

Penerapan Data Mining Dengan Menggunakan Metode K-Means Clustering Dalam Pengelompokan Data Nilai Pada SMA YKPP PENDOPO Untuk Menentukan Jurusan Ipa Dan Ips

Wira Kokoh Andriyan¹⁾

¹⁾Program Studi Teknik Informatika, Universitas Bina Darma Palembang, Jalan Jenderal A. Yani No. 3 Palembang Sumatera Selatan, Indonesia.
e-mail: *wirakokoh92@gmail.com

Abstrak

Dengan perkembangan ilmu pengetahuan dan teknologi didalam bidang pendidikan, teknologi sangat berperan penting mendukung sistem pembelajaran pada kurikulum 2013. Penelitian ini bertujuan untuk mengetahui jumlah siswa yang akan masuk dikelas IPA dan kelas IPS sesuai dengan penjurusaannya agar dapat mengembangkan potensi diri dan meningkatkan nilai akademik siswa, dengan itu penulis membutuhkan data nilai siswa/siswi pada SMA YKPP PENDOPO dan diharapkan dapat membantu mengambil keputusan dengan menerapkan metode k-means clustering. algoritma k-means ialah dengan membagi data kemudian mengelompokkannya kedalam beberapa cluster yang memiliki karakteristik data dan membagi setiap cluster berdasarkan perbedaan karakter antar cluster. Clustering adalah 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. Hasil dari pengujian di lakukan dalam penerapan data mining dengan metode k-means clustering dalam pengelompokan data nilai siswa untuk penjurusan ipa dan ips, dapat disimpulkan bahwa k-means dapat digunakan untuk mengelompokkan penentuan jurusan dengan hasil yang optimal dan tidak memakan waktu yang lama. pengujian juga menggunakan tools Rapid Miner dan menggunakan data skunder, dimana data ini berisi nilai-nilai raport pada saat pendaftaran. Berdasarkan hasil pengujian menunjukan bahwa, di antara 58 siswa yang terdaftar di SMA YKPP Pendopo, ada 15 siswa yang terdaftar di kelas IPS 1 di Cluster 0, 11 siswa yang terdaftar di kelas IPA 1 di Cluster 1, 15 siswa yang terdaftar di kelas IPS 2 di Cluster 2, dan 13 siswa yang terdaftar di kelas IPA 2 di Cluster 3.

Kata kunci—Data, K-Means, Clustering, RapidMiner

Abstract

With the development of science and technology in the field of education, technology plays an important role in supporting the learning system in the 2013 curriculum. This study aims to determine the number of students who will enter science and social classes in accordance with their majors in order to develop their potential and improve students' academic grades, with that the author needs data on student grades at YKPP PENDOPO High School and is expected to help make decisions by applying the k-means clustering method. The k-means algorithm is to divide the data then group it into several clusters that have data characteristics and divide each cluster based on differences in character between clusters. Clustering is a process where grouping and dividing into several data sets to form similar patterns and grouped in the same cluster and separate themselves by forming different patterns in different clusters. The results of the tests carried out in the application of data mining with

the k-means clustering method in grouping student score data for the majors of science and ips, it can be concluded that k-means can be used to group the determination of majors with optimal results and does not take a long time. The test also uses Rapid Miner tools and uses secondary data, where this data contains report card values at the time of registration. Based on the test results show that, among the 58 students enrolled in YKPP Pendopo High School, there are 15 students enrolled in social studies class 1 in Cluster 0, 11 students enrolled in science class 1 in Cluster 1, 15 students enrolled in social studies class 2 in Cluster 2, and 13 students enrolled in science class 2 in Cluster 3.

Keywords— *Data, K-Means, Clustering, RapidMiner*

1. PENDAHULUAN

Pesatnya perkembangan ilmu pengetahuan dan teknologi telah merambah kedalam semua aspek kehidupan seperti dibidang bisnis, sosial, atau pun dibidang pendidikan. Teknologi sangat berperan penting pada sektor pendidikan yang mendukung sistem pembelajaran dikurikulum 2013 [1].

Dimana penjurusan untuk Sekolah Menengah Atas (SMA) dilaksanakan ketika akan memasuki jenjang SMA yaitu pada waktu pendaftaran awal masuk sekolah. Penentuan jurusan bagi siswa/siswi merupakan proses memfokuskan siswa pada bidang konsentrasi tertentu, hal ini dilakukan agar setiap siswa dapat mempelajari lebih dalam tentang pelajaran yang sesuai dengan konsentrasi yang telah ditentukan, serta mengarahkan setiap siswa agar bisa mengembangkan kemampuan diri dan minat yang dimiliki [2].

Akibat kesalahan dalam pemilihan jurusan yang tidak sesuai dengan yang seharusnya diterima siswa akan menyebabkan menurunnya motivasi belajar dan rendahnya prestasi karena jurusan yang dipilih ternyata tidak sesuai dengan kemampuan dan keterampilan siswa/siswi [3]. Dengan penjurusan ini siswa diharapkan dapat memaksimalkan potensi, bakat atau keterampilan, dengan demikian akan memaksimalkan nilai akademis siswa/siswi.

Penentuan jurusan yang dilakukan selama ini mempunyai banyak kelemahan, antara lain berdasarkan keinginan siswa tanpa melihat latar belakang nilai akademisnya dan tidak ada acuan yang tepat dalam melakukan pengelompokkan jurusan siswa, Sehingga jurusan yang dipilih terkadang menjadi masalah bagi siswa di kemudian hari. Hal ini sejalan dengan Permendikbud 81 A tahun 2013 tentang implementasi kurikulum 2013 menjelaskan bahwa pendidikan merupakan proses yang sistematis manusia secara holistik yang memungkinkan potensi diri berkembang secara optimal [4].

Peminatan atau penjurusan ini juga membantu memberi gambaran akan bidang yang nantinya akan ditekuni oleh siswa-siswi setelah lulus sekolah nanti. Namun dalam penentuan jurusan untuk SMA sering sekali menjadi permasalahan di sekolah, antara lain sulitnya dalam menentukan siswa mana yang akan masuk untuk menempati jurusan tertentu. Kriteria penentuan jurusan siswa hanya berdasarkan nilai akademik, penentuan jurusan dengan sistem yang lama ditentukan berdasarkan nilai akhir saja maksudnya bahwa penentuan hanya berdasarkan nilai rata-rata USBN (Ujian Sekolah Berstandar Nasional) tanpa menjabarkan setiap nilai dari mata pelajaran yang mewakili jurusan tersebut.

Salah satu cara untuk mempermudah penentuan penjurusan SMA adalah dengan cara mengelompokkan (clustering) data siswa yaitu, untuk clustering IPA dikelompokkan menjadi 2 kelompok antara lain IPA 1 dan IPA 2. Untuk clustering IPS dikelompokkan menjadi 2 kelompok antara lain IPS 1 dan IPS 2.

Untuk mengelompokkan siswa ini, metode yang digunakan adalah *K-Means Clustering*. Metode ini merupakan metode untuk mengelompokkan objek atau data yang mempunyai atribut dan mempunyai jumlah data yang banyak ke dalam bentuk satu atau lebih kelompok, sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster/kelompok yang sama [5].

algoritma clustering berusaha mensegmentasi seluruh kumpulan data ke dalam subkelompok-subkelompok atau cluster-cluster homogen secara relatif. Dimana kesamaan record dalam cluster dimaksimalkan dalam kesamaan dengan record diluar cluster ini diminimalkan. Untuk kriteria yang digunakan dalam penelitian ini adalah kriteria nilai akademik, nilai IQ dan minat siswa. Data Mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (machine learning) untuk menganalisis dan mengekstrasi pengetahuan (knowledge) secara otomatis. Berdasarkan kondisi tersebut, penulis ingin menerapkan Data mining untuk mengelompokkan jurusan tersebut menggunakan metode *K-means Clustering* pada SMA YKPP PENDOPO. Hal ini diharapkan dapat mempermudah dan meningkatkan kinerja guru dalam pengolahan data untuk menentukan jurusan bagi siswa/siswi yang sesuai dengan kemampuan dan minat dari siswa itu sendiri.

2. METODE PENELITIAN

2.1 Populasi

Tahapan Populasi dalam penelitian ini adalah data nilai siswa/siswi pada saat pendaftaran yang dapat menukung dalam penentuan jurusan untuk masuk ke kelas 10 SMA YKPP PENDOPO.

2.2 Data

Tahapan ini Data adalah segala fakta, angka, atau teks yang dapat diproses oleh komputer. Saat ini, akumulasi pertumbuhan jumlah data berjalan dengan cepat dalam format dan basis data yang berbeda. Informasi adalah pola, asosiasi atau hubungan anantara semua data yang dapat memberikan informasi [6].

Data penelitian ini adalah data nilai semester terakhir atau semester 6 SMP siswa/siswi pada saat waktu pendaftaran pada SMA YKPP PENDOPO pada tahun ajaran 2021 - 2022, dengan jumlah sebanyak 58 siswa dan siswi yang mendaftar.

2.3 Tahapan Pengumpulan Data

Teknik pengumpulan data merupakan suatu cara yang dilakukan oleh peneliti untuk memperoleh data data yang diperlukan. Dalam penelitian ini teknik yang digunakan antara lain sebagai berikut:

2.4 Observasi

Melalui observasi penulis melakukan tinjauan langsung pada SMA YKPP PENDOPO untuk mendapatkan informasi mengenai profil sekolah dan data-data yang akan diperlukan dalam penelitian ini.

2.5 Wawancara

Wawancara dilakukan pada kepala sekolah dapat mendalami permasalahan secara keseluruhan sehingga penelitian ini memiliki konsep untuk memecahkan masalah.

2.6 Studi Lilelature

Mempelajari teori - teori tentang penerapan data mininng dalam pengelompokan data nilai siswa menggunakan k-means clustering, pada pemilihan jurusan serta beberapa metode pengolahan data melalui jurnal-jurnal dan buku yang berkaitan

2.7 Jenis Data

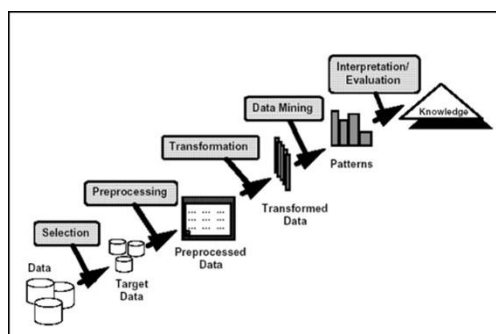
Jenis yang digunakan dalam penelitian ini adalah jenis data sekunder. Data sekunder adalah data yang diperoleh peneliti dari sumber atau data yang sudah ada. Data sekunder yang

digunakan dalam penelitian ini adalah dataset dari nilai raport pada saat pendaftaran masuk SMA YKPP Pendopo.

2.8 Data Mining

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan [7].

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



Gambar 1. Alur Knowledge Discovery in Database (KDD)

- 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 YKPP Pendopo 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 YKPP Pendopo 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) merupakan visualisi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Penelitian

Data cleaning, proses ini bertujuan untuk membersihkan data yang tidak sesuai atau memiliki duplikasi pada data, metode ini tidak hanya membersihkan data tetapi juga dapat memperbaiki kesalahan ketik. Dalam penelitian ini, data telah dibersihkan sehingga disimpulkan bahwa data tersebut layak untuk proses clustering.

Pre-processing, proses yang digunakan untuk menyaring data yang dibutuhkan dalam proses clustering, data diperoleh dari pendaftar yang diterima di SMA YKPP Pendopo. Variabel yang digunakan dalam penelitian ini adalah nama dan nilai setiap mata pelajaran pada semester 6.

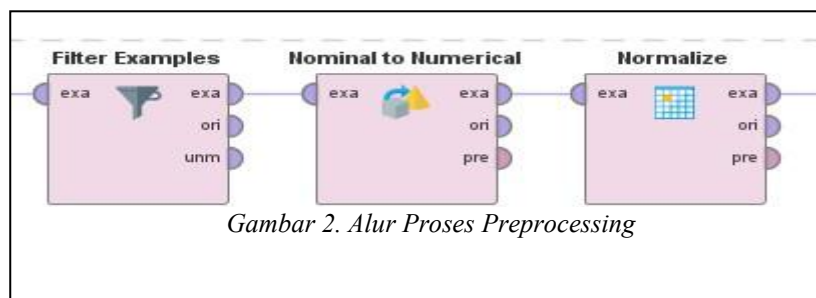
Data Transforamiton, adalah suatu teknik dalam suatu desain logika database atau basis data untuk mengelompokkan atribut-atribut dari suatu relasi sedemikian rupa sehingga membentuk struktur relasi yang baik tanpa redundansi. Hal ini bertujuan untuk mengorganisasikan data ke dalam tabel-tabel untuk memenuhi kebutuhan pengguna dan memfasilitasi pengembangan data.

a. *Filter Examples*

Operator *filter examples* ini digunakan untuk memilih contoh dari *ExampleSet* mana yang disimpan dan *ExampleSet* mana yang harus dihapus.

b. *Nominal to Numerical*

Operator ini digunakan untuk mengubah jenis atribut non numerik yang dipilih menjadi



Gambar 2. Alur Proses Preprocessing

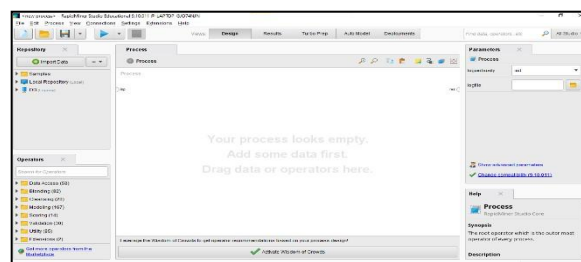
jenis numerik. Dan operator ini juga digunakan untuk memetakan semua nilai atribut ke nilai numerik

c. *Normalize*

Operator *normalize* ini digunakan untuk menormalkan nilai atribut yang telah dipilih.

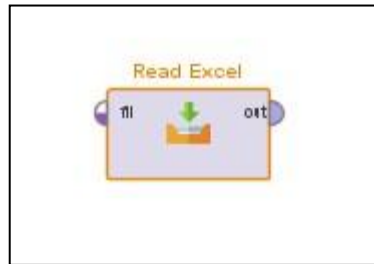
3.2 Pembahasan

Pada bagian ini data akan diterapkan proses data mining yang didapat dari data SMA YKPP Pendopo dengan jumlah data sebanyak 58 siswa yang kemudian akan diolah menggunakan *software RapidMiner* dengan menggunakan metode *K-Means Clustering*. Pada tahapan ini juga akan dijelaskan proses pemodelan, pengolahan data menggunakan metode K-Means clustering dengan *RapidMiner 9.10* adalah sebagai berikut.



Gambar 3. Tampilan awal RapidMiner

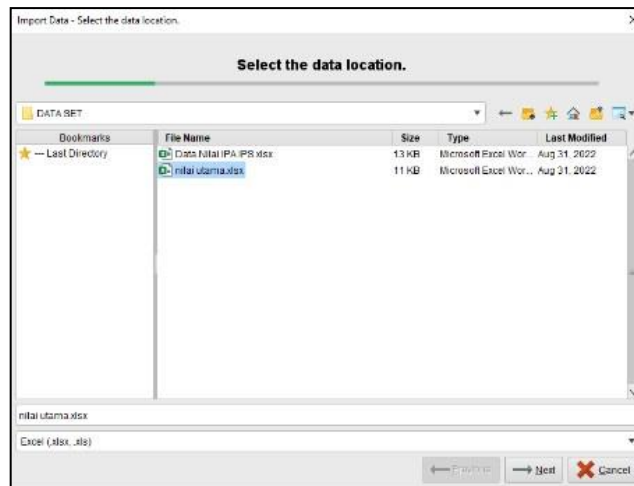
Gambar di atas adalah tampilan Home page atau halaman utama pada *software RapidMiner*, halaman ini yang akan digunakan dalam pengolahan data pengelompokan jurusan.



Gambar 4. Operator Read Excel

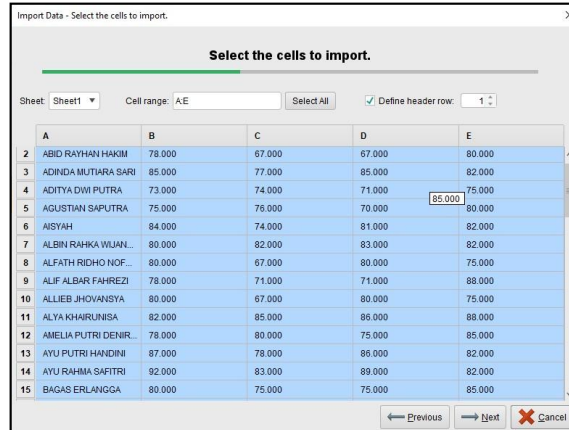
Read Excel pada Gambar diatas bertujuan untuk memasukan dataset yang akan diproses untuk pengolahan data, dalam mengelompokan siswa untuk menentukan jurusan. Untuk mengimport data pada aplikasi *RapidMiner* terdiri dari beberapa tahap, berikut langkah langkah yang harus lakukan:

3.3 Tahap Mengimport Data 1-3



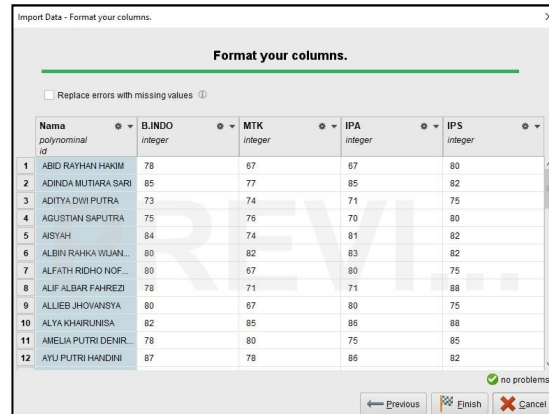
Gambar 5. Alur import data 1

3.4 Tahap Mengimport Data 2-3



Gambar 6. Alur import data 2

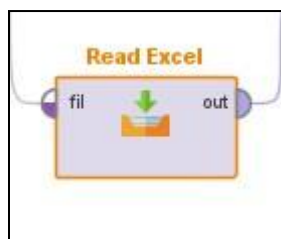
3.5 Tahap Mengimport Data 3-3



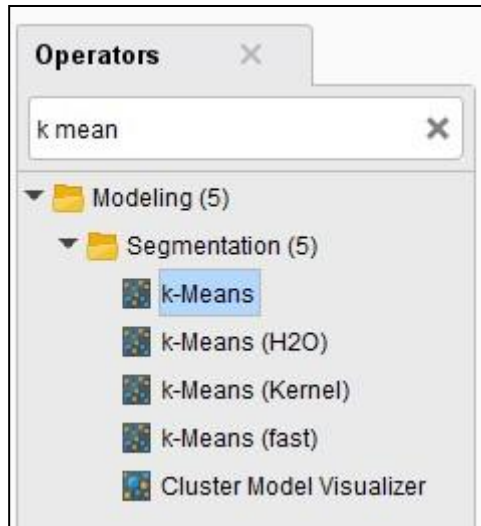
Gambar 7. Alur import data 3

Tahapan import data selesai setelah mengklik tombol finish, maka akan terlihat operator baru *Read Excel* yang sudah berisi file data nilai utama.xlsx dan data tersebut telah siap untuk diuji yang telah melalui proses pre-processing. Gambar 8. Operator Data *Read Excel* adalah operator yang sudah berisi data yang akan siap di uji.

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

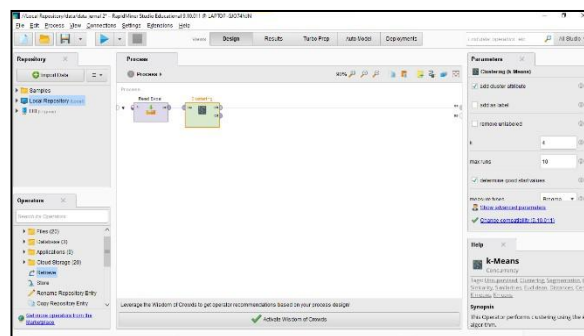


Gambar 8. Operator *Read Excel* yang telah terisi file dataset

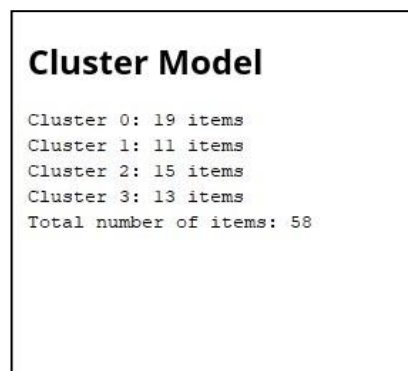


Gambar 9. Operator K-Means

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



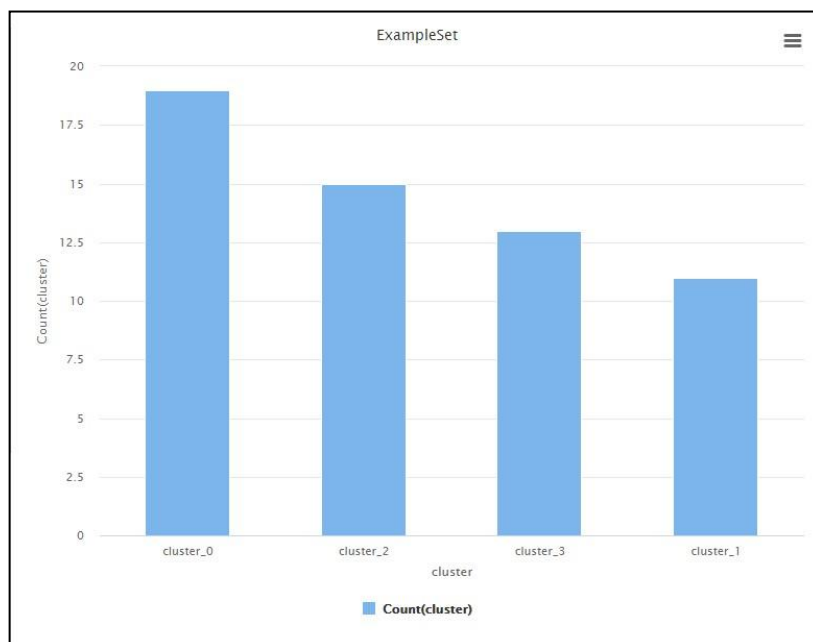
Gambar 10. Tampilan Operator Read Excel dan K Means Clustering



Gambar 11. Cluster model

Cluster model pada gambar diatas adalah hasil dari proses *Start execution* pada *tools rapidminer* dengan jumlah cluster sebanyak 4 cluster.

Tahap berikutnya adalah visualisasi data, visualisasi data berfungsi sebagai penyampaian informasi dari hasil data yang di miliki dengan menggunakan diagram atau chart, sehingga penyampaian data hasil lebih mudah di pahami, informatif dan mendalam.



Gambar 12. Visualisasi Hasil Cluster Pada Chart Bar

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan dalam penerapan data mining dalam pengelompokan penentuan jurusan dengan menggunakan metode *k-means clustering* dapat disimpulkan bahwa metode *k-means clustering* dapat digunakan untuk mengelompokkan penentuan jurusan dengan hasil yang optimal dan tidak memakan waktu yang lama. Menurut hasil pengujian menunjukan bahwa, di antara 58 siswa yang terdaftar di SMA YKPP Pendopo, ada 19 siswa yang terdaftar di kelas IPS 1 di Cluster 0, 11 siswa yang terdaftar di kelas IPA 1 di Cluster 1, 15 siswa yang terdaftar di kelas IPS 2 di Cluster 2, dan 13 siswa yang terdaftar di kelas IPA 2 di Cluster 3.

5. SARAN

Penelitian saat ini terbatas pada penerapan data mining dengan algoritma *k-means clustering* menggunakan RapidMiner untuk mengelompokkan jurusan siswa di SMA YKPP Pendopo dengan jurusan IPA dan IPS. Karena keterbatasan tersebut, penulis berharap pada penelitian selanjutnya algoritma *k-means* ini nantinya dapat diimplementasikan pada aplikasi atau website sehingga lebih mempermudah para guru dalam menentukan jurusan dengan hasil yang terperinci.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada semua pihak yang telah mendukung terlaksananya penelitian ini, terutama kepada SMA YKPP Pendopo yang telah bersedia sebagai

objek penelitian dan juga membantu memberikan informasi penting terkait dalam penelitian, dan juga kepada ibu Vivi Sahfitri, M.Kom., M.M selaku pembimbing saya yang telah sabar membimbing saya dalam penelitian ini. Saya ucapkan terima kasih banyak.

DAFTAR PUSTAKA

- [1] F. Firza and S. Sarjono, "Penerapan Algoritma K-Means Dalam Metode Clustering Untuk Peminatan Jurusan Bagi Siswa Swasta Pelita Raya Kota Jambi," *J. Manaj. Sist. Inf.*, vol. 5, no. 3, pp. 371–382, 2020, [Online]. Available: <http://ejournal.stikom-db.ac.id/index.php/manajemensisteminformasi/article/view/907>.
- [2] S. W. Nengsih, I. Alfian, D. Aji, and S. Anwar, "Analisis Pengelompokan Penentuan Jurusan Siswa Sma Menggunakan Metode K-Means Clustering," *J. Betrik*, no. 03, pp. 242–248, 2021.
- [3] Y. Lase and E. Panggabean, "Implementasi Metode K-Means Clustering Dalam Sistem Pemilihan Jurusan Di SMK Swasta Harapan Baru," *J. Teknol. dan Ilmu Komput. Prima*, vol. 2, no. 2, p. 43, 2019, doi: 10.34012/jutikomp.v2i2.723.
- [4] S. Bombana and M. S. Said, "PENERAPAN ALGORITMA K-MEANS DALAM PENENTUAN JURUSAN SISWA," vol. 5, no. 2, 2020.
- [5] 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. 2, pp. 228–233, 2018.
- [6] B. Hasmaulina, "Penerapan Data Mining Untuk Membentuk Kelompok Belajar Menggunakan Metode Clustering Di SMK Negeri 3 Seluma Pendahuluan Studi Literatur," *J. ILMU Komput. DAN Inform.*, vol. 4, pp. 57–71, 2021.
- [7] A. Asroni, H. Fitri, and E. Prasetyo, "Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik)," *Semesta Tek.*, vol. 21, no. 1, pp. 60–64, 2018, doi: 10.18196/st.211211.