan nilai centroid awal yang telah ditentukan secara random. Kluster tertinggi (C1) diperoleh dari nilai maximum pada rata-rata data, kluster rendah (C2) diperoleh dari nilai minimum pada rata-rata data.

Berikut adalah nilai centroid data awal pada iterasi 1 :

Tabel 5.3: Nilai Centroid Awal (Iterasi 1)

C1	Maximum	69,53
C2	Minimum	28,01

- Setelah data nilai pusat kluster ditentukan, selanjutnya menghitung jarak setiap data terhadap pusat kluster dengan menggunakan rumus

sebagai berikut yang dilakukan dengan titik pusat (centroid) pada kluster pertama. Berikut adalah perhitungannya:

$$D(1.1) = \sqrt{(69,53 - 43,72)^2} = 25,81$$

$$D(1.2) = \sqrt{(69,53 - 48,11)^2} = 21,42$$

Lakukan sampai dengan D (1.33)

Perhitungan rata-rata data pada titik pusat centroid pada kluster 2 :

$$D(2.1) = \sqrt{((28,01-43,72))^2} = 15,71$$

$$D(2.2) = \sqrt{((28,01-48,11))^2} = 20,10$$

Lakukan sampai dengan D (2.33)

Berikut adalah hasil dari perhitungan rata-rata data pada titik pusat centroid pada setiap kluster :

Tabel 5.4: Hasil Perhitungan Centroid pada Setiap Kluster (Iterasi 1)

Provinsi	Rata-Rata	C1	C2
Aceh	43,72	25,81	15,71
Sumatera Utara	48,11	21,42	20,10
Sumatera Barat	50,45	19,09	22,43
Riau	55,11	14,43	27,09
Jambi	65,40	4,13	37,39
Sumatera Selatan	64,84	4,69	36,83
Bengkulu	69,05	0,49	41,03
Lampung	66,96	2,57	38,95
Kep, Bangka Belitung	65,70	3,83	37,69
Kep, Riau	51,29	18,24	23,28
Dki Jakarta	55,49	14,04	27,48
Jawa Barat	63,77	5,76	35,76
Jawa Tengah	62,69	6,85	34,67
Di Yogyakarta	60,50	9,04	32,49
Jawa Timur	62,86	6,67	34,85
Banten	60,68	8,85	32,67
Bali	65,14	4,39	37,13

Provinsi	Rata-Rata	C1	C2
Nusa Tenggara Barat	57,11	12,42	29,10
Nusa Tenggara Timur	39,11	30,42	11,10
Kalimantan Barat	64,51	5,02	36,50
Kalimantan Tengah	69,53	0,00	41,52
Kalimantan Selatan	67,80	1,74	39,79
Kalimantan Timur	58,68	10,86	30,66
Sulawesi Utara	67,71	1,83	39,69
Sulawesi Tengah	58,33	11,20	30,32
Sulawesi Selatan	47,78	21,76	19,77
Sulawesi Tenggara	50,34	19,19	22,33
Gorontalo	63,22	6,32	35,20
Sulawesi Barat	42,23	27,31	14,22
Maluku	36,71	32,82	8,70
Maluku Utara	48,39	21,14	20,38
Papua Barat	33,61	35,92	5,60
Papua	28,01	41,52	0,00

Lalu hitung jarak terdekat dengan menggunakan Euclidean Distance.

Tabel 5.5: Hasil Perhitungan Jarak Data pada Iterasi 1

No	Jarak Terpendek	Hasil	C1	C2
1	15,71	C2		1
2	20,10	C2		1
3	19,09	C1	1	
4	14,43	C1	1	
5	4,13	C1	1	
6	4,69	C1	1	
7	0,49	C1	1	
8	2,57	C1	1	
9	3,83	C1	1	
10	18,24	C1	1	
11	14,04	C1	1	

No	Jarak Terpendek	Hasil	C1	C2
12	5,76	C1	1	
13	6,85	C1	1	
14	9,04	C1	1	
15	6,67	C1	1	
16	8,85	C1	1	
17	4,39	C1	1	
18	12,42	C1	1	
19	11,10	C2		1
20	5,02	C1	1	
21	0,00	C1	1	
22	1,74	C1	1	
23	10,86	C1	1	
24	1,83	C1	1	
25	11,20	C1	1	
26	19,77	C2		1
27	19,19	C1	1	
28	6,32	C1	1	
29	14,22	C2		1
30	8,70	C2		1
31	20,38	C2		1
32	5,60	C2		1
33	0,00	C2		1

Berikut ini adalah hasil dari perhitungan jarak data ke titik pusat cluster pada iterasi 1 :

Tabel 5.6: Hasil Kluster pada Iterasi 1

Kluster	Hasil
C1	24
C2	9

- Selanjutnya lakukan kembali langkah 4 dan 5 jika nilai centroid hasil dari iterasi pertama dengan nilai centroid selanjutnya bernilai sama

ataupun nilai centroid sudah optimal serta posisi kluster tidak mengalami perubahan lagi maka proses iterasi berhenti. Namun jika posisi kluster masih berubah-ubah, maka proses iterasi terus berlanjut pada iterasi berikutnya sampai hasil cluster bernilai sama.

6. Menghitung titik pusat baru menggunakan hasil dari setiap anggota pada masing-masing cluster. Berikut contoh perhitungan titik pusat cluster baru pada iterasi ke dua (x) :

$$\begin{aligned}
 C1x &= \frac{50,54 + 55,11 + 65,40 + 64,84 + 69,05 + 66,96 + 65,70 + 51,29 + 55,49 + 63,77 + 62,69 + 60,50 + 62,86 + 60,68 + 65,14 + 57,11 + 64,51 + 69,53 + 67,80 + 58,68 + 67,71 + 58,33 + 50,34 + 63,22}{24} \\
 &= 61,55 \\
 C2x &= \frac{43,72 + 48,11 + 39,11 + 47,78 + 42,23 + 36,71 + 48,39 + 33,61 + 28,01}{9} \\
 &= 40,85
 \end{aligned}$$

Tabel 5.7: Nilai Centroid Baru (Iterasi 2)

C1	Maximum	61,55
C2	Minimum	40,85

Selanjutnya melakukan sebuah perhitungan nilai centroid baru tersebut dengan nilai rata-rata data sebagai berikut :

Perhitungan rata-rata data pada titik pusat centroid pada cluster 1 :

$$E (1.1) = \sqrt{(61,55 - 43,72)^2} = 17,82$$

$$E (1.2) = \sqrt{(61,55 - 48,11)^2} = 13,44$$

Lakukan sampai dengan E (1.33)

Perhitungan rata-rata data pada titik pusat centroid pada cluster 2 :

$$E (2.1) = \sqrt{(40,85 - 43,72)^2} = 2,78$$

$$E (2.2) = \sqrt{(40,85 - 48,11)^2} = 7,26$$

Lakukan sampai dengan E (2.33)

Berikut adalah hasil dari perhitungan rata-rata data pada titik pusat centroid pada setiap kluster :

Tabel 5.8: Hasil Perhitungan Centroid pada Setiap Kluster (Iterasi 2)

Provinsi	Rata-Rata	C1	C2
Aceh	43,72	17,82	2,87
Sumatera Utara	48,11	13,44	7,26
Sumatera Barat	50,45	11,10	9,59
Riau	55,11	6,44	14,25
Jambi	65,40	3,86	24,55
Sumatera Selatan	64,84	3,29	23,99
Bengkulu	69,05	7,50	28,19
Lampung	66,96	5,41	26,11
Kep, Bangka Belitung	65,70	4,15	24,85
Kep, Riau	51,29	10,26	10,44
Dki Jakarta	55,49	6,06	14,64
Jawa Barat	63,77	2,23	22,92
Jawa Tengah	62,69	1,14	21,83
Di Yogyakarta	60,50	1,05	19,64
Jawa Timur	62,86	1,31	22,00
Banten	60,68	0,86	19,83
Bali	65,14	3,60	24,29
Nusa Tenggara Barat	57,11	4,44	16,26
Nusa Tenggara Timur	39,11	22,43	1,74
Kalimantan Barat	64,51	2,96	23,66
Kalimantan Tengah	69,53	7,99	28,68
Kalimantan Selatan	67,80	6,25	26,94
Kalimantan Timur	58,68	2,87	17,82
Sulawesi Utara	67,71	6,16	26,85
Sulawesi Tengah	58,33	3,22	17,48
Sulawesi Selatan	47,78	13,77	6,92
Sulawesi Tenggara	50,34	11,20	9,49
Gorontalo	63,22	1,67	22,36
Sulawesi Barat	42,23	19,32	1,37
Maluku	36,71	24,84	4,14
Maluku Utara	48,39	13,15	7,54

Provinsi	Rata-Rata	C1	C2
Papua Barat	33,61	27,94	7,24
Papua	28,01	33,54	12,84

Lalu hitung jarak terdekat dengan centroid dengan menggunakan Euclidean Distance.

Tabel 5.9: Hasil Perhitungan jarak Data pada Iterasi 2

No	Jarak Terpendek	Hasil	C1	C2
1	2,87	C2		1
2	7,26	C2		1
3	9,59	C2		1
4	6,44	C1	1	
5	3,86	C1	1	
6	3,29	C1	1	
7	7,50	C1	1	
8	5,41	C1	1	
9	4,15	C1	1	
10	10,26	C1	1	
11	6,06	C1	1	
12	2,23	C1	1	
13	1,14	C1	1	
14	1,05	C1	1	
15	1,31	C1	1	
16	0,86	C1	1	
17	3,60	C1	1	
18	4,44	C1	1	
19	1,74	C2		1
20	2,96	C1	1	
21	7,99	C1	1	
22	6,25	C1	1	
23	2,87	C1	1	
24	6,16	C1	1	
25	3,22	C1	1	

No	Jarak Terpendek	Hasil	C1	C2
26	6,92	C2		1
27	9,49	C2		1
28	1,67	C1	1	
29	1,37	C2		1
30	4,14	C2		1
31	7,54	C2		1
32	7,24	C2		1
33	12,84	C2		1

Berikut adalah hasil dari perhitungan jarak data ke titik pusat kluster pada iterasi 2

Tabel 5.10: Hasil Kluster pada Iterasi 2

Kluster	Hasil
C1	22
C2	11

Selanjutnya menghitung titik pusat cluster baru menggunakan hasil dari setiap anggota data pada masing-masing cluster. Berikut perhitungan titik pusat cluster baru iterasi ke tiga (x) :

$$\begin{aligned}
 C1x &= \frac{55,11+65,40+64,84+69,05+66,96+65,70+51,26+55,49+63,77+62,69+60,50+62,86+60,68+65,14+57,11+64,51+69,53+67,80+58,68+67,71+58,33+63,22}{22} \\
 &= 62,56
 \end{aligned}$$

$$\begin{aligned}
 C2x &= \frac{43,72 + 48,11 + 50,45 + 39,11 + 47,78 + 50,34 + 42,23 + 33,71 + 48,39 + 33,61 + 28,01}{11} \\
 &= 42,59
 \end{aligned}$$

Tabel 5.11: Nilai Centroid Baru (Iterasi 3)

C1	Maximum	62,56
C2	Minimum	42,59

Perhitungan rata-rata data pada titik pusat centroid pada kluster 1 :

$$F (1.1) = \sqrt{(62,56 - 43,72)^2} = 18,84$$

$$F (1.2) = \sqrt{(42,59 - 48,11)^2} = 14,45$$

Lakukan sampai dengan F (1.33)

Perhitungan rata-rata data pada titik pusat centroid pada kluster 2 :

$$F (2.1) = \sqrt{(62,56 - 42,72)^2} = 1,14$$

$$F (2.2) = \sqrt{(42,59 - 48,11)^2} = 5,52$$

Lakukan sampai dengan F (2.33)

Berikut adalah hasil dari perhitungan rata-rata data pada titik pusat centroid pada setiap kluster :

Tabel 5.12: Hasil Perhitungan Centroid pada Setiap Kluster (Iterasi 3)

Provinsi	Rata-Rata	C1	C2
Aceh	43,72	18,84	1,14
Sumatera Utara	48,11	14,45	5,52
Sumatera Barat	50,45	12,12	7,86
Riau	55,11	7,46	12,52
Jambi	65,40	2,84	22,81
Sumatera Selatan	64,84	2,28	22,25
Bengkulu	69,05	6,48	26,46
Lampung	66,96	4,40	24,37
Kep, Bangka Belitung	65,70	3,14	23,11
Kep, Riau	51,29	11,27	8,70
Dki Jakarta	55,49	7,07	12,90
Jawa Barat	63,77	1,21	21,19
Jawa Tengah	62,69	0,12	20,10
Di Yogyakarta	60,50	2,07	17,91
Jawa Timur	62,86	0,30	20,27
Banten	60,68	1,88	18,10
Bali	65,14	2,58	22,55
Nusa Tenggara Barat	57,11	5,45	14,52
Nusa Tenggara Timur	39,11	23,45	3,47
Kalimantan Barat	64,51	1,95	21,92
Kalimantan Tengah	69,53	6,97	26,94

Provinsi	Rata-Rata	C1	C2
Kalimantan Selatan	67,80	5,23	25,21
Kalimantan Timur	58,68	3,89	16,09
Sulawesi Utara	67,71	5,14	25,12
Sulawesi Tengah	58,33	4,23	15,74
Sulawesi Selatan	47,78	14,78	5,19
Sulawesi Tenggara	50,34	12,22	7,76
Gorontalo	63,22	0,65	20,63
Sulawesi Barat	42,23	20,33	0,36
Maluku	36,71	25,85	5,88
Maluku Utara	48,39	14,17	5,80
Papua Barat	33,61	28,95	8,98
Papua	28,01	34,55	14,58

Lalu hitung jarak terdekat dengan centroid dengan menggunakan Euclidean Distance.

Tabel 5.13: Hasil Perhitungan jarak Data pada Iterasi 3

No	Jarak Terpendek	Hasil	C1	C2
1	1,14	C2		1
2	5,52	C2		1
3	7,86	C2		1
4	7,46	C1	1	
5	2,84	C1	1	
6	2,28	C1	1	
7	6,48	C1	1	
8	4,40	C1	1	
9	3,14	C1	1	
10	8,70	C2		1
11	7,07	C1	1	
12	1,21	C1	1	
13	0,12	C1	1	
14	2,07	C1	1	
15	0,30	C1	1	
16	1,88	C1	1	

No	Jarak Terpendek	Hasil	C1	C2
17	2,58	C1	1	
18	5,45	C1	1	
19	3,47	C2		1
20	1,95	C1	1	
21	6,97	C1	1	
22	5,23	C1	1	
23	3,89	C1	1	
24	5,14	C1	1	
25	4,23	C1	1	
26	5,19	C2		1
27	7,76	C2		1
28	0,65	C1	1	
29	0,36	C2		1
30	5,88	C2		1
31	5,80	C2		1
32	8,98	C2		1
33	14,58	C2		1

Berikut adalah hasil dari perhitungan jarak data ke titik pusat kluster pada iterasi 3

Tabel 5.14: Hasil Kluster pada Iterasi 3

Kluster	Hasil
C1	21
C2	12

Selanjutnya menghitung titik pusat kluster baru menggunakan hasil dari setiap anggota data pada masing-masing kluster. Berikut perhitungan titik pusat kluster baru iterasi ke empat (x) :

$$\begin{aligned}
 C1x &= \frac{55,11+65,40+64,84+69,05+66,96+65,70+ \\
 &\quad 55,49+63,77+62,69+60,50+62,86+60,68+65,14+ \\
 &\quad 57,11+64,51+69,53+67,80+58,68+67,71+58,33+}{21} \\
 &= 63,10
 \end{aligned}$$

$$C2x = \frac{43,72 + 48,11 + 50,45 + 51,29 + 39,11 + 47,78 + 50,34 + 42,23 + 36,71 + 48,39 + 33,61 + 28,01}{12}$$

$$= 43,31$$

Tabel 5.15: Nilai Centroid Baru (Iterasi 4)

C1	Maximum	63,10
C2	Minimum	43,31

Perhitungan rata-rata data pada titik pusat centroid pada kluster 1 :

$$F (1.1) = \sqrt{(63,10 - 43,72)^2} = 19,37$$

$$F (1.2) = \sqrt{(43,31 - 48,11)^2} = 14,99$$

Lakukan sampai dengan F (1.33)

Perhitungan rata-rata data pada titik pusat centroid pada kluster 2 :

$$F (2.1) = \sqrt{(63,10 - 42,72)^2} = 0,41$$

$$F (2.2) = \sqrt{(43,31 - 48,11)^2} = 4,80$$

Lakukan sampai dengan F (2.33)

Berikut adalah hasil dari perhitungan rata-rata data pada titik pusat centroid pada setiap kluster :

Tabel 5.16: Hasil Perhitungan Centroid pada Setiap Kluster (Iterasi 4)

Provinsi	Rata-Rata	C1	C2
Aceh	43,72	19,37	0,41
Sumatera Utara	48,11	14,99	4,80
Sumatera Barat	50,45	12,65	7,13
Riau	55,11	7,99	11,79
Jambi	65,40	2,30	22,09
Sumatera Selatan	64,84	1,74	21,53
Bengkulu	69,05	5,95	25,73
Lampung	66,96	3,86	23,65
Kep. Bangka Belitung	65,70	2,60	22,39
Kep. Riau	51,29	11,81	7,98
Dki Jakarta	55,49	7,61	12,18
Jawa Barat	63,77	0,68	20,46

Provinsi	Rata-Rata	C1	C2
Jawa Tengah	62,69	0,41	19,37
Di Yogyakarta	60,50	2,60	17,18
Jawa Timur	62,86	0,24	19,54
Banten	60,68	2,41	17,37
Bali	65,14	2,04	21,83
Nusa Tenggara Barat	57,11	5,99	13,80
Nusa Tenggara Timur	39,11	23,98	4,20
Kalimantan Barat	64,51	1,41	21,20
Kalimantan Tengah	69,53	6,43	26,22
Kalimantan Selatan	67,80	4,70	24,48
Kalimantan Timur	58,68	4,42	15,36
Sulawesi Utara	67,71	4,61	24,39
Sulawesi Tengah	58,33	4,77	15,02
Sulawesi Selatan	47,78	15,32	4,46
Sulawesi Tenggara	50,34	12,75	7,03
Gorontalo	63,22	0,12	19,90
Sulawesi Barat	42,23	20,87	1,09
Maluku	36,71	26,39	6,60
Maluku Utara	48,39	14,71	5,08
Papua Barat	33,61	29,49	9,70
Papua	28,01	35,09	15,30

Lalu hitung jarak terdekat dengan centroid dengan menggunakan Euclidean Distance.

Tabel 5.17: Hasil Perhitungan jarak Data pada Iterasi 4

No	Jarak Terpendek	Hasil	C1	C2
1	0,41	C2		1
2	4,80	C2		1
3	7,13	C2		1
4	7,99	C1	1	
5	2,30	C1	1	
6	1,74	C1	1	
7	5,95	C1	1	
8	3,86	C1	1	
9	2,60	C1	1	

No	Jarak Terpendek	Hasil	C1	C2
10	7,98	C2		1
11	7,61	C1	1	
12	0,68	C1	1	
13	0,41	C1	1	
14	2,60	C1	1	
15	0,24	C1	1	
16	2,41	C1	1	
17	2,04	C1	1	
18	5,99	C1	1	
19	4,20	C2		1
20	1,41	C1	1	
21	6,43	C1	1	
22	4,70	C1	1	
23	4,42	C1	1	
24	4,61	C1	1	
25	4,77	C1	1	
26	4,46	C2		1
27	7,03	C2		1
28	0,12	C1	1	
29	1,09	C2		1
30	6,60	C2		1
31	5,08	C2		1
32	9,70	C2		1
33	15,30	C2		1

Berikut adalah hasil dari perhitungan jarak data ke titik pusat kluster pada iterasi 4:

Tabel 5.18: Hasil Kluster pada Iterasi 4

Kluster	Hasil
C1	21
C2	12

Didapatkan sebuah hasil akhir yang dimana pada iterasi ke 3 dan 4 hasil yang sama. Hasil dari kedua iterasi tersebut bernilai C1 = 21 dan C2 = 12 dan pada

data tersebut tidak mengalami perubahan lagi, maka proses iterasi pun berhenti dan dapat disimpulkan sebagai berikut :

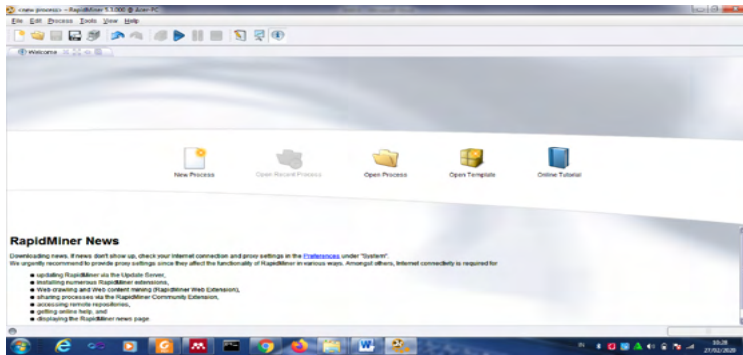
1. Cluster Tinggi (C1) yaitu dengan jumlah data sebanyak 21 provinsi yaitu : Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep.Bangka Belitung, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Sulawesi Utara, Sulawesi Tengah, Gorontalo.
2. Cluster Rendah (C2) yaitu dengan jumlah data sebanyak 12 provinsi yaitu : Aceh, Sumatera Utara, Sumatera Barat, Kep.Riau, Nusa Tenggara Timur, Sulawesi Selatan, Sulawesi Tenggara, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua.

5.2 Pengujian dengan RapidMiner

Pada bagian ini akan menjelaskan bagaimana proses uji coba perhitungan data yang telah dihasilkan pada bagian pengolahan data diatas. Pengujian data pada penelitian ini yaitu dengan menggunakan perangkat lunak RapidMiner yang mana hasil dari uji perhitungan tersebut akan menghasilkan keluaran yang sesuai dengan perhitungan yang manual yang telah dihitung dengan menggunakan perangkat lunak Microsoft Excel.

5.2.1 Tampilan Menu Utama RapidMiner

Berikut ini akan dijelaskan bagian dari menu awal pada tools RapidMiner, yaitu tampilan halaman awal ketika akan memulai uji coba dengan menggunakan menu “New process”.

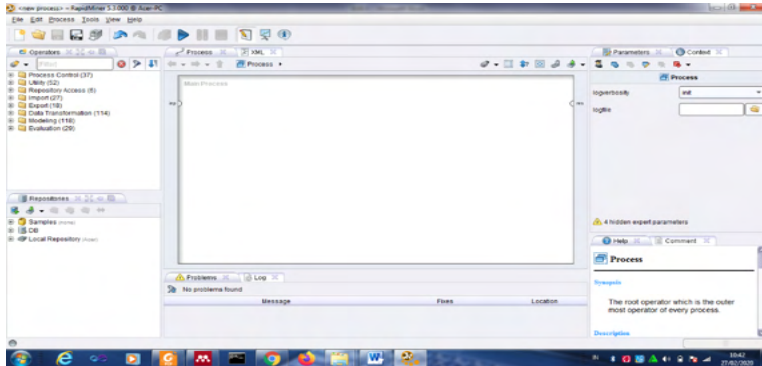


Gambar 5.1: Tampilan Awal RapidMiner 5.3

1. **“New Process”** menu ini berisikan tampilan awal jika kita akan membuat proses baru dengan data yang baru yang belum pernah ada dalam tools ini.
2. **“Open Recent Process”** berfungsi untuk menampilkan dan membuka proses yang baru saja dibuat dengan jangka waktu tertentu, biasanya akan langsung muncul pada tampilan awal tools tanpa harus mencari pada penyimpanan lokal.
3. **“Open Process”** menu ini berfungsi untuk membuka proses yang telah dibuat sebelumnya yang telah kita simpan di penyimpanan lokal.
4. **“Open Template”** menu ini berisikan pilihan-pilihan proses lain yang telah disediakan oleh tools RapidMiner.
5. **“Outline Tutorial”** menu ini digunakan untuk memulai tutorial atau tahapan secara *online*. Tutorial berfungsi untuk memberikan pengenalan dan beberapa konsep data mining.

5.2.2 Tampilan Menu Excel Rapidminer

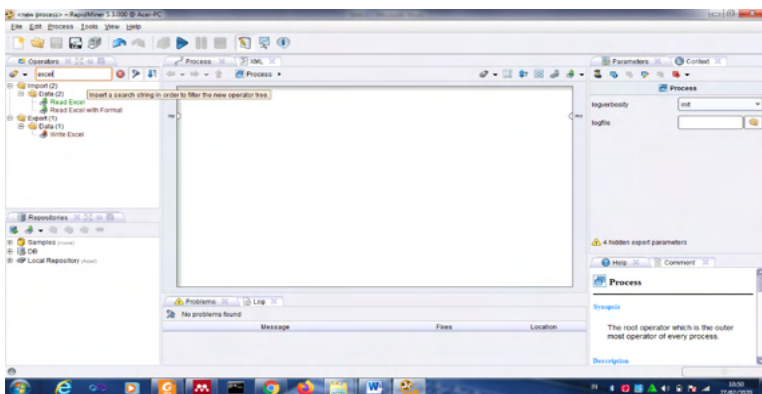
Dikarenakan data yang akan diuji adalah data baru, maka pilih New Process. Maka akan muncul jendela seperti gambar berikut :



Gambar 5.2: “New Process” untuk Import Data Excel RapidMiner

5.2.3 Tampilan Read Excel RapidMiner

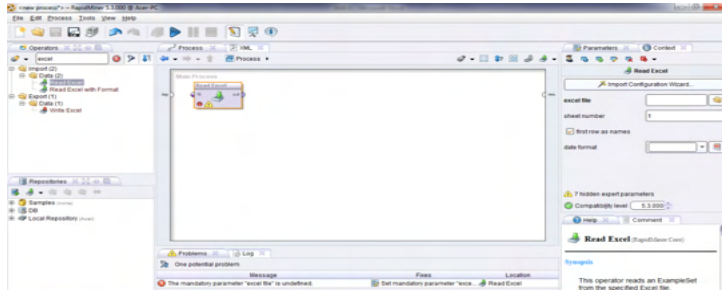
Sistem ini menjelaskan sebuah cara memasukkan data baru yang akan dieksekusi lebih lanjut, pada bagian ini data yang akan dieksekusi adalah data excel. Tahapan nya dapat dilihat dari gambar di bawah ini :



Gambar 5.3: Input Data Read Excel RapidMiner

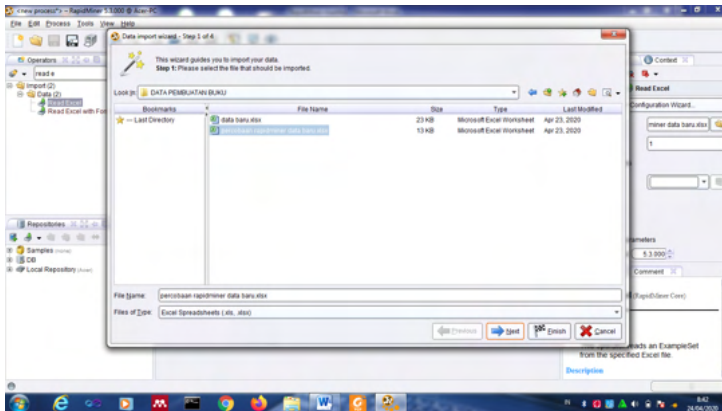
Klik pada pojok atas bagian kiri tab operators lalu pilih **“Read Excel”**.

Kemudian akan muncul tampilan seperti gambar di bawah ini :



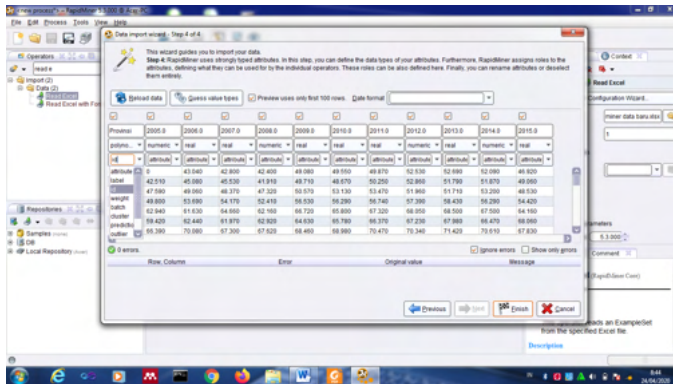
Gambar 5.4: Tampilan Read Excel RapidMiner

Setelah itu pilih bagian pojok kanan atas pilih gambar folder **“excel file”** dan pilih file data yang akan digunakan. Setelah file terpilih, lalu klik pada bagian pojok kanan atas, yaitu **“import configuration wizard”** kemudian pilih file data yang akan digunakan seperti tampilan di bawah ini :



Gambar 5.5: Tampilan Data Import awal RapidMiner

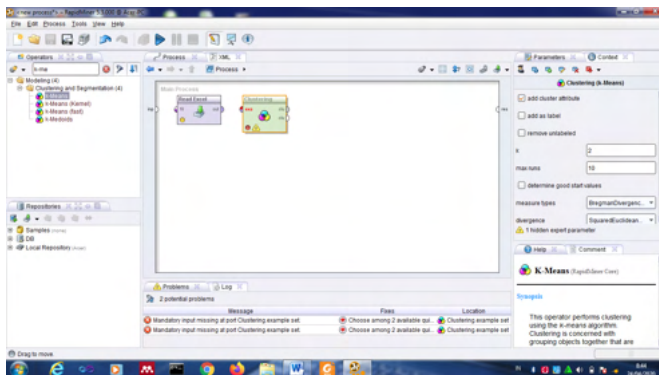
Gambar di atas menunjukkan proses pengambilan data yang akan diproses, setelah itu pilih next dan next lagi dan lakukan pemilihan tipe data di mana pada bagian *attribute text* diubah menjadi **“id”** lalu pilih finish pada data tersebut sehingga akan muncul tampilan seperti gambar di bawah ini :



Gambar 5.6: Tampilan Data Import kedua RapidMiner

5.2.4 Pemrosesan RapidMiner

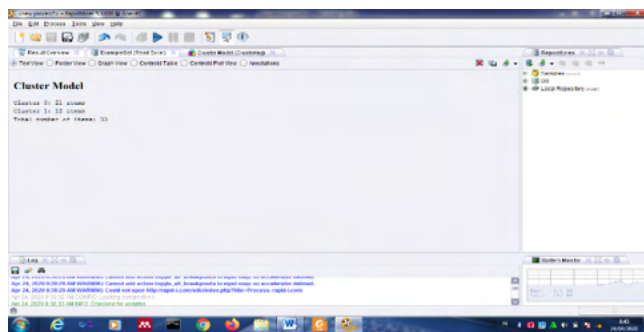
Pada bagian ini akan menjelaskan tahapan-tahapan dalam k-means pada data yang telah di import dalam RapidMiner. Tahapan pertama pilih pada pojok atas bagian kiri tab operators lalu pilih “k-means” dan ketika muncul hubungkan “Read excel dan k-means” dan ubah nilai K sebanyak 2 kluster. Kemudian akan muncul tampilan seperti gambar di bawah ini :



Gambar 5.7: Pemrosesan Data

5.2.5 Tampilan Keluaran Sistem

Setelah melakukan pemrosesan data selanjutnya untuk mendapatkan sebuah hasil pengelompokan maka pada tahap selanjutnya yaitu meng-klik tanda play panah biru pada bagian toolbar atas untuk menjalankan pemrosesan data sehingga mendapatkan hasil akhir dalam sebuah keluaran sistem. Berikut tampilan hasil akhir dari pemrosesan data di atas :

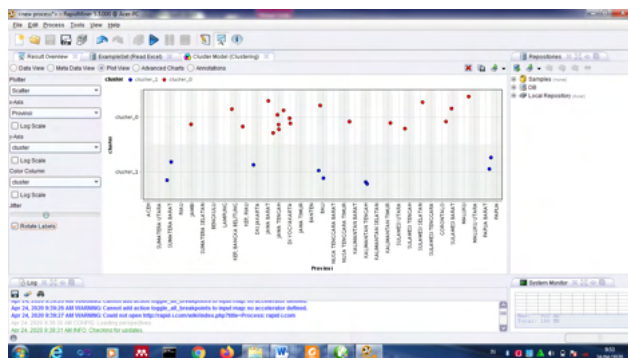


Gambar 5.8: Nilai Kluster Model RapidMiner

Keterangan :

1. Jumlah Cluster 0 (Tinggi) berjumlah 21 items
2. Jumlah Cluster 1 (Rendah) berjumlah 12 items
3. Jumlah keseluruhan items adalah 33

Sehingga dapat diketahui hasil pengelompokan dari RapidMiner 5.3 dapat dilihat dari gambar di bawah ini :



Gambar 5.9: Hasil Akhir Pengelompokan RapidMiner

5.2.6 Validasi data

Hasil perhitungan cluster menggunakan algoritma k-means dengan perhitungan menggunakan aplikasi RapidMiner.

Tabel 5.19: Hasil Perhitungan K-Means dengan Aplikasi RapidMiner

cal	Provinsi	<i>K-Means</i>	<i>RapidMiner</i>
1	Aceh	cluster_1	Cluster 2
2	Sumatera Utara	cluster_1	Cluster 2
3	Sumatera Barat	cluster_1	Cluster 2
4	Riau	cluster_0	Cluster 1
5	Jambi	cluster_0	Cluster 1
6	Sumatera Selatan	cluster_0	Cluster 1
7	Bengkulu	cluster_0	Cluster 1
8	Lampung	cluster_0	Cluster 1
9	Kep, Bangka Belitung	cluster_0	Cluster 1
10	Kep, Riau	cluster_1	Cluster 2
11	Dki Jakarta	cluster_0	Cluster 1
12	Jawa Barat	cluster_0	Cluster 1
13	Jawa Tengah	cluster_0	Cluster 1
14	Di Yogyakarta	cluster_0	Cluster 1
15	Jawa Timur	cluster_0	Cluster 1
16	Banten	cluster_0	Cluster 1
17	Bali	cluster_0	Cluster 1
18	Nusa Tenggara Barat	cluster_0	Cluster 1
19	Nusa Tenggara Timur	cluster_1	Cluster 2
20	Kalimantan Barat	cluster_0	Cluster 1
21	Kalimantan Tengah	cluster_0	Cluster 1
22	Kalimantan Selatan	cluster_0	Cluster 1
23	Kalimantan Timur	cluster_0	Cluster 1
24	Sulawesi Utara	cluster_0	Cluster 1
25	Sulawesi Tengah	cluster_0	Cluster 1
26	Sulawesi Selatan	cluster_1	Cluster 2

cal	Provinsi	<i>K-Means</i>	<i>RapidMiner</i>
27	Sulawesi Tenggara	cluster_1	Cluster 2
28	Gorontalo	cluster_0	Cluster 1
29	Sulawesi Barat	cluster_1	Cluster 2
30	Maluku	cluster_1	Cluster 2
31	Maluku Utara	cluster_1	Cluster 2
32	Papua Barat	cluster_1	Cluster 2
33	Papua	cluster_1	Cluster 2

Bab 6

Penerapan Algoritma K-Medoids Clustering

6.1 Analisis Data Manual K-Medoids

Beberapa tahapan dalam proses perhitungan manual menggunakan Algoritma K-Medoids pada penduduk buta huruf di Indonesia adalah :

Tabel 6.1: Data Awal

Provinsi									
	2009	2010	2011	2012	2013	2014	2015	2016	2017
Aceh	3.13	2.74	3.66	3.38	2.98	2.28	2.09	2.26	1.85
Sumatera Utara	2.47	2.40	2.85	2.36	1.95	1.31	1.18	1.12	1.04
Sumatera Barat	2.79	2.60	3.41	2.97	2.37	1.47	1.29	1.19	1.07
Riau	1.69	1.49	2.19	2.01	1.93	1.16	1.06	0.93	0.79
Jambi	3.93	3.67	3.99	3.57	2.91	1.99	1.92	1.99	1.72
Sumatera Selatan	2.47	2.34	3	2.81	2.51	1.71	1.58	1.54	1.35
Bengkulu	4.46	4.15	4.32	3.80	3.18	2.27	2.10	2.25	1.96
Lampung	4.95	4.75	4.44	4.35	3.78	3.12	2.98	3.22	2.79
Kep.Bangka Belitung	4.13	4.12	3.98	3.73	3.27	2.16	2.11	2.34	2.03
Kep. Riau	3.54	2.51	2.16	2.01	1.90	1.23	1.08	1.16	1.12
Dki Jakarta	0.99	0.81	1.10	0.86	0.82	0.43	0.39	0.36	0.32
Jawa Barat	3.56	3.38	3.62	3.39	2.95	1.84	1.78	1.78	1.62
Jawa Tengah	9.36	8.98	8.58	8.45	7.82	6.27	6.18	6.70	6.01
Di Yogyakarta	9.02	8.38	7.76	7.30	6.62	5.12	5.05	5.41	4.91
Jawa Timur	10.99	10.53	10.32	9.65	8.88	7.77	7.70	8.41	7.47
Banten	3.56	3.40	3.35	3.07	3.05	2.51	2.33	2.45	2.21
Bali	11.52	10.51	9.73	8.81	8.28	6.70	6.52	7.18	6.41
Nusa Tenggara Barat	17.20	16.51	14.56	14.14	13.34	11.36	11.34	12.94	11.47
Nusa Tenggara	10.34	9.84	10.80	9.70	8.37	7.62	7.27	8.48	7.25

Menentukan jumlah cluster(k) dari n objek adalah 3 cluster.

1. Mengasumsikan centroid awal yang telah ditentukan seperti pada tabel 6.2

Tabel 6.2: Medoids Awal

Sulawesi Tengah	3.75	3.50	5.03	4.58	3.71	2.76	2.49	2.47	2.17
Sulawesi Selatan	11.33	10.84	10.52	9.90	8.68	7.73	7.67	8.47	7.47
Sulawesi Tenggara	7.34	7.10	7.59	7.38	6.47	5.24	5.09	5.74	4.95

2. Menempatkan objek-objek non medoids ke dalam kluster yang paling dekat dengan medoids berdasarkan jarak Euclidean. Berikut adalah perhitungan jarak pada jumlah penyakit Masyarakat Pukesdes Bukit Maraja.

$$D_{aceh,c1} = \sqrt{(3,13 - 3,75)^2 + (2,75 - 3,50)^2 + (3,66 - 5,03)^2 + (3,38 - 4,58)^2 + (2,98 - 3,71)^2 + (2,28 - 2,76)^2 + (2,09 - 2,40)^2 + (2,26 - 2,49)^2 + (1,85 - 2,17)^2}$$

$$= 2,30$$

$$D_{aceh,c2} = \sqrt{(3,13 - 11,33)^2 + (2,74 - 10,84)^2 + (3,66 - 10,52)^2 + (3,38 - 9,90)^2 + (2,98 - 8,68)^2 + (2,28 - 7,73)^2 + (2,09 - 7,67)^2 + (2,26 - 8,48)^2 + (1,85 - 7,47)^2}$$

$$= 19,65$$

Hasil dari keseluruhan dapat dilihat pada tabel 6.3 sebagai berikut :

Tabel 6.3: Hasil Perhitungan Algoritma K-Medoids Iterasi ke-1

Provinsi	C1	C2	C3	Jarak Terdekat	Kluster
Aceh	2.30	19.65	10.94	2.30	1
Sumatera Utara	4.73	22.13	13.46	4.73	1
Sumatera Barat	3.81	21.29	12.61	3.81	1
Riau	5.90	23.37	14.67	5.90	1
Jambi	2.02	19.06	10.42	2.02	1
Sumatera Selatan	3.88	21.32	12.62	3.88	1

Provinsi	C1	C2	C3	Jarak Terdekat	Kluster
Bengkulu	1.66	18.11	9.49	1.66	1
Lampung	2.19	16.20	7.58	2.19	1
Kep. Bangka Belitung	1.74	18.34	9.70	1.74	1
Kep. Riau	5.11	22.09	13.50	5.11	1
Dki Jakarta	8.36	25.75	17.08	8.36	1
Jawa Barat	2.46	19.68	11.02	2.46	1
Jawa Tengah	12.85	4.86	3.99	3.99	2
Di Yogyakarta	10.18	7.72	2.16	2.16	2
Jawa Timur	17.28	0.60	8.38	0.60	2
Banten	2.38	19.11	10.46	2.38	1
Bali	15.58	2.70	6.89	2.70	2
Nusa Tenggara Barat	31.22	13.61	22.32	13.61	2
Nusa Tenggara Timur	16.55	1.55	7.69	1.55	2
Kalimantan Barat	13.27	4.68	4.47	4.47	2
Kalimantan Tengah	5.07	22.50	13.83	5.07	1
Kalimantan Selatan	2.74	19.61	11.02	2.74	1
Kalimantan Timur	4.62	21.90	13.25	4.62	1
Kalimantan Utara	10.42	24.96	16.94	10.42	1
Sulawesi Utara	8.45	25.90	17.21	8.45	1
Sulawesi Tengah	0.00	17.63	8.91	0.00	1
Sulawesi Selatan	17.63	0.00	8.75	0.00	2
Sulawesi Tenggara	8.91	8.75	0.00	0.00	2
Gorontalo	1.71	18.87	10.23	1.71	1
Sulawesi Barat	15.66	2.69	6.97	2.69	2
Maluku	4.82	22.32	13.63	4.82	1
Maluku Utara	3.27	20.31	11.71	3.27	1
Papua Barat	3.62	14.82	6.45	3.62	1
Papua	77.14	59.89	68.36	59.89	3

Setelah didapatkan hasil jarak dari setiap objek (cost) pada iterasi ke-1 maka lanjut ke iterasi ke-2. Kandidat medoid baru (non-medoid) pada iterasi ke-2 dapat dilihat pada tabel 6.3.

Hitung kembali jarak dari setiap objek pada iterasi ke-2 dengan menggunakan medoid baru pada tabel 6.3, maka didapatkan hasil keseluruhannya dari iterasi ke-2 dapat dilihat pada tabel 6.4 sebagai berikut:

Tabel 6.4: Hasil Perhitungan Algoritma K-Medoids Iterasi ke-2

Provinsi	C1	C2	C3	Jarak Terdekat	Kluster
Aceh	2.05	17.76	2.84	2.05	1
Sumatera Utara	3.64	20.19	0.57	0.57	1
Sumatera Barat	2.65	19.32	1.07	1.07	1
Riau	4.94	21.47	1.40	1.40	1
Jambi	1.23	17.11	3.31	1.23	1
Sumatera Selatan	3.05	19.42	1.26	1.26	1
Bengkulu	1.45	16.15	4.29	1.45	3
Lampung	3.23	14.35	6.20	3.23	3
Kep. Bangka Belitung	1.57	16.41	4.05	1.57	1
Kep. Riau	4.01	20.19	1.77	1.77	1
Dki Jakarta	7.31	23.87	3.66	3.66	1
Jawa Barat	1.56	17.74	2.70	1.56	1
Jawa Tengah	14.09	3.40	17.53	3.40	3
Di Yogyakarta	11.29	5.84	14.76	5.84	3
Jawa Timur	18.55	2.66	21.98	2.66	2
Banten	2.37	17.28	3.49	2.37	1
Bali	16.72	1.80	20.20	1.80	3
Nusa Tenggara Barat	32.44	15.74	35.90	15.74	2
Nusa Tenggara Timur	17.83	2.13	21.28	2.13	3
Kalimantan Barat	14.61	3.84	17.94	3.84	3
Kalimantan Tengah	3.95	20.56	0.47	0.47	1
Kalimantan Selatan	1.60	17.64	2.93	1.60	1
Kalimantan Timur	3.53	19.97	1.00	1.00	1
Kalimantan Utara	10.46	23.63	8.36	8.36	1
Sulawesi Utara	7.40	24.01	3.75	3.75	1
Sulawesi Tengah	1.71	15.66	4.82	1.71	3
Sulawesi Selatan	18.87	2.69	22.32	2.69	2
Sulawesi Tenggara	10.23	6.97	13.63	6.97	3
Gorontalo	0.00	16.82	3.68	0.00	1
Sulawesi Barat	16.82	0.00	20.36	0.00	3
Maluku	3.68	20.36	0.00	0.00	1
Maluku Utara	1.93	18.32	2.30	1.93	1
Papua Barat	4.51	12.71	7.94	4.51	3
Papua	78.47	61.90	81.93	61.90	2

3. Hitung Total Simpangan (S)

Setelah didapatkan nilai jarak antara iterasi ke-1 dan iterasi ke-2, hitung total simpangan (S) dengan mencari selisih dari nilai total cost baru- nilai total cost lama. Dengan ketentuan jika $S < 0$, maka tukar nilai objek dengan menentukan medoid baru.

$$S = \text{Total cost baru} - \text{Total cost lama}$$

$$3063.56 - 1225.31$$

$$= 1,8338,25$$

Dengan nilai $S > 0$ proses kluster dihentikan, sehingga diperoleh anggota tiap kluster dapat dilihat pada tabel 6.5.

Tabel 6.5: Medoid Baru (Non-Medoid 2) Iterasi ke-3

Provinsi	C1	C2	C3	Jarak Terdekat	Cluster
Aceh	79.21	1.94	5.58	1.94	1
Sumatera Utara	81.77	2.10	7.81	2.10	1
Sumatera Barat	80.91	1.38	6.91	1.38	1
Riau	82.93	3.53	9.09	3.53	1
Jambi	78.76	1.37	4.86	1.37	1
Sumatera Selatan	80.89	1.92	7.13	1.92	1
Bengkulu	77.83	2.23	3.99	2.23	1
Lampung	75.91	4.27	3.06	3.06	3
Kep. Bangka Belitung	78.05	2.11	4.38	2.11	2
Kep. Riau	81.84	2.21	7.86	2.21	2
Dki Jakarta	85.36	5.73	11.46	5.73	2
Jawa Barat	79.35	1.05	5.49	1.05	2
Jawa Tengah	64.60	15.53	10.15	10.15	3
Di Yogyakarta	67.54	12.67	7.35	7.35	3
Jawa Timur	60.18	19.99	14.51	14.51	3
Banten	78.75	2.17	5.32	2.17	2
Bali	62.25	18.11	12.64	12.64	3
Nusa Tenggara Barat	46.75	33.86	28.35	28.35	3
Nusa Tenggara Timur	60.79	19.32	13.78	13.78	3
Kalimantan Barat	64.18	16.01	10.65	10.65	3
Kalimantan Tengah	82.14	2.43	8.14	2.43	2

Provinsi	C1	C2	C3	Jarak Terdekat	Cluster
Kalimantan Selatan	79.35	0.75	5.31	0.75	2
Kalimantan Timur	81.58	1.87	7.60	1.87	2
Kalimantan Utara	83.53	9.43	13.10	9.43	2
Sulawesi Utara	85.47	5.88	11.60	5.88	2
Sulawesi Tengah	77.14	3.27	3.62	3.27	2
Sulawesi Selatan	59.89	20.31	14.82	14.82	3
Sulawesi Tenggara	68.36	11.71	6.45	6.45	3
Gorontalo	78.47	1.93	4.51	1.93	2
Sulawesi Barat	61.90	18.32	12.71	12.71	3
Maluku	81.93	2.30	7.94	2.30	2
Maluku Utara	80.04	0.00	5.97	0.00	2
Papua Barat	74.41	5.97	0.00	0.00	3
Papua	0.00	80.04	74.41	0.00	1

Dengan langkah yang sama seperti sebelumnya untuk menentukan jarak dari setiap objek pada iterasi ke-3.

$$\begin{aligned}
 D_{aceh,e1} &= \sqrt{(3,13 - 3,78)^2 + 2,74 - 3,48)^2 + (3,66 - 3,65)^2} \\
 &\quad + (3,38 - 3,12)^2 + (2,98 - 2,34)^2 + (2,28 - 1,54)^2 \\
 &\quad + (2,09 - 1,39)^2 + (2,26 - 1,33)^2 + (1,85 - 1,22)^2 \\
 &= 19,94
 \end{aligned}$$

$$\begin{aligned}
 D_{aceh,e2} &= \sqrt{(3,13 - 6,40)^2 + 2,74 - 4,41)^2 + (3,66 - 6,99)^2} \\
 &\quad + (3,38 - 4,80)^2 + (2,98 - 4,31)^2 + (2,28 - 3,37)^2 \\
 &\quad + (2,09 - 2,83)^2 + (2,26 - 2,95)^2 + (1,85 - 2,62)^2 \\
 &= 5,58
 \end{aligned}$$

Tabel 6.6: Data Perhitungan Algoritma K-Medoids Hasil Iterasi ke-3

Provinsi	C1	C2	C3	Jarak Terdekat	Kluster
Aceh	79.21	1.94	5.58	1.94	1
Sumatera Utara	81.77	2.10	7.81	2.10	1
Sumatera Barat	80.91	1.38	6.91	1.38	1
Riau	82.93	3.53	9.09	3.53	1
Jambi	78.76	1.37	4.86	1.37	1
Sumatera Selatan	80.89	1.92	7.13	1.92	1

Provinsi	C1	C2	C3	Jarak Terdekat	Kluster
Bengkulu	77.83	2.23	3.99	2.23	1
Lampung	75.91	4.27	3.06	3.06	3
Kep. Bangka Belitung	78.05	2.11	4.38	2.11	2
Kep. Riau	81.84	2.21	7.86	2.21	2
Dki Jakarta	85.36	5.73	11.46	5.73	2
Jawa Barat	79.35	1.05	5.49	1.05	2
Jawa Tengah	64.60	15.53	10.15	10.15	3
Di Yogyakarta	67.54	12.67	7.35	7.35	3
Jawa Timur	60.18	19.99	14.51	14.51	3
Banten	78.75	2.17	5.32	2.17	2
Bali	62.25	18.11	12.64	12.64	3
Nusa Tenggara Barat	46.75	33.86	28.35	28.35	3
Nusa Tenggara Timur	60.79	19.32	13.78	13.78	3
Kalimantan Barat	64.18	16.01	10.65	10.65	3
Kalimantan Tengah	82.14	2.43	8.14	2.43	2
Kalimantan Selatan	79.35	0.75	5.31	0.75	2
Kalimantan Timur	81.58	1.87	7.60	1.87	2
Kalimantan Utara	83.53	9.43	13.10	9.43	2
Sulawesi Utara	85.47	5.88	11.60	5.88	2
Sulawesi Tengah	77.14	3.27	3.62	3.27	2
Sulawesi Selatan	59.89	20.31	14.82	14.82	3
Sulawesi Tenggara	68.36	11.71	6.45	6.45	3
Gorontalo	78.47	1.93	4.51	1.93	2
Sulawesi Barat	61.90	18.32	12.71	12.71	3
Maluku	81.93	2.30	7.94	2.30	2
Maluku Utara	80.04	0.00	5.97	0.00	2
Papua Barat	74.41	5.97	0.00	0.00	3
Papua	0.00	80.04	74.41	0.00	1

Setelah didapatkan nilai jarak iterasi ke-3, hitung total simpangan (S) dengan mencari selisih dari nilai total cost baru- nilai total cost lama. Dengan ketentuan jika $S < 0$, maka tukar nilai objek dengan menentukan medoid baru.

$$S = \text{Total cost baru} - \text{Total cost lama}$$

$$= 3063.56 - 1225.31$$

$$= 1838,25$$

Dengan nilai $S > 0$ proses kluster dihentikan, sehingga diperoleh anggota tiap kluster dapat dilihat pada tabel 6.7.

Tabel 6.7: Hasil Pengklasteran dengan K-Medoids Clustering

Provinsi	C1	C2	C3	Jarak Terdekat	Cluster
Aceh	79.21	1.94	5.58	1.94	1
Sumatera Utara	81.77	2.10	7.81	2.10	1
Sumatera Barat	80.91	1.38	6.91	1.38	1
Riau	82.93	3.53	9.09	3.53	1
Jambi	78.76	1.37	4.86	1.37	1
Sumatera Selatan	80.89	1.92	7.13	1.92	1
Bengkulu	77.83	2.23	3.99	2.23	1
Lampung	75.91	4.27	3.06	3.06	3
Kep. Bangka Belitung	78.05	2.11	4.38	2.11	2
Kep. Riau	81.84	2.21	7.86	2.21	2
Dki Jakarta	85.36	5.73	11.46	5.73	2
Jawa Barat	79.35	1.05	5.49	1.05	2
Jawa Tengah	64.60	15.53	10.15	10.15	3
Di Yogyakarta	67.54	12.67	7.35	7.35	3
Jawa Timur	60.18	19.99	14.51	14.51	3
Banten	78.75	2.17	5.32	2.17	2
Bali	62.25	18.11	12.64	12.64	3
Nusa Tenggara Barat	46.75	33.86	28.35	28.35	3
Nusa Tenggara Timur	60.79	19.32	13.78	13.78	3
Kalimantan Barat	64.18	16.01	10.65	10.65	3
Kalimantan Tengah	82.14	2.43	8.14	2.43	2
Kalimantan Selatan	79.35	0.75	5.31	0.75	2
Kalimantan Timur	81.58	1.87	7.60	1.87	2
Kalimantan Utara	83.53	9.43	13.10	9.43	2
Sulawesi Utara	85.47	5.88	11.60	5.88	2
Sulawesi Tengah	77.14	3.27	3.62	3.27	2
Sulawesi Selatan	59.89	20.31	14.82	14.82	3
Sulawesi Tenggara	68.36	11.71	6.45	6.45	3
Gorontalo	78.47	1.93	4.51	1.93	2
Sulawesi Barat	61.90	18.32	12.71	12.71	3
Maluku	81.93	2.30	7.94	2.30	2

Provinsi	C1	C2	C3	Jarak Terdekat	Cluster
Maluku Utara	80.04	0.00	5.97	0.00	2
Papua Barat	74.41	5.97	0.00	0.00	3
Papua	0.00	80.04	74.41	0.00	1

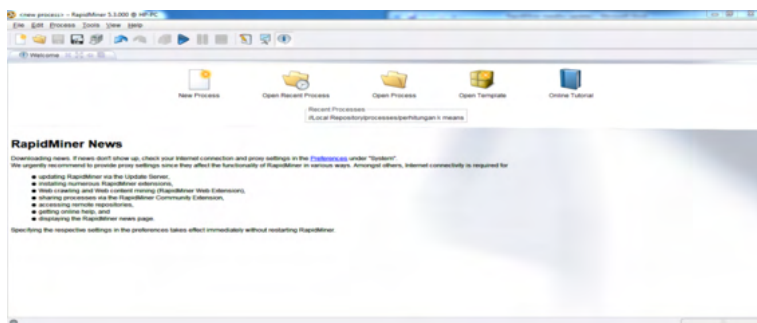
6.2 Pengujian Dengan RapidMiner

Pada tampilan awal tools Rapidminer 5.3 terdapat beberapa menu pilihan seperti NewProcess, Open Recent Process, Open Process, Open Template dan OnlineTutorial dapat dilihat pada gambar 6.1.

6.3 Prosedur Kerja RapidMiner

Prosedur kerja dari perangkat lunak RapidMiner sebagai berikut :

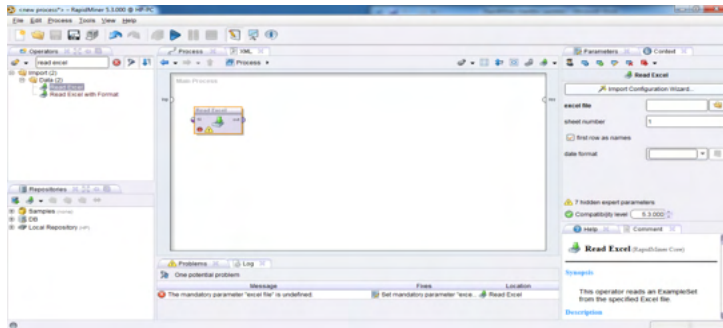
1. Klik **“New Process”** untuk membuat sebuah lembar pekerjaan yang baru untuk memulai sebuah proses pengolahan data



Gambar 6.1: Tampilan Awal Rapidminer 5.3

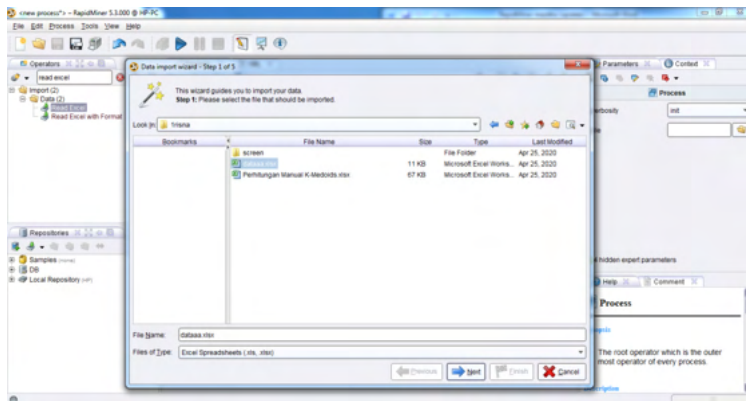
2. Lalu selanjutnya pilih pada bagian pojok kiri atas bagian **“Operators”** dan ketik **“Read Excel”** dikolom tersebut, maka secara otomatis akan tampil seperti berikut ini :

54 Data Mining : Penerapan Algoritma K-Means Clustering dan K-Medoids Clustering



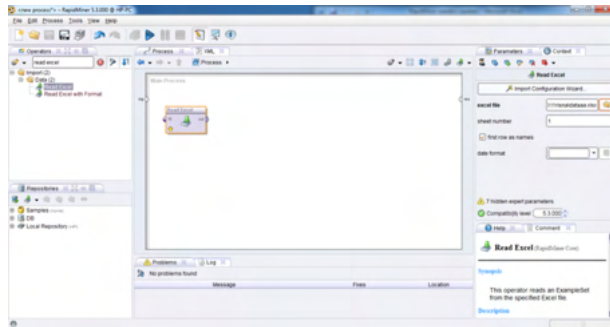
Gambar 6.2: Tampilan New Process

3. Selanjutnya pilih file yang akan diolah pada **“excel file”** dan klik **“Open”**



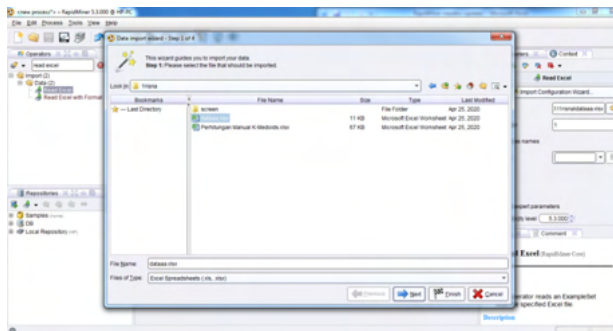
Gambar 6.3: Input Data Excel

4. Selanjutnya setelah file terpilih yang akan dilakukan selanjutnya yaitu memilih data yang akan dikluster, dengan cara memilih **“Import Configuration Wizard”** pada bagian pojok kanan atas



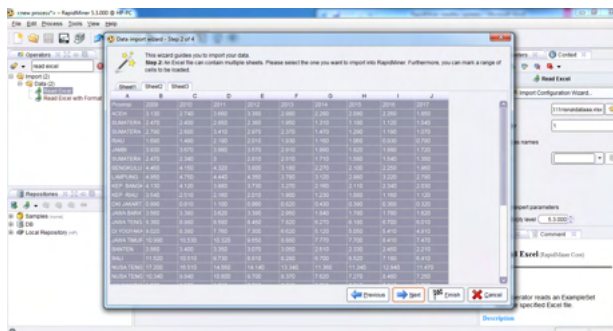
Gambar 6.4: Pemrosesan Data Tahap 1

5. Setelah data dipilih selanjutnya klik “Next”



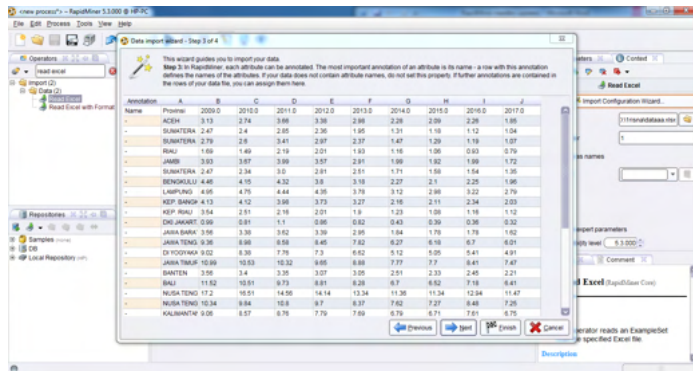
Gambar 6.5: Alur proses import step 1

6. Lalu blok data yang akan di cluster dan pilih “Next” lagi



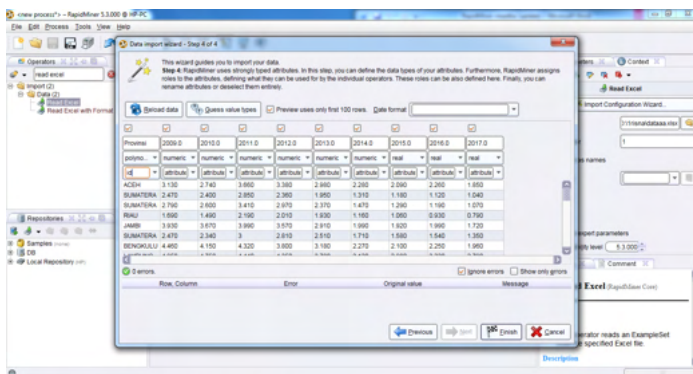
Gambar 6.6: Data import data tahap 2

Tampilan akan tampak seperti gambar di bawah ini, lalu pilih “Next” lagi



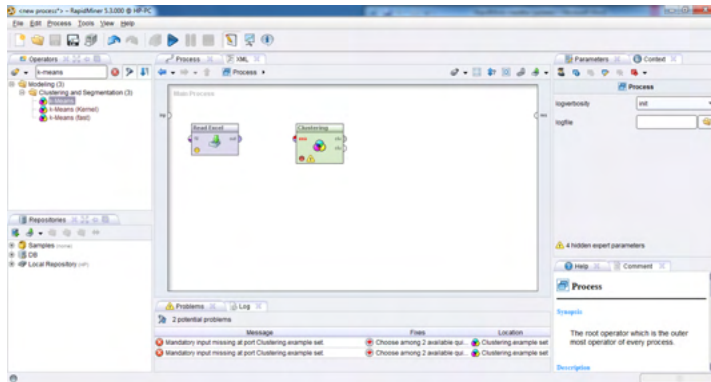
Gambar 6.7: Data import tahap 3

7. Berikutnya yaitu merubah kolom pertama dengan type data menjadi “id” lalu pilih “finish”.



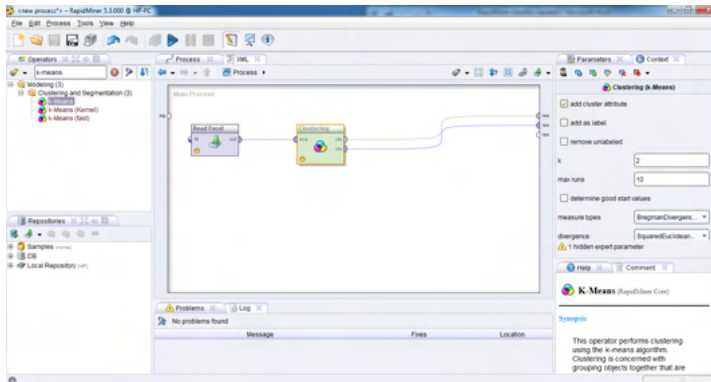
Gambar 6.8: Data import Tahap 4

8. Selanjutnya yaitu memilih Metode K-Means yang akan digunakan untuk mengkluster data yaitu dengan memilih pada bagian operators “K-Means” dan ubah nilai “k” menjadi beberapa kluster yang akan digunakan.



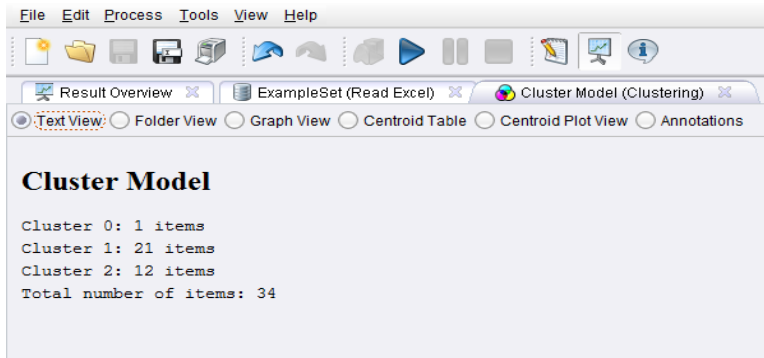
Gambar 6.9: Pemrosesan Data Tahap 1

9. Hubungkan operator “excel” dan “K-Means” lalu tekan tombol “Play” pada bagian toolbar.



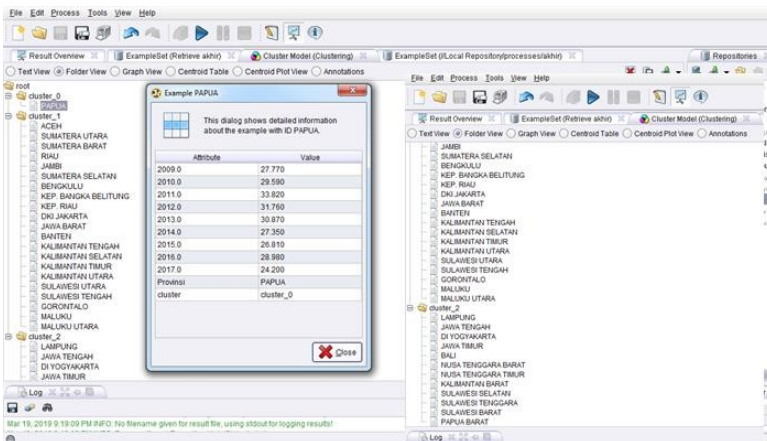
Gambar 4.10: Pemrosesan Data Tahap 1

10. Berikut adalah hasil kluster yang telah didapat dengan menggunakan sebanyak 3 nilai kluster.



Gambar 6.11: Nilai kluster Model RapidMiner

Selanjutnya adalah tampilan “**Folder View**” yang ada pada perangkat lunak RapidMiner 5.3.



Gambar 6.12: Folder view RapidMiner

Keterangan :

1. Jumlah kluster 0 (Tinggi) berjumlah 1 items
2. Jumlah kluster 1 (Rendah) berjumlah 21 items
3. Jumlah kluster 3(sedang) berjumlah 12 items
4. Jumlah keseluruhan items adalah 34

6.4 Validasi Data

Pada data yang telah dianalisa disimpulkan bahwa data yang digunakan adalah valid. Ini dibuktikan dengan hasil akhir perhitungan manual dengan hasil akhir dari perangkat lunak RapidMiner 5.3 dapat menampilkan hasil akhir yang sama. Berikut adalah tampilan kluster model berupa teks yang ada dalam aplikasi RapidMiner 5.3. Dimana tampilan ini adalah penjelasan dari kluster yang terbentuk.

Tabel 6.8: Hasil Perhitungan K-Means dengan Perangkat Lunak RapidMiner

Cal	Provinsi	K-Medoid	Rapidminer
1	Aceh	Cluster_1	Cluster 1
2	Sumatera Utara	Cluster_1	Cluster 1
3	Sumatera Barat	Cluster_1	Cluster 1
4	Riau	Cluster_1	Cluster 1
5	Jambi	Cluster_1	Cluster 1
6	Sumatera Selatan	Cluster_1	Cluster 1
7	Bengkulu	Cluster_1	Cluster 1
8	Lampung	Cluster_3	Cluster 2
9	Kepulauan Bangka Belitung	Cluster_1	Cluster 1
10	Kepulauan Riau	Cluster_1	Cluster 1
11	Dki Jakarta	Cluster_1	Cluster 1
12	Jawa Barat	Cluster_1	Cluster 1
13	Jawa Tengah	Cluster_3	Cluster 2
14	Di Yogyakarta	Cluster_3	Cluster 2
15	Jawa Timur	Cluster_3	Cluster 2
16	Banten	Cluster_1	Cluster 1
17	Bali	Cluster_3	Cluster 2
18	Nusa Tenggara Barat	Cluster_3	Cluster 2
19	Nusa Tenggara Timur	Cluster_3	Cluster 2
20	Kalimantan Barat	Cluster_3	Cluster 2

Cal	Provinsi	K-Medoid	Rapidminer
21	Kalimantan Tengah	Cluster_1	Cluster 1
22	Kalimantan Selatan	Cluster_1	Cluster 1
23	Kalimantan Timur	Cluster_1	Cluster 1
24	Kalimantan Utara	Cluster_1	Cluster 1
25	Sulawesi Utara	Cluster_1	Cluster 1
26	Sulawesi Tengah	Cluster_1	Cluster 1
27	Sulawesi Selatan	Cluster_3	Cluster 2
28	Sulawesi Tenggara	Cluster_3	Cluster 2
29	Gorontalo	Cluster_1	Cluster 1
30	Sulawesi Barat	Cluster_3	Cluster 2
31	Maluku	Cluster_1	Cluster 1
31	Maluku Utara	Cluster_1	Cluster 1
32	Papua Barat	Cluster_3	Cluster 2
33	Papua	Cluster_0	Cluster 0

Daftar Pustaka

- Angga Ginanjar Mabur, L. R. (2012) 'Penerapan Data Mining Untuk Memprediksi Kriteria Nasabah Kredit', *Jurnal Komputer dan Informatika (KOMPUTA)*, 1(1), pp. 53–57.
- Darmi, Y. and Setiawan, A. (2016) 'Penerapan Metode Clustering K-Means Dalam', *Y. Darmi, A. Setiawan*, 12(2), pp. 148–157.
- Eska, J. (2016) 'Penerapan Data Mining Untuk Prekdiksi Penjualan Wallpaper Menggunakan Algoritma C4.5 STMIK Royal Ksiran', *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, 2, pp. 9–13.
- Indriani, I. et al. (2019) 'Algoritma K-Medoids untuk Mengelompokkan Desa yang Memiliki Fasilitas Sekolah di Indonesia', (September), pp. 520–527.
- Maulana, A. and Fajrin, A. A. (2018) 'Penerapan Data Mining Untuk Analisis Pola Pembelian Konsumen Dengan Algoritma Fp-Growth Pada Data Transaksi Penjualan Spare Part Motor', *Klik - Kumpulan Jurnal Ilmu Komputer*, 5(1), p. 27. doi: 10.20527/klik.v5i1.100.
- Mustafa, M. S., Ramadhan, M. R. and Thenata, A. P. (2018) 'Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier', *Creative Information Technology Journal*, 4(2), p. 151. doi: 10.24076/citec.2017v4i2.106.
- Ramdhani, F. and Hoyyi, A. (2015) 'Pengelompokan Provinsi Di Indonesia Berdasarkan Karakteristik Kesejahteraan Rakyat Menggunakan Metode K-Means Cluster', *Jurnal Gaussian*, 4(4), pp. 875–884.
- Rofiqi, A. Y. (2017) 'Clustering Berita Olahraga Berbahasa Indonesia Menggunakan Metode K-Medoid Bersyarat', *Jurnal SimanteC*, 6(1), pp. 25–32.
- Sundari, S. et al. (2019) 'Analisis K-Medoids Clustering Dalam Pengelompokkan Data Imunisasi Campak Balita di Indonesia', *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 1(September), p. 687. doi: 10.30645/senaris.v1i0.75.

- Syahdan, S. Al and Sindar, A. (2018) 'Data Mining Penjualan Produk Dengan Metode Apriori Pada Indomaret Galang Kota', Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI), 1(2). doi: 10.32672/jnkti.v1i2.771.
- Tiwari, M. and Singh, R. (2012) 'Comparative Investigation of K-Means and K-Medoid Algorithm on Iris Data', International Journal of Engineering Research and Development, 4(8), pp. 69–72. Available at: www.ijerd.com.
- Witanto, S., Ratnawati, D. E. and Anam, S. (2019) 'Pengelompokan Fungsi Aktif Senyawa Data SMILES (Simplified Molecular Input Line Entry System) Menggunakan Metode K-Means Dengan Inisialisasi Pusat Klaster Menggunakan Metode Heuristic $O(N \log N)$ ', Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 3(1), pp. 702–707.

Biodata Penulis

Dr. Mochamad Wahyudi, MM, M.Kom., M.Pd.



Lahir di Jakarta. Menyelesaikan pendidikan formal STMIK Budi Luhur di Jakarta Program Studi Teknik Komputer (S-1) Tahun 1997, Universitas Budi Luhur di Jakarta Program Studi Magister Manajemen (S-2) Konsentrasi Manajemen Sistem Informasi Tahun 2003 Universitas Budi Luhur di Jakarta Program Studi Magister Ilmu Komputer (S-2) Konsentrasi Rekayasa e-Bisnis Tahun 2008, Universitas Muhammadiyah Prof. Dr. Hamka (UHAMKA), Jakarta. Program Studi Magister Administrasi Pendidikan (S-2) Tahun 2012, Universitas Negeri Jakarta (UNJ), Jakarta di Program Studi Doktor Manajemen Pendidikan/Administrasi Pendidikan (S-3) Tahun 2014 dan saat sedang menyelesaikan studi di Universitas Sumatera Utara (USU) di Medan Program Studi Doktor Ilmu Komputer (S3) Tahun 2018 s/d sekarang. Praktisi di Bidang IT ini saat ini bekerja sebagai dosen di Universitas Bina Sarana Informatika, Program Studi Magister Ilmu Komputer (S2) STMIK Nusa Mandiri dan Program Doktor (S3) Universitas Negeri Jakarta. Dosen dengan jabatan fungsional Lektor Kepala (Associate Profesor) ini memiliki pangkat penyetaraan Pembina Utama Muda (IV/c) dan saat ini menjabat sebagai Rektor di Universitas Bina Sarana Informatika di Jakarta. Memiliki berbagai sertifikasi kompetensi dari Badan Nasional Sertifikasi Profesi (BNSP) antara lain untuk Bidang Keahlian ICT Project Manager, Bidang Keahlian Programmer dan Bidang Keahlian Network Administrator. Memiliki sertifikasi internasional seperti : Certified CCNA dari Cisco, Certified Hacking Forensic Investigator (CHFI) dari EC-Council, USA dan Certified Ethical Hacking and Countermeasures (CEH). EC-Council, USA. Banyak publikasi yang telah

dihasilkan dalam Bidang Informatika dan Komputer (Ilmu Komputer) dan diterbitkan pada jurnal-jurnal nasional maupun internasional.

Masitha



Lahir di Marihat Butar III pada 23 Juli 1998. Mahasiswa semester akhir STIKOM Tunas Bangsa Pematangsiantar, Program Studi Sistem Informasi.

Risna Saragih



Lahir di Marubun betania, 28 Mei 1998. Mahasiswa semester akhir STIKOM Tunas Bangsa Pematangsiantar, Program Studi Sistem Informasi.

Solikhun, S.Kom., M.Kom.



Lahir di Pemalang pada tanggal 04 Januari 1983. Ia menyelesaikan kuliah dan mendapat gelar Sarjana Komputer pada 23 November 2008. Ia merupakan alumnus Program Studi Teknik Informatika Sekolah Tinggi Manajemen Informatika Dan Komputer Kristen Neumann Indonesia. Pada tahun 2011 mengikuti Program Magister Teknik Informatika Universitas Putra Indonesia “YPTK” Padang dan lulus pada tahun 2012. Pada tahun 2018 mengikuti Program Doktorat Ilmu Komputer Universitas Sumatera Utara. Pada tahun 2008 diangkat menjadi Dosen AMIK Tunas Bangsa dan ditempatkan di Program Studi Manajemen Informatika.

DATA MINING

Penerapan Algoritma K-Means Clustering dan K-Medoids Clustering

Buku ini membahas tentang penerapan Algoritma K-Means dan Algoritma K-Medoids dalam mengelompokkan data.

Buku ini membahas :

Bab 1 : Data Mining

Bab 2 : Konsep Algoritma K-Means Clustering

Bab 3 : Konsep Algoritma K-Medoids Clustering

Bab 4 : Rapidminer dan Pemasangan Perangkat Lunak

Bab 5 : Penerapan Algoritma K-Means Clustering

Bab 6 : Penerapan Algoritma K-Medoids Clustering



YAYASAN KITA MENULIS
press@kitamenulis.id
www.kitamenulis.id

ISBN 978-623-7645-81-8

