



ISSN : 2339 - 1871

## JURNAL ILMIAH BETRIK

Besemah Teknologi Informasi dan Komputer

Editor Office : LPPM Sekolah Tinggi Teknologi Pagar Alam, Jln. Masik Siagim No. 75  
Simpang Mbacang, Pagar Alam, SUM-SEL, Indonesia  
Phone : +62 852-7901-1390.

Email : [betrik@sttpagaralam.ac.id](mailto:betrik@sttpagaralam.ac.id) | [admin.jurnal@sttpagaralam.ac.id](mailto:admin.jurnal@sttpagaralam.ac.id)

Website : <https://ejournal.sttpagaralam.ac.id/index.php/betrik/index>

### ANALISIS PENGELOMPOKAN PENENTUAN JURUSAN SISWA SMA MENGUNAKAN METODE K-MEANS CLUSTERING

Sri Wahyuni Nengsih<sup>1</sup>, Ibnu Alfian<sup>2</sup>, Dimas Aji<sup>3</sup>, Saeful Anwar<sup>4</sup>

<sup>1,2,3,4</sup>Program Studi Teknik Informatika Sekolah Tinggi Manajemen Informatika Komputerisasi  
(STMIK IKMI Cirebon)

Jl. Perjuangan No. 10B Kec. Kesambi Kota Cirebon, Jawa Barat 45131

Sur-el : [sriwahyuni12@gmail.com](mailto:sriwahyuni12@gmail.com)<sup>1</sup>, [alfianibnu@gmail.com](mailto:alfianibnu@gmail.com)<sup>2</sup>, [ajidimas10@gmail.com](mailto:ajidimas10@gmail.com)<sup>3</sup>,  
[saefula21@gmail.com](mailto:saefula21@gmail.com)<sup>4</sup>

**Abstrak:** Pesatnya perkembangan ilmu pengetahuan dan teknologi telah merubah kedalam semua aspek kehidupan seperti dalam bisnis, sosial, ataupun dalam aspek pendidikan. Teknologi dalam peran pendidikan tentunya sangat penting dalam mendukung sistem pembelajaran dikurikulum 2013. Salah satu cara untuk mempermudah penentuan jurusan SMA adalah dengan menggunakan metode K-Means Clustering. Algoritma K-Means merupakan algoritma yang paling sederhana jika dibandingkan dengan yang lain, algoritma ini mempunyai kelebihan mudah diterapkan dan dijalankan, mudah diadaptasi dan banyak digunakan untuk dipraktekan dalam tugas data minin. K-Means membagi data kemudian mengelompokannya kedalam beberapa *cluster* yang memiliki kemiripan dan membagi setiap cluster berdasarkan perbedaan antar *cluster*. *Clustering* adalah proses pengempokan data serupa kedalam kelompok yang berbeda, lebih tepatnya partisi dari sebuah data set kedalam subset sehingga data data disetiap subset memiliki manfaat. *Clustering* juga dikatakan suatu proses dimana mengelompokan dan membagi menjadi beberapa data set hingga membentuk pola yang serupa dan dikelompokan pada cluster yang sama dan memisahkan diri dengan membentuk pola yang berbeda di cluster yang berbeda. hasil dari penelitian yang sudah dilakukan dalam menganalisa pengelompokan penentuan jurusan menggunakan metode k-means dapat disimpulkan bahwa k-means dapat digunakan untuk mengelompokan penentuan jurusan dengan hasil yang optimal dan tidak memakan waktu yang lama. Berdasarkan hasil pengujian menunjukan bahwa cluster\_0 ada 5 siswa yang ditolak di dua jurusan, cluster\_1 142 siswa masuk jurusan IPS dan cluster\_2 177 siswa masuk dalam jurusan IPA. Membandingkan dengan hasil dari penjurusan siswa yang dilakukan SMA 21 Bandung menggunakan *excel* sangat berbeda dengan hasil yang sudah diuji menggunakan *RapidMiner*, terdapat 144 siswa jurusan

**Kunci Utama:** Penentuan Jurusan, k-means.

**Abstract:** The 2013 school year curriculum for admission of new students has undergone changes where the determination of the majors is usually done at the time of advancement of class XI, now prospective students are required to choose a major at the time of entering high school registration. Determining the department of SMA Negeri 21 Bandung using student report cards grades from semester 1 to 5. This of course will take a long time in the process of determining the majors, as well as the curriculum section which still uses manual methods to make mistakes in determining majors. The cause of the students being wrong in determining the major is not recognizing the characteristics of each department, not recognizing their interests

*and talents. Based on the problems that have been described, a solution can be proposed to group the direction using the K-Means Clustering method. This method is used to classify and analyze data into machine learning and produce information from a data analysis. The purpose of this study is to facilitate the curriculum section in determining majors and there are no mistakes for students in administering majors. The results of grouping in determining the majors of SMA Negeri 21 Bandung using the K-Means method get 3 clusters. Cluster\_0 5 students do not enter 2 majors, cluster\_1 142 social studies students and cluster\_2 177 science students and this greatly facilitates the curriculum in determining majors. Solution for the part of SMA Negeri 21 Bandung to be more careful in determining the majors.*

**Keywords :** *Determination of Department, k-means.*

## I. PENDAHULUAN

Pesatnya perkembangan ilmu pengetahuan dan teknologi telah merubah kedalam semua aspek kehidupan seperti dalam bisnis, sosial, ataupun dalam aspek pendidikan. Teknologi dalam peran pendidikan tentunya sangat penting dalam mendukung sistem pembelajaran dikurikulum 2013. Dimana penjurusan untuk Sekolah Menengah Atas (SMA) dalam menentukan penjurusan dilaksanakan ketika akan memasuki jenjang SMA atau bisa disebut waktu pendaftaran awal SMA dilakukan penjurusan. Penentuan penjurusan dilakukan berdasarkan kriteria-kriteria dari nilai akademik yang menjadi ciri dari masing-masing jurusan, dengan demikian dapat diperoleh kemungkinan siswa memenuhi semua jurusan atau sebagian dari kriteria jurusan [8].

Penentuan jurusan bagi siswa/siswi merupakan proses dalam memfokuskan siswa dibidang konsentrasi tertentu, hal ini dilakukan agar setiap individu dapat mempelajari lebih dalam tentang pelajaran yang sesuai dengan konsentrasi yang telah ditentukan dan mengarahkan setiap individu agar bisa mengembangkan kemampuan diri dan minat yang dimiliki. Penjurusan diharapkan dapat memaksimalkan setiap potensi atau bakat yang dimiliki setiap individu.

Dapat pula memaksimalkan nilai akademis. Kesalahan dalam pemilihan jurusan yang tidak sesuai dengan yang seharusnya siswa terima, hal itu akan menyebabkan menurunnya motivasi belajar dan rendahnya prestasi karena

jurusan yang dipilih ternyata tidak sesuai dengan bakat yang dimiliki.

Penjurusan siswa dibagi menjadi 2 yaitu jurusan IPA dan jurusan IPS, penjurusan ini memiliki kriterianya masing-masing. Kriteria dari penjurusan IPA itu sendiri adalah fisika, matematika, kimia, biologi, sedangkan penjurusan untuk IPS terdiri dari sejarah, geografi, ekonomi dan sosiologi [1]. Sesuai dengan kurikulum ajaran tahun 2013 penjurusan dilakukan pada saat pendaftaran masuk kelas 10, hal itu bertujuan agar pelajaran yang akan diberikan kepada siswa/siswi menjadi lebih terarah karena akan lebih mendalami mata pelajaran yang telah mereka pilih sesuai dengan minat dan kemampuannya [2]. Proses penjurusan pada kurikulum ajaran 2013 di SMA ini cukup rumit, karena penjurusan yang akan dilakukan menggunakan data rapot Sekolah Menengah Pertama (SMP) dari hasil belajar siswa mulai dari semester 1 sampai dengan semester 5 yang dilampirkan siswa [3].

Salah satu cara untuk mempermudah penentuan jurusan SMA adalah dengan menggunakan metode *K-Means Clustering*. Algoritma *K-Means* merupakan algoritma yang paling sederhana jika dibandingkan dengan yang lain, algoritma ini mempunyai kelebihan mudah diterapkan dan dijalankan, mudah diadaptasi dan banyak digunakan untuk dipraktikkan dalam tugas data minin. *K-Means* membagi data kemudian mengelompokkannya kedalam beberapa cluster yang memiliki kemiripan dan membagi setiap cluster berdasarkan

perbedaan antar cluster [4]. Clustering adalah proses pengempokan data serupa kedalam kelompok yang berbeda, lebih tepatnya partisi dari sebuah data set kedalam subset sehingga data data disetiap subset memiliki manfaat [5]. Clustering juga dikatakan suatu proses dimana mengelompokkan dan membagi menjadi beberapa data set hingga membentuk pola yang serupa dan dikelompokkan pada cluster yang sama dan memisahkan diri dengan membentuk pola yang berbeda di cluster yang berbeda [6].

Berdasarkan pada permasalahan diatas, maka dibuat analisis tentang pengelompokan penentuan jurusan untuk siswa yang bertujuan untuk memberikan kemudahan bagi siswa dalam memilih jurusan yang sesuai dengan kemampuan dan minat dari siswa itu sendiri. Selain itu penjurusan menggunakan metode k-means clustering ini bertujuan juga untuk memudahkan dan meningkatkan kinerja guru dalam pengolahan data.

## II. METODE PENELITIAN

### 2.1 Populasi dan Sampel

Populasi merupakan wilayah generaliasi yang terdiri dari objek atau subjek dan mempunyai kualitas serta memiliki karakteristik tertentu yang telah ditetapkan oleh peneliti dan ditarik kesimpulannya. Populasi yang dimaksud dalam penelitian ini adalah siswa tahun ajaran 2020/2021 SMA Negeri 21 Bandung. Sedangkan untuk sampel sendiri adalah bagian dari jumlah karakteristik tersebut dan bila populasi itu besar, maka peneliti dapat menggunakan sampel yang diambil dari populasi tersebut.

### 2.2 Metode Pengumpulan Data

Dalam penelitian ini teknik yang digunakan untuk pengumpulan data yaitu dengan cara:

### 2.3 Observasi

Melalui observasi penulis melakukan tinjauan langsung pada SMA Negeri 21 Bandung untuk mendapatkan informasi mengenai profil sekolah dan data-data yang akan diperlukan dalam pembuatan bahan skripsi. Data yang digunakan dalam penelitian ini adalah data nilai raport semester 1s/d 5 dari siswa pendaftar SMA Negeri 21 Bandung.

### 2.4 Studi Literatur

Mengkaji dan mempelajari teori analisis data mengenai pengelompokan *k-means clustering* dan pemilihan jurusan serta beberapa metode pengolahan data melalui jurnal-jurnal dan buku yang berkaitan dengan jurnal pengelompokan penentuan jurusan menggunakan metode k-means.

### 2.5 Jenis Data

Jenis data terdiri dari jenis data primer dan jenis data sekunder. Berikut penjelasan dari data primer dan data skunder.

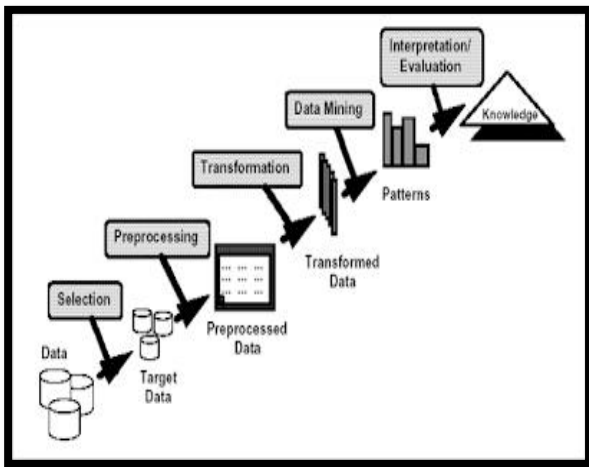
#### a. Data Primer

Data primer adalah data yang diperoleh peneliti secara langsung. Pada penelitian ini sumber yang dijadikan objek adalah siswa pendaftar SMA Negeri 21 Bandung sebanyak 324 siswa.

#### b. Data Sekunder

Data sekunder adalah data yang diperoleh peneliti dari sumber atau data yang sudah ada. Data sekunder yang digunakan dalam penelitian ini adalah dataset dari nilai raport pendaftar SMA Negeri 21 Bandung.

Alur penelitian yang digunakan dalam penelitian ini menggunakan tahapan *Knowledge Discovery in Database (KDD)* yang terdiri dari beberapa proses, berikut



proses tahapan KDD yang dilakukan dalam menganalisis data seperti pada gambar 1

**Gambar 1. Proses *Knowledge Discovery in Database***

Data yang digunakan pada penelitian ini ialah data pendaftar SMA Negeri 21 Bandung tahun ajaran 2020/2021, dan data yang dipakai sebanyak 324 siswa.<sup>3</sup>

Data yang digunakan terdiri dari:

- a. *Data selection* (Seleksi Data) Data yang ada dalam database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk analisis yang akan diambil dari database. Data yang diperoleh dari SMA Negeri 21 memiliki isian yang tidak sempurna seperti data yang belum terisi nilai raport, data yang tidak *valid* atau data yang salah penulisan nama lengkap. Pembersihan data akan mempengaruhi sistem data mining karena data yang akan ditangani akan berkurang jumlah dan kompleksitasnya.
- b. *Preprocessing* Proses pendahuluan dan pembersihan data merupakan operasi dasar seperti penghapusan nois dilakukan. Sebelum proses data mining dapat dilakukan, perlu adanya proses cleaning pada data yang menjadi fokus KDD. Proses cleaning itu sendiri terdiri

dari membuang duplikasi data, memperbaiki kesalahan pada data seperti pada kesalahan penulisan nama lengkap siswa atau salah penulisan pada asal sekolah siswa.

- c. *Transformation* Data diubah atau digabung kedalam format yang sesuai untuk diproses dalam data mining.
- d. *Data Mining* Proses Data mining yaitu proses mencari pola atau informasi menarik dalam data terpilih pada dataset dari SMA Negeri 21 Bandung dengan menggunakan teknik atau metode tertentu. Tahap ini merupakan inti dari tahapan *knowledge discovery in databases (KDD)* yang dilakukan untuk menganalisis data yang telah dibersihkan.
- e. *Interpretation/Evaluation* Untuk mengidentifikasi pola-pola menarik kedalam *knowledge based* yang ditemukan. *Knowledge Presentation* (Presentasi Pola yang untuk menghasilkan aksi)

### III. HASIL DAN PEMBAHASAN

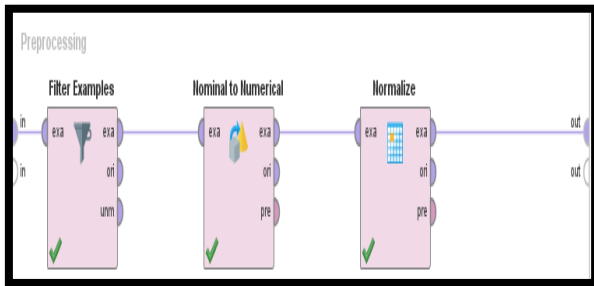
#### 3.1 Hasil Penelitian

*Data cleaning*, proses ini bermaksud untuk membersihkan data tidak sesuai, atau duplikat data pada metode ini tak hanya membersihkan data akan tetapi bisa juga untuk membetulkan kekeliruan dalam pengetikan. Pada penelitian ini data sudah terlepas dari ketidak konsistenan sehingga disimpulkan bahwa data sudah layak digunakan untuk proses *clustering*.

*Preprocessing*, proses yang digunakan untuk menyaring data yang diperlukan dalam proses *clustering*, data didapat dari pendaftar yang diterima di SMA Negeri 21 Bandung. Variabel yang digunakan dalam penelitian ini ialah nama, dan nilai setiap mata pelajaran dari semester 1 s/d 5

Normalisasi Data, merupakan sebuah teknik dalam logical desain sebuah basis data atau database, teknik mengelompokkan atribut dari suatu relasi sehingga membentuk struktur relasi yang baik tanpa redundasi. Tujuannya adalah untuk mengorganisasikan data ke dalam tabel-tabel untuk memenuhi kebutuhan pemakai, menghilangkan kerangkapan data, mengurangi kompleksitas dan mempermudah modifikasi data. Proses normalisasi yaitu dengan cara:

- Data diuraikan dalam bentuk tabel, selanjutnya dianalisis berdasarkan persyaratan tertentu ke beberapa tingkat.
- Apabila tabel yang diuji belum memenuhi persyaratan tertentu maka tabel tersebut perlu dipecah menjadi beberapa tabel yang lebih sederhana sampai memenuhi bentuk yang optimal.

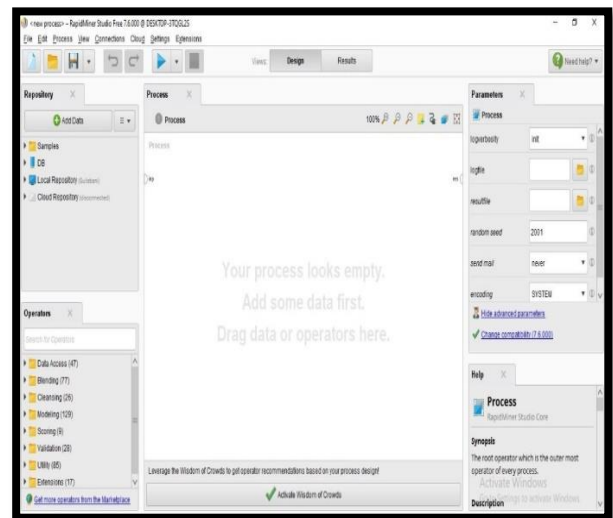


**Gambar 2. Proses Preprocessing Nominal to Numeric**

- Filter Examples*, operator *filter examples* ini digunakan untuk memilih contoh dari *ExampleSet* mana yang disimpan dan *ExampleSet* mana yang harus dihapus.
- Nominal to Numerical*, operator *Nominal to Numerical* digunakan untuk mengubah jenis atribut non numerik yang dipilih menjadi jenis numerik. Dan operator ini juga digunakan untuk memetakan semua nilai atribut ke nilai numerik.
- Normalize*, operator *normalize* ini digunakan untuk menormalkan nilai atribut yang telah dipilih.

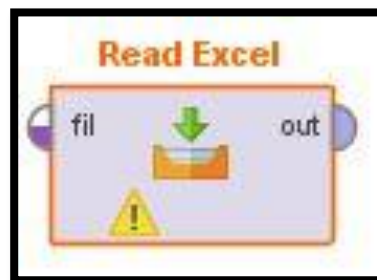
### 3.2 Pembahasan

Pada bagian ini data akan dianalisa secara deskriptif, data didapat dari data sekolah SMA Negeri 21 Bandung dengan jumlah data sebanyak 324 siswa yang kemudian akan diolah menggunakan *software RapidMiner* dengan menggunakan metode *K-Means Clustering*. Pada bagian ini pula dibahas bagaimana proses pemodelan pengolahan data menggunakan metode k-means dengan *RapidMiner 7.6* adalah sebagai berikut:



**Gambar 3. Halaman Utama RapidMiner**

Gambar di atas adalah halaman utama dari tampilan software *RapidMiner*, halaman yang akan digunakan dalam pengolahan data pengelompokan jurusan siswa SMA.



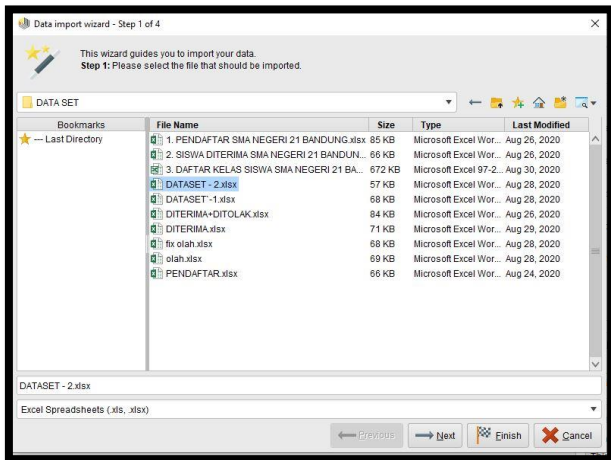
**Gambar 4. Operator Data Read Excel**

*Read Excel* pada Gambar diatas bertujuan untuk meload atau memasukan dataset yang akan diproses untuk pengolahan data dalam mengelompokkan siswa untuk menentukan jurusan. Untuk mengimport



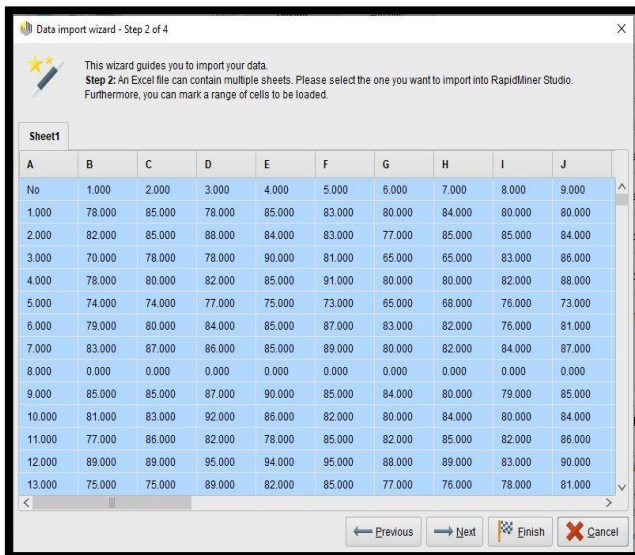
data pada aplikasi RapidMiner terdiri dari beberapa tahap, berikut langkah-langkah yang harus lakukan:

### 3.3 Data Import Wizard Step-1 of 4



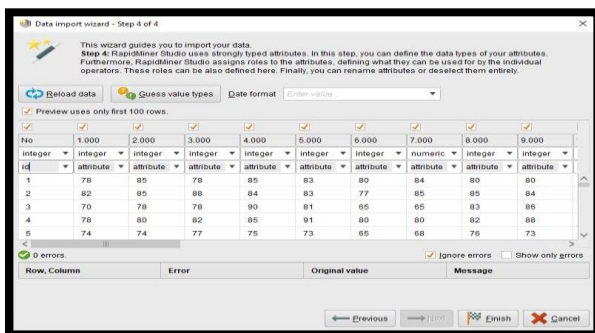
Gambar 5. Data Import Wizard Step 1 of 4

### 3.4 Data Import Wizard Step-2 of 4



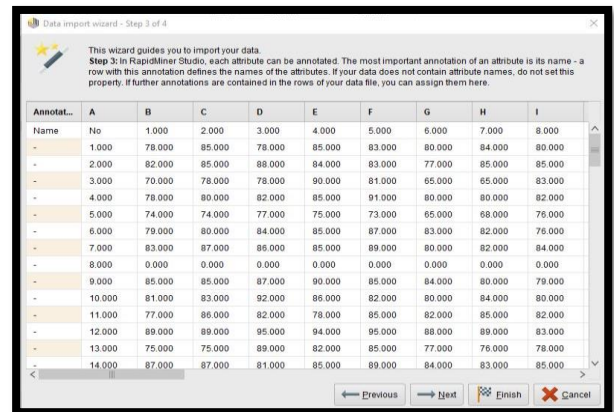
Gambar 6. Data Import Wizard Step 2 of 4

### 3.5 Data Import Wizard Step-3 of 4



Gambar 7. Data Import Wizard Step 3 of 4

### 3.6 Data Import Wizard Step-4 of 4



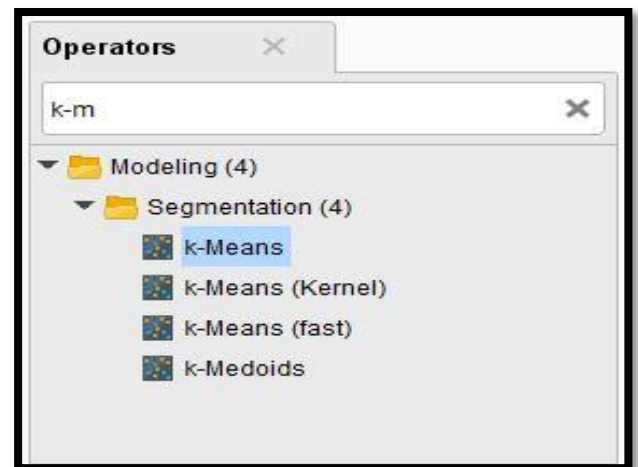
Gambar 8 Data Import Wizard Step 4 of 4

Tahapan *import* sesuai setelah mengklik finish, maka akan terlihat operator baru *Read Excel* yang sudah berisi file *dataset-2.xlsx* dan data tersebut telah siap untuk diuji. Gambar 4.8 Operator *Data Read Excel* adalah operator yang sudah berisi data yang akan di uji.



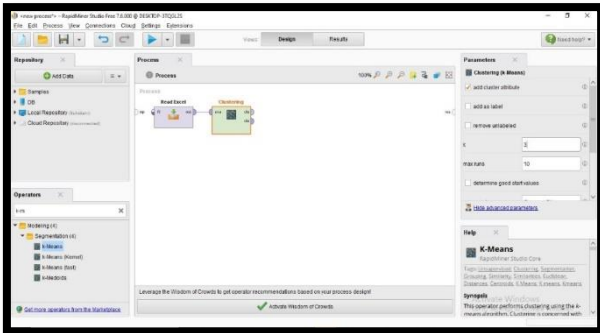
Gambar 9 Operator Data Read Excel Dengan File dataset -2

Langkah berikutnya adalah dengan memasukkan operator *K-Means* dengan memilih menu *K-Means*.



Gambar 10 Operator K-Means

Operator *K-Means* pada Gambar diatas adalah operator yang berfungsi untuk mengelompokan data siswa dalam menentukan jurusan dengan memaksimalkan kemiripan data antar setiap *cluster*.



**Gambar 11 Tampilan Operator Read Excel dan K-Means**

#### IV. KESIMPULAN

Berdasarkan hasil dari penelitian yang sudah dilakukan dalam menganalisa pengelompokan penentuan jurusan menggunakan metode *k-means* dapat disimpulkan bahwa *k-means* dapat digunakan untuk mengelompokan penentuan jurusan dengan hasil yang optimal dan tidak memakan waktu yang lama. Berdasarkan hasil pengujian menunjukkan bahwa *cluster\_0* ada 5 siswa yang ditolak di dua jurusan, *cluster\_1* 142 siswa masuk jurusan IPS dan *cluster\_2* 177 siswa masuk dalam jurusan IPA. Membandingkan dengan hasil dari penjurusan siswa yang dilakukan SMA 21 Bandung menggunakan excel sangat berbeda dengan hasil yang sudah diuji menggunakan *RapidMiner*, terdapat 144 siswa jurusan IPS dan 180 siswa jurusan IPA.

#### DAFTAR RUJUKAN

- [1] M. D. Afriansyah, F. Teknik, I. Komputer, and U. B. Darma, "Klasifikasi Menentukan Penjurusan Siswa Pada SMA 18 Palembang," no. 1, pp. 2683–2686.
- [2] P. Teknik, I. Universitas, and D. Nuswantoro, "Penentuan Jurusan

- Siswa Sekolah Menengah Atas Disesuaikan Dengan Minat Siswa Menggunakan Algoritma Fuzzy C-Means" vol. 13, pp. 57–68, 2017.
- [3] C. Purnamaningsih and A. Aziz, "Pemanfaatan Metode K-Means Clustering dalam Penentuan Penjurusan Siswa SMA," vol. 3, no. 1, pp. 27–33, 2014.
- [4] P. Smk, N. Kota, J. Tandy, and S. Assegaff, "Analisis Dan Perancangan Clustering Siswa Baru Menggunakan Metode K-Means Pada Smk Negeri 1 Kota Jambi," vol. 4, no. 4, pp. 389–399, 2019.
- [5] N. Putu, E. Merliana, and A. J. Santoso, "Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means," pp. 978–979.
- [6] S. Muklis, "Sistem Pendukung Keputusan Minat Pemilihan Jurusan Sma Dengan Metode K-Means Cluster Analysis," pp. 6–8, 2015.
- [7] Y. Syahra, "Penerapan Data Mining Dalam Pengelompokan Data Nilai Siswa Untuk Penentuan Jurusan Siswa Pada SMA Tamora Menggunakan Algoritma K-Means Clustering," vol. 17, no. SAINTIKOM, pp. 228–233, 2018.
- [8] T. Syahputra, J. Halim, and E. P. Sintho, "Penerapan Data Mining Dalam Menentukan Pilihan Jurusan Bidang Studi Sma Menggunakan Metode Clustering Dengan Teknik Single Linkage," JURTEKSI (Jurnal Teknol. dan Sist. Informasi), vol. IV, no. 2, pp. 205–208, 2018.